A stochastic model of gene expression with polymerase recruitment and pause release

Z. Cao, T. Filatova, D. A. Oyarzún and R. Grima

July 19, 2020

Abstract

Transcriptional bursting is a major source of noise in gene expression. The telegraph model of gene expression, whereby transcription switches between "on" and "off" states, is the dominant model for bursting. Recently it was shown that the telegraph model cannot explain a number of experimental observations from perturbation data. Here we study an alternative model that is consistent with the data and which explicitly describes RNA polymerase recruitment and polymerase pause release, two steps necessary for mRNA production. We derive the exact steady-state distribution of mRNA numbers and an approximate steady-state distribution of protein numbers which are given by generalized hypergeometric functions. The theory is used to calculate the relative sensitivity of the coefficient of variation of mRNA fluctuations for thousands of genes in mouse fibroblasts. This indicates that the size of fluctuations is mostly sensitive to the rate of burst initiation and the mRNA degradation rate. Furthermore we show that (i) the time-dependent distribution of mRNA numbers is accurately approximated by a modified telegraph model with a Michaelis-Menten like dependence of the effective transcription rate on RNA polymerase abundance. (ii) the model predicts that if the polymerase recruitment rate is comparable or less than the pause release rate, then upon gene replication the mean number of RNA per cell remains approximately constant. This gene dosage compensation property has been experimentally observed and cannot be explained by the telegraph model with constant rates.

Statement of Significance

The random nature of gene expression is well established experimentally. Mathematical modelling provides a means of understanding the factors leading to the observed stochasticity. There is evidence that the classical two-state model of stochastic mRNA dynamics (the telegraph model) cannot describe perturbation experiments and a new model that includes polymerase dynamics has been proposed. In this paper, we present the first detailed study of this model, deriving an exact solution for the mRNA distribution in steady-state conditions, an approximate time-dependent solution and showing the model can explain gene dosage compensation. As well, we use the theory together with transcriptomic data, to deduce which parameters when perturbed lead to a maximal change in the size of mRNA fluctuations.

1 Introduction

There is widespread evidence that mammalian genes are expressed in bursts: infrequent periods of transcriptional activity that produce a large number of messenger RNA (mRNA) transcripts within a short period of time [1–3]. This is in contrast to constitutive expression where mRNAs are produced in random, uncorrelated events, with a time-independent probability [4]. The size and frequency of transcriptional bursts affect the magnitude of temporal fluctuations in mRNA and protein content of a cell, and thus constitute an important source of intracellular noise [5].

A large number of studies have sought to elucidate the mechanisms leading to bursting and by constructing simple stochastic models that can explain the data. The simplest of these models is the telegraph model whereby (i) a gene is in two states, an ON state where mRNA is expressed and an OFF state where there is no expression. (ii) mRNA degrades in the cytoplasm. These first-order reactions are effective since each encapsulates the effect of a large number of underlying biochemical reactions. The chemical master equation

of this model has been solved exactly to obtain the probability distribution of mRNA numbers as a function of time [6]. For parameter conditions consistent with bursty expression, the steady-state distribution is well approximated by a negative binomial that fits some of the experimental data [7].

Recent studies have extended the telegraph model in various directions (see [8] for a recent review). Mammalian cells have been shown to display complex promoter dynamics during the switch from transcriptionally inactive to active states. Such dynamics cannot be described by a single reaction step whose time is exponentially distributed [2], as assumed by the telegraph model. In [9] this complexity is accounted for by deriving analytical expressions linking the Fano factor of mRNA distributions to the general waiting-time distribution of the time to switch from inactive to active states. In contrast, other works [10–12] have sought to describe promoter dynamics with transitions between a number of discrete promoter states, only some of which are active; in special cases of such models, the steady-state distribution of mRNA fluctuations can be derived analytically. Moreover, dynamic regulation of eve stripe 2 expression in living Drosophila [13] suggests the occurrence of multiple rates of Pol II loading, which argues in favour of the multistate model rather than the simpler telegraph model. Another study, based on live cell imaging of the amoeba Dictyostelium, postulates a continuum of transcriptional states [14] rather than discrete states. All these models share a common property with the telegraph model, namely that when a transcript is produced, the gene state is unchanged.

Bartman et al. [15] recently argued that it is unclear how polymerase recruitment and pause release, two well-known steps in mRNA production, map onto the active and inactive states assumed by the telegraph model. This argument also applies to the various multistate variants of the telegraph model. In particular, in these models one cannot tell whether the initiation of a burst permits polymerase recruitment to occur or whether it permits release from the paused state. In [15], the telegraph model and several possible models of transcription were considered that incorporated bursting (burst initiation and termination steps) together with polymerase recruitment and pause release steps. Using stochastic simulations in conjunction with RNA FISH and Pol II ChIP-seq measurements, they showed that the only model compatible with the data is one in which (i) polymerase recruitment follows after burst initiation and (ii) only one polymerase is permitted to bind each promoter-proximal region at a time, and this bound polymerase has to undergo pause release before a second polymerase can be recruited to a gene copy (in line with the findings in [16, 17]). We emphasize that while this model has three effective gene states, it is not a special case of the multistate gene models studied in [10–12]. These models assume that the gene state does not change upon production of mRNA because they model the production of a mature transcript without detailed modelling of the steps between transcriptional initiation and termination. However the model expounded in [15] models transcription at a finer level of detail which requires that the production of nascent mRNA results in a change of gene state, a property that is crucial to capture property (ii) above. Note the number of nascent mRNA molecules, irrespective of their length, is equal to the number of polymerases currently transcribing the gene [18]. An interesting recent review discussing the assumptions behind common gene expression models including those with polymerase dynamics can be found in [19].

In this article, we present the first detailed study of the model proposed by Bartman et al [15]. The article is organized as follows. In Section 2.1 we introduce the chemical master equation formulation of the model, in Section 2.2 we obtain an an exact steady-state solution of this model and in Section 2.3 we use the theoretical results and transcriptomic data to investigate the sensitivity of the size of mRNA fluctuations to the five parameters. In Section 2.4 we show that by mapping the model onto an effective telegraph model, we can obtain an approximate time-dependent solution. In Section 2.5 we show that while our model has three effective promoters states, it is not the same as the refractory model of gene expression devised by Naef and co-workers [2]. In Section 2.6 we show that the protein number distribution can also be obtained in the limit of fast mRNA decay and that this is generally different than that obtained using the conventional three-stage model of gene expression [20]. We finish with a discussion of the biological implications of our results in Section 3.

2 Results

2.1 Model

We consider a stochastic transcriptional bursting model (recently introduced in [15] and henceforth referred to as the multi-scale model; see Fig. 1A), whereby a gene fluctuates between three states: two permissive states (D_{10} and D_{11}) and a non-permissive state (D_0).

The transition from D_0 to D_{10} (burst initiation) is mediated by transcription factor binding with rate

constant σ_u which is reversible with rate constant σ_b (this transition may alternatively represent other processes such as nucleosome remodeling). Subsequently the binding of RNA polymerase II (Pol II) to D_{10} with rate constant λ (which is proportional to Pol II abundance) leads to D_{11}. This represents a state in which Pol II is paused, and models the experimental observation that Pol II pauses downstream of the transcription initiation site preceding productive elongation [17]. The polymerase is released from this state with rate constant ρ leading to two simultaneous processes: (i) since now the polymerase can actively transcribe RNA, it implies the production of nascent mRNA (denoted as N) with rate ρ; (ii) the gene state changes from D_{11} to D_{10}. This step models the experimental observation that unless the polymerase is unpaused, there is no binding of new Pol II [16,17]. In the paused state D_{11}, both the polymerase and the transcription factor can unbind from the gene and lead to the non-permissive state D_0 (burst termination). Both reversible switches operate at different timescales (hours versus minutes) with $\max\{\sigma_b, \sigma_u\} \ll \min\{\rho, \lambda\}$, leading to multi-scale transcriptional bursting [15, 21]. After termination, the nascent mRNA becomes a mature mRNA (denoted by M); this occurs with rate r. Subsequently the mature mRNA decays with rate constant d. Note that we assume all reactions to be first-order, characterized by exponentially distributed waiting times between successive reactions.

In what follows, for simplicity, we assume that the lifetime of nascent mRNA is very short, i.e. r is large, such that the reaction $D_{11} \to D_{10} + N$, $N \to M$ can be approximated by the single reaction step $D_{11} \to D_{10} + M$. In the next section, we derive the steady-state distribution of mature mRNA (simply called mRNA henceforth).

