Joint distribution of nuclear and cytoplasmic mRNA levels in stochastic models of gene expression: analytical results and parameter inference

Yiling Wang,[1] Juraj Szavits-Nossan,[2] Zhixing Cao,[1,3,*] and Ramon Grima[2,†]

[1] *State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China*

[2] *School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, U. K.*

[3] *Department of Chemical Engineering, Queen's University, Kingston, Canada K7L 3N6*

(Dated: October 15, 2024)

Common stochastic models of gene expression predict the analytical distribution of the total mRNA level per cell but not at subcellular resolution. Here, for a wide class of models of transcription initiation, we obtain an exact steady-state solution for the joint distribution of nuclear and cytoplasmic mRNA levels per cell. Correcting the solution for extrinsic noise and fitting to single human cell data, we precisely quantify the extent of bursty expression in thousands of genes and associate it with their biological functions.

Gene expression is an inherently noisy process [1–4] resulting in significant variation of mRNA and protein numbers between cells. Stochastic models have been studied to shed light on the biophysical processes leading to this variability [5–7] and to fit data and thereby estimate the parameter values specific to each gene [8–11].

The vast majority of models of stochastic gene expression are based on the telegraph model whereby a gene is in one of two states (active and inactive), mRNA is produced from the active state and subsequently decays [12, 13]. The mRNA in this model has often been interpreted as the total mRNA in a cell. However, nowadays mRNA levels can be measured simultaneously in both the nucleus and the cytoplasm [14–17]. Hence, more refined stochastic models of gene expression are needed which can fit this type of data and to also understand how the spatial partitioning of molecular systems may play a role in the reduction of noise [18].

Models involving more than one type of mRNA, i.e. describing all or a subset of nascent, mature nuclear and mature cytoplasmic mRNA levels, have been constructed and solved either numerically [15, 19] or analytically for various special cases of interest [20–27]. The analytical solutions provide expressions for either the moments or the marginal distributions of mRNA levels, except Ref. [20] which additionally provides the joint distribution of nuclear and cytoplasmic mRNA. The model in Ref. [20] is a special case of the telegraph model in which the gene is active for an infinitesimally short time during which a random number of mRNAs (a burst) are produced. This model is problematic because the assumption of instantaneous bursting is debatable for eukaryotic genes since many of these spend many minutes in the active state [28]. As well, the assumption of exponentially distributed off times is in contrast to experimental evidence for some mammalian genes [29].

In this paper, we consider a general stochastic model of gene expression that does not make the aforementioned assumptions and derive the exact joint distribution for mRNA levels measured at two different points of the mRNA lifecycle —these two types of mRNA can be interpreted as nascent and mature nuclear mRNA or unspliced and spliced mRNA or mature nuclear and cytoplasmic mRNA. We verify the accuracy of the solution using Monte Carlo simulations and finally use it to devise an accurate and computationally efficient method for the inference of gene-specific kinetic parameters from commonly available data.

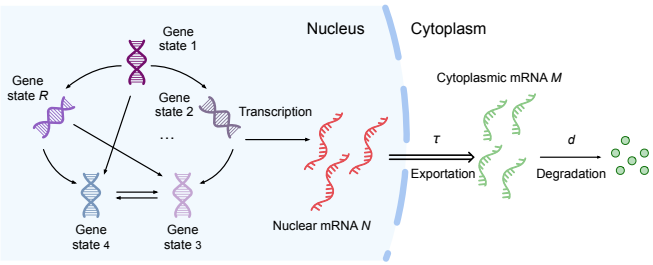

FIG. 1. Illustration of a gene expression model describing initiation (gene switching between a finite number R of inactive and active states), transcription from an active state and export after a finite time τ of mRNA from the nucleus to the cytoplasm where it is finally degraded. This is an example of a wide class of systems in which production of nuclear mRNA is a renewal process; for such systems we derive the steady-state joint distribution of nuclear and cytoplasmic mRNA counts.

General theory. We consider a general stochastic model of nuclear and cytoplasmic mRNA fluctuations in which the production of nuclear mRNA constitutes a renewal process [30], i.e. we assume that the interarrival times of successive nuclear mRNA production events are independent and identically distributed random variables. This assumption is supported by experimental evidence [31]. We denote the probability density function of the interarrival times by $f(t)$ and assume that the mean interarrival time $\alpha = \int_0^\infty t f(t) \mathrm{d}t$ is finite. After a time delay τ, nuclear mRNAs (N) are then exported to the cytoplasm, thus becoming cytoplasmic mRNA (M). These molecules then degrade at a rate d. For now, we assume that τ is deterministic because this makes the analytic computations tractable. In accordance with

previous work [32], we do not model the degradation of nuclear mRNA because bona fide coding transcripts are specifically protected from RNA decay enzymes in the nucleus [33]. The whole model can be summarised by the following reaction scheme,

$$\varnothing \rightsquigarrow N, \quad N \overset{\tau}{\Rightarrow} M, \quad M \overset{d}{\rightarrow} \varnothing, \tag{1}$$

where $\rightsquigarrow$ denotes a reaction that takes a random amount of time whose probability density function is $f(t)$ and $\Rightarrow$ denotes a reaction of fixed duration τ. In Fig. 1 we illustrate a specific example of a gene expression model of the type of Eq. (1).

To compute the steady-state joint distribution of nuclear and cytoplasmic mRNA, we employ queueing theory, whose application to gene expression has been recently reviewed in Ref. [34]. Queueing theory is a mathematical theory that describes customers arriving to a facility where they receive a service and leave [35]. Queueing systems are typically described using Kendall's shorthand notation $A/S/n$, where A denotes the arrival process, S the service process and n is the number of servers [36]. Eq. (1) is mathematically equivalent to a queueing system consisting of two queues connected in a series. The first queue describes the production (arrival) of nuclear mRNA (customers) and its processing (service) into cytoplasmic mRNA. It corresponds to the $G/D/\infty$ queue, where G denotes renewal arrivals with general (arbitrary) interarrival time distribution and D denotes deterministic (fixed-time) service. The output of this queue feeds into the second queue, which describes the degradation (service) of cytoplasmic mRNA. The second queue corresponds to the $G/M/\infty$ queue, where M denotes Markov or memoryless service. Since the nuclear processing is deterministic, both queues have the same interarrival time distribution. We assume that both types of mRNA are available for processing immediately after they are created, hence both queues have infinitely many servers.

