Supplemental Material

The Stochastic Kinetics of Nascent RNA

Heng Xu[1,2,3,*], Samuel O. Skinner[1,2,3], Anna Marie Sokac[3] and Ido Golding[1,2,3,*]

[1]Center for Theoretical Biological Physics, Rice University, Houston, Texas, USA

[2]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[3]Verna & Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas, USA

* hengx@bcm.edu, golding@bcm.edu

Supplemental Material

Table of Contents:

Supplemental Material

1. Model assumptions

Our model describes transcription in four steps: Gene activation, transcription initiation, nascent RNA synthesis (elongation), and release (**Fig. 1a**). Regarding the gene activation step, we assume that the gene can randomly switch between an active state (state 1) and an inactive state (state 0), with switching in both directions as a Poisson process. This assumption is often used in models of cellular RNAs [1-3], to explain the observed bursty kinetics of RNA production and the measured RNA copy-number statistics [1,4,5]. Recently, it has been shown that, in some instances, more than two gene states are needed to describe the observed RNA copy-number statistics [6,7]. To accommodate for such scenarios, our model can be extended to the more general case of N gene states (see **SM Section 14.1**).

In the active state, we model the initiation of new transcripts as a Poisson process with rate k_{INI}. The assumption of Poisson initiation follows the standard model for cellular RNAs [1-3]. The model can be extended to include more complex initiation kinetics, as described e.g. in [8] (see **SM Section 14.2**). Following initiation, the synthesis of a nascent RNA molecule from the gene is described as a continuous elongation process with a constant speed V_{EL}. The approximation of constant elongation speed is justified by the fact that, even if each elongation step is highly stochastic, when examined at the resolution of nascent RNA measurements (~10-100 fluorescent molecules per gene [1,9,10]), the variability in speed is significantly diminished [8,11,12].

In the model, we neglect the interaction between adjacent RNA polymerases (RNAPs). To justify this, we note that a single transcribing RNAP covers a region of DNA sequence of size $\sim 25\text{-}40\,\mathrm{base\ pairs\ (bp)}$ [13,14]. For the *hb* gene analyzed in this work, we found a maximum number of $60\text{-}80\,\mathrm{RNAPs}$ per gene (corresponding to $k_{\mathrm{INI}} \sim 20\text{-}30\ \mathrm{min}^{-1}$). Considering the length of the *hb* gene ($3.6\ \mathrm{kbp}$), the average center-to-center distance between neighboring RNAPs is ~ $45\text{-}60\ \mathrm{bp}$. Thus, transcribing RNAPs are not expected to overlap each other. Consistent with this picture, a recent numerical study suggested that RNAP interaction can still be neglected for $k_{\mathrm{INI}} \lesssim 30\ \mathrm{min}^{-1}$ [8].

However, it is also possible that the stochastic kinetics of RNAP movement during elongation will lead to RNAP-RNAP interaction. This will be the case, for example, if RNAP pauses or

backtracks for durations longer than the ~1 sec it takes to cover the inter-RNAP distance [15-17]. Even without direct contact, RNAPs may interact indirectly by modifying the local supercoiling state of transcribed DNA [18,19]. A number of previous works modeled RNAP interaction [8,19-21], and their approaches can in the future be incorporated into our modeling framework.

Following nascent RNA synthesis, we assume that the completed transcript stays at the gene for a certain time period T_S before being released. This step in the model is aimed at capturing the post-elongation phase, where multiple RNA processes take place, e.g. splicing, capping and polyadenylation [22]). The duration and stochasticity of these processes may depend heavily on the specific gene and organism: Experimentally measured T_S range from much shorter [23,24] to much longer [25-27] than the RNA elongation time (L/V_{EL}). We use a fixed delay time (T_S) to describe the post elongation residence, corresponding to a deterministic release kinetics. However, stochastic release kinetics can be included in the model (see **SM Section 14.3**).

2. Comparison with previous models of nascent RNA kinetics

A number of models have been developed to describe nascent RNA kinetics. Among them, a few focus on one or two specific features of transcription (e.g. RNAP pausing and traffic jam [20,21], sequence-dependent motion of RNAP [28]), while treating all other aspects of the transcription process in a highly simplified manner (e.g. neglecting gene activation). More recently, in parallel with the emergence of single-cell nascent RNA measurements, more comprehensive models of transcription kinetics were introduced, aimed at explaining the experimentally-measured statistics of nascent RNA. Zenklusen et al. [5] proposed a model where gene activation and transcription initiation are described as Poisson processes, and RNA synthesis is described as a fixed time interval. Using Gillespie simulations, they examined the distribution of RNAP numbers on the gene. The authors did not write down the master equation and did not calculate the corresponding probability distribution. In addition, the model only described the copy-number statistics of RNAP rather than of partially synthesized nascent RNA, thus limiting the ability to directly compare model predictions with experimental smFISH data. Such a comparison required an extra step of rescaling smFISH signals to the equivalent number of RNAPs [7]. More recently, Choubey et al. [8] introduced a different model, where in addition to

stochastic gene activation and transcription initiation, RNAP movement on the gene is also described as a chain of stochastic hopping from one base pair to the next. This formalism allowed modeling varying lengths of nascent RNA. The model also considered complex multi-step kinetics for transcription initiation. By writing the master equation for the model, Choubey et al. derived the mean and variance of RNAP numbers at the gene. However, the single-base resolution of the model made the master equation too complicated to solve, even numerically.

In the formalism presented in this work, we explicitly model the observed nascent RNA signal by considering the length of partially-synthesized RNA and the specific detection scheme (e.g. an smFISH probe set). We describe RNA synthesis as a continuous, deterministic elongation process. This is similar to the approach of [5], and can be thought of as the continuous limit of Choubey et al.'s model [8]. Applying this simplification allows us to directly solve the master equation for nascent RNA and investigate the shape of the distribution function, steps which were not possible using previous models. At the same time, our modeling framework is flexible enough that we can, in a straightforward manner, extend the model to include additional features such as multi-step gene activation, non-Poissonian initiation, stochastic RNA release, finite smFISH labeling efficiency, etc. (see **SM Section 14**).

3. Constructing the master equation for *P*(*n*,*m*)

Following **Eq. (1)** (*Note: red pointers indicate equations in the main text*), which describes $P(\nu,\mu)$, we now derive an equation explicitly for $P(n,m)$. As the first step, the distribution of the system state (n,m) at time t can be formally written as

$$\bar{\mathbf{P}}_t = \mathbf{U}(0,-T_{\mathrm{RES}})\bar{\mathbf{P}}_{t,\tau=-T_{\mathrm{RES}}}. \tag{S1}$$

Here, we have rewritten **Eq. (1)** in a full matrix form, combining different μ (or m) values. Specifically, (1) $\bar{\mathbf{P}}_t \equiv \left[\mathbf{P}_t(m)\right]^{\mathrm{T}}$ is the vectorized probability distribution of the system state (n,m) at time t. Since $\mu = m$ for $\tau = 0$, $\bar{\mathbf{P}}_t$ is equivalent to the vectorized probability distribution of (ν,μ) for $\tau = 0$, i.e. $\bar{\mathbf{P}}_t = \bar{\mathbf{P}}_{t,\tau=0} \equiv \left[\mathbf{P}_{t,\tau=0}(\mu)\right]^{\mathrm{T}}$. (2) $\bar{\mathbf{P}}_{t,\tau=-T_{\mathrm{RES}}} \equiv \left[\mathbf{P}_{t,\tau=-T_{\mathrm{RES}}}(\mu)\right]^{\mathrm{T}}$ is the

vectorized initial condition. Since $m = 0$ for $\tau = -T_{\mathrm{RES}}$, only the $m = 0$ components of $\bar{\mathbf{P}}_{t,\tau=-T_{\mathrm{RES}}}$ are nonzero, and equal to the marginal distribution of the gene state n at time $t - T_{\mathrm{RES}}$, i.e. $\bar{\mathbf{P}}_{t,\tau=-T_{\mathrm{RES}}} = \left[\delta(m)\int_0^\infty \mathbf{P}_{t-T_{\mathrm{RES}}}(m)dm\right]^{\mathrm{T}}$. (3) $\mathbf{U} = \mathrm{T}\left[\exp\left(\int_{-T_{\mathrm{RES}}}^{0}\left(\bar{\mathbf{K}} + \bar{\mathbf{K}}_{\mathrm{INI}}(\tau)\right)d\tau\right)\right]$ is the propagator from $\tau = -T_{\mathrm{RES}}$ to $\tau = 0$, with $\bar{\mathbf{K}} = \left[\mathbf{K}\delta(m-m)\right]$ and $\bar{\mathbf{K}}_{\mathrm{INI}} = \left[\mathbf{K}_{\mathrm{INI}}\left(\delta(m-m-g(\tau)) - \delta(m-m)\right)\right]$ the full matrix versions of $\mathbf{K}$ and $\mathbf{K}_{\mathrm{INI}}$ combining different m and m values.

Eq. (S1) relates the distribution of the system state (n,m) at time t to the marginal distribution of the gene state n at time $t - T_{\mathrm{RES}}$ for any given t. Therefore, to calculate how (n,m) evolves from t to $t + dt$, we only need to know how the marginal distribution of n (the only nonzero component of $\bar{\mathbf{P}}_{t,\tau=-T_{\mathrm{RES}}}$) evolves from $t - T_{\mathrm{RES}}$ to $t - T_{\mathrm{RES}} + dt$. Since transitions between different n's are simply described by $\mathbf{K}$, we construct the master equation for $P(n,m)$ as

$$\dot{\bar{\mathbf{P}}}_t = \mathbf{M}\bar{\mathbf{P}}_t = \mathbf{U}\bar{\mathbf{K}}\mathbf{U}^{-1}\bar{\mathbf{P}}_t, \tag{S2}$$

where $\mathbf{M} \equiv \mathbf{U}\bar{\mathbf{K}}\mathbf{U}^{-1}$ is the transition matrix for the system state (n,m). Notice that **Eq. (S2)** is analogous to the Heisenberg representation in quantum mechanics. In other words, $\mathbf{M}$ is the solution of the following Heisenberg equation for $\mathcal{M}$ at $\tau = 0$:

$$\dot{\mathcal{M}} = [\bar{\mathbf{K}} + \bar{\mathbf{K}}_{\mathrm{INI}}(\tau), \mathcal{M}], \tag{S3}$$

where the initial condition is $\mathcal{M}_{\tau=-T_{\mathrm{RES}}} = \bar{\mathbf{K}}$.