2.2 Exact solution

Let $P_\theta(n, t)$ ($\theta = 0, 10, 11$) denote the probability of a cell being in state D_θ with n mRNAs at time t (arguments n and t are hereafter omitted for brevity). The dynamics of probability P_θ are described by the

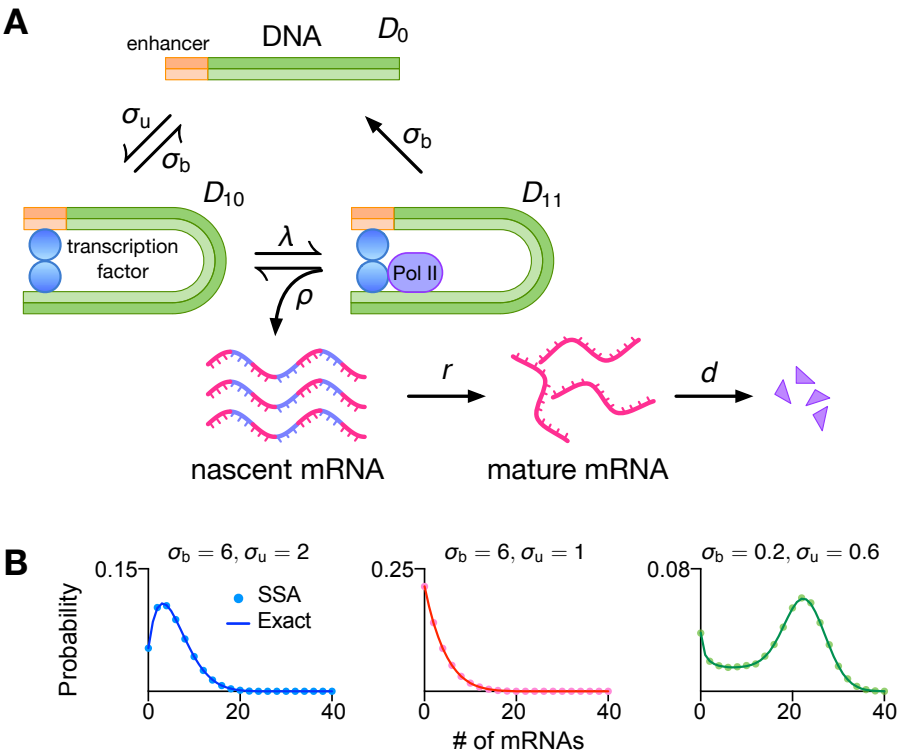

Figure 1: (A) Schematic of the stochastic multi-scale transcriptional bursting model. (B) Analytical distribution for mature mRNA numbers (under the assumption of short-lived nascent mRNA) is given by Eq. (6) and agrees with stochastic simulations using the SSA. The kinetic parameters are $\rho = 60$, $\lambda = 40$, $d = 1$; other parameters are indicated in each panel.

set of coupled master equations

$$\begin{aligned}\partial_t P_0 =&(\mathbb{E}^1-1)dnP_0-\sigma_{\rm u}P_0+\sigma_{\rm b}(P_{10}+P_{11}),\\ \partial_t P_{10} =&(\mathbb{E}^1-1)dnP_{10}-(\sigma_{\rm b}+\lambda)P_{10}+\sigma_{\rm u}P_0+\rho\mathbb{E}^{-1}P_{11},\\ \partial_t P_{11} =&(\mathbb{E}^1-1)dnP_{11}-(\rho+\sigma_{\rm b})P_{11}+\lambda P_{10},\end{aligned} \tag{1}$$

where the step operator $\mathbb{E}^i$ acts on a general function $g(n)$ as $\mathbb{E}^i g(n)=g(n+i)$ [22]. To solve Eq. (1), we use the generating function method and define $G_\theta(z)=\sum_n z^n P_\theta(n)$ for $\theta=0,10,11$ so that Eq. (1) can be recast as a set of coupled partial differential equations

$$\partial_t G_0+d(z-1)\partial_z G_0=-\sigma_{\rm u}G_0+\sigma_{\rm b}G_{10}+\sigma_{\rm b}G_{11}, \tag{2a}$$

$$\partial_t G_{10}+d(z-1)\partial_z G_{10}=\rho z G_{11}-(\sigma_{\rm b}+\lambda)G_{10}+\sigma_{\rm u}G_0, \tag{2b}$$

$$\partial_t G_{11}+d(z-1)\partial_z G_{11}=-\rho G_{11}-\sigma_{\rm b}G_{11}+\lambda G_{10}, \tag{2c}$$

wherein the variable z is dropped for brevity. By setting $z=1$ and the time derivatives to zero (considering steady-state conditions), we can deduce that the probability of being in the non-permissive state D_0 is $G_0(1)=\sigma_b/(\sigma_u+\sigma_b)$ and the probability of being in one of the two permissive states D_{10} or D_{11} is $G_{10}(1)+G_{11}(1)=\sigma_u/(\sigma_u+\sigma_b)$.

In order to solve Eq. (2) for $G_0(z),G_{10}(z),G_{11}(z)$ in steady-state conditions, we set $\partial_t G_\theta=0$, solve G_{10} from Eq. (2c) as a function of G_{11}, and combine the yielded result to solve G_0 from Eq. (2b) as a function of G_{11} so that Eq. (2a) consequently becomes a differential equation with G_{11} being the only variable

$$d^3u^2\partial_u^3G_{11}+(3d+\gamma_1+\gamma_2)d^2u\partial_u^2G_{11}+[(d+\gamma_1)(d+\gamma_2)-\rho\lambda u]d\partial_u G_{11}-(d+\sigma_{\rm u})\rho\lambda G_{11}=0, \tag{3}$$

with $u=z-1$, $\gamma_1=\sigma_{\rm b}+\sigma_{\rm u}$ and $\gamma_2=\rho+\lambda+\sigma_{\rm b}$. By defining a new variable $x=\rho\lambda u/d^2$, Eq. (3) can be further simplified to

$$x^2\partial_x^3G_{11}+\Big(1+\frac{\gamma_1+d}{d}+\frac{\gamma_2+d}{d}\Big)x\partial_x^2G_{11}+\Big(\frac{\gamma_1+d}{d}\frac{\gamma_2+d}{d}-x\Big)\partial_x G_{11}-\frac{\sigma_{\rm u}+d}{d}G_{11}=0,$$

which is in the canonical form of the differential equation for the generalized hypergeometric function

$$x^2\partial_x^3f(x)+(1+b_1+b_2)x\partial_x^2f(x)+(b_1b_2-x)\partial_x f(x)-a_1f(x)=0,$$

admitting the solution $f(x)=C_1F_2(a_1;b_1,b_2;x)$ with C being an integration constant. Hence, the solution for G_{11} is in terms of the generalized hypergeometric function

$$G_{11}=C\cdot{}_1F_2\left(\frac{\sigma_{\rm u}+d}{d};\frac{\gamma_1+d}{d},\frac{\gamma_2+d}{d};\frac{\rho\lambda}{d^2}u\right). \tag{4}$$

On the other hand, summing Eqs. (2a)-(2c) and denoting $G=\sum_\theta G_\theta$, one can get $\partial_u G=\rho G_{11}/d$, which together with Eq. (4) leads to

$$G(u)=C_2\cdot{}_1F_2\left(\frac{\sigma_{\rm u}}{d};\frac{\sigma_{\rm b}+\sigma_{\rm u}}{d},\frac{\sigma_{\rm b}+\rho+\lambda}{d};\frac{\rho\lambda}{d^2}u\right).$$

Note that in the last step we made use of the general relation $\partial_z\,{}_1F_2(a;b,c;z)=\frac{a}{bc}\cdot{}_1F_2(a+1;b+1,c+1;z)$. The integration constant C_2 is found to be 1 by using the normalization condition $G(0)=1$. Hence, the exact solution for the generating function is

$$G(u)={}_1F_2\left(\frac{\sigma_{\rm u}}{d};\frac{\gamma_1}{d},\frac{\gamma_2}{d};\frac{\rho\lambda}{d^2}u\right). \tag{5}$$

Hence it follows that the marginal probability of finding n mRNAs in a cell is

$$P(n)=\frac{1}{n!}\frac{{\rm d}^nG(u)}{{\rm d}u^n}\bigg|_{u=-1}=\frac{1}{n!}\left(\frac{\rho\lambda}{d^2}\right)\frac{(\frac{\sigma_{\rm u}}{d})_n}{(\frac{\sigma_{\rm b}+\sigma_{\rm u}}{d})_n(\frac{\sigma_{\rm b}+\rho+\lambda}{d})_n}{}_1F_2\left(\frac{\sigma_{\rm u}}{d}+n;\frac{\sigma_{\rm b}+\sigma_{\rm u}}{d}+n,\frac{\sigma_{\rm b}+\rho+\lambda}{d}+n;-\frac{\rho\lambda}{d^2}\right), \tag{6}$$

where $(\cdot)_n$ is the Pochhammer symbol. In Fig. 1B we show that distributions obtained from Eq. (6) as well as the corresponding modality (a phenotypic signature [23]) are indistinguishable from distributions produced using the stochastic simulation algorithm (SSA) [24]. Note that here we have solved for the mature mRNA

distribution, under the assumption that nascent mRNA is short lived. In cases where this assumption is not physiologically meaningful and one is interested in the nascent mRNA distribution, then the latter is given by Eq. 6 with d replaced by r (the rate at which nascent mRNA changes to mature mRNA due to the termination of transcription).

2.2.1 Special case of bursty transcription.

It can be further shown by perturbation theory in Appendix A that when ρ, λ and σ_b are much greater than the rest of the parameters, the exact solution Eq. (6) reduces to the negative binomial distribution $P(n) = \mathrm{NB}(\frac{\sigma_\mathrm{u}}{d}, \frac{\rho}{\rho+\alpha})$ with $\alpha = \sigma_\mathrm{b}\gamma_2/\lambda$. Note the constraint on the parameters leads to time-series with large and short-lived bursts of transcription (since ρ, λ, and σ_b are large), separated by long silent intervals (since σ_u is small). Such bursty trancription is common in mammalian cells [3].