Let $N(t)$ and $M(t)$ denote the number of nuclear and cytoplasmic mRNA at time t, respectively. We are interested in the joint distribution $P(n,m)$ of $N(t)=n$ and $M(t)=m$ in the limit of $t\to\infty$. Let $\xi(t)$ denote the residual time from t until the next cytoplasmic mRNA is produced, and $f_m(x)$ the probability density function of $\xi(t)=x$ conditioned on $M(t)=m$ in the stationary limit (i.e. when $t\to\infty$). We note that because nuclear mRNA is processed deterministically, $N(t)$ is equal to the number of nuclear mRNA arrivals between $t-\tau$ and t, of which the first arrival occurs at time $t-\tau+\xi(t)$. For $\xi(t)\le\tau$, let $Y(\tau-\xi(t))$ denote the number of nuclear RNA arrivals from $t-\tau+\xi(t)$ until t, not including the arrival at $t-\tau+\xi(t)$, and let $K_{n-1}(\tau-\xi(t))$ denote the probability of $Y(\tau-\xi(t))=n-1$. From these definitions, it follows that the stationary joint distribution $P(n,m)$ can be written as

$$\frac{P(n,m)}{P_M(m)} = \begin{cases} \int_\tau^\infty f_m(x)\mathrm{d}x, & n=0, \\ \int_0^\tau f_m(x)K_{n-1}(\tau-x)\mathrm{d}x, & n\ge 1, \end{cases} \tag{2}$$

where $P_M(m)$ is the stationary probability of $M(t)=m$. The interpretation of Eq. (2) is the following. The event of $N(t)=n$ and $M(t)=m$ is equivalent of having: (i) $\xi(t)=x$ and $M(t)=m$, whose joint probability is $f_m(x)\mathrm{d}xP_M(m)$, and (ii) $Y(\tau-\xi(t))=n-1$, whose probability is $K_{n-1}(\tau-x)$. The product of these two probabilities is then integrated over all x from 0 until τ. For $n=0$, there are no arrivals between $t-\tau$ and τ, hence we integrate $f_m(x)\mathrm{d}xP_M(m)$ over $x\ge\tau$.

In the context of queueing theory, $P_M(m)$ is the stationary queue-length distribution of the $G/M/\infty$ queue, which has been computed in Ref. [37] for an arbitrary interarrival time distribution with finite mean α. The probability $K_{n-1}(\tau-x)$ is known from renewal theory [30] and reads

$$K_{n-1}(\tau-x) = F_{n-1}(\tau-x) - F_n(\tau-x),$$

where $F_n(t)=\int_0^t f^{*n}(t')\mathrm{d}t'$, and f^{*n} is the n-fold convolution of f. Finally, the probability density function $f_m(x)$ has been computed in Ref. [37] and reads

$$f_m(x) = \frac{1}{\alpha P_M(m)} \sum_{j=m-1}^{\infty} P_j \cdot \binom{j+1}{m} \times \int_0^\infty \mathrm{e}^{-mdy}\left(1-\mathrm{e}^{-dy}\right)^{j+1-m} f(x+y)\mathrm{d}y, \tag{3}$$

where α is the mean interarrival time and P_j is given by

$$P_j = \sum_{r=j}^{\infty}(-1)^{r-j}\binom{r}{j}C_r, \quad j\ge 0.$$

In the above expression, coefficients C_r are defined as

$$C_0 = 1, \quad C_r = \prod_{i=1}^{r}\frac{\phi(id)}{1-\phi(id)}, \quad r\ge 1, \tag{4}$$

and $\phi(s)=\int_0^\infty \mathrm{e}^{-st}f(t)\mathrm{d}t$.

The joint probability $P(n,m)$ is most conveniently computed via the probability generating function (PGF) $G(z_1,z_2)=\sum_{n=0}^\infty\sum_{m=0}^\infty P(n,m)z_1^n z_2^m$. Using Eq. (2), $G(z_1,z_2)$ can be compactly written as

$$G(z_1,z_2) = [G_N(z_1)-1] + G_M(z_2) + \Psi(z_1,z_2), \tag{5}$$

where $G_N(z_1)$ and $G_M(z_2)$ are the probability generating functions of nuclear and cytoplasmic mRNA distributions, respectively. The expression for $G_M(z_2)$ is known explicitly [37],

$$G_M(z_2) = 1 + \sum_{m=1}^{\infty}\frac{C_{m-1}}{\alpha dm}(z_2-1)^m. \tag{6}$$

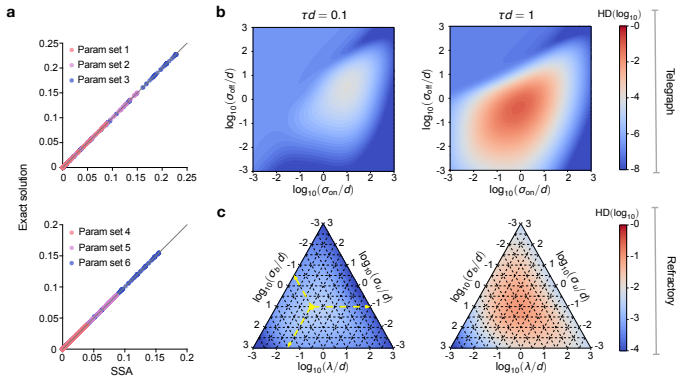

FIG. 2. (a) P-P plots of the joint distributions computed from the solutions derived using queueing theory and the delay SSA. The exact solutions for Model I Eq. (9) (upper) are comprised of Eqs. (10) and (11), while those for Model II Eq. (12) (lower) are given by Eqs. (S55) and (11). The six sets of kinetic parameters are summarised in SM Table S1. (b) Hellinger distance (HD) between the joint distributions of the non-Markovian model (Model I) and its Markovian counterpart (Model III) as a function of the normalised gene state switching rates $\sigma_{\rm on}/d$ and $\sigma_{\rm off}/d$ (or $\sigma_{\rm u}/d$, $\sigma_{\rm b}/d$ and λ/d), and the normalised export time τd (ρ is fixed to 5). (c) same as (b) but comparing Models II and IV. See SM Fig. S1 for the case $\tau d = 10$. Note that the yellow dashed arrows in (c) indicate how to read the ternary plot. In this figure the sum of the three logarithmic variables is equal to -3.