Since $\bar{\mathbf{K}}$, $\bar{\mathbf{K}}_{\mathrm{INI}}$, and $\mathbf{U}$ all have infinite dimensions, it is difficult to solve the actual form of $\mathbf{M}$ from **Eq. (S3)**. Instead, by transforming $P(n,m)$ into its characteristic function $\Psi(n,\omega)$ [29], we can decouple different dimensions, and convert **Eq. (S2)** into a two-dimensional equation

$$\dot{\boldsymbol{\Psi}}(\omega) = \tilde{\mathbf{M}}\boldsymbol{\Psi}(\omega) = \tilde{\mathbf{U}}\mathbf{K}\tilde{\mathbf{U}}^{-1}\boldsymbol{\Psi}(\omega), \tag{S4}$$

where $\tilde{\mathbf{U}} = \mathrm{T}\left[\exp\left(\int_{-T_{\mathrm{RES}}}^{0}\left(\mathbf{K} + (e^{i\omega g(\tau)} - 1)\mathbf{K}_{\mathrm{INI}}\right)d\tau\right)\right]$ is the propagator of **Eq. (2)**, and $\tilde{\mathbf{M}} \equiv \tilde{\mathbf{U}}\mathbf{K}\tilde{\mathbf{U}}^{-1}$ is the transition matrix for (n,ω). The 2×2 transition matrix $\tilde{\mathbf{M}}$ is therefore the solution of the Heisenberg equation for $\mathcal{M}$

$$\dot{\tilde{\mathcal{M}}} = [\mathbf{K} + (e^{i\omega g(\tau)} - 1)\mathbf{K}_{\mathrm{INI}}, \tilde{\mathcal{M}}] \tag{S5}$$

at $\tau = 0$, with initial condition $\tilde{\mathcal{M}}_{\tau=-T_{\mathrm{RES}}} = \mathbf{K}$.

Eqs. (S2) and **(S4)** can be used as an alternative to **Eqs. (1)** and **(2)** for solving $P(n,m)$. This can be done either at steady state (as described in the paper), or for the time-dependent behavior (as mentioned in the **Conclusion**).

4. Numerical methods for solving Eq. 1

If analytical solutions for **Eq. 1** are not available, numerical solutions can be obtained using the Finite State Projection (FSP) method [6,9,30]. The basic idea is to simplify **Eq. 1** so that it includes only a finite set of m values. We first discretize the nascent RNA signal to $m = 0, \Delta m, 2\Delta m, \cdots$ Here, the bin size Δm must match the smallest change in the contribution function g. For example, if $g = 1$, m is naturally discretized, and $\Delta m = 1$. If $g = -\tau$, m is continuous, so the bin size needs to be much smaller than one ($\Delta m \ll 1$). We write the discretized version of **Eq. 1** [9]:

$$\dot{\bar{\mathbf{P}}} = \left(\bar{\mathbf{K}} + \bar{\mathbf{K}}_{\mathrm{INI}}(\tau)\right)\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{K}-\mathbf{K}_{\mathrm{INI}} & 0 & 0 & 0 \\ 0 & \mathbf{K}-\mathbf{K}_{\mathrm{INI}} & 0 & \cdots \\ 0 & 0 & \mathbf{K}-\mathbf{K}_{\mathrm{INI}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \mathbf{K}_{\mathrm{INI}} & 0 & 0 & \ddots \\ 0 & \mathbf{K}_{\mathrm{INI}} & 0 & \ddots \\ 0 & 0 & \mathbf{K}_{\mathrm{INI}} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{P}(0) \\ \mathbf{P}(\Delta m) \\ \mathbf{P}(2\Delta m) \\ \vdots \\ \mathbf{P}(g(\tau)) \\ \mathbf{P}(g(\tau)+\Delta m) \\ \mathbf{P}(g(\tau)+2\Delta m) \\ \vdots \end{bmatrix}, \tag{S6}$$

where

$$\bar{\mathbf{K}} = \begin{bmatrix} \mathbf{K} & 0 & 0 & \cdots \\ 0 & \mathbf{K} & 0 & \cdots \\ 0 & 0 & \mathbf{K} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \bar{\mathbf{K}}_{\mathrm{INI}}(\tau) = \begin{bmatrix} -\mathbf{K}_{\mathrm{INI}} & 0 & 0 & \cdots \\ 0 & -\mathbf{K}_{\mathrm{INI}} & 0 & \cdots \\ \vdots & 0 & -\mathbf{K}_{\mathrm{INI}} & \cdots \\ \mathbf{K}_{\mathrm{INI}} & \vdots & 0 & \cdots \\ 0 & \mathbf{K}_{\mathrm{INI}} & \vdots & \ddots \\ 0 & 0 & \mathbf{K}_{\mathrm{INI}} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad \text{(S7)}$$

Eq. (S6) is solved by propagating the initial state $\bar{\mathbf{P}}_{\tau=-T_{\mathrm{RES}}}$ through the time window $-T_{\mathrm{RES}} \le \tau \le 0$, i.e.

$$\bar{\mathbf{P}}_{\tau=0} = \left(\mathbf{I} + \bar{\mathbf{K}}\Delta\tau + \bar{\mathbf{K}}_{\mathrm{INI}}(-\Delta\tau)\Delta\tau\right)\cdots\left(\mathbf{I} + \bar{\mathbf{K}}\Delta\tau + \bar{\mathbf{K}}_{\mathrm{INI}}(-T_{\mathrm{RES}})\Delta\tau\right)\bar{\mathbf{P}}_{\tau=-T_{\mathrm{RES}}}, \quad \text{(S8)}$$

where $\mathbf{I}$ is the unit matrix, and $\Delta\tau \ll T_{\mathrm{RES}}$ is the discretized time step. By setting an upper limit of $\mathscr{m}$, and truncating $\bar{\mathbf{P}}$, $\bar{\mathbf{K}}$, and $\bar{\mathbf{K}}_{\mathrm{INI}}$ into finite dimensional vector and matrices, we are able to numerically calculate the steady-state distribution of $P(n,m)$ [9]. Clearly, the truncated range of $\mathscr{m}$ must be large enough to cover the main portion of the distribution [6,30]. Following [9], we set the upper limit of $\mathscr{m}$ to be 60.

We implemented the FSP method using MATLAB (MathWorks Inc.). According to **Eq. 1**, $P(n,m)$ is a function of $k_{01}\cdot T_{\mathrm{RES}}$, $k_{10}\cdot T_{\mathrm{RES}}$, $k_{\mathrm{INI}}\cdot T_{\mathrm{RES}}$. When solving **Eq. (S8)**, we always set $T_{\mathrm{RES}}=1$ and replace $\{k_{01}, k_{10}, k_{\mathrm{INI}}\}$ with their products with the actual value of T_{RES} (see **SM Section 11** for the estimation of T_{RES}). To balance accuracy and speed, we used the following parameter values: $\Delta\mathscr{m}=0.001$ and $\Delta\tau=0.0001$ for theoretical distributions shown in **Fig. 4(a)** and **Fig. S1(b)**; $\Delta\mathscr{m}=0.1$ and $\Delta\tau=0.001$ for fitting of experimental data shown in **Fig. 3**; $\Delta\mathscr{m}=1/72$ and $\Delta\tau=0.001$ for theoretical distributions shown in **Fig. 4(c)**.

Stochastic simulation of the model was performed following the Gillespie algorithm [5,9,31].

Supplemental Material

5. Deriving the moments of *P*(*m*)

In addition to the full distribution function $P(m)$, the moments of that distribution can also be used for interpreting experimental data [8,32,33]. In general, the N th raw moment $\mu_N \equiv \langle m^N \rangle$ can be calculated from the characteristic function Ψ as [29]

$$\mu_N = \mathbf{u} \cdot (-i)^N \left. \frac{d^N \mathbf{\Psi}}{d\omega^N} \right|_{\tau=0,\omega=0}, \tag{S9}$$

with $\mathbf{u} = (1,1)$. Hence, by taking the derivative of **Eq. (2)** with respect to ω, we obtain

$$\frac{d}{d\tau}\left(\left. \frac{d^N \mathbf{\Psi}}{d\omega^N} \right|_{\omega=0} \right) = \mathbf{K} \left. \frac{d^N \mathbf{\Psi}}{d\omega^N} \right|_{\omega=0} + \mathbf{K}_{\mathrm{INI}} \sum_{j=0}^{N-1} \binom{N}{j} (ig)^{N-j} \left. \frac{d^j \mathbf{\Psi}}{d\omega^j} \right|_{\omega=0}. \tag{S10}$$

This is a set of first order linear differential equations for $\left. \frac{d^N \mathbf{\Psi}}{d\omega^N} \right|_{\omega=0}$, whose solutions are well known. For $\tau = 0$, we have

$$\mu_N = \mathbf{u} \cdot \sum_{I=1}^{N} \sum_{\underbrace{0=k_0<k_i<k_I=N}_{i=1,\cdots,I-1}} \int_{-T_{\mathrm{RES}}}^{0} d\tau_1 \cdots \int_{-T_{\mathrm{RES}}}^{\tau_{I-1}} d\tau_I \mathrm{T}\left[\prod_{i=1}^{I} \binom{k_i}{k_{i-1}} g(\tau_i)^{k_i - k_{i-1}} \mathbf{W}(\tau_i) \right] \mathbf{\Psi}_{\tau=-T_{\mathrm{RES}}}(0), \tag{S11}$$

where $\mathbf{W}(t) = e^{-\mathbf{K}t} \mathbf{K}_{\mathbf{INI}} e^{\mathbf{K}t}$. Specifically, we find the mean (μ_1) and the variance ($\bar{\mu}_2 = \mu_2 - \mu_1^2$) of m to be

$$\mu_1 = \mathbf{u} \cdot \int_{-T_{\mathrm{RES}}}^{0} d\tau_1 g(\tau_1) \mathbf{W}(\tau_1) \mathbf{\Psi}_{\tau=-T_{\mathrm{RES}}}(0) = \frac{k_{01} k_{\mathrm{INI}}}{k_{01} + k_{10}} \int_{-T_{\mathrm{RES}}}^{0} g(\tau_1) d\tau_1, \tag{S12}$$

$$\begin{aligned} \bar{\mu}_2 &= \mathbf{u} \cdot \left\{ \int_{-T_{\mathrm{RES}}}^{0} d\tau_1 g(\tau_1)^2 \mathbf{W}(\tau_1) + \int_{-T_{\mathrm{RES}}}^{0} d\tau_1 \int_{-T_{\mathrm{RES}}}^{\tau_1} d\tau_2 g(\tau_1) g(\tau_2) \left[\mathbf{W}(\tau_1), \mathbf{W}(\tau_2) \right] \right\} \mathbf{\Psi}_{\tau=-T_{\mathrm{RES}}}(0) \\ &= \frac{k_{01} k_{\mathrm{INI}}}{k_{01} + k_{10}} \int_{-T_{\mathrm{RES}}}^{0} g(\tau_1)^2 d\tau_1 + \frac{k_{01} k_{10} k_{\mathrm{INI}}^2}{k^2} \int_{-T_{\mathrm{RES}}}^{0} d\tau_1 \int_{-T_{\mathrm{RES}}}^{\tau_1} d\tau_2 2 g(\tau_1) g(\tau_2) \sinh\left(k(\tau_2 - \tau_1) \right), \end{aligned} \tag{S13}$$

where [·] is the commutation operator and $k = k_{01} + k_{10}$. The first term on the RHS of **Eq. (S13)** represents the contribution from Poissonian transcription initiation when the gene is in the active

state, whereas the second term denotes the added fluctuations caused by transitions between gene states [32,33].