2.2.2 Relationship to the telegraph model.

It can also be shown that that in the limit of large ρ, the exact solution Eq. (6) reduces to the confluent hypergeometric solution of the telegraph model (see Appendix B). This is equivalent to the steady-state solution of the two state system $D_0 \underset{\sigma_\mathrm{b}}{\overset{\sigma_\mathrm{u}}{\rightleftharpoons}} D_{10} \xrightarrow{\lambda} D_{10} + M, M \xrightarrow{d} \varnothing$. The reduction to a two state model results from genes spending a short time in state D_{11} due to the large value of ρ. The production of an mRNA molecule involves the slow reaction step from D_{10} to D_{11} with rate λ followed by a very fast reverse step with rate ρ. Hence the rate of mRNA production is determined by the reaction rate of the slowest reaction, i.e. it is equal to λ. By similar reasoning, we can deduce that in the limit of large λ, the gene spends short time in the state D_{10} and the multi-scale model reduces to the two state telegraph model with a rate of mRNA production equal to ρ.

2.3 Sensitivity analysis

The exact solution in Eq. (5) allows us to examine the stochastic properties of the multi-scale model over large swathes of parameter space. We investigate the relative sensitivity of the coefficient of variation of mRNA fluctuations, $\mathrm{CV} = \sqrt{\mathrm{Var}(n)}/\langle n\rangle$, which is typically employed as a measure of the magnitude of transcriptional noise. To this end, we calculate the first two central moments, ($\langle n\rangle$ and $\mathrm{Var}(n)$), from Eq. (5) using $\langle n\rangle = \partial_u G|_{u=0}$ and $\mathrm{Var}(n) = \partial_u^2 G|_{u=0} + \langle n\rangle - \langle n\rangle^2$. The mean and CV are then given by

$$\langle n\rangle = \frac{\sigma_\mathrm{u}\lambda\rho}{d\gamma_1\gamma_2}, \tag{7a}$$

$$\mathrm{CV}^2 = \frac{1}{\langle n\rangle} + \frac{d}{\sigma_\mathrm{u}} \cdot \frac{\gamma_1}{\gamma_1 + d} \cdot \frac{\gamma_2}{\gamma_2 + d}. \tag{7b}$$

Note that since the parameters ρ and λ appear symmetrically in Eq. (7), for simplicity we enforce the constraint $\rho = \lambda$ (we will relax this constraint later). Hence, the relative sensitivity of the quantity $\overline{\mathrm{CV}} = \mathrm{CV}|_{\rho=\lambda}$, which can serve as a gauge of transcriptional noise, is insightful to study and defined as $\Lambda_p = (p/\overline{\mathrm{CV}})\partial\overline{\mathrm{CV}}/\partial p$ for a model parameter p, meaning that 1% change in p leads to a Λ_p% change in $\overline{\mathrm{CV}}$. The parameter values for the sensitivity analysis were sampled from experimental distributions recently inferred for 3575 genes of CAST allele in mouse fibroblasts [3], using the telegraph model. To obtain values for ρ and λ, we equate the mean of the telegraph model (with ON switching rate σ_b, OFF switching rate σ_u, transcription rate ρ_u and degradation rate d) $\langle n\rangle_\mathrm{tel} = \sigma_\mathrm{u}\rho_\mathrm{u}/\gamma_1 d$ with the mean of the multi-scale model (Eq. (7a)) under the constraint $\rho = \lambda$, giving

$$\rho = \rho_\mathrm{u}\left(1 + \sqrt{1 + \frac{\sigma_\mathrm{b}}{\rho_\mathrm{u}}}\right). \tag{8}$$

Distributions for each parameter in the dataset are presented in Fig. 2A and the box plots in Fig. 2B show the relative sensitivity for each parameter. The parameters in order of most sensitive first are σ_u, d, σ_b and $\rho = \lambda$. This order is the same as obtained by ranking parameters according to the inverse of their mean experimental values (the mean of the distributions in Fig. 2A) implying that changes to the CV are most easily accomplished by perturbations to the slowest reactions. Given the vectors Λ_{p_1} and Λ_{p_2} for any pair $p_1 \neq p_2$ and p_1, p_2 in the set $\{\rho, \lambda, \sigma_\mathrm{b}, \sigma_\mathrm{u}, d\}$ where each entry is a different gene, in Fig. 2C we calculate the Pearson correlation coefficient between the vectors and the corresponding joint distributions. This shows that $(\sigma_\mathrm{u}, \sigma_\mathrm{b})$ is the least dependent pairing and hence they constitute a quasi-orthogonal decomposition of

the sensitivity. In other words, a change in the CV due to a change in σ_u is practically uncorrelated with a change in the CV due to a change in σ_b, and hence these two parameters can be seen as independent "control knobs" to change the CV; this is of interest in synthetic biology, where an engineering design approach is taken to modify a biological system for improved functionality [25, 26]. The same set of parameters ranked by sensitivity are obtained, if instead of setting $\lambda = \rho$, we consider $\rho \gg \lambda$, or $\lambda \gg \rho$, and hence it appears that our results in this section are robust and invariant with respect to the ratio λ/ρ.

2.4 Effective telegraph model

Earlier we showed that in the limit of large ρ or large λ, the solution of the multi-scale model tends to the solution of the telegraph model. Next we use the first passage time method to reduce the multi-scale model into an effective telegraph model, without making the aforementioned assumptions. To this end, we consider the transcription motif of the multi-scale model, $D_{10} \xrightarrow{\lambda} D_{11} \xrightarrow{\rho} D_{10} + M$, whose corresponding master equations for producing newborn mRNA starting from state D_{10} are

$$\begin{aligned} \partial_t P_{10} &= -\lambda P_{10}, \\ \partial_t P_{11} &= \lambda P_{10} - \rho P_{11}, \\ \partial_t P_{\mathrm{M}} &= \rho P_{11}, \end{aligned} \tag{9}$$

where P_{10}, P_{11} and P_{M} represent the probability of staying in states D_{10}, D_{11} or producing a new mRNA respectively. We remark that the reaction $D_{11} \to D_0$ is absent from the motif due to its relatively small reaction rate σ_b compared to ρ and λ. The initial conditions for Eq. (9) are $P_{10}|_{t=0} = 1$, $P_{11}|_{t=0} = P_{\mathrm{M}}|_{t=0} = 0$. Solving for P_{M} in Eq. (9), we can calculate the mean first passage time for mRNA production

$$\langle t_f \rangle = \int_0^\infty t P_f \mathrm{d}t = \frac{\rho + \lambda}{\lambda \rho}, \tag{10}$$

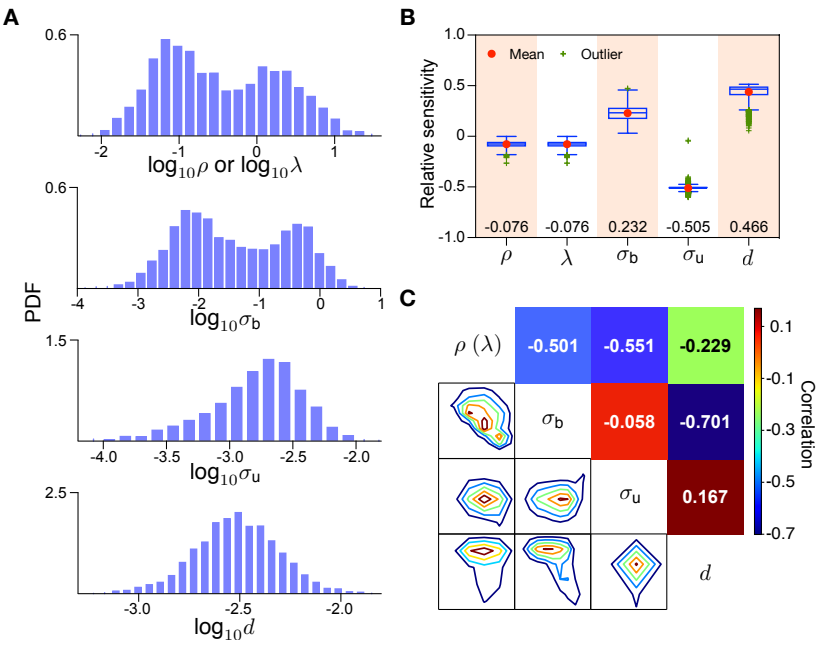

Figure 2: Relative sensitivity analysis of the coefficient variation $\overline{\mathrm{CV}}$ of mRNA noise over 5 kinetic parameters for 3575 genes of CAST allele data for mouse fibroblasts.(A) Distributions of the kinetic parameters in the dataset (obtained from [3]); values of ρ or λ are calculated using Eq. (8). (B) Box plots indicate the median (values shown at bottom), the 25%, 75% quantiles, and mean and outliers of relative sensitivity. (C) Joint distributions and Pearson correlation between the relative sensitivity vectors for each pair of parameters suggest that (σ_b, σ_u) and (σ_u, d) are the least dependent pairs.

where $P_f = \partial_t P_{\mathrm{M}}$ is the first-passage time distribution [27]. Since the effective transcription rate is the inverse of the mean first passage time, it immediately follows that the effective telegraph model is

$$D_0 \underset{\sigma_{\mathrm{b}}}{\overset{\sigma_{\mathrm{u}}}{\rightleftharpoons}} D_{10} \xrightarrow{\rho_{\mathrm{u}}=\frac{\lambda\rho}{\lambda+\rho}} D_{10} + M, \quad M \xrightarrow{d} \varnothing. \tag{11}$$

Alternatively, one can obtain this result by equating the means of our model Eq. (7a) and of the telegraph model $\langle n \rangle_{\mathrm{tel}} = \rho_{\mathrm{u}}\sigma_{\mathrm{u}}/\gamma_1 d$ and solving for the effective production rate ρ_{u}, giving $\rho_{\mathrm{u}} = \lambda\rho/\gamma_2 \simeq \frac{\lambda\rho}{\lambda+\rho}$ since typically $\rho, \lambda \gg \sigma_{\mathrm{b}}$.