On the other hand, $G_N(z_1)$ [38] and $\Psi(z_1, z_2)$ can be computed by inverting their Laplace transforms with respect to the nuclear processing time τ,

$$\mathcal{L}_\tau\{G_N(z_1)\}(s) = \frac{1}{s} + \frac{(z_1-1)[1-\phi(s)]}{\alpha s^2[1-z_1\phi(s)]}, \tag{7}$$

$$\mathcal{L}_\tau\{\Psi(z_1,z_2)\}(s) = \frac{z_1-1}{\alpha s[1-z_1\phi(s)]} \times \sum_{r=1}^{\infty} C_r \frac{\mathcal{L}_x\{\phi(rd,x)\}(s)}{\phi(rd)}(z_2-1)^r. \tag{8}$$

In the last expression, $\phi(rd,x) = \int_0^\infty \mathrm{e}^{-rdy} f(x+y)\mathrm{d}y$ and $\mathcal{L}_x\{\phi(rd,x)\}(s)$ is the Laplace transform of $\phi(rd,x)$ with respect to x. The derivation of Eqs. (5) and (8) is presented in Supplemental Material (SM) I.

Model I. All results so far have been derived for an arbitrary (general) interarrival time distribution. To specify a particular model of nuclear mRNA production, we need to specify the form of $f(t)$. Typically, nuclear mRNA production is modelled as a Markov process in which the gene switches between a number of discrete states and produces mRNA from one or more of these states [39]. The most widely used biophysical model is the telegraph model [12], whose two compartment extension is given by

$$G \underset{\sigma_{\rm on}}{\overset{\sigma_{\rm off}}{\rightleftharpoons}} G^*, \quad G \xrightarrow{\rho} G+N, \quad N \overset{\tau}{\Rightarrow} M, \quad M \xrightarrow{d} \varnothing, \tag{9}$$

where G^* and G are the inactive and active gene states, respectively. Henceforth Eq. (9) will be referred to as Model I. Applying the general result obtained using queueing theory (Eq. (5)) we obtain an analytical expression for $G(z_1, z_2)$ of Model I (for details see SM II)

$$G = \frac{\nu_+ e^{-\nu_-\tau} - \nu_- e^{-\nu_+\tau}}{d\theta} M\left(\frac{\sigma_{\rm on}}{d}, \frac{\sigma_{\rm on}+\sigma_{\rm off}}{d}, \frac{\rho u_2}{d}\right) + \rho u_1 \frac{e^{-\nu_-\tau} - e^{-\nu_+\tau}}{d\theta} G_{M1}. \tag{10}$$

Here $G_{M1} = \alpha M(1+\sigma_{\rm on}/d, 1+\sigma_{\rm off}/d+\sigma_{\rm on}/d, x_2)$ with $\alpha = \sigma_{\rm on}/(\sigma_{\rm off}+\sigma_{\rm on})$, $x_2 = \rho u_2/d$ and $M(a,b,x)$ being the Kummer function. Additionally, $\nu_\pm = d(r \pm \theta - 1)/2$, $r = 1 - x_1 + (\sigma_{\rm off}+\sigma_{\rm on})/d$, $\theta = [(x_1 - \sigma_{\rm off}/d - \sigma_{\rm on}/d)^2 + 4\sigma_{\rm on}x_1/d]^{\frac{1}{2}}$, $x_1 = \rho u_1/d$ and $u_{1,2} = z_{1,2} - 1$. An alternative (non-queueing theory based) derivation of the solution from the delay master equation can be found in SM III. We note that the generating function marginalized to include only information about nuclear mRNA, i.e. $G(z_1, z_2 = 1)$, has been previously derived [38, 40]; note that in these papers the species N is interpreted as the bound RNA polymerase rather than mature nuclear mRNA. Finally, the joint distribution solution is constructed using the relation

$$P(n,m) = \frac{1}{n!m!}\frac{\partial^n}{\partial u_1^n}\frac{\partial^m}{\partial u_2^m} G\bigg|_{u_1=-1,u_2=-1}, \tag{11}$$

which can be efficiently calculated by a symbolic computation program. We emphasize that unlike current solutions of the joint distribution [20], our solution does not assume that the time spent in state G is infinitesimally short (instantaneous bursts).

In Fig. 2a upper panel, we use P-P plots (probability–probability plots) to show the accuracy of the exact solution given by Eqs. (10) and (11) by comparing it with the joint distributions ($\hat{P}(n,m)$) computed from 10^6 independent runs of the delay stochastic simulation algorithm (delay SSA, the rejection method in Ref. [41]). In this plot, we show all the pairs of conditional cumulative probabilities $\sum_{i=1}^{n} F(i,m)$ ($F \in \{P, \hat{P}\}$) for any possible values of $n, m \in [0, N_{\rm max}]$, where $N_{\rm max}$ is the largest mRNA count recorded in the SSA simulations. All the points fall on the line $y = x$, thereby verifying the accuracy of the analytical solution.

Model II. Another common stochastic transcriptional model is the refractory model, a three-state model in which burst initiation requires two steps [29]. It was developed to account for the experimental observation that for some mammalian genes the distribution of "off" intervals is not exponential but instead peaks at a nonzero value. By modifying it to a two-compartment version, the non-Markovian refractory model (referred to as Model II) consists of the following set of reactions

$$G_1 \xrightarrow{\sigma_{\rm b}} G_2 \xrightarrow{\sigma_{\rm u}} G_3 \xrightarrow{\lambda} G_1, \quad G_1 \xrightarrow{\rho} G_1 + N, \quad N \overset{\tau}{\Rightarrow} M, \quad M \xrightarrow{d} \varnothing. \tag{12}$$

Applying the queueing theory result (Eq. (5)), we find the generating function solution of the stationary joint distribution is given by the sum of four ${}_2F_2$ hypergeometric functions (SM IV). Combining with Eq. (11), the joint distributions of mRNA numbers of Model II can be calculated, which is found to be in excellent agreement with the delay SSA (Fig. 2a lower panel).