6. Deriving the steady-state distribution for $g = 1$

For $g=1$, **Eq. 1** becomes

$$\begin{cases} \dot{P}(0,m) = -k_{01}P(0,m) + k_{10}P(1,m) \\ \dot{P}(1,m) = k_{01}P(0,m) - k_{10}P(1,m) - k_{\mathrm{INI}}P(1,m) + k_{\mathrm{INI}}P(1,m-1), \end{cases} \tag{S14}$$

and the initial conditions are (for $T_{\mathrm{RES}}=1$)

$$\begin{cases} P_{\tau=-1}(0,m) = \dfrac{k_{10}\delta(m)}{k_{01}+k_{10}} \\ P_{\tau=-1}(1,m) = \dfrac{k_{01}\delta(m)}{k_{01}+k_{10}}. \end{cases} \tag{S15}$$

Since m only takes integer values, we define the generating function $F_n(z,\tau) \equiv \sum_{m=0}^{\infty} z^m P_\tau(n,m)$, and transform **Eq. (S14)** to obtain

$$\begin{cases} \dot{F}_0 = -k_{01}F_0 + k_{10}F_1 \\ \dot{F}_1 = k_{01}F_0 - k_{10}F_1 + (z-1)k_{\mathrm{INI}}F_1, \end{cases} \tag{S16}$$

with the initial conditions

$$\begin{cases} F_0\big|_{\tau=-1} = \dfrac{k_{10}}{k_{01}+k_{10}} \\ F_1\big|_{\tau=-1} = \dfrac{k_{01}}{k_{01}+k_{10}}. \end{cases} \tag{S17}$$

By further defining the marginal generating function $F(z,\tau) \equiv F_0(z,\tau) + F_1(z,\tau) = \sum_{m=0}^{\infty} z^m P_\tau(m)$ corresponding to the marginal probability $P_\tau(m) \equiv P_\tau(0,m) + P_\tau(1,m)$, we convert **Eq. (S16)**

to a 2nd order differential equation (**Eq. (6)**),

$$\ddot{F}+\left[k_{10}+k_{01}+(1-z)k_{\mathrm{INI}}\right]\dot{F}+(1-z)k_{\mathrm{INI}}k_{01}F=0, \tag{S18}$$

with the initial conditions

$$\begin{cases} F\big|_{\tau=-1}=1 \\ \dot{F}\big|_{\tau=-1}=(z-1)\dfrac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}. \end{cases} \tag{S19}$$

Next, using the ansatz $F=e^{\omega\tau}$, we obtain, from **Eq. (S18)**,

$$\omega^2+\left[k_{10}+k_{01}+(1-z)k_{\mathrm{INI}}\right]\omega+(1-z)k_{\mathrm{INI}}k_{01}=0, \tag{S20}$$

whose solutions are

$$\omega_{1,2}=\frac{-\left[k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)\right]\mp\sqrt{\left[k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)\right]^2-4k_{01}k_{\mathrm{INI}}(1-z)}}{2}. \tag{S21}$$

The general solution of **Eq. (S18)** is $F=Ae^{\omega_1\tau}+Be^{\omega_2\tau}$, where the parameters A and B are determined by the initial conditions in **Eq. (S19)**. With $T_{\mathrm{RES}}=1$, we obtain the steady-state generating function,

$$F(z,\tau=0)=\frac{\omega_2e^{\omega_1}-\omega_1e^{\omega_2}}{\omega_2-\omega_1}+\frac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}(z-1)\frac{e^{\omega_2}-e^{\omega_1}}{\omega_2-\omega_1}. \tag{S22}$$

The steady-state distribution is calculated using

$$P(m)=\frac{1}{m!}\frac{\partial^m}{\partial z^m}F(z,0)\big|_{z=0}. \tag{S23}$$

To apply this formula, we consider the explicit dependence of F on z. Specifically, we divide $\omega_{1,2}$ into two parts, $\omega_{1,2}=\dfrac{\eta_1\mp\eta_2}{2}$, where $\eta_1=-\left[k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)\right]$ is a linear function of z,

and $\eta_2^2=\left[k_{\mathrm{INI}}(1-z)+\kappa_1\right]\left[k_{\mathrm{INI}}(1-z)+\kappa_2\right]$ is the product of two linear functions of z, with $\kappa_{1,2}=k_{10}-k_{01}\pm 2i\sqrt{k_{01}k_{10}}$. Consequently, **Eq. (S22)** can be re-written as

$$F(z,0)=\left[\frac{2k_{01}k_{\mathrm{INI}}}{k_{01}+k_{10}}(z-1)-\eta_1\right]\frac{e^{\eta_1/2}}{2}E_1+e^{\eta_1/2}E_0, \tag{S24}$$

with $E_{\mathrm{s}}=\sum_{l=0}^{\infty}\frac{\left[k_{\mathrm{INI}}(1-z)+\kappa_1\right]^l\left[k_{\mathrm{INI}}(1-z)+\kappa_2\right]^l}{(2l+s)!2^{2l}}$. The ith derivative of E_{s} at $z=0$ obeys

$$\begin{aligned}\left.\frac{\partial^i E_s}{\partial z^i}\right|_{z=0}&=\sum_{2l\geq i}\sum_{w=\max(0,i-l)}^{\min(l,i)}\binom{l}{w}\binom{l}{i-w}\frac{(-1)^i\,i!}{(2l+s)!}\left(\frac{k_{\mathrm{INI}}+\kappa_1}{2}\right)^{l-w}\left(\frac{k_{\mathrm{INI}}+\kappa_2}{2}\right)^{l-i+w}\left(\frac{k_{\mathrm{INI}}}{2}\right)^{i}\\&=M_{s,i}\left(\frac{k_{\mathrm{INI}}}{2}\right)^{i},\end{aligned} \tag{S25}$$

where

$$M_{s,i}=\sum_{2l\geq i}\sum_{w=\max(0,i-l)}^{\min(l,i)}\binom{l}{w}\binom{l}{i-w}\frac{(-1)^i\,i!}{(2l+s)!}\left(\frac{k_{\mathrm{INI}}+\kappa_1}{2}\right)^{l-w}\left(\frac{k_{\mathrm{INI}}+\kappa_2}{2}\right)^{l-i+w}. \tag{S26}$$

Hence, the probability distribution $P(m)$ is given by (**Eq. (7)**)

$$\begin{aligned}P(m)=\frac{e^{-\frac{k_{01}+k_{10}+k_{\mathrm{INI}}}{2}}}{m!}\left(\frac{k_{\mathrm{INI}}}{2}\right)^m\Bigg\{&\left[\frac{k_{01}+k_{10}+k_{\mathrm{INI}}}{2}-\frac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}\right]\sum_{i=0}^{m}\binom{m}{i}M_{1,i}\\&+\sum_{i=0}^{m}\binom{m}{i}M_{0,i}+m\frac{k_{01}-k_{10}}{k_{01}+k_{10}}\sum_{i=0}^{m-1}\binom{m-1}{i}M_{1,i}\Bigg\}.\end{aligned} \tag{S27}$$

7. An approximation for *P*(*m*) in the limit of slow gene-state switching

In the limit where $(k_{01}\,\&\,k_{10})\ll k_{\mathrm{INI}}$ and $(k_{01}\text{or}\,k_{10})\ll 1$, the square root term in **Eq. (S21)** can be approximated by $\eta_2\approx -k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)$, which leads to

Supplemental Material

$$\begin{cases} \omega_1 = -k_{10} - k_{\mathrm{INI}}(1-z) \\ \omega_2 = -k_{01}. \end{cases} \tag{S28}$$

Hence, **Eq. (S22)** is simplified to

$$\begin{aligned} F(z,0) &\approx \frac{-k_{01} - \dfrac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}(z-1)}{-k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)} e^{-k_{10}-k_{\mathrm{INI}}} e^{k_{\mathrm{INI}}z} + \frac{k_{10} - \dfrac{k_{\mathrm{INI}}k_{10}}{k_{01}+k_{10}}(z-1)}{-k_{01}+k_{10}+k_{\mathrm{INI}}(1-z)} e^{-k_{10}} \\ &= \frac{I_1(z)}{J(z)} e^{-k_{10}-k_{\mathrm{INI}}} e^{k_{\mathrm{INI}}z} + \frac{I_2(z)}{J(z)} e^{-k_{10}}, \end{aligned} \tag{S29}$$

where

$$\begin{cases} I_1(z) = -k_{01} - \dfrac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}(z-1) \\ I_2(z) = k_{10} - \dfrac{k_{\mathrm{INI}}k_{10}}{k_{01}+k_{10}}(z-1) \\ J(z) = -k_{01}+k_{10}+k_{\mathrm{INI}}(1-z) \end{cases} \tag{S30}$$

are linear functions of z. Using the following identities [2]:

$$\frac{\partial^n}{\partial z^n}\frac{I(z)}{J(z)} = n!\sum_{k=0}^{n}\frac{\partial^{n-k}}{\partial z^{n-k}} I(z) \sum_{j=0}^{k}\frac{(-1)^j (k+1) J(z)^{-j-1}}{(j+1)!(n-k)!(k-j)!}\frac{\partial^k}{\partial z^k} J(z)^j, \tag{S31}$$

and applying **Eq. (S23)**, we derive the steady-state distribution of m as

$$\begin{aligned} P(m) = {} & \frac{k_{10}e^{-k_{01}}}{k_{01}+k_{10}}\delta_{m,0} + \frac{k_{01}e^{-k_{10}-k_{\mathrm{INI}}}}{k_{01}+k_{10}}\frac{k_{\mathrm{INI}}^m}{m!} \\ & + \left(\frac{k_{\mathrm{INI}}}{k_{\mathrm{INI}}+k_{10}-k_{01}}\right)^{m+1} \frac{2k_{01}k_{10}e^{-k_{01}}}{(k_{01}+k_{10})k_{\mathrm{INI}}}\left[1 - e^{-k_{\mathrm{INI}}-k_{10}+k_{01}}\sum_{i=0}^{m}\frac{(k_{\mathrm{INI}}+k_{10}-k_{01})^i}{i!}\right]. \end{aligned} \tag{S32}$$

Since $(k_{01}\text{or } k_{10}) \ll 1$, **Eq. (S32)** can be approximated as

$$P(m) = \frac{k_{10}}{k_{01}+k_{10}}\delta_{m,0} + \frac{k_{01}}{k_{01}+k_{10}}\frac{k_{\mathrm{INI}}^m}{m!}e^{-k_{\mathrm{INI}}}, \tag{S33}$$

which is the weighed sum of two Poisson distributions, with rates 0 and k_{INI}.