In Fig. 3, we show the high accuracy of the effective telegraph model approximation from Eq. (11). In particular, Fig. 3A shows a heatmap of the distance between the distributions of mRNA numbers predicted by the effective telegraph model and the multi-scale model. As a distance measure, we use the Hellinger distance (HD), a Euclidean distance based metric normalized to the interval between 0 and 1. The effective telegraph model is naturally a more accurate description to the multi-scale model when there is one rate limiting step (large difference between ρ and λ) rather than when there are two rate limiting steps ($\rho = \lambda$).

Since the time-dependent distribution of the telegraph model is known in closed-form [6, 28], it follows that by the effective model in Eq. (11) we have an approximation for the time-dependent distribution of the multi-scale model too. The accuracy of this approximation is shown in Fig. 3B where it is compared to the time-dependent distributions computed using the SSA for the multi-scale model. The parameters here correspond to those of Point I in Fig. 3A (the largest HD). Differences between the distributions of the two models are negligible except near time $t = 0$. We further investigate how burst initiation and termination rates $(\sigma_{\mathrm{u}}, \sigma_{\mathrm{b}})$ affect the approximation error with a heatmap of HD as a function of σ_{u} and σ_{b} (Fig. 3C), and a stochastic bifurcation diagram for the number of modes of the effective telegraph and multi-scale model distributions (Fig. 3D) at steady state. The point of maximum HD in Fig. 3C (Point II) displays distributions that are not that different from each other – see upper right inset of Fig. 3D. The two

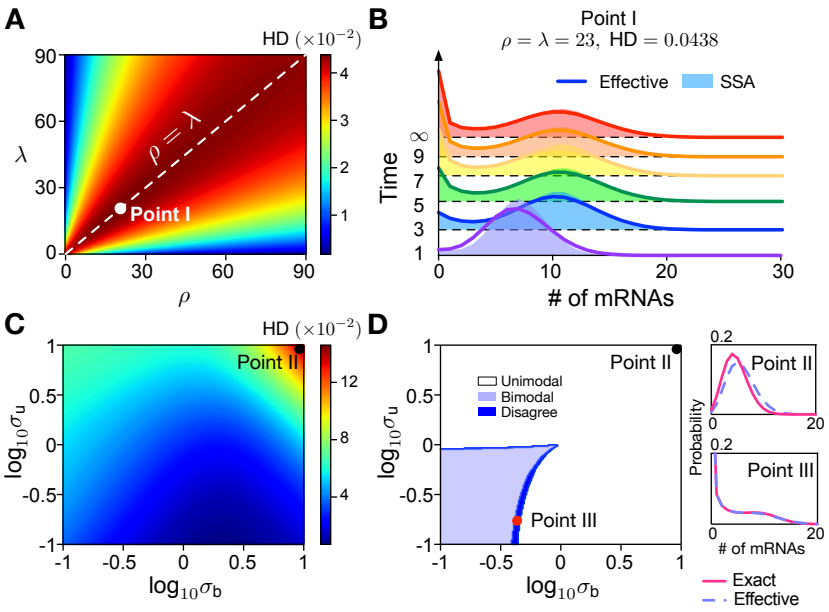

Figure 3: An effective telegraph model (given by reaction scheme (11)) approximates the distribution of mRNA numbers of the multi-scale model. (A) Hellinger Distance (HD) between steady-state distributions of mRNA numbers for the effective telegraph model and the multi-scale model as a function of ρ and λ with $\sigma_{\mathrm{u}} = 0.2$, $\sigma_{\mathrm{b}} = 0.1$ and $d = 1$. The discrepancy between the two distributions grows as ρ and λ approach the line $\rho = \lambda$. (B) Shows the time-dependent distributions for Point I in (A) (the point with the largest HD) predicted by the effective model compared to those computed by the SSA for the multi-scale model. (C) Heatmap of HD between both distributions as a function of σ_{b} and σ_{u} with $\rho = \lambda = 23$ and $d = 1$. (D) Stochastic bifurcation diagram for the number of modes of the steady-state distributions predicted by the two models. The small dark blue region is where modality of both models disagree. Insets show distributions corresponding to the points marked in (C,D).

models display the same number of modes in all regions of parameter space except for a narrow region where modality detection is challenging because the distributions have a broad plateau – see lower right inset of Fig. 3D (Point III). This again confirms the high accuracy of the effective telegraph model approximation. The biological implications of the Michaelis-Menten dependence of the transcription rate ρ_{u} in Eq. (11) on λ and ρ is discussed in the Conclusion; in particular there we argue how this special feature of our model can explain gene dosage compensation observed in experiments.

2.5 Connection to the refractory model

Besides the telegraph model, another prevalent stochastic transcriptional model is the refractory model [2] (a three-state model, see Fig. 4A left), wherein the burst initiation requires two steps. This model was devised to explain the experimental observation that the distribution of "off" intervals is not exponential but rather has a peak at a non-zero value. To understand the connection between our model and the refractory model, we first exactly solve the refractory model for the steady-state distribution of mRNA numbers.

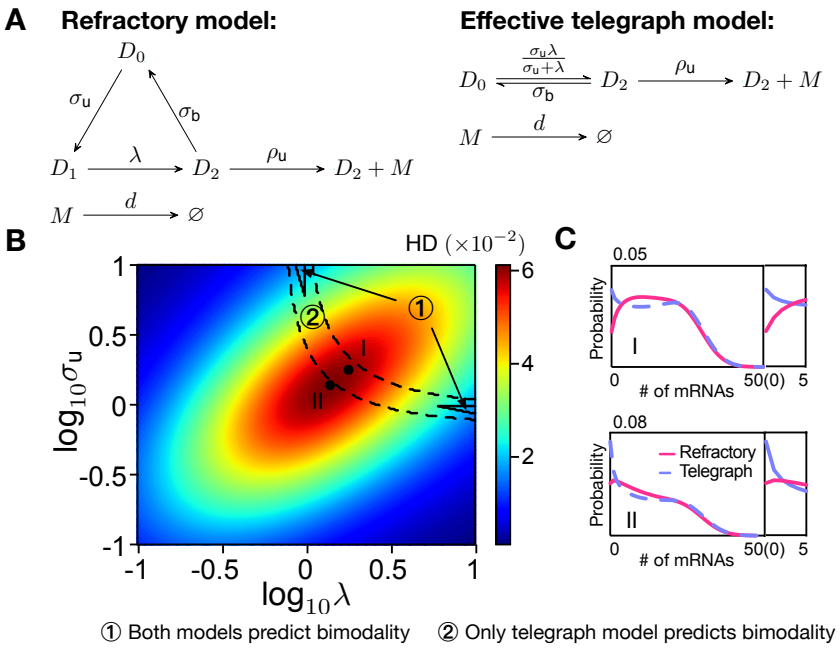

Figure 4: Effective telegraph model approximation for the refractory model. (A) Schematics of both models. (B) Hellinger distance between the steady-state distributions of mRNA numbers predicted by both models, and a bifurcation diagram of their number of modes(black lines) as a function of σ_{u} and λ with $\sigma_{\mathrm{b}} = 0.8$, $\rho_{\mathrm{u}} = 30$ and $d = 1$. (C) Distributions for Points I and II in (B), showing significant disagreement in the height of the zero mode (insets show a zoom at the mode at zero).

Given the reaction scheme illustrated in Fig. 4A, it follows that the temporal evolution of probability $P_\theta(n)$ of finding n mRNAs and gene state D_θ ($\theta = 0$, 1 or 2) can be described by the following master equations

$$\begin{cases} \partial_t P_0(n) = (\mathbb{E}^1 - 1)dnP_0(n) + \sigma_{\mathrm{b}}P_2(n) - \sigma_{\mathrm{u}}P_0(n), \\ \partial_t P_1(n) = (\mathbb{E}^1 - 1)dnP_1(n) + \sigma_{\mathrm{u}}P_0(n) - \lambda P_1(n), \\ \partial_t P_2(n) = (\mathbb{E}^1 - 1)dnP_2(n) + (\mathbb{E}^{-1} - 1)\rho_{\mathrm{u}}P_2(n) + \lambda P_1(n) - \sigma_{\mathrm{b}}P_2(n). \end{cases}$$