Comparison with Markovian models. Alternative competing models to the ones we solved are those where the export reaction ($N \overset{\tau}{\Rightarrow} M$) is modelled by a first-order reaction ($N \xrightarrow{\gamma} M$), i.e. export time is an exponentially distributed random variable and hence the dynamics are Markovian [32, 42]. We will henceforth refer to the Markovian equivalent of Models I and II as Models III and IV, respectively. To meaningfully compare the two types of models, we set $\gamma = 1/\tau$ so that the mean nuclear export time is the same in both. In Fig. 2b we show the Hellinger distance between the steady-state joint distributions of Models I and III whereas in Fig. 2c we show the same but for Models II and IV. Note that the solution of the Markovian models is not known analytically and hence is computed numerically using finite state projection (FSP [43]). We find that the differences between the two types of models are very small when $d\tau$ is small (nuclear export is fast compared to cytoplasmic degradation). This is further substantiated by a theoretical analysis of the first two moments of mRNA fluctuations (SM V). An analysis of experimental data in Ref. [32] shows that about 80% of genes in mammalian cells have $d\tau < 1$ (SM VI and SM Fig. S2) and hence suggests that in most cases, independent of the actual mechanism describing nuclear export, our queueing theory based solutions of Models I and II are accurate.

A PGF-based parameter inference method. In principle, given the exact joint distribution of molecule numbers Eq. (11), one can use the method of maximum likelihood to estimate the transcriptional parameters from measurements of the nuclear and cytoplasmic mRNA in each cell. In practice this is not computationally efficient because this process involves a search over parameter space and for each point one needs to compute a large number of derivatives symbolically. To circumvent this issue, we develop an inference method that uses only the generating function solution. This consists of (i) constructing an empirical generating function from the data and (ii) finding the parameter set which minimizes the distance between the latter function and the analytical generating function. Given measurements of (n_i, m_i) counts of nuclear-cytoplasmic mRNAs for each cell in a population of n_c cells, it can be shown that the corresponding empirical generating function is given by

$$\bar{G}(z_1, z_2) = \sum_{i=1}^{n_c} \frac{1}{n_c} z_1^{n_i} z_2^{m_i}. \tag{13}$$

The inferred transcriptional parameters are then obtained by minimizing the following loss function:

$$J(\phi) = \int_0^1 \int_0^1 |\mathcal{G}(z_1, z_2, \phi) - \bar{G}(z_1, z_2)|^2 \mathrm{d}z_1 \mathrm{d}z_2, \tag{14}$$

where $\mathcal{G}(z_1, z_2, \phi)$ is the analytical generating function of a user-defined gene expression model and $\hat{\phi} = \arg\min_\phi J(\phi)$ is the inferred parameter vector. Further details of the inference method are summarised in SM VII.

To understand the performance (including accuracy and computational efficiency) of the proposed inference method in the real experimental setup, we tested it on simulated data. Specifically, to closely mimic the experimental data collected from mammalian cells, we account for extrinsic noise due to the coupling of transcription to cell volume [44, 45] by choosing the transcription rate ρ to be proportional to the normalized cell volume β (cell volume divided by the average cell volume), i.e., the transcription reaction in the non-Markovian Model I reads $G \xrightarrow{\bar{\rho}\beta} G + N$. This implies that the burst size $(\rho/\sigma_{\text{off}})$ increases with the cell volume, as has been shown by various experimental studies in eukaryotic cells [46, 47]. Furthermore, since eukaryotic cells have multiple copies of each gene, for simplicity we assume that each cell has two independent copies [48]. Our model is a simplification of more complex models of growing and dividing cells [49]. Given a parameter set and a volume distribution, one can then simulate Model I using the delay SSA [41] to generate n_c tuples of nuclear-cytoplasmic mRNA and cell volume observations (n_i, m_i, β_i) which mimics experimental measurements from n_c cells. In Fig. 3a we show a cartoon that illustrates the setup. The model's generating function solution (modified to account for extrinsic noise due to volume and two independent gene copies) is

$$\mathcal{G}(z_1, z_2, \phi) = \int_0^\infty G^2(z_1, z_2, \phi, \beta)\hat{p}(\beta)\mathrm{d}\beta, \tag{15}$$

where the parameter vector is $\phi = \{\bar{\rho}, \sigma_{\text{on}}, \sigma_{\text{off}}, \tau\}$. Note that the distribution of the normalized cell volume $\hat{p}(\beta)$ is estimated by fitting β_i with a kernel density estimation model (KDE) [50]. The inference method modified to account for extrinsic noise then consists of the optimization problem defined by Eq. (14) with $\mathcal{G}(z_1, z_2, \phi)$ given by Eq. (15). For implementation details see SM VIII A.

In Fig. 3b we show by the purple bars the relative errors of this inference method, given as input 15 different datasets obtained by simulation of Model I (SM Table S2) with the normalised cell volume distribution Gamma(7.5,0.11) [51]. The relative errors are small and comparable in size to those obtained by the conventional method of maximizing the likelihood (the joint probability distribution solution of the delay CME) numerically computed using the FSP method (SM VIII B) which are shown by the blue bars. However, the PGF method is remarkably faster, showing a speedup of several thousand