Eq. (S33) is identical to the solution for the commonly used two-state model for cellular RNA kinetics [1-3,5] in the same limit defined above ($(k_{01}\,\&\,k_{10}) \ll k_{\mathrm{INI}}$ and $(k_{01}\,\mathrm{or}\,k_{10}) \ll 1$), as long as we replace the residence time T_{RES} with the RNA degradation rate k_{D} (**Fig. 2(a)** panels I, III, and IV). To compare the two models outside this limit, we measured the relative difference d between their distributions P_{nas} and P_{cell}, defined as $d = \max\left(\left|P_{\mathrm{nas}} - P_{\mathrm{cell}}\right|\right) / \max\left(P_{\mathrm{nas}}\right)$ (**Fig. S1**). We found that $d \sim 1$ for $(k_{01}\,\&\,k_{10}) \sim 1$, indicating that the probability distributions predicted by the nascent and cellular RNA models are very different under this condition (**Fig. 2(a)** panel II).

8. Deriving the steady-state characteristic function for $g = -\tau$

For $g = -\tau$, **Eq. 2** becomes

$$\begin{cases} \dot{\Psi}(0,\omega) = -k_{01}\Psi(0,\omega) + k_{10}\Psi(1,\omega) \\ \dot{\Psi}(1,\omega) = k_{01}\Psi(0,\omega) - k_{10}\Psi(1,\omega) - k_{\mathrm{INI}}\Psi(1,\omega) + k_{\mathrm{INI}}e^{-i\omega\tau}\Psi(1,\omega), \end{cases} \tag{S34}$$

with the initial conditions

$$\begin{cases} \Psi_{\tau=-1}(0,\omega) = \dfrac{k_{10}}{k_{01}+k_{10}} \\ \Psi_{\tau=-1}(1,\omega) = \dfrac{k_{01}}{k_{01}+k_{10}}. \end{cases} \tag{S35}$$

We define a marginal characteristic function $\Psi(\omega) \equiv \Psi(0,\omega) + \Psi(1,\omega) = \int_0^\infty e^{im\omega} P(m)dm$. **Eq. (S34)** then becomes

$$\begin{cases} \dot{\Psi} = (e^{-i\omega\tau} - 1)k_{\mathrm{INI}}\Psi_1 \\ \dot{\Psi}_1 = k_{01}\Psi - \left[k_{10} + k_{01} + (1 - e^{-i\omega\tau})k_{\mathrm{INI}}\right]\Psi_1, \end{cases} \tag{S36}$$

where $\Psi_1 = \Psi(1,\omega)$. Combining the two equations in **Eq. (S36)** results in a 2nd order differential equation for Ψ_1 (**Eq. (8)**),

$$\ddot{\Psi}_1 + \left[k_{01} + k_{10} + k_{\mathrm{INI}}\left(1 - e^{-i\omega\tau}\right)\right]\dot{\Psi}_1 - \left[k_{\mathrm{INI}}\left(k_{01} - i\omega\right)e^{-i\omega\tau} - k_{\mathrm{INI}}k_{01}\right]\Psi_1 = 0, \tag{S37}$$

with the initial conditions

$$\begin{cases} \Psi_1\big|_{\tau=-1} = \dfrac{k_{01}}{k_{01}+k_{10}} \\ \dot{\Psi}_1\big|_{\tau=-1} = (e^{i\omega}-1)\dfrac{k_{\mathrm{INI}}k_{01}}{k_{01}+k_{10}}. \end{cases} \tag{S38}$$

To solve **Eq. (S37)**, we substitute variables using $z = e^{-i\omega\tau}$, and obtain

$$\left\{-\omega^2\hat{\theta}^2 - i\omega\left[k_{01} + k_{10} + k_{\mathrm{INI}}\left(1 - z\right)\right]\hat{\theta} - \left[k_{\mathrm{INI}}\left(k_{01} - i\omega\right)z - k_{\mathrm{INI}}k_{01}\right]\right\}\Psi_1 = 0, \tag{S39}$$

where $\hat{\theta} = z\dfrac{\partial}{\partial z}$. **Eq. (S39)** is similar to the confluent hypergeometric differential equation [34], except for an extra term $k_{\mathrm{INI}}k_{01}\Psi_1$. To eliminate this term, we use the substitution $\Psi_1 \mapsto z^s W$, and rewrite **Eq. (S39)** as

$$\left\{\theta^2 + \frac{k_{01}+k_{10}+k_{\mathrm{INI}}-2is\omega}{-i\omega}\theta - \frac{k_{\mathrm{INI}}}{-i\omega}z\left(\theta - \frac{k_{01}-i(s+1)\omega}{-i\omega}\right)\right\}W = 0, \tag{S40}$$

where s obeys

$$(-is\omega)^2 + (k_{01}+k_{10}+k_{\mathrm{INI}})(-is\omega) + k_{01}k_{\mathrm{INI}} = 0. \tag{S41}$$

Defining $\kappa = -is\omega$, we write the solution of **Eq. (S40)** as

$$\begin{aligned} W = {} & A_0(\omega)\,{}_1F_1\left(1 + i\frac{k_{01}+\kappa}{\omega}, 1 + i\frac{k_{01}+k_{10}+k_{\mathrm{INI}}+2\kappa}{\omega}, \frac{ik_{\mathrm{INI}}}{\omega}z\right) \\ & + B_0(\omega)U\left(1 + i\frac{k_{01}+\kappa}{\omega}, 1 + i\frac{k_{01}+k_{10}+k_{\mathrm{INI}}+2\kappa}{\omega}, \frac{ik_{\mathrm{INI}}}{\omega}z\right), \end{aligned} \tag{S42}$$

where κ is obtained by solving **Eq. (S41)**,

$$\kappa_{1,2}=\frac{-\left(k_{01}+k_{10}+k_{\mathrm{INI}}\right)\pm\sqrt{\left(k_{01}+k_{10}+k_{\mathrm{INI}}\right)^2-4k_{01}k_{\mathrm{INI}}}}{2}. \tag{S43}$$

Here, ${}_1F_1(a,b,x)$ and $U(a,b,x)$ are the confluent hypergeometric functions of the first and second kind [34,35]. The coefficients $A_0(\omega)$ and $B_0(\omega)$ still need to be determined.

Converting W back to Ψ_1 and using the relation between ${}_1F_1$ and U [34,35] results in a more symmetric-looking expression,

$$\begin{aligned}\Psi_1 = {} & A(\omega)e^{\kappa_1\tau}\,{}_1F_1\left(1+i\frac{k_{01}+\kappa_1}{\omega},1+i\frac{\kappa_1-\kappa_2}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{-i\omega\tau}\right)\\ & +B(\omega)e^{\kappa_2\tau}\,{}_1F_1\left(1+i\frac{k_{01}+\kappa_2}{\omega},1+i\frac{\kappa_2-\kappa_1}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{-i\omega\tau}\right),\end{aligned} \tag{S44}$$

where $A(\omega)$ and $B(\omega)$ are new coefficients.

Applying the initial conditions (**Eq. (S38)**), we find the coefficients $A(\omega)$ and $B(\omega)$ to be

$$\begin{cases} A(\omega)=\dfrac{k_{01}e^{\kappa_1}}{\left(k_{01}+k_{10}\right)M}\left[\left(-i\omega+k_{01}+\kappa_2\right){}_1F_1\left(2+i\dfrac{k_{01}+\kappa_2}{\omega},1+i\dfrac{\kappa_2-\kappa_1}{\omega},\dfrac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right)\right. \\ \qquad\left.-\left((e^{i\omega}-1)k_{\mathrm{INI}}-k_{01}\right){}_1F_1\left(1+i\dfrac{k_{01}+\kappa_2}{\omega},1+i\dfrac{\kappa_2-\kappa_1}{\omega},\dfrac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right)\right] \\ B(\omega)=\dfrac{k_{01}e^{\kappa_2}}{\left(k_{01}+k_{10}\right)M}\left[\left(-i\omega+k_{01}+\kappa_1\right){}_1F_1\left(2+i\dfrac{k_{01}+\kappa_1}{\omega},1+i\dfrac{\kappa_1-\kappa_2}{\omega},\dfrac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right)\right. \\ \qquad\left.-\left((e^{i\omega}-1)k_{\mathrm{INI}}-k_{01}\right){}_1F_1\left(1+i\dfrac{k_{01}+\kappa_1}{\omega},1+i\dfrac{\kappa_1-\kappa_2}{\omega},\dfrac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right)\right], \end{cases} \tag{S45}$$

with

$$\begin{aligned}M = {} & \left(-i\omega+k_{01}+\kappa_2\right){}_1F_1\left(2+i\frac{k_{01}+\kappa_2}{\omega},1+i\frac{\kappa_2-\kappa_1}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right){}_1F_1\left(1+i\frac{k_{01}+\kappa_1}{\omega},1+i\frac{\kappa_1-\kappa_2}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right)\\ & -\left(-i\omega+k_{01}+\kappa_1\right){}_1F_1\left(2+i\frac{k_{01}+\kappa_1}{\omega},1+i\frac{\kappa_1-\kappa_2}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right){}_1F_1\left(1+i\frac{k_{01}+\kappa_2}{\omega},1+i\frac{\kappa_2-\kappa_1}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}e^{i\omega}\right).\end{aligned} \tag{S46}$$