The corresponding generating function equations are given by

$$\partial_t G_0 + d(z-1)\partial_z G_0 = \sigma_{\mathrm{b}}G_2 - \sigma_{\mathrm{u}}G_0, \tag{12a}$$

$$\partial_t G_1 + d(z-1)\partial_z G_1 = \sigma_{\mathrm{u}}G_0 - \lambda G_1, \tag{12b}$$

$$\partial_t G_2 + d(z-1)\partial_z G_2 = \rho_{\mathrm{u}}(z-1)G_2 + \lambda G_1 - \sigma_{\mathrm{b}}G_2, \tag{12c}$$

where $G_\theta = \sum_n z^n P_\theta(n)$. We intend to solve Eqs. (12) at steady state and thus set $\partial_t G_\theta = 0$. Then, we

solve G_1 as a function of G_2 from Eq. (12c), subsequently substitute it into Eq. (12b) and solve G_0 as a function of G_2. Following that, Eq. (12a) becomes an ordinary differential equation with G_2 being the only variable to be solved

$$u^2\partial_u^3 G_2 + (3+\tilde{\lambda}+\tilde{\sigma}_\mathrm{b}+\tilde{\sigma}_\mathrm{u}-\tilde{\rho}_\mathrm{u}u)u\partial_u^2 G_2 + [1+\tilde{\sigma}_\mathrm{b}+\tilde{\sigma}_\mathrm{u}+\tilde{\sigma}_\mathrm{b}\tilde{\sigma}_\mathrm{u}-\tilde{\rho}_\mathrm{u}(3+\tilde{\sigma}_\mathrm{u})u+\tilde{\lambda}(1+\tilde{\sigma}_\mathrm{b}+\tilde{\sigma}_\mathrm{u}-\tilde{\rho}_\mathrm{u}u)]\partial_u G_2 \\ -(1+\tilde{\lambda})(1+\tilde{\sigma}_\mathrm{u})\tilde{\rho}_\mathrm{u}G_2 = 0, \tag{13}$$

where $\tilde{\rho}_\mathrm{u}$, $\tilde{\lambda}$, $\tilde{\sigma}_\mathrm{b}$ and $\tilde{\sigma}_\mathrm{u}$ are the kinetic parameters normalized with respect to d and $u = z - 1$. Eq. (13) is the canonical form of the differential equation for the generalized hypergeometric function ${}_2F_2$, admitting the solution

$$G_2(u) = C \cdot {}_2F_2(\tilde{\lambda}+1, \tilde{\sigma}_\mathrm{u}+1; \beta_1-\beta_2+1, \beta_1+\beta_2+1; \tilde{\rho}_\mathrm{u}u), \tag{14}$$

where C is an integration constant, and β_1 and β_2 denote

$$\beta_1 = \frac{\tilde{\sigma}_\mathrm{u}+\tilde{\sigma}_\mathrm{b}+\tilde{\lambda}}{2}, \qquad \beta_2 = \frac{1}{2}\sqrt{\tilde{\lambda}^2 - 2\tilde{\lambda}\tilde{\sigma}_\mathrm{b} + \tilde{\sigma}_\mathrm{b}^2 - 2\tilde{\lambda}\tilde{\sigma}_\mathrm{u} - 2\tilde{\sigma}_\mathrm{b}\tilde{\sigma}_\mathrm{u} + \tilde{\sigma}_\mathrm{u}^2}.$$

Summing Eqs. (12) leads to $\partial_u G = \partial_u(\sum_\theta G_\theta) = \tilde{\rho}_\mathrm{u}G_2$, one can obtain G from Eq. (14) in the form of the generalized hypergeometric function:

$$G(u) = C_2 \cdot {}_2F_2\left(\tilde{\lambda}, \tilde{\sigma}_\mathrm{u}; \beta_1-\beta_2, \beta_1+\beta_2; \tilde{\rho}_\mathrm{u}u\right), \tag{15}$$

and C_2 is found to be 1 by the normalization condition $G(0) = 1$. Eq. (15) together with $P(n) = \frac{1}{n!}\frac{\mathrm{d}^n G}{\mathrm{d}u^n}|_{u=-1}$ defines the distribution of mRNA numbers for the refractory model in steady-state conditions. A similar solution is also known for a generalization of the refractory model [29].

The next step is to map the refractory model onto an effective telegraph model by matching the mean mRNA numbers

$$\langle n\rangle_\mathrm{ref} = \frac{\lambda\sigma_\mathrm{u}\rho_\mathrm{u}}{d(\lambda\sigma_\mathrm{u}+\lambda\sigma_\mathrm{b}+\sigma_\mathrm{u}\sigma_\mathrm{b})}, \quad \langle n\rangle_\mathrm{tel} = \frac{\rho_\mathrm{u}\bar{\sigma}_\mathrm{u}}{d(\sigma_\mathrm{b}+\bar{\sigma}_\mathrm{u})}.$$

leading to an effective burst initiation rate $\bar{\sigma}_\mathrm{u} = \frac{\sigma_\mathrm{u}\lambda}{\sigma_\mathrm{u}+\lambda}$ and the corresponding effective model shown in Fig. 4A right. Note that while the multi-scale model is approximately equivalent to an effective telegraph model with a renormalized mRNA production rate, the refractory model's telegraph approximation leads to a renormalized rate of switching to the active state.

We then compare the steady-state distributions of the refractory model and its effective telegraph model. A heatmap of HD quantifying their distributional difference and a modality diagram (marked as black lines) of the two distributions are illustrated in Fig. 4B. Both the regions of high HD and Region 2 where only the telegraph model predicts bimodality are significantly large; also Region 1 where both predict bimodality is small. This shows that the refractory model, in general, is not well approximated by the telegraph model, particularly the latter's probability for low mRNA numbers is not accurate – see Fig. 4C. Given the telegraph model's excellent approximation to the multi-scale model, it is clear that the multi-scale model and refractory model can be distinguished.

2.6 Protein dynamics

Finally, for completeness, we extend the multi-scale model to provide analytic steady-state distributions of protein numbers. This allows interpretations of single-cell data of protein expression (see for example [30]). We consider the network in Fig. 1A with two additional reactions: (i) a first-order reaction modelling the translation of mRNA to proteins with rate constant k. (ii) a first-order reaction modelling the decay of protein with rate constant d_p. It is shown in Appendix C that under the classic short-lived mRNA assumption ($d \gg d_\mathrm{p}$) [20], the generating function corresponding to the steady-state distribution of protein numbers is given by

$$G(v) = {}_3F_2(a_1, a_2, a_3; b_1, b_2; bv), \tag{16}$$

with $b_1 = (\sigma_\mathrm{b} + \sigma_\mathrm{u})/d_\mathrm{p}$, $b_2 = (\sigma_\mathrm{b} + \lambda + \rho)/d_\mathrm{p}$, the mean translational burst size $b = k/d$, and the parameters a_1, a_2 and a_3 being solutions of the equations

$$\begin{aligned} &a_1 a_2 a_3 = \sigma_\mathrm{u} \lambda \rho / d_\mathrm{p}^3, \\ &a_1 + a_2 + a_3 = b_1 + b_2, \\ &a_1 a_2 + a_1 a_3 + a_2 a_3 = b_1 b_2 + \lambda \rho / d_\mathrm{p}^2. \end{aligned}$$

In the limit of large λ or ρ, we show in Appendix C that Eq. (16) reduces to the Gaussian hypergeometric function (${}_2F_1$), which was reported in Ref. [20] for the classical three-stage model of gene expression in the limit of fast mRNA decay.

3 Conclusion

Here we performed the first detailed analytical study of a multi-scale model of bursty gene expression based on recent experimental data from mammalian cells [15]. The conventional telegraph model does not include an independently regulated pause release step and hence cannot differentiate the effects of changing polymerase pause release versus polymerase recruitment rates. Whereas, the multi-scale model studied here can distinguish these effects. While our model has three effective gene states (one of which regulating pause release), it is not a special case of existing multistate models because in our model, the gene state changes upon production of new nascent mRNAs to model the experimental observation that, unless the polymerase is unpaused (and nascent mRNA starts being actively transcribed by this polymerase), there can be no binding of new Pol II. In contrast, current models assume the gene state does not change upon production of mRNA because they model the production of a mature transcript without detailed modelling of the steps between transcriptional initiation and termination.

We have derived simple closed-form expressions for the approximate time evolution of the mRNA numbers and used the theory to understand which reactions contribute mostly to fluctuations. We also showed that (i) this model can be distinguished from the refractory model, another three gene state model popular in the literature. (ii) a number of previous models in the literature are special cases of our model, valid only in certain parameter regimes. Specifically the mRNA and protein distributions of the conventional three-stage model of gene expression provide a good approximation to the multi-scale bursting model in certain regions of parameter space as shown in Appendices B and C.

The simplicity of the equations for the mean and the variance allow the inference of rate parameters from single-cell data using maximum likelihood methods [31]. Potential extensions include (i) the impact of cell cycle effects such as binomial partitioning and variability in the cell cycle duration; (ii) introducing a detailed description of polymerase movement along the gene during elongation. The use of the recently developed linear mapping approximation [32] appears to be a promising means to extend the analytical solution of the present model to include feedback loops via DNA-protein interactions [33, 34].

An important result of the paper is that the time dependent mRNA distribution of the multi-scale model with polymerase dynamics and three states can be accurately approximated by the two state telegraph model, modified with a Michaelis-Menten like dependence of the effective transcription rate on polymerase abundance. Specifically, by Eq. (11) the transcription rate of a gene locus is $\rho_\mathrm{u} = \lambda\rho/(\lambda + \rho)$ where λ is the binding rate of Pol II (see Fig. 1A) which is proportional to the local number of Pol II molecules at the gene locus with active transcription [35]. This equation implies that the transcription rate is proportional to the local number of Pol II molecules if λ is approximately less than ρ, i.e. if the Pol II binding rate is less than or equal to the rate at which Pol II is unpaused. In contrast, if unpausing is the rate limiting step ($\rho \ll \lambda$) then the transcription rate is practically independent of the local Pol II number.