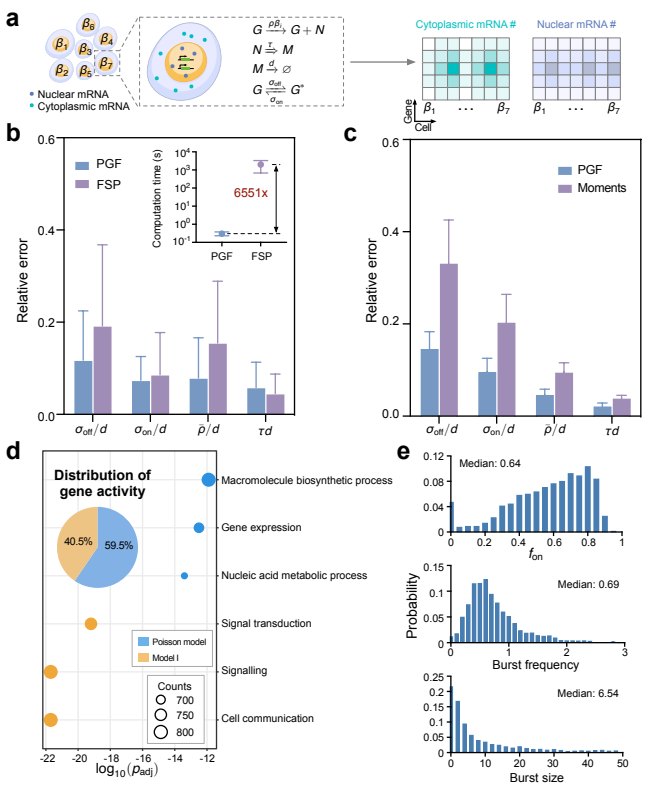

FIG. 3. (a) Illustration of the simulated experimental setup used to test the PGF-based inference method. Each cell's expression is described by an augmented version of Model I with two gene copies and a transcription rate that depends on the normalised cell volume β sampled from a gamma distribution. Delay SSA simulations generate joint count data for 15 batches of 2000 cells each with a different set of kinetic parameters. (b) Use of simulated data to compare the PGF-based inference method with maximum likelihood based inference using FSP shows comparable accuracy but higher computational efficiency: 0.30s (PGF) vs 1991.52s (FSP). The relative error is averaged over the 15 datasets and error bars show the standard error of the mean. (c) The PGF-based inference method is also more accurate than the method of moments. (d) Application of the PGF-based model selection method identifies the fraction of genes best fit by the Poisson model and Model I. This is followed by an ontology analysis which suggests the biological process functions associated with each of the two clusters of genes. (e) Application of the PGF-based inference method to fit the augmented version of Model I to nuclear and cytoplasmic MERFISH data for 2184 mammalian genes; these genes are better fitted by Model I than by the Poisson model and pass a fitting quality test on the moments (see Fig. S5). The inference output is shown by distributions of the fraction of time spent in the active state $f_{\rm on}$, the burst frequency (normalised by the degradation rate) and the burst size across the genome.

times over FSP (inset of Fig. 3b) — this enhanced efficiency is particularly useful when fitting stochastic models to single-cell transcriptomic data for large numbers of genes, especially if the data is at sub-cellular resolution [52]. We also find that the PGF method is typically more accurate than another popular method based on matching the model's first and second moments of mRNA fluctuations to those computed from data (Fig. 3c, SM VIII C and SM Table S3).

Application of the PGF-based parameter inference method to experimental data. We consider single-cell nuclear and cytoplasmic mRNA count data for 6724 mammalian genes measured in 1368 human osteosarcoma cells [14] using multiplexed error-robust fluorescence in situ hybridization (MERFISH)(SM IX). Due to the absence of cell volume data, we used the total number of transcripts detected for each cell as a proxy for the cell volume (this is a reasonably good approximation as discussed in Ref. [53]); this is further divided by its average across the cell population to obtain the values of β_i. Note that while we accounted for extrinsic noise due to volume, we did not account for technical noise; for this dataset the latter is minimal because transcript detection efficiency is high (80%) [14]. Since we did not *a priori* know if Model I is the best model to describe gene expression, we first fit the PGF of 2 competing models of gene expression to the empirical PGF of the data and then utilised 10-fold cross validation to determine the best fitting model (SM X). These 2 models are the Poisson model (constitutive expression; gene is always in a permissive state from which transcription can occur) and Model I (bursty expression; gene can switch from an off state to a permissive state from which transcription can occur). The former model is defined by the reaction schemes:

$$\text{Poisson:}\quad G \xrightarrow{\rho} G + N, \quad N \overset{\tau}{\Longrightarrow} M, \quad M \xrightarrow{d} \varnothing. \qquad (16)$$

As before, extrinsic noise due to coupling of transcription to volume is introduced through ρ for Model I and the Poisson model (SM VIII). We found 4002 and 2722 genes that are best fit by the Poisson model and Model I, respectively. The corresponding fractions are summarised by the piechart in Fig. 3d inset. The high percentage of genes (about 60%) that are better fit by the Poisson model after controlling for extrinsic noise is in line with other recent studies [54–56]. Note that we did not fit Model II because experiments with simulated data showed that for a sample of size of few thousand cells it is not possible to easily tell apart Models I and II (Fig. S4). We further used the Database for Annotation, Visualization and Integrated Discovery (DAVID) [57] to identify biological functions related to the identified gene clusters. In Fig. 3d, we show that macromolecule production and signaling are associated with those genes best fit by the Poisson model and Model I, respectively. In Fig. 3e we show the inferred distributions of the burst size, burst frequency (normalised by the degradation rate) and the fraction of time spent in the active state $f_{\rm on} = \sigma_{\rm on}/(\sigma_{\rm on} + \sigma_{\rm off})$ of the genes that are best fit by Model I and which satisfy a reasonable fitting quality constraint (the average relative error across

the first- and second-order moments of counts is less than 35%; see Fig. S5). The median f_{on} is 0.64, the median burst size is 6.83 and the median burst frequency (normalised by the degradation rate) is 0.69.

In SM Table S5 we repeat the model selection and parameter inference procedure for models with different assumptions and type of input data. Specifically we vary the number of gene copies assumed for Model I (4 instead of 2), vary the type of data (total counts per cell instead of joint count data) and the type of noise (intrinsic noise only instead of both intrinsic and extrinsic noise). We find that if we do not correct for extrinsic noise, the inference procedure leads to a large overestimation of the number of genes that are best fit by Model I ($\approx 90\%$ rather than $\approx 40\%$ estimated earlier). For those genes which are best fit by Model I, the use of total count data rather than joint data leads to a significant underestimation of f_{on} and overestimation of the burst frequency. This makes a strong case for the use of transcriptomic data at sub-cellular resolution (joint data) for accurate parameter inference.

In summary, we have used queueing theory to derive an exact solution of the steady-state joint distribution of nuclear and cytoplasmic mRNA levels in a model of gene expression with general initiation mechanisms involving an arbitrary number of gene states. We have also devised a PGF-based parameter inference and model selection method that fits the exact PGF to the empirical PGF computed from transcriptomic data. Its application to a large experimental dataset led to a simple relationship between the type of gene expression and biological function.