Consequently, the characteristic function at time $\tau = 0$ is

$$\begin{aligned}\Psi\big|_{\tau=0} &= \frac{1}{k_{01}}\dot{\Psi}_1\big|_{\tau=0} + \frac{k_{01}+k_{10}}{k_{01}}\Psi_1\big|_{\tau=0} \\ &= A(\omega)\frac{\left(-i\omega+k_{01}+\kappa_1\right)}{k_{01}}{}_1F_1\left(2+i\frac{k_{01}+\kappa_1}{\omega},1+i\frac{\kappa_1-\kappa_2}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}\right) \\ &+B(\omega)\frac{\left(-i\omega+k_{01}+\kappa_2\right)}{k_{01}}{}_1F_1\left(2+i\frac{k_{01}+\kappa_2}{\omega},1+i\frac{\kappa_2-\kappa_1}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}\right) \\ &+A(\omega)\frac{2k_{01}+k_{10}}{k_{01}}{}_1F_1\left(1+i\frac{k_{01}+\kappa_1}{\omega},1+i\frac{\kappa_1-\kappa_2}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}\right) \\ &+B(\omega)\frac{2k_{01}+k_{10}}{k_{01}}{}_1F_1\left(1+i\frac{k_{01}+\kappa_2}{\omega},1+i\frac{\kappa_2-\kappa_1}{\omega},\frac{ik_{\mathrm{INI}}}{\omega}\right).\end{aligned} \tag{S47}$$

Transforming $\Psi(\omega)$ back to an analytical form of $P(m)$ is challenging. In this paper, we calculate $P(m)$ using the finite state projection method [6,7,9](**SM Section 4** above). **Figure S2(a)** depicts $P(m)$ for several parameter values.

9. Estimating the modality of *P*(*m*)

It has been shown that the modality of a distribution strongly affects the relation between the third standardized moment (or skewness, defined as $\gamma = \bar{\mu}_3/\sigma^3$, where $\bar{\mu}_3$ is the third moment around the mean) and fourth standardized moment (or kurtosis, defined as $\kappa = \bar{\mu}_4/\sigma^4$, where $\bar{\mu}_4$ is the fourth moment around the mean) [36]. An important relation between these two quantities is [37]

$$\kappa - \gamma^2 \geq 1. \tag{S48}$$

Typically, bimodal (or multimodal) distributions result in lower values of $\kappa-\gamma^2$ compared to unimodal distributions. Specifically, $\kappa-\gamma^2=1$ corresponds to the two-point Bernoulli distribution (bimodal), whereas the highest value of $\kappa-\gamma^2$ (infinity) corresponds to a single-peak delta distribution (unimodal). Previous studies have demonstrated the use of $\kappa-\gamma^2$ as a

measure for the modality of distributions [36]. Here, we define the reciprocal of this quantity as the bimodality coefficient β, i.e.

$$\beta = \frac{1}{\kappa - \gamma^2}, \tag{S49}$$

which is bounded between 0 and 1. The β value for a unimodal distribution is closer to 0, whereas that of a bimodal distribution is closer to 1. Using the uniform distribution ($\beta_{\text{th}} = 5/9$) as a critical case, we distinguished bimodal distributions from unimodal ones (**Fig. 2(b)** and **Fig. S2(b)**). Using the $N = 0$ component of $P(m)$, i.e. $P_{N=0} \equiv \int_0^\infty P_{N=0}(m)dm = \mathbf{u} \cdot e^{(\mathbf{K}-\mathbf{K}_{\text{INI}})} \boldsymbol{\Psi}_{\tau=-1}$ (see **Eq. 4** and **SM Section 12**), The unimodal phase can be further divided into two parts according to the position of the peak (**Fig. 2(b)** and **Fig. S2(b)**): If $P_{N=0} > 0.5$, $P(m)$ is interpreted to have the major peak at $m = 0$; otherwise, its major peak is interpreted to be at $m > 0$.

In this work, the calculation of β is based on **Eq. 5**. The analytical expression is derived using Mathematica and the Symbolic Math Toolbox of MATLAB.

10. Experimental details

smFISH experiments were performed as described in [9]. Briefly, fruit fly embryos were collected and chemically fixed ~2 hours after fertilization. Sets of smFISH probes were hybridized against their target RNA in the embryo. Specifically, to measure *hb* transcription (**Fig. 3(a)**), wild type (Oregon-R (OreR)) embryos were labeled using a set of 48 TAMRA-conjugated probes [9]. To compare the effect of different contribution functions (**Fig. 3(b)**), a transgenic fly strain (*bcd3-lacZ*) was used, in which a *lacZ* reporter gene is driven by three high-affinity Bcd binding sites fused to a minimal promoter [9]. A set of 30 TAMRA-conjugated probes and a set of 30 Alexa 647-conjugated probes (corresponding to *lacZ* probes #1-30 and #43-72 used in [9], respectively) were simultaneously applied to label the 5' and 3' portions of the *lacZ* gene. To demonstrate the discontinuity in $P(m)$ (**Fig. 4(c)**), *bcd3-lacZ* embryos were labeled using a set of 72 TAMRA-conjugated probes (from [9]). Following smFISH labeling, embryos were stained

using Hoechst 33342 (to label cell nuclei), mounted on glass slides, and imaged using a confocal microscope.

The nuclear cleavage cycle of each embryo was determined as in [9]. Embryos at cycles 11-13 were used for analysis. RNA signals were quantified using custom MATLAB scripts [9]. Nascent RNA was distinguished from mature RNA as described in [9], using a threshold of $m = 3$.

11. Fitting the experimental distribution of nascent RNA

For each imaged embryo, the anterior-posterior (AP) position (in units of embryo length, EL) of each cell nucleus was measured. In **Fig. 3(a)**, nuclei within the following three representative AP position ranges were studied (corresponding to pink regions in **Fig. 3(a)**): 0.25-0.35 EL, 0.45-0.55 EL, 0.65-0.75 EL. In **Fig. 3(b)**, nuclei within the range 0.2-0.3 EL were studied. In **Fig. 4(d)**, nuclei within the range 0.1-0.3 EL were studied. For the given AP position range, the histogram of nascent RNA (per gene copy) was constructed. To fit the experimental histograms to our model, we applied maximum likelihood estimation (MLE) [3,9]. Briefly, we looked for the parameter set $\mathbf{K} = \{k_{01}, k_{10}, k_{\text{INI}}\}$ that maximizes the likelihood

$$L(M \mid \mathbf{K}) = \prod_i P(m_i \mid \mathbf{K}) \tag{S50}$$

for the set of nascent RNA signals $M = \{m_1, m_2, \cdots\}$. Here, $P(m_i \mid \mathbf{K})$ is the probability of observing the nascent RNA signal m_i given the parameter set $\mathbf{K}$. Its value was calculated using the FSP method (see **SM Section 4**), where the contribution function $g(\tau)$ was constructed based on the locations of smFISH probes on the gene (**Table S1**). The parameter search was done using a combination of simplex and simulated annealing methods (MATLAB function: "simulannealbnd"). To increase accuracy, the search was applied 24 times, and the result with the highest likelihood was chosen.

We note that, according to **Eq. 1**, $P(n,m)$ is a function of $k_{01} \cdot T_{\text{RES}}$, $k_{10} \cdot T_{\text{RES}}$, $k_{\text{INI}} \cdot T_{\text{RES}}$. Hence, fitting the nascent RNA distribution does not provide independent estimates of the kinetic

parameters and of T_{RES}, but rather their products. To determine the actual values of $\{k_{01}, k_{10}, k_{\mathrm{INI}}\}$, we evaluate T_{RES} based on the gene length L and the experimentally reported V_{EL}. Specifically, in the *Drosophila* embryo, V_{EL} was measured to be 1.54 ± 0.14 kbp/min [38]. Consequently, we estimated T_{RES} to be 2.5 minutes for the *hb* gene ($L = 3.6$ kbp) and 2 minutes for the *lacZ* gene ($L = 3.0$ kbp). We set T_{S} to be 0 [10,38].

12. Calculating the discontinuity in *P*(*m*)

We consider the case $T_{\mathrm{RES}} = 1$ and $g = -\tau$. According to **Eq. 4**, the probability of observing m nascent RNAs given N initiation events in the time interval $-T_{\mathrm{RES}} \le \tau \le 0$ is

$$\mathbf{P}_N(m) = \begin{cases} \delta(m) e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})} \boldsymbol{\Psi}_{\tau=-1} & , N = 0 \\ \int_{-1}^{0} d\tau_1 \cdots \int_{-1}^{\tau_{N-1}} d\tau_N \delta\left(m + \sum_{i=1}^{N} \tau_i \right) \mathbf{f}(\tau_1, \cdots, \tau_N) & , N > 0, \end{cases} \tag{S51}$$

where $\mathbf{f}(\tau_1, \cdots, \tau_N) = \mathrm{T}\left[\prod_{i=1}^{N} \mathbf{V}(\tau_i) \right] e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})} \boldsymbol{\Psi}_{\tau=-1}$ is the probability density for a specific time series of initiation events at τ_1, $\cdots$, τ_N. Here, we consider the marginal distribution $P_N(m) = \mathbf{u} \cdot \mathbf{P}_N(m)$. According to **Eq. (S51)**, for $N > 0$, $P_N(m)$ and its derivatives up to the $(N-1)^{\mathrm{th}}$ order are all continuous within the range $0 < m < N$, and $P_N(m) = 0$ for $m > N$.

At $m = 0$, $P_0(0)$ is infinite. Since $P_1(0) = \mathbf{u} \cdot \mathbf{K}_{\mathbf{INI}} e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})} \boldsymbol{\Psi}_{\tau=-1}$ and $P_{N>1}(0) = 0$, $P(0^+) = \sum_{N=1}^{\infty} P_N(0)$ is finite. Therefore, $P(m)$ has an infinite discontinuity at $m = 0$.