Now when the number of gene copies doubles during replication, the local number of Pol II molecules will correspondingly decrease due to increased sharing of Pol II. Hence if we are in the regime $\lambda \lesssim \rho$, the transcription rate per gene copy decreases; thus the total transcription rate for a gene per cell post-replication will be consequently slower than twice the total transcription rate pre-replication. This implies that the mean number of RNA per cell is not significantly affected by replication; indeed this "dosage compensation" has been observed experimentally for some genes in mouse embryonic stem cells [36] though a different explanation than above was suggested. In one study [37] it was estimated that for 6 yeast genes (RPB2, RPB3, TAF5, TAF6, TAF12, KAP104), the formation of the pre-initiation complex at the promoter (λ) is approximately equal to the rate at which the RNA polymerase escapes the promoter (ρ); hence gene dosage compensation via polymerase sharing, as implied by our model, maybe common. In contrast if we are in the regime $\rho \ll \lambda$, the transcription rate per gene copy before and after replication is the same, and

hence the total transcription rate for a gene per cell post-replication will be twice the total transcription rate pre-replication. This is also what is predicted by the telegraph model with constant burst initiation and termination rates, and observed experimentally for a reporter gene expressed from a strong synthetic promoter [36]. Note that since the mean burst size is the mean number of RNAs transcribed when the gene is on, by our reasoning above it also follows that when $\lambda \lesssim \rho$, the mean burst size is altered upon gene replication. The idea that the number of RNA polymerases is the limiting factor in transcription has been recently hypothesized [38], and has implications for the mitigation of burden imposed by gene circuits in synthetic biology [39]. Our model here goes one step further by deriving the explicit relationship between the transcription rate and the number of RNA polymerases. Generally our model supports the observation that there are differences in transcriptional activity between different stages of the cell cycle [40] that cannot be explained by the conventional telegraph model.

Author Contributions

Z.C. formulated the research question, performed the calculations, produced the figures and wrote an initial draft of the manuscript. T.F. performed some of the calculations for protein distributions. D.O. supervised the research and edited the manuscript. R.G. formulated the research question, supervised the research and wrote the manuscript with assistance from the co-authors.

Acknowledgements

Z.C. gratefully acknowledges support of the UK Research Councils' Synthetic Biology for Growth programme and of the BBSRC, EPSRC and MRC (B/M018040/1) and careful proofreading by J. Holehouse. R.G. acknowledges support from BBSRC grant no. BB/M025551/1. D.O. acknowledges support from the Human Frontier Science Program (grant no. RGY076/2015).

Appendix

A Analytic distribution for mRNA numbers when ρ, λ and σ_b are large

Given the large values of ρ, λ and σ_b, we implement the following parametrization

$$\sigma_\mathrm{b} \mapsto \sigma_\mathrm{b}\delta, \qquad \rho \mapsto \rho\delta, \qquad \lambda \mapsto \lambda\delta,$$

where δ is a large real number.

By means of the method of characteristics, solving Eq. (2) is tantamount to seeking a solution to the ODE system

$$\begin{aligned}
&\partial_s t = 1 \qquad\qquad \Rightarrow \quad t = s \\
&\partial_s z = d(z-1) \quad \Rightarrow \quad z - 1 = r\mathrm{e}^{ds} \\
&\partial_s G = \rho\delta(z-1)G_{11}, && (17\mathrm{a}) \\
&\partial_s G_{10} = \rho\delta z G_{11} - \sigma_\mathrm{b}\delta G_{10} - \lambda\delta G_{10} + \sigma_\mathrm{u}(G - G_{10} - G_{11}), && (17\mathrm{b}) \\
&\partial_s G_{11} = -\rho\delta G_{11} - \sigma_\mathrm{b}\delta G_{11} + \lambda\delta G_{10}. && (17\mathrm{c})
\end{aligned}$$

Dividing δ on both sides of Eqs. (17a)–(17c), one obtains a singular system consisting of

$$\begin{cases}
\epsilon\partial_s G = \rho(z-1)G_{11}, \\
\epsilon\partial_s G_{10} = \rho z G_{11} - \sigma_\mathrm{b} G_{10} - \lambda G_{10} + \epsilon\sigma_\mathrm{u}(G - G_{10} - G_{11}), \\
\epsilon\partial_s G_{11} = -\rho G_{11} - \sigma_\mathrm{b} G_{11} + \lambda G_{10},
\end{cases} \tag{18}$$

with $\epsilon = 1/\delta \simeq 0$. Expanding G, G_{10} and G_{11} in Eq. (18) as a series in powers of ϵ,

$$G = G^{(0)} + \epsilon G^{(1)} + \mathcal{O}(\epsilon^2), \qquad G_{10} = G_{10}^{(0)} + \epsilon G_{10}^{(1)} + \mathcal{O}(\epsilon^2), \qquad G_{11} = G_{11}^{(0)} + \epsilon G_{11}^{(1)} + \mathcal{O}(\epsilon^2),$$

and matching the orders of ϵ, we have

$$\text{Order of } \epsilon^0 : \begin{cases} \rho(z-1)G_{11}^{(0)} = 0 & \Rightarrow \quad G_{11}^{(0)} = 0 \\ \rho z G_{11}^{(0)} - \sigma_{\mathrm{b}} G_{10}^{(0)} - \lambda G_{10}^{(0)} = 0 & \Rightarrow \quad G_{10}^{(0)} = 0 \end{cases}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases} \partial_s G^{(0)} = \rho(z-1)G_{11}^{(1)} & \\ \partial_s G_{10}^{(0)} = \rho z G_{11}^{(1)} - \sigma_{\mathrm{b}} G_{10}^{(1)} - \lambda G_{10}^{(1)} + \sigma_{\mathrm{u}}(G^{(0)} - G_{10}^{(0)} - G_{11}^{(0)}) & \Rightarrow \quad \rho z G_{11}^{(1)} - \sigma_{\mathrm{b}} G_{10}^{(1)} - \lambda G_{10}^{(1)} + \sigma_{\mathrm{u}} G^{(0)} = 0 \\ \partial_s G_{11}^{(0)} = -\rho G_{11}^{(1)} - \sigma_{\mathrm{b}} G_{11}^{(1)} + \lambda G_{10}^{(1)} & \Rightarrow \quad -\rho G_{11}^{(1)} - \sigma_{\mathrm{b}} G_{11}^{(1)} + \lambda G_{10}^{(1)} = 0. \end{cases}$$

Then, we have

$$\partial_s G^{(0)} = -\frac{\rho u \sigma_{\mathrm{u}}}{\rho u - \alpha} G^{(0)},$$

where $\alpha = \sigma_{\mathrm{b}}\gamma_2/\lambda$ and $u = z - 1 = r\mathrm{e}^{ds}$. Its solution immediately follows as

$$G^{(0)} = C(r)(\rho r \mathrm{e}^{ds} - \alpha)^{-\frac{\sigma_{\mathrm{u}}}{d}}, \tag{19}$$

with $C(r)$ being a function of r to be determined from the initial condition. Suppose that the initial condition for this process is $g(u) = G^{(0)}|_{t=0}$, which is known *a priori*. For instance, say the initial distribution of n mRNA molecules is $P(n) = p_n$, then $g(u) = \sum_n p_n (u+1)^n$. Letting s be equal to 0 (or equivalently $t = 0$), it follows $u = r$ and $g(u) = g(r)$, and we can establish the following relation

$$g(r) = C(r)(\rho r - \alpha)^{-\frac{\sigma_{\mathrm{u}}}{d}},$$

from which we can solve $C(r)$ as

$$C(r) = g(r)(\rho r - \alpha)^{\frac{\sigma_{\mathrm{u}}}{d}}.$$

Substituting the latter back into Eq. (19) and replacing $r = u\mathrm{e}^{-dt}$, we can calculate the leading-order solution of G from (19) as

$$G(u) = g(u\mathrm{e}^{-dt}) \left(\frac{\rho u \mathrm{e}^{-dt} - \alpha}{\rho u - \alpha} \right)^{\frac{\sigma_{\mathrm{u}}}{d}}. \tag{20}$$

At steady state, the leading-order solution in (20) becomes

$$G(z) = \left(\frac{\alpha}{\alpha - \rho(z-1)} \right)^{\frac{\sigma_{\mathrm{u}}}{d}},$$

and the corresponding distribution of mRNA numbers is a negative binomial distribution $\mathrm{NB}(\frac{\sigma_{\mathrm{u}}}{d}, \frac{\rho}{\rho+\alpha})$.