Acknowledgment.— Z.C. acknowledges support from NSFC Grant (62073137), Shanghai Action Plan for Technological Innovation Grants (22ZR1415300, 22511104000, 23S41900500), the Natural Science and Engineering Research Council of Canada's (NSERC's) Discovery Grant (RGPIN-2024-06015) and Shanghai Center of Biomedicine Development. R.G. and J.S.-N. acknowledge support from a Leverhulme Trust grant (RPG-2020-327).

* z.cao@queensu.ca
† ramon.grima@ed.ac.uk

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, Science **297**, 1183 (2002).
[2] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, Noise in eukaryotic gene expression, Nature **422**, 633 (2003).
[3] J. Rodriguez, G. Ren, C. R. Day, K. Zhao, C. C. Chow, and D. R. Larson, Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity, Cell **176**, 213 (2019).
[4] A. Sanchez and I. Golding, Genetic determinants and cellular constraints in noisy gene expression, Science **342**, 1188 (2013).
[5] N. Kumar, T. Platini, and R. V. Kulkarni, Exact distributions for stochastic gene expression models with bursting and feedback, Physical Review Letters **113**, 268105 (2014).
[6] N. Friedman, L. Cai, and X. S. Xie, Linking stochastic dynamics to population distribution: an analytical framework of gene expression, Physical Review Letters **97**, 168302 (2006).
[7] L. Ham, R. D. Brackston, and M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression, Physical Review Letters **124**, 108101 (2020).
[8] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings, Nature Methods **11**, 197 (2014).
[9] A. J. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Genomic encoding of transcriptional burst kinetics, Nature **565**, 251 (2019).
[10] X. Fu, H. P. Patel, S. Coppola, L. Xu, Z. Cao, T. L. Lenstra, and R. Grima, Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions, Elife **11**, e82493 (2022).
[11] Z. Kilic, M. Schweiger, C. Moyer, D. Shepherd, and S. Pressé, Gene expression model inference from snapshot RNA data using bayesian non-parametrics, Nature Computational Science **3**, 174 (2023).
[12] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis, Theoretical Population Biology **48**, 222 (1995).
[13] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS Biology **4**, e309 (2006).
[14] C. Xia, J. Fan, G. Emanuel, J. Hao, and X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression, Proceedings of the National Academy of Sciences **116**, 19490 (2019).
[15] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation, Proceedings of the National Academy of Sciences **115**, 7533 (2018).
[16] G. Li and G. Neuert, Multiplex RNA single molecule FISH of inducible mRNAs in single yeast cells, Scientific Data **6**, 94 (2019).
[17] D. E. Weidemann, J. Holehouse, A. Singh, R. Grima, and S. Hauf, The minimal intrinsic stochasticity of constitutively expressed eukaryotic genes is sub-Poissonian, Science Advances **9**, eadh5138 (2023).
[18] T. Stoeger, N. Battich, and L. Pelkmans, Passive noise filtering by cellular compartmentalization, Cell **164**, 1151 (2016).
[19] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, Single-cell analysis of transcription kinetics across the cell cycle, Elife **5**, e12175 (2016).
[20] A. Singh and P. Bokes, Consequences of mRNA transport on stochastic variability in protein levels, Biophysical Journal **103**, 1087 (2012).
[21] M. Sturrock, S. Li, and V. Shahrezaei, The influence

of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression, Journal of Theoretical Biology **424**, 55 (2017).

[22] G. Gorin and L. Pachter, Modeling bursty transcription and splicing with the chemical master equation, Biophysical Journal **121**, 1056 (2022).

[23] G. Gorin, S. Yoshida, and L. Pachter, Assessing markovian and delay models for single-nucleus RNA sequencing, Bulletin of Mathematical Biology **85**, 114 (2023).

[24] T. Filatova, N. Popovic, and R. Grima, Statistics of nascent and mature RNA fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination, Bulletin of Mathematical Biology **83**, 1 (2021).

[25] T. Filatova, N. Popović, and R. Grima, Modulation of nuclear and cytoplasmic mRNA fluctuations by time-dependent stimuli: Analytical distributions, Mathematical Biosciences **347**, 108828 (2022).

[26] J. Szavits-Nossan and R. Grima, Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle, Physical Review E **105**, 014410 (2022).

[27] C. Shi, X. Yang, J. Zhang, and T. Zhou, Stochastic modeling of the mRNA life process: A generalized master equation, Biophysical Journal **122**, 4023 (2023).

[28] Time spent in the active state can be estimated from data in Table 1 of Ref [58] since the mean active time is the inverse of σ_b.

[29] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science **332**, 472 (2011).

[30] D. Cox, Renewal theory (Methuen, London, 1967).

[31] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Real-time observation of transcription initiation and elongation on an endogenous yeast gene, Science **332**, 475 (2011).

[32] K. B. Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, and S. Itzkovitz, Nuclear retention of mRNA in mammalian tissues, Cell Reports **13**, 2653 (2015).

[33] S. Bresson and D. Tollervey, Surveillance-ready transcription: nuclear RNA decay as a default fate, Open Biology **8**, 170270 (2018).

[34] J. Szavits-Nossan and R. Grima, Solving stochastic gene-expression models using queueing theory: A tutorial review, Biophysical Journal **123**, 1034 (2024).

[35] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, Fundamentals of queueing theory, 4th ed. (Wiley-Interscience, Hoboken, New Jersey, 2008).

[36] D. G. Kendall, Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain, The Annals of Mathematical Statistics **24**, 338 (1953).

[37] L. Takács, On a coincidence problem concerning telephone traffic, Acta Mathematica Academiae Scientiarum Hungarica **9**, 45 (1958).

[38] J. Szavits-Nossan and R. Grima, Steady-state distributions of nascent rna for general initiation mechanisms, Phys. Rev. Res. **5**, 013064 (2023).

[39] In queueing theory, this type of Markov process is known as the Markovian arrival process (MAP) [59]. Since not all MAPs are renewal processes [34], one must be careful to use a renewal MAP, in which case the interarrival time distribution becomes a phase-type distribution [60].

[40] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Stochastic kinetics of nascent RNA, Physical Review Letters **117**, 128101 (2016).

[41] X. Cai, Exact stochastic simulation of coupled chemical reactions with delays, The Journal of Chemical Physics **126** (2007), 10.1063/1.2710253.