At $m = 1$, because $P_1(m = 1^-) = \mathbf{u} \cdot e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})} \mathbf{K}_{\mathbf{INI}} \boldsymbol{\Psi}_{\tau=-1} > 0$ and $P_1(m = 1^+) = 0$, $P(m)$ has a jump discontinuity with magnitude

$$\Delta P_{N=1} = P(m)\big|_{m=1^-} - P(m)\big|_{m=1^+} = P_1(m = 1^-) - P_1(m = 1^+) = \mathbf{u} \cdot e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})} \mathbf{K}_{\mathbf{INI}} \boldsymbol{\Psi}_{\tau=-1}. \tag{S52}$$

Supplemental Material

For $N>1$, to calculate $P_N(m=N^-)$, we define $\bar{F}_N(m)\equiv\int_m^N \mathbf{u}\cdot\mathbf{P}_N(m)dm$ as the complementary cumulative distribution given N initiation events. For m smaller but very close to N, we have

$$\bar{F}_N(m\to N^-)\approx\frac{(N-m)^N}{(N!)^2}\mathbf{u}\cdot\mathbf{f}(-1,\cdots,-1), \tag{S53}$$

and

$$P_N(m\to N^-)=-\frac{d\bar{F}_N(m\to N^-)}{dm}\approx\frac{(N-m)^{N-1}}{N!(N-1)!}\mathbf{u}\cdot\mathbf{f}(-1,\cdots,-1). \tag{S54}$$

Eq. (S54) shows that, when m is approaches N from the left, $P_N(m)$ drops to zero following a power law of order $(N-1)$. This indicates that $P_N(m=N^-)$ and its derivatives up to the $(N-2)$ [th] order are all zero, whereas

$$\left.\frac{d^{N-1}P_N(m)}{dm^{N-1}}\right|_{m=N^-}=\frac{(-1)^{N-1}}{N!}\mathbf{u}\cdot\mathbf{f}(-1,\cdots,-1)=\frac{(-1)^{N-1}}{N!}\mathbf{u}\cdot e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})}\mathbf{K}_{\mathrm{INI}}^N\boldsymbol{\Psi}_{\tau=-1} \tag{S55}$$

is nonzero. Since, $\left.\frac{d^{N-1}P_N(m)}{dm^{N-1}}\right|_{m=N^+}=0$, the $(N-1)$ [th] order derivative of $P(m)$ has a jump discontinuity at $m=N$ with magnitude (**Eq. (9)**)

$$\Delta P_N=\left.\frac{d^{N-1}P(m)}{dm^{N-1}}\right|_{m=N^-}-\left.\frac{d^{N-1}P(m)}{dm^{N-1}}\right|_{m=N^+}=\frac{(-1)^{N-1}}{N!}\mathbf{u}\cdot e^{(\mathbf{K}-\mathbf{K}_{\mathrm{INI}})}\mathbf{K}_{\mathrm{INI}}^N\boldsymbol{\Psi}_{\tau=-1}. \tag{S56}$$

13. Experimental observation of the discontinuity in *P*(*m*)

To experimentally detect the discontinuity in $P(m)$ (**Fig. 4(c)**), *bcd3-lacZ* embryos were labeled using smFISH (see **SM Section 10**). For each imaged embryo, the nascent RNA signal m from the anterior part of the embryo (AP positions 0.1-0.3 EL) was recorded. The histogram of m was fitted to the theoretical model (see **SM Section 11**), to estimate the kinetic parameters k_{01}, k_{10}, k_{INI}.

To visualize the effect of discontinuity in $P(m)$, we needed to focus on small m values, where the magnitude of the jump (ΔP_N) is high. Considering the threshold used to distinguish nascent RNAs from mature ones (see **SM Section 10**), we chose a range $3.5 \le m < 6.5$ for the analysis. The amount of data points in this range is small, due to the low $P(m)$ value (estimated using k_{01}, k_{10}, k_{INI}). To increase the density of data, we defined a new variable $m_0 = m - [m]$ ([·] denotes the nearest integer function) to map m values around different integers into a common range from −0.5 to 0.5. To obtain an accurate histogram of m_0, we collected ~500 data points from 23 embryos. This histogram was then compared with the theoretical prediction using the kinetic parameters (k_{01}, k_{10}, k_{INI}) estimated above. In performing this comparison, we incorporated two additional factors: (1) the finite accuracy in nascent RNA quantification, caused by the imprecise estimation of the intensity of a single RNA [5,9,39], and (2) the probe binding probability p_0 [9,40] (see **SM Section 14.5**). We then applied MLE (see **SM Section 11**), allowing a small fold change f_0 for all m values ($m \mapsto f_0 m$, factor #1) and $p_0 < 1$ (factor #2). We found that the histogram matches the theoretical prediction best for $f_0 = 0.9$ and $p_0 = 0.8$.

14. Possible extensions of the model

Below we describe possible extensions of the model, allowing the incorporation of more complex transcription kinetics and additional experimental details.

14.1. Modeling *N* gene states

The model can be extended to describe a general case of N gene states [6]. In such a case, **Eqs. (1)-(5)** and **Eq. (9)** still hold, but the dimensions of vectors and matrices involved will increase to N. For example, for a three-state model ($N = 3$), we have

$$\mathbf{P}(m) = \begin{bmatrix} P(0,m) \\ P(1,m) \\ P(2,m) \end{bmatrix}, \mathbf{K} = \begin{bmatrix} -k_{01} - k_{02} & k_{10} & k_{20} \\ k_{01} & -k_{10} - k_{12} & k_{21} \\ k_{02} & k_{12} & -k_{20} - k_{21} \end{bmatrix}, \mathbf{K}_{\mathrm{INI}} = \begin{bmatrix} k_{\mathrm{INI},0} & 0 & 0 \\ 0 & k_{\mathrm{INI},2} & 0 \\ 0 & 0 & k_{\mathrm{INI},2} \end{bmatrix}, \quad \text{(S57)}$$

where k_{ij} $(i, j = 0, 1, 2)$ is the transition rate between gene states i and j, and $k_{\mathrm{INI},i}$ is the rate of transcription initiation for gene state i. In a general case, the initial condition $\mathbf{P}_{\tau=-T_{\mathrm{RES}}}(m)$ for the steady-state solution satisfies

$$\begin{cases} \mathbf{K}\mathbf{P}_{\tau=-T_{\mathrm{RES}}}(0) = 0 \\ \mathbf{P}_{\tau=-T_{\mathrm{RES}}}(m > 0) = 0. \end{cases} \tag{S58}$$

14.2. Modeling multi-step transcription initiation

Previous work suggested that, in some instances, transcription initiation may involve multiple (S) steps [8], each described as a Poisson process with rate $k_{\mathrm{I},s}$ ($s = 1, 2, \cdots, S$). In that case, following [8], we can describe the system (originally 2-state) using a special $(S+1)$-state model, where states n = 2, …, S+1 represent different initiation steps. Here, transition rates between states $n = 2, \cdots, S+1$ are replaced by $k_{\mathrm{I},s}$, and $\mathbf{K}_{\mathrm{INI}}$ becomes off-diagonal, indicating the transition of system state from $n = S+1$ to $n = 2$ after initiation. For example, for a two-step initiation model ($S = 2$), we have

$$\mathbf{K} = \begin{bmatrix} -k_{01} & k_{10} & 0 \\ k_{01} & -k_{10} - k_{\mathrm{I},1} & k_{\mathrm{I},2} \\ 0 & k_{\mathrm{I},1} & -k_{\mathrm{I},2} \end{bmatrix}, \mathbf{K}_{\mathrm{INI}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & k_{\mathrm{I},2} \\ 0 & 0 & 0 \end{bmatrix}, \tag{S58}$$

Eqs. (1)-(5) and **Eq. (9)** still hold.

14.3. Modeling the stochastic release of nascent RNA

According to previous studies, the post-elongation dwell time of nascent RNA on the gene may be long compared to the elongation time [25-27], and the release kinetics is not well understood. A reasonable alternative to the deterministic release hypothesis used in the paper is that the RNA release is a Poisson process, with rate k_{R}. To solve the model in this case, we note that the nascent RNA signal is composed of two species: (1) The elongating nascent RNA, which corresponds to initiation events happening at $-T_{\mathrm{EL}} \equiv -L/V_{\mathrm{EL}} \le \tau \le 0$, and is described by **Eq. 1**. (2) The post-elongation nascent RNA, which corresponds to initiation events happening at $\tau < -T_{\mathrm{EL}}$. This species follows the same kinetics as cellular RNA [1-3,5], if we equate the release

rate k_{R} with the RNA degradation rate k_{D}. To include this species in our model, we generalize **Eq. (S1)**, such that $P(n,m)$ now receives contributions from all $\tau \le 0$, i.e.

$$\bar{\mathbf{P}}_{t,\tau=0} = \mathbf{U}(0,-\infty)\bar{\mathbf{P}}^{0}_{t,\tau=-\infty} = \mathbf{U}(0,-T_{\mathrm{EL}})\mathbf{U}(-T_{\mathrm{EL}},-\infty)\bar{\mathbf{P}}^{0}_{t,\tau=-\infty} = \mathbf{U}(0,-T_{\mathrm{EL}})\bar{\mathbf{P}}^{*}_{t,\tau=-T_{\mathrm{EL}}}, \tag{S59}$$

where $\mathbf{U}(\tau_2,\tau_1)$ is the propagator from τ_1 to τ_2. Here, we have divided the entire history $\tau \in (-\infty, 0]$ into the two parts, corresponding to the different nascent RNA species. **Eq.** Error! Reference source not found. shows that the existence of post-elongation nascent RNA modifies the distribution $P(n,m)$ at $\tau = -T_{\mathrm{EL}}$. The new distribution, denoted as $\bar{\mathbf{P}}^{*}_{t,\tau=-T_{\mathrm{EL}}}$, can be solved from the cellular RNA model of [1]. Specifically, at steady-state, we have

$$\begin{cases} P^{*}_{\tau=-T_{\mathrm{EL}}}(0,m) = \dfrac{1}{m!}\dfrac{\dfrac{k_{01}}{k_{\mathrm{R}}}\left(\dfrac{k_{01}}{k_{\mathrm{R}}}\right)^{(m)}}{\left(\dfrac{k_{01}}{k_{\mathrm{R}}}+\dfrac{k_{10}}{k_{\mathrm{R}}}\right)^{(m+1)}}\left(\dfrac{k_{\mathrm{INI}}}{k_{\mathrm{R}}}\right)^{m} {}_1F_1\left(\dfrac{k_{01}}{k_{\mathrm{R}}}+m;\dfrac{k_{01}}{k_{\mathrm{R}}}+\dfrac{k_{10}}{k_{\mathrm{R}}}+m+1;-\dfrac{k_{\mathrm{INI}}}{k_{\mathrm{R}}}\right)\sum_{i=0}^{\infty}\delta(m-i) \\ P^{*}_{\tau=-T_{\mathrm{EL}}}(1,m) = \dfrac{1}{m!}\dfrac{\left(\dfrac{k_{01}}{k_{\mathrm{R}}}\right)^{(m+1)}}{\left(\dfrac{k_{01}}{k_{\mathrm{R}}}+\dfrac{k_{10}}{k_{\mathrm{R}}}\right)^{(m+1)}}\left(\dfrac{k_{\mathrm{INI}}}{k_{\mathrm{R}}}\right)^{m} {}_1F_1\left(\dfrac{k_{01}}{k_{\mathrm{R}}}+m+1;\dfrac{k_{01}}{k_{\mathrm{R}}}+\dfrac{k_{10}}{k_{\mathrm{R}}}+m+1;-\dfrac{k_{\mathrm{INI}}}{k_{\mathrm{R}}}\right)\sum_{i=0}^{\infty}\delta(m-i), \end{cases} \tag{S60}$$

where $x^{(m)} = x(x+1)\cdots(x+m-1)$ denotes the rising factorial, and the sum of delta functions is to ensure that m only takes integer values at $\tau = -T_{\mathrm{EL}}$. The distribution at $\tau = 0$ is solved by substituting **Eq.** Error! Reference source not found. into **Eq.** Error! Reference source not found..