B Convergence to telegraph model for large ρ

To this end, we parametrize ρ as $\rho \mapsto \rho\delta$, where δ is a large real number. As such, Eq. (2) can be recast as

$$\partial_t G_0 + d(z-1)\partial_z G_0 = -\sigma_{\mathrm{u}} G_0 + \sigma_{\mathrm{b}} G_{10} + \sigma_{\mathrm{b}} G_{11}, \tag{21a}$$

$$\partial_t G_{10} + d(z-1)\partial_z G_{10} + (\sigma_{\mathrm{b}} + \lambda) G_{10} - \sigma_{\mathrm{u}} G_0 = \rho\delta z G_{11}, \tag{21b}$$

$$\partial_t G_{11} + d(z-1)\partial_z G_{11} + \sigma_{\mathrm{b}} G_{11} - \lambda G_{10} = -\rho\delta G_{11}. \tag{21c}$$

Dividing both sides of Eqs. (21b)-(21c) by δ and setting $\epsilon = \delta^{-1}$, we have that

$$\epsilon \Big(\partial_t G_{10} + d(z-1)\partial_z G_{10} + (\sigma_{\mathrm{b}} + \lambda) G_{10} - \sigma_{\mathrm{u}} G_0 \Big) = \rho z G_{11}, \tag{22a}$$

$$\epsilon \Big(\partial_t G_{11} + d(z-1)\partial_z G_{11} + \sigma_{\mathrm{b}} G_{11} - \lambda G_{10} \Big) = -\rho G_{11}. \tag{22b}$$

Again using the same method as before, we expand G_0, G_{10} and G_{11} in Eqs. (21a) and (22) as a series in powers of ϵ, collect the terms for ϵ^0 and ϵ^1 and obtain

$$\text{Order of } \epsilon^0 : \begin{cases} \partial_t G_0^{(0)} + d(z-1)\partial_z G_0^{(0)} = -\sigma_\mathrm{u} G_0^{(0)} + \sigma_\mathrm{b} G_{10}^{(0)} + \sigma_\mathrm{b} G_{11}^{(0)}, \\ \rho z G_{11}^{(0)} = 0, \\ \rho G_{11}^{(0)} = 0, \end{cases} \tag{23}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases} \partial_t G_{10}^{(0)} + d(z-1)\partial_z G_{10}^{(0)} + (\sigma_\mathrm{b} + \lambda) G_{10}^{(0)} - \sigma_\mathrm{u} G_0^{(0)} = \rho z G_{11}^{(1)}, \\ \partial_t G_{11}^{(0)} + d(z-1)\partial_z G_{11}^{(0)} + \sigma_\mathrm{b} G_{11}^{(0)} - \lambda G_{10}^{(0)} = -\rho G_{11}^{(1)}. \end{cases} \tag{24}$$

From Eq. (23), we can solve that $G_{11}^{(0)} = 0$, with which we can further get $\lambda G_{10}^{(0)} = \rho G_{11}^{(1)}$ from Eq. (24). Given both results, Eqs. (23) and (24) can be simplified to

$$\begin{cases} \partial_t G_0^{(0)} + d(z-1)\partial_z G_0^{(0)} = -\sigma_\mathrm{u} G_0^{(0)} + \sigma_\mathrm{b} G_{10}^{(0)}, \\ \partial_t G_{10}^{(0)} + d(z-1)\partial_z G_{10}^{(0)} = \lambda(z-1) G_{10}^{(0)} - \sigma_\mathrm{b} G_{10}^{(0)} + \sigma_\mathrm{u} G_0^{(0)}, \end{cases}$$

which are exactly the generating function equations of the telegraph model (See Eqs. (A2) and (A3) in [28]), thus showing that the multi-scale transcriptional bursting model converges to the telegraph model when $\rho \to \infty$. A similar proof can be constructed to show that the telegraph model is also obtained in the limit $\lambda \to \infty$.

C Analytic marginal distribution for protein numbers for the multi-scale model in the limit of fast mRNA decay

To the reaction scheme illustrated in Fig. 1A, we add two reactions: (i) a first-order reaction modelling the translation of mRNA to proteins with rate constant k. (ii) a first-order reaction modelling the decay of protein with rate constant d_p. The following coupled master equations describe the time evolution of the probability $P_\theta(n,m)$ of finding n mRNAs, m proteins and gene state D_θ ($\theta = 0, 10, 11$) in a cell:

$$\begin{cases} \partial_t P_0(n,m) = & d(n+1)P_0(n+1,m) - dnP_0(n,m) + d_\mathrm{p}(m+1)P_0(n,m+1) - d_\mathrm{p} m P_0(n,m) \\ & + knP_0(n,m-1) - knP_0(n,m) - \sigma_\mathrm{u} P_0(n,m) + \sigma_\mathrm{b} P_{10}(n,m) + \sigma_\mathrm{b} P_{10}(n,m), \\ \partial_t P_{10}(n,m) = & d(n+1)P_{10}(n+1,m) - dnP_{10}(n,m) + d_\mathrm{p}(m+1)P_{10}(n,m+1) - d_\mathrm{p} m P_{10}(n,m) \\ & + knP_{10}(n,m-1) - knP_{10}(n,m) + \sigma_\mathrm{u} P_0(n,m) - (\sigma_\mathrm{b} + \lambda)P_{10}(n,m) + \rho P_{11}(n-1,m), \\ \partial_t P_{11}(n,m) = & d(n+1)P_{11}(n+1,m) - dnP_{11}(n,m) + d_\mathrm{p}(m+1)P_{11}(n,m+1) - d_\mathrm{p} m P_{11}(n,m) \\ & + knP_{11}(n,m-1) - knP_{11}(n,m) + \lambda P_{10}(n,m) - (\rho + \sigma_\mathrm{b})P_{11}(n,m). \end{cases} \tag{25}$$

By defining $G_\theta = \sum_n \sum_m {z_\mathrm{m}}^n {z_\mathrm{p}}^m P_\theta(n,m)$, solving Eq. (25) is tantamount to seeking solutions to the set of differential equations

$$\begin{cases} \partial_t G_0 + [d(z_\mathrm{m}-1) - k(z_\mathrm{p}-1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_0 + d_\mathrm{p}(z_\mathrm{p}-1)\partial_{z_\mathrm{p}} G_0 = -\sigma_\mathrm{u} G_0 + \sigma_\mathrm{b} G_{10} + \sigma_\mathrm{b} G_{11}, \\ \partial_t G_{10} + [d(z_\mathrm{m}-1) - k(z_\mathrm{p}-1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_{10} + d_\mathrm{p}(z_\mathrm{p}-1)\partial_{z_\mathrm{p}} G_{10} = \sigma_\mathrm{u} G_0 - (\sigma_\mathrm{b} + \lambda) G_{10} + \rho z_\mathrm{m} G_{11}, \\ \partial_t G_{11} + [d(z_\mathrm{m}-1) - k(z_\mathrm{p}-1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_{11} + d_\mathrm{p}(z_\mathrm{p}-1)\partial_{z_\mathrm{p}} G_{11} = \lambda G_{10} - (\rho + \sigma_\mathrm{b}) G_{11}. \end{cases} \tag{26}$$

By means of the method of characteristics, Eq. (26) is equivalently represented as

$$\partial_s t = 1, \qquad \partial_s z_\mathrm{m} = d(z_\mathrm{m}-1) - k(z_\mathrm{p}-1)z_\mathrm{m}, \qquad \partial_s z_\mathrm{p} = d_\mathrm{p}(z_\mathrm{p}-1),$$

and

$$\begin{cases} \partial_s G_0 = -\sigma_\mathrm{u} G_0 + \sigma_\mathrm{b} G_{10} + \sigma_\mathrm{b} G_{11}, \\ \partial_s G_{10} = \sigma_\mathrm{u} G_0 - (\sigma_\mathrm{b} + \lambda) G_{10} + \rho z_\mathrm{m} G_{11}, \\ \partial_s G_{11} = \lambda G_{10} - (\rho + \sigma_\mathrm{b}) G_{11}. \end{cases}$$

Assuming that mRNA decays much faster than protein such that $\partial_s z_{\rm m} \simeq 0$ [20], we get that

$$z_{\rm m} = \frac{1}{1-bv}, \qquad \text{and} \qquad v = z_{\rm p} - 1, \tag{27}$$

and $b = k/d$ is the mean translational burst size. Using Eq. (27) we can reduce Eq. (26) to

$$v\partial_v G_0 = -\tilde{\sigma}_{\rm u} G_0 + \tilde{\sigma}_{\rm b} G_{10} + \tilde{\sigma}_{\rm b} G_{11}, \tag{28a}$$

$$v\partial_v G_{10} = \tilde{\sigma}_{\rm u} G_0 - (\tilde{\sigma}_{\rm b} + \tilde{\lambda}) G_{10} + \frac{\tilde{\rho}}{1-bv} G_{11}, \tag{28b}$$

$$v\partial_v G_{11} = \tilde{\lambda} G_{10} - (\tilde{\rho} + \tilde{\sigma}_{\rm b}) G_{11}, \tag{28c}$$

where $\tilde{\sigma}_{\rm b}$, $\tilde{\sigma}_{\rm u}$, $\tilde{\rho}$ and $\tilde{\lambda}$ are kinetic parameters normalized with respect to protein degradation rate $d_{\rm p}$. It follows from summing Eqs. (28) that

$$G_{11} = \frac{(1-bv)\partial_v G}{\tilde{\rho} b}. \tag{29}$$

Using the definitions $b_1 = \tilde{\sigma}_{\rm b} + \tilde{\sigma}_{\rm u}$ and $b_2 = \tilde{\sigma}_{\rm b} + \tilde{\lambda} + \tilde{\rho}$ and plugging Eq. (29) into Eqs. (28b) and (28c), it gives us that

$$(1-bv)v^2\partial_v^3 G + [1 + b_1 + b_2 - bv(3 + b_1 + b_2)]v\partial_v^2 G + \{b_1 b_2 - bv[(1+b_1)(1+b_2) + \tilde{\lambda}\tilde{\rho}]\}\partial_v G - b\tilde{\sigma}_{\rm u}\tilde{\lambda}\tilde{\rho} G = 0,$$

which admits a solution

$$G(v) = {}_3F_2(a_1, a_2, a_3; b_1, b_2; bv), \tag{30}$$

with a_1, a_2 and a_3 being roots of

$$\begin{cases} a_1 a_2 a_3 = \tilde{\sigma}_{\rm u}\tilde{\lambda}\tilde{\rho}, \\ a_1 + a_2 + a_3 = b_1 + b_2, \\ a_1 a_2 + a_1 a_3 + a_2 a_3 = b_1 b_2 + \tilde{\lambda}\tilde{\rho}. \end{cases}$$

Hence summarizing, Eq. (30) and $P(m) = \frac{{\rm d}^m G(v)}{m! {\rm d}v^m}|_{v=-1}$ define the steady-state distribution of protein numbers, which is

$$P(m) = \frac{b^m}{m!}\frac{(a_1)_n (a_2)_n (a_3)_n}{(b_1)_n (b_2)_n} {}_3F_2(a_1+n, a_2+n, a_3+n; b_1+n, b_2+n; -b),$$

given that mRNA is short-lived.