[42] M. M. Hansen, R. V. Desai, M. L. Simpson, and L. S. Weinberger, Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability, Cell Systems **7**, 384 (2018).

[43] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, The Journal of Chemical Physics **124** (2006), 10.1063/1.2145882.

[44] C. Jia and R. Grima, Coupling gene expression dynamics to cell size dynamics and cell cycle events: Exact and approximate solutions of the extended telegraph model, Iscience **26** (2023), 10.1016/j.isci.2022.105746.

[45] P. Thomas and V. Shahrezaei, Coordination of gene expression noise with cell size: analytical results for agent-based models of growing cell populations, Journal of the Royal Society Interface **18**, 20210274 (2021).

[46] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms, Molecular Cell **58**, 339 (2015).

[47] S. Berry and L. Pelkmans, Mechanisms of cellular mrna transcript homeostasis, Trends in Cell Biology **32**, 655 (2022).

[48] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells, Science **343**, 193 (2014).

[49] Our modelling approach does not explicitly capture fluctuations in mRNA counts due to the the cell-cycle phase dependence of kinetic parameters, cell-cycle duration variability, doubling of gene copy number during replication and cell division which some other published models account for [44, 58]. A major reason for excluding such biological detail is that these models are time-dependent which makes the exact solution of the joint distribution of nuclear and cytoplasmic mRNA a difficult technical problem. Another reason is that these complex models have a considerably large number of parameters which makes it very challenging to reliably infer them using mRNA count data from a few thousand cells of one type (typical sample size).

[50] Y.-C. Chen, A tutorial on kernel density estimation and recent advances, Biostatistics & Epidemiology **1**, 161 (2017).

[51] A gamma distribution or a similar unimodal distribution (such as a lognormal distribution) is a good approximation to the cell volume distribution when the cell-cycle duration is highly noisy; see Fig. 2A in Ref. [61]. Other papers have used such distributions of parameters to model extrinsic noise on the rates of the telegraph model [7, 62].

[52] M. Zhang, X. Pan, W. Jung, A. R. Halpern, S. W. Eichhorn, Z. Lei, L. Cohen, K. A. Smith, B. Tasic, Z. Yao, et al., Molecularly defined and spatially resolved cell atlas of the whole mouse brain, Nature **624**, 343 (2023).

[53] A. Riba, A. Oravecz, M. Durik, S. Jiménez, V. Alunni, M. Cerciat, M. Jung, C. Keime, W. M. Keyes, and N. Molina, Cell cycle gene regulation dynamics revealed by rna velocity and deep-learning, Nature Communications **13**, 2865 (2022).

[54] R. Foreman and R. Wollman, Mammalian gene expression variability is explained by underlying cell state, Molecular Systems Biology **16**, e9146 (2020).

[55] R. Ietswaart, S. Rosa, Z. Wu, C. Dean, and M. Howard, Cell-size-dependent transcription of flc and its antisense long non-coding rna coolair explain cell-to-cell expression variation, Cell Systems **4**, 622 (2017).

[56] N. Battich, T. Stoeger, and L. Pelkmans, Control of transcript variability in single mammalian cells, Cell **163**, 1596 (2015).

[57] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, David: database for annotation, visualization, and integrated discovery, Genome Biology **4**, 1 (2003).

[58] Z. Cao and R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, Proceedings of the National Academy of Sciences **117**, 4682 (2020).

[59] M. F. Neuts, A versatile markovian point process, Journal of Applied Probability **16**, 764–779 (1979).

[60] C. Commault and J.-P. Chemla, An invariant of representations of phase-type distributions and some applications, Journal of Applied Probability **33**, 368 (1996).

[61] C. Jia, A. Singh, and R. Grima, Cell size distribution of lineage data: analytic results and parameter inference, Iscience **24** (2021), 10.1016/j.isci.2021.102220.

[62] R. Grima and P.-M. Esmenjaud, Quantifying and correcting bias in transcriptional parameter inference from single-cell data, Biophysical Journal **123**, 4 (2024).

[63] W. R. Inc., Mathematica, Version 14.1 (2024), Champaign, IL.

[64] A. Leier and T. T. Marquez-Lago, Delay chemical master equation: direct and closed-form solutions, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471**, 20150049 (2015).

[65] Q. Jiang, X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima, Neural network aided approximation and parameter inference of non-Markovian models of gene expression, Nature Communications **12**, 2618 (2021).

[66] N. G. Van Kampen, Stochastic processes in physics and chemistry, Vol. 1 (Elsevier, 1992).

[67] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Stochasticity of gene products from transcriptional pulsing, Physical Review E **79**, 031911 (2009).

[68] NIST digital library of mathematical functions, `https://dlmf.nist.gov/`, Release 1.1.10 of 2023-06-15.

[69] W. R. Inc., HypergeometricPFQ (2024), Champaign, IL.

Supplemental Material

Contents

I. Derivation of the joint probability generating function for an arbitrary (general) interarrival time distribution 2

II. Derivation of Eq. (10) using queueing theory 3

III. Derivation of Eq. (10) using the delay master equation 4
- A. Derivation of the delay master equations 4
- B. Solution conditional on the gene state 7
- C. Solution independent of the gene state 9

IV. Stochastic description of reaction system Eq. (12) using queueing theory 9

V. The first and second-order moments of Models I and III converge under fast export conditions 12

VI. Estimation of the ratio of export to degradation times from experimental data 13

VII. Parameter inference using the generating function solution for synthetic data without extrinsic noise and for one gene copy 13

VIII. Parameter inference for data with extrinsic noise and multiple gene copies 14
- A. Generating-function-based parameter inference 14
- B. Finite-state-projection-based parameter inference 15
- C. Method-of-moments-based parameter inference 16