Figure S3 depicts the distributions calculated from **Eq.** Error! Reference source not found. (with $g = 1$), for different parameter values. Stochastic simulations of the same model, also shown, agree with the analytical calculation. We compare the distributions to those calculated for the case of deterministic release with the same average release time, i.e. $T_{\mathrm{S}} = 1/k_{\mathrm{R}}$. As mentioned in the main text, in the limit $(k_{01}\,\&\,k_{10}) \ll k_{\mathrm{INI}}$ and $(k_{01}\,\mathrm{or}\,k_{10}) \ll 1$, the solution for the nascent RNA model with $g = 1$ is identical to that for a cellular RNA model with the same average residence time. Hence, in this limit, the stochastic release can be mapped to a deterministic

release with $T_S = 1/k_R$, i.e. $\bar{\mathbf{P}}^*_{\tau=-T_{EL}} \simeq \mathbf{U}(-T_{EL}, -T_{EL}-T_S)\bar{\mathbf{P}}^0_{t,\tau=-T_{EL}-T_S}$. In **Fig. S3**, the three left columns describe this scenario. Outside this limiting case, the two release kinetics can result in quite different distributions (**Fig. S3**, right column).

14.4. Modeling both nascent and mature RNA

Using smFISH, nascent and mature RNA can be measured simultaneously in the same cell [1,5,41,42]. Studying the joint distribution of the two species can help elucidate the relation between transcription and downstream processes such as RNA degradation and partition. In our model (with deterministic release), the observed nascent RNA m_n all comes from initiation events happening at $-T_{RES} \le \tau \le 0$, whereas the observed mature RNA m_m corresponds to initiation events happening at $\tau < -T_{RES}$. The two species are almost independent, and are only related through the gene state ν at $\tau = -T_{RES}$.

Specifically, m_m and $\nu_{\tau=-T_{RES}}$ are described by the cellular RNA model [1-3,5]. If the probability distribution obtained by solving the cellular RNA model is $P^{\text{mature}}(\nu_{\tau=-T_{RES}}, m_m)$, then for a given pair $(\nu_{\tau=-T_{RES}}, m_m)$, we can use its $\nu_{\tau=-T_{RES}}$ component as an initial condition when solving **Eq. 1** for the corresponding probability distribution of nascent RNA m_n. Since $\nu_{\tau=-T_{RES}}$ can be either 0 or 1, we have two possible initial conditions, i.e. $\bar{\mathbf{P}}^{\nu=0}_{\tau=-T_{RES}} = \delta(m_n)\sum_m P^{\text{mature}}(0,m)\begin{pmatrix}1\\0\end{pmatrix}$ and $\bar{\mathbf{P}}^{\nu=1}_{\tau=-T_{RES}} = \delta(m_n)\sum_m P^{\text{mature}}(1,m)\begin{pmatrix}0\\1\end{pmatrix}$. Following **Eq. (S1)**, we calculate the corresponding probability distribution of (ν, m_n) at $\tau = 0$ for each of these initial conditions as

$$\begin{cases} \bar{\mathbf{P}}^{\nu_{\tau=-T_{RES}}=0}_{\tau=0} = \mathbf{U}(0,-T_{RES})\bar{\mathbf{P}}^{\nu=0}_{\tau=-T_{RES}} \\ \bar{\mathbf{P}}^{\nu_{\tau=-T_{RES}}=1}_{\tau=0} = \mathbf{U}(0,-T_{RES})\bar{\mathbf{P}}^{\nu=1}_{\tau=-T_{RES}} \end{cases}. \tag{S61}$$

Consequently, the joint distribution of the two RNA species is

$$P(n, m_n, m_m) = \sum_{i=0}^{1} P^{\nu_{\tau=-T_{RES}}=i}_{\tau=0}(n, m_n) P^{\text{mature}}(i, m_m). \tag{S62}$$

Supplemental Material

14.5. Modeling the stochastic binding of smFISH probes

In smFISH experiments, the observed RNA signal m is affected by the stochastic binding of fluorescent oligonucleotide probes [9,40]. To include this process in our model, we consider that each RNA molecule is targeted by a set of N_0 fluorescent probes, each directed at a specific RNA sequence (a binding site). With the proper contribution function g, we solve **Eq. 1** and then write the probability distribution of the total number of available binding sites (i_0) on all nascent RNAs present at the gene as

$$\mathbf{P}_{\text{site}}(i_0) = \mathbf{P}(i_0 / N_0) / N_0, \tag{S63}$$

where $\mathbf{P}(\cdot)$ is the solution of **Eq. 1**. Assuming that probe binding follows a binomial distribution with probability p_0, the probability distribution of the number of bound probes (i_{b}) is

$$\mathbf{P}_{\text{bound}}(i_{\text{b}}) = \sum_{i_0=i_{\text{b}}}^{\infty} \binom{i_0}{i_{\text{b}}} p_0^{i_{\text{b}}} (1-p_0)^{i_0-i_{\text{b}}} \mathbf{P}_{\text{site}}(i_0). \tag{S64}$$

In the last step, we convert i_{b} to the observed signal m using the typical signal from a single RNA molecule ($I_1 = p_0 \cdot N_0$), i.e. $m = i_{\text{b}} / I_1$. The probability distribution of m is thus

$$\mathbf{P}_{\text{ob}}(m) = I_1 \mathbf{P}_{\text{bound}}(m \cdot I_1). \tag{S65}$$

14.6. Calculating the joint distribution for multiple smFISH probe sets

In some smFISH experiments, a single gene is labeled using multiple smFISH probe sets, each carrying a different fluorescent dye (see **Fig. 3(b)**). Assuming $p_0 = 1$ and following a similar derivation to that of **Eq. 1**, we can write the master equation for the joint distribution of multiple smFISH signals (m_s, $s = 1, 2, \cdots$):

$$\frac{d\mathbf{P}(m_1, m_2, \cdots)}{d\tau} = (\mathbf{K} - \mathbf{K}_{\text{INI}})\mathbf{P}(m_1, m_2, \cdots) + \mathbf{K}_{\text{INI}}\mathbf{P}(m_1 - g_1(\tau), m_2 - g_2(\tau), \cdots), \tag{S66}$$

with m_s and g_s the pseudo-observable and contribution function for signal s, respectively.

Supplemental Material

In this paper, we examined the joint distribution of two probe sets, targeting the 5' and 3' portions of the *lacZ* reporter gene in transgenic fly embryos (see **Fig. 3(b)** and **SM Section 10**). Since solving **Eq.** Error! Reference source not found. is computationally intensive, we instead performed a stochastic simulation (see **SM Section 4**). The simulated joint distribution agreed well with the experimental one ($R^2 > 0.98$).

Supplemental Material

References:

[1] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).
[2] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. U. S. A. **105**, 17256 (2008).
[3] B. Munsky, G. Neuert, and A. van Oudenaarden, Science **336**, 183 (2012).
[4] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[5] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).
[6] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, Science **339**, 584 (2013).
[7] A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq, Cell. Rep. (2014).
[8] S. Choubey, J. Kondev, and A. Sanchez, PLoS Comp. Biol. **11**, e1004345 (2015).
[9] H. Xu, L. A. Sepulveda, L. Figard, A. M. Sokac, and I. Golding, Nat. Methods **12**, 739 (2015).
[10] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).
[11] K. Adelman, A. La Porta, T. J. Santangelo, J. T. Lis, J. W. Roberts, and M. D. Wang, Proc. Natl. Acad. Sci. U. S. A. **99**, 13538 (2002).
[12] L. Rosenfeld, E. Kepten, S. Yunger, Y. Shav-Tal, and Y. Garini, Phys. Rev. E **92**, 032715 (2015).
[13] N. Korzheva, A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb, and S. A. Darst, Science **289**, 619 (2000).
[14] C. P. Selby, R. Drapkin, D. Reinberg, and A. Sancar, Nucleic Acids Res. **25**, 787 (1997).
[15] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, Nature **446**, 820 (2007).
[16] E. Roldan, A. Lisica, D. Sanchez-Taltavull, and S. W. Grill, Phys. Rev. E **93**, 062411 (2016).
[17] A. Lisica, C. Engel, M. Jahnel, E. Roldan, E. A. Galburt, P. Cramer, and S. W. Grill, Proc. Natl. Acad. Sci. U. S. A. **113**, 2946 (2016).
[18] S. Chong, C. Chen, H. Ge, and X. S. Xie, Cell **158**, 314 (2014).
[19] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo, Phys. Rev. Lett. **117**, 018101 (2016).
[20] S. Klumpp and T. Hwa, Proc. Natl. Acad. Sci. U. S. A. **105**, 18159 (2008).
[21] M. Voliotis, N. Cohen, C. Molina-Paris, and T. B. Liverpool, Biophys. J. **94**, 334 (2008).
[22] D. L. Bentley, Nat. Rev. Genet. **15**, 163 (2014).
[23] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Science **332**, 475 (2011).
[24] R. M. Martin, J. Rino, C. Carvalho, T. Kirchhausen, and M. Carmo-Fonseca, Cell. Rep. **4**, 1144 (2013).
[25] A. Coulon, M. L. Ferguson, V. de Turris, M. Palangat, C. C. Chow, and D. R. Larson, eLife **3** (2014).
[26] N. P. Hoyle and D. Ish-Horowicz, Proc. Natl. Acad. Sci. U. S. A. **110**, E4316 (2013).
[27] J. Singh and R. A. Padgett, Nat. Struct. Mol. Biol. **16**, 1128 (2009).
[28] L. Bai, A. Shundrovsky, and M. D. Wang, J. Mol. Biol. **344**, 335 (2004).
[29] N. G. v. Kampen, *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam ; Boston, 2007).
[30] B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
[31] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).