Next we will show the solution Eq. (30) converges to the Gaussian hypergeometric function (${}_2F_1$) for the three-stage gene expression model [20] when ρ is large. To this end, we parameterize $\tilde{\rho}$ in Eqs. (28b)-(28c) as $\tilde{\rho} \mapsto \tilde{\rho}\delta$ where δ is a large number. Dividing both sides of Eqs. (28b)-(28c) by δ, we have

$$\epsilon\big(v\partial_v G_{10} - \tilde{\sigma}_{\rm u} G_0 + (\tilde{\sigma}_{\rm b} + \tilde{\lambda}) G_{10}\big) = \frac{\tilde{\rho}}{1-bv} G_{11}, \tag{31a}$$

$$\epsilon\big(v\partial_v G_{11} - \tilde{\lambda} G_{10} + \tilde{\sigma}_{\rm b} G_{11}\big) = -\tilde{\rho} G_{11}, \tag{31b}$$

where $\epsilon = 1/\delta \simeq 0$. Again similarly, we expand G_0, G_{10} and G_{11} in Eqs. (28a) and (31) as a series in powers of ϵ, collect the terms for ϵ^0 and ϵ^1 and obtain

$$\text{Order of } \epsilon^0 : \begin{cases} v\partial_v G_0^{(0)} = -\tilde{\sigma}_{\rm u} G_0^{(0)} + \tilde{\sigma}_{\rm b} G_{10}^{(0)} + \tilde{\sigma}_{\rm b} G_{11}^{(0)}, \\ \dfrac{\tilde{\rho}}{1-bv} G_{11}^{(0)} = 0, \\ \tilde{\rho} G_{11}^{(0)} = 0, \end{cases} \tag{32}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases} v\partial_v G_{10}^{(0)} - \tilde{\sigma}_{\rm u} G_0^{(0)} + (\tilde{\sigma}_{\rm b} + \tilde{\lambda}) G_{10}^{(0)} = \dfrac{\tilde{\rho}}{1-bv} G_{11}^{(1)}, \\ v\partial_v G_{11}^{(0)} - \tilde{\lambda} G_{10}^{(0)} + \tilde{\sigma}_{\rm b} G_{11}^{(0)} = -\tilde{\rho} G_{11}^{(1)}, \end{cases} \tag{33}$$

From Eqs. (32), we get $G_{11}^{(0)} = 0$, which is used to reduce Eqs. (33) and the first equation in Eq. (32) to

$$\begin{cases} v\partial_v G_0^{(0)} = -\tilde{\sigma}_{\mathrm{u}} G_0^{(0)} + \tilde{\sigma}_{\mathrm{b}} G_{10}^{(0)}, \\ v\partial_v G_{10}^{(0)} = \tilde{\sigma}_{\mathrm{u}} G_0^{(0)} - \tilde{\sigma}_{\mathrm{b}} G_{10}^{(0)} + \dfrac{\tilde{\lambda} bv}{1-bv} G_{10}^{(0)}. \end{cases} \tag{34}$$

Note that Eq. (34), which is the leading order of Eqs. (28), is exactly the same as the generating functions of the three-stage gene expression model reported in [20] (See Eqs. (68)-(69) in SI thereof). By means of similar arguments, one can show the reduction of our model when λ is large.

References

[1] Halpern, K. B. *et al.* Bursty gene expression in the intact mammalian liver. *Mol. Cell.* **58**, 147–156 (2015).

[2] Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474 (2011).

[3] Larsson, A. J. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251 (2019).

[4] Zenklusen, D., Larson, D. R. & Singer, R. H. Single-rna counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).

[5] Sanchez, A. & Golding, I. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–1193 (2013).

[6] Peccoud, J. & Ycart, B. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).

[7] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mrna synthesis in mammalian cells. *PLOS Biol.* **4**, e309 (2006).

[8] Tunnacliffe, E. & Chubb, J. R. What is a transcriptional burst? *Trends in Genetics* (2020).

[9] Kumar, N. & Kulkarni, R. V. Constraining the complexity of promoter dynamics using fluctuations in gene expression. *Physical Biology* **17**, 015001 (2019).

[10] Zhou, T. & Zhang, J. Analytical results for a multistate gene model. *SIAM Journal on Applied Mathematics* **72**, 789–818 (2012).

[11] Rodriguez, J. *et al.* Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* **176**, 213–226 (2019).

[12] Zhang, J. & Zhou, T. Stationary moments, distribution conjugation and phenotypic regions in stochastic gene transcription. *Mathematical biosciences and engineering: MBE* **16**, 6134–6166 (2019).

[13] Bothma, J. P. *et al.* Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living drosophila embryos. *Proceedings of the National Academy of Sciences* **111**, 10598–10603 (2014).

[14] Corrigan, A. M., Tunnacliffe, E., Cannon, D. & Chubb, J. R. A continuum model of transcriptional bursting. *Elife* **5**, e13051 (2016).

[15] Bartman, C. R. *et al.* Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol. Cell.* **73**, 519–532 (2019).

[16] Shao, W. & Zeitlinger, J. Paused rna polymerase ii inhibits new transcriptional initiation. *Nature genetics* **49**, 1045 (2017).

[17] Gressel, S. *et al.* Cdk9-dependent rna polymerase ii pausing controls transcription initiation. *Elife* **6**, e29736 (2017).

[18] Xu, H., Skinner, S. O., Sokac, A. M. & Golding, I. Stochastic kinetics of nascent rna. *Phys. Rev. Lett.* **117**, 128101 (2016).

[19] Phillips, R. *et al.* Figure 1 theory meets figure 2 experiments in the study of gene expression. *Annual review of biophysics* **48**, 121–163 (2019).

[20] Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA* **105**, 17256–17261 (2008).

[21] Tantale, K. *et al.* A single-molecule view of transcription reveals convoys of rna polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).

[22] Van Kampen, N. G. *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam, 1992).

[23] Thomas, P., Popović, N. & Grima, R. Phenotypic switching in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **111**, 6994–6999 (2014).

[24] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

[25] Mannan, A. A., Liu, D., Zhang, F. & Oyarzún, D. A. Fundamental Design Principles for Transcription-Factor-Based Metabolite Biosensors. *ACS Synthetic Biology* **6**, 1851–1859 (2017).

[26] Arpino, J. A. *et al.* Tuning the dials of synthetic biology. *Microbiology* **159**, 1236 (2013).

[27] Redner, S. *A guide to first-passage processes* (Cambridge University Press, 2001).

[28] Iyer-Biswas, S., Hayot, F. & Jayaprakash, C. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E* **79**, 031911 (2009).

[29] Zhou, T. & Zhang, J. Analytical results for a multistate gene model. *SIAM J. Appl. Math.* **72**, 789–818 (2012).

[30] Bothma, J. P., Norstad, M. R., Alamos, S. & Garcia, H. G. Llamatags: a versatile tool to image transcription factor dynamics in live embryos. *Cell* **173**, 1810–1822 (2018).

[31] Cao, Z. & Grima, R. Accuracy of parameter estimation for auto-regulatory transcriptional feedback loops from noisy data. *J. Royal Soc. Interface* **16**, 20180967 (2019).

[32] Cao, Z. & Grima, R. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nat. Commun.* **9**, 3305 (2018).

[33] Holehouse, J. & Grima, R. Revisiting the reduction of stochastic models of genetic feedback loops with fast promoter switching. *Biophys. J.* **117**, 1311–1330 (2019).

[34] Holehouse, J., Cao, Z. & Grima, R. Stochastic modeling of auto-regulatory genetic feedback loops: a review and comparative study. *Biophysical Journal* **118**, 1517–1525 (2020).

[35] Cisse, I. I. *et al.* Real-time dynamics of rna polymerase ii clustering in live human cells. *Science* **341**, 664–667 (2013).

[36] Skinner, S. O. *et al.* Single-cell analysis of transcription kinetics across the cell cycle. *Elife* **5**, e12175 (2016).

[37] Choubey, S., Kondev, J. & Sanchez, A. Deciphering transcriptional dynamics in vivo by counting nascent rna molecules. *PLoS computational biology* **11** (2015).

[38] Lin, J. & Amir, A. Homeostasis of protein and mrna concentrations in growing cells. *Nature communications* **9**, 1–11 (2018).

[39] Nikolados, E. M., Weiße, A. Y., Ceroni, F. & Oyarzún, D. A. Growth defects and loss-of-function in synthetic gene circuits. *ACS Synthetic Biology* **8**, 1231–1240 (2019).

[40] Zopf, C., Quinn, K., Zeidman, J. & Maheshri, N. Cell-cycle dependence of transcription dominates noise in gene expression. *PLoS computational biology* **9** (2013).