[32] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[33] L. H. So, A. Ghosh, C. Zong, L. A. Sepulveda, R. Segev, and I. Golding, Nat. Genet. **43**, 554 (2011).
[34] F. W. J. Olver and National Institute of Standards and Technology (U.S.), *NIST handbook of mathematical functions* (Cambridge University Press : NIST, Cambridge ; New York, 2010).
[35] P. Dennery and A. Krzywicki, *Mathematics for physicists* (Dover Publications, Mineola, N.Y., 1996).
[36] T. R. Knapp, J. Mod. App. Stat. Meth. **6**, 3 (2007).
[37] L. R. Shenton and K. O. Bowman, J. Amer. Statist. Assoc. **72**, 206 (1977).
[38] H. G. Garcia, M. Tikhonov, A. Lin, and T. Gregor, Curr. Biol. **23**, 2140 (2013).
[39] S. O. Skinner, L. A. Sepulveda, H. Xu, and I. Golding, Nat. Protoc. **8**, 1100 (2013).
[40] E. Lubeck and L. Cai, Nat. Methods **9**, 743 (2012).
[41] A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, Science **280**, 585 (1998).
[42] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, eLife **5** (2016).

Supplemental Material

TABLE S1. Contribution functions for different smFISH probe sets

smFISH probe set	Contribution function
48 *hb* probes	$g(\tau)=\begin{cases} 1 & , -1<\tau\le -\frac{3}{4} \\ \frac{-8\tau-1}{5} & , -\frac{3}{4}<\tau\le -\frac{1}{8} \\ 0 & , -\frac{1}{8}<\tau\le 0 \end{cases}$
30 *lacZ* probes (5' portion)	$g(\tau)=\begin{cases} 1 & , -1<\tau\le -0.37 \\ \frac{-\tau}{0.37} & , -0.37<\tau\le 0 \end{cases}$
30 *lacZ* probes (3' portion)	$g(\tau)=\begin{cases} \frac{-\tau+0.54}{0.46} & , -1<\tau\le -0.54 \\ 0 & , -0.54<\tau\le 0 \end{cases}$
72 *lacZ* probes	$g(\tau)=-\tau$

Supplemental Material

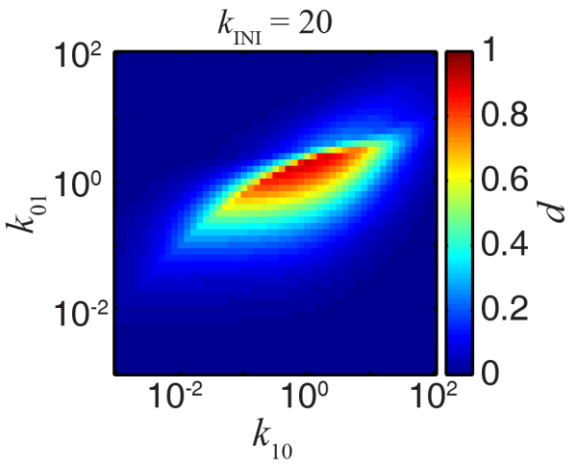

FIG. S1. The difference in P(m) between number of transcribing RNAPs and total cellular RNA.

The relative difference between the theoretical distributions of number of transcribing RNAPs and of total cellular RNA (d, see **SM Section 7**) is plotted as a function of k_{01} and k_{10}, for $k_{INI} = 20$.

Supplemental Material

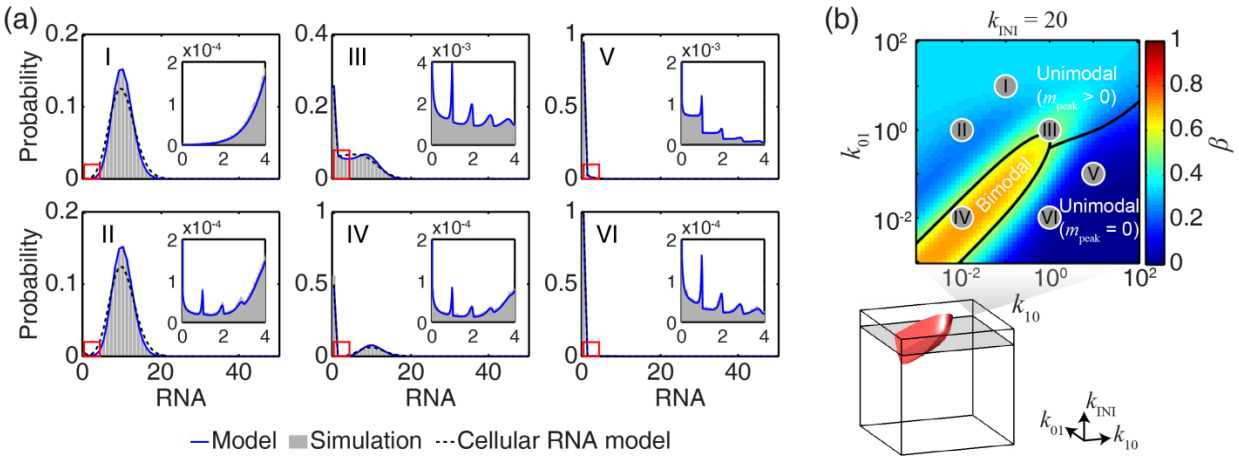

FIG. S2. The probability distribution of nascent RNA at the gene.

(a) The probability distributions $P(m)$ for $g=-\tau$ (blue) for a few parameter values. Also shown are the results of stochastic simulations (gray, bin width = 1), and the solution for a two-state model of cellular RNA kinetics with k_{INI} half of the original magnitude (dashed line). Insets, $P(m)$ at the low range of m (red boxes, bin width = 0.01).

(b) The bimodality coefficient β for $g=-\tau$ as a function of k_{01}, k_{10} and k_{INI} was calculated and thresholded ($\beta_{\mathrm{th}}=5/9$, bottom, red surface) to classify $P(m)$ as either bimodal or unimodal. The unimodal distributions were further classified based on the peak position. Parameter values corresponding to panel **a** are marked as gray circles.

Supplemental Material

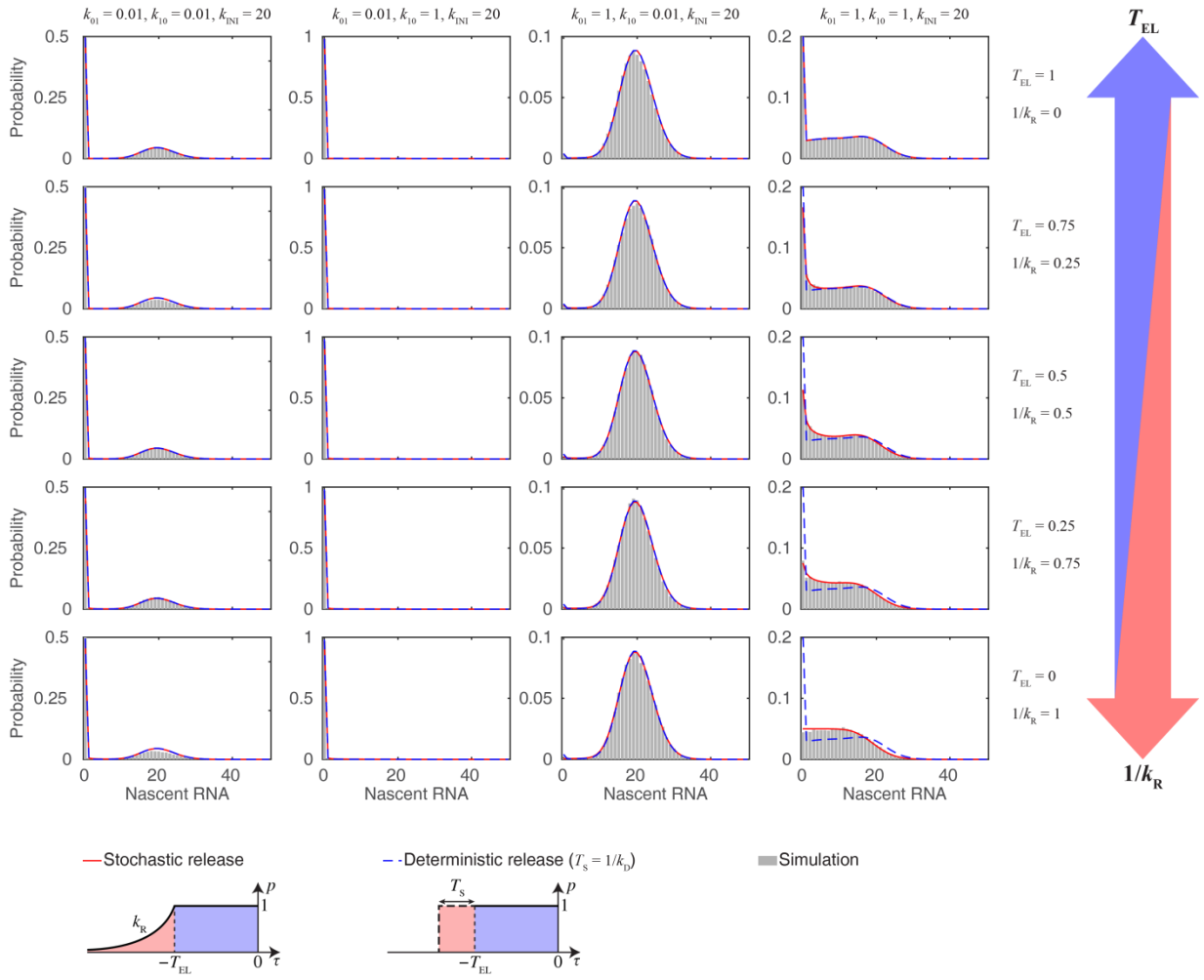

FIG. S3. The probability distribution for the number of RNAPs at the gene in the case of stochastic release.

The solution for stochastic release kinetics (binned to integer values, red, see schematic on the bottom) for different parameter values. Also shown are the results of stochastic simulations (gray), and the solution for deterministic release kinetics with $T_S = 1/k_R$ (dashed blue line, see schematic on the bottom). $T_{EL} + 1/k_R = 1$ in all cases.