IX. Processing of MERFISH data 17

X. Generating-function-based model selection using 10-fold cross validation and ontology analysis 17

XI. Code and data availability 18

I. Derivation of the joint probability generating function for an arbitrary (general) interarrival time distribution

To compute $G(z_1, z_2)$, we multiply $f_m(x)$ in Eq. (3) by $P_M(m)z_2^m$ and sum it over all $m \geq 0$, which gives

$$\sum_{m=0}^{\infty} f_m(x)P_M(m)z_2^m = \frac{1}{\alpha}\int_0^{\infty} \left[1 + (z_2 - 1)\mathrm{e}^{-\beta y}\right] U\left(1 + (z_2 - 1)\mathrm{e}^{-\beta y}\right) f(x + y)\mathrm{d}y,$$

where $U(z)$ is defined as

$$U(z) = \sum_{j=0}^{\infty} P_j z^j.$$

Using the following result derived in Ref. [37],

$$U(z) = \sum_{r=0}^{\infty} C_r (z - 1)^r,$$

we get that

$$\sum_{m=0}^{\infty} f_m(x)P_M(m)z_2^m = f_\infty(x) + \sum_{r=1}^{\infty} \frac{C_{r-1} + C_r}{\alpha}\left[\int_0^{\infty} \mathrm{e}^{-rdy} f(x + y)\mathrm{d}y\right](z_2 - 1)^r, \tag{S1}$$

where $f_\infty(t)$ is the probability density function of the forward recurrence time (the time until the next arrival measured from an arbitrary time point) in the stationary limit [30],

$$f_\infty(t) = \frac{1 - F(t)}{\alpha}, \tag{S2}$$

and $F(t)$ is the cummulative probability function of the interarrival time, $F(t) = \int_0^t f(t')\mathrm{d}t'$.

Next, we multiply $K_{n-1}(\tau - x)$ by z_1^n and sum it over all $n \geq 1$, which gives

$$\sum_{n=1}^{\infty} K_{n-1}(\tau - x)z_1^n = 1 + (z_1 - 1)\sum_{n=0}^{\infty} F_n(\tau - x)z_1^n, \tag{S3}$$

where

$$F_0(t) = 1, \quad F_n(t) = \int_0^t f^{*n}(t')\mathrm{d}t',\ n \geq 1,$$

and f^{*n} is the n-fold convolution of f. Let us denote by $P_N(n)$ the probability distribution of nuclear RNA in the stationary limit. According to Ref. [38], $P_N(n)$ is given by

$$P_N(n) = \begin{cases} \int_\tau^\infty f_\infty(x)\mathrm{d}x, & n = 0, \\ \int_0^\tau f_\infty(x)K_{n-1}(\tau - x)\mathrm{d}x, & n \geq 1. \end{cases} \tag{S4}$$

If we combine Eqs. (S1)-(S4), we get

$$G(z_1, z_2) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} P(n, m)z_1^n z_2^m = [G_N(z_1) - 1] + G_M(z_2) + \Psi(z_1, z_2),$$

where $G_N(z_1)$ is the stationary probability generating function of nuclear mRNA number,

$$G_N(z_1) = \sum_{n=0}^{\infty} P_N(z_1)z_1^n,$$

and $\Psi(z_1, z_2)$ is defined as

$$\Psi(z_1, z_2) = (z_1 - 1) \sum_{n=0}^{\infty} \sum_{r=1}^{\infty} \frac{C_{r-1} + C_r}{\alpha} \left[\int_0^{\tau} \mathrm{d}x \int_0^{\infty} \mathrm{d}y \, \mathrm{e}^{-rdy} f(x+y) F_n(\tau - x) \right] z_1^n (z_2 - 1)^r.$$

Let us denote by $\phi(rd, x)$ the following integral,

$$\phi(rd, x) = \int_0^{\infty} \mathrm{e}^{-rdy} f(x+y) \mathrm{d}y.$$

The expression for $\Psi(z_1, z_2)$ can now be written more compactly as

$$\Psi(z_1, z_2) = \frac{(z_1 - 1)}{\alpha} \sum_{n=0}^{\infty} \sum_{r=1}^{\infty} (C_{r-1} + C_r) \left[\int_0^{\tau} \mathrm{d}dx \, \phi(rd, x) F_n(\tau - x) \right] z_1^n (z_2 - 1)^r.$$

By taking the Laplace transform of $\Psi(z_1, z_2)$ with respect to τ we get

$$\begin{aligned} \mathcal{L}_\tau\{\Psi(z_1, z_2)\}(s) &= \frac{z_1 - 1}{\alpha s[1 - z_1 \phi(s)]} \sum_{r=1}^{\infty} (C_{r-1} + C_r) \mathcal{L}_x\{\phi(rd, x)\}(s)(z_2 - 1)^r, \\ &= \frac{z_1 - 1}{\alpha s[1 - z_1 \phi(s)]} \sum_{r=1}^{\infty} C_r \frac{\mathcal{L}_x\{\phi(rd, x)\}(s)}{\phi(rd)} (z_2 - 1)^r, \end{aligned}$$

where we used that $C_{r-1} + C_r = C_r/\phi(rd)$ [37] and $\mathcal{L}_x\{\phi(rd, x)\}(s)$ is the Laplace transform of $\phi(rd, x)$ with respect to x,

$$\mathcal{L}_x\{\phi(rd, x)\}(s) = \int_0^{\infty} \mathrm{d}x \mathrm{e}^{-sx} \phi(rd, x) = \int_0^{\infty} \mathrm{d}x \mathrm{e}^{-sx} \int_0^{\infty} dy \, \mathrm{e}^{-rdy} f(x+y).$$

II. Derivation of Eq. (10) using queueing theory

The probability density of the time between two nuclear mRNA production events for the reaction system Eq. (9) is a phase-type distribution of the form

$$f(t) = \kappa \mathrm{e}^{-D_0 t}(-D_0 \mathbf{1}^\top),$$

with $\kappa = [0, \ 1]^\top$, $\mathbf{1} = [1, \ 1]^\top$ and

$$D_0 = \begin{bmatrix} -\sigma_{\text{on}} & \sigma_{\text{on}} \\ \sigma_{\text{off}} & -\sigma_{\text{off}} - \rho \end{bmatrix}.$$

The corresponding Laplace transform of $f(t)$ is

$$\phi(s) = \kappa(sI - D_0)^{-1}(-D_0 \mathbf{1}^\top) = \frac{\rho(s + \sigma_{\text{on}})}{(s + \sigma_{\text{off}} + \rho)(s + \sigma_{\text{on}}) - \sigma_{\text{on}}\sigma_{\text{off}}}. \tag{S5}$$

The mean interarrival time α follows from $\alpha = -\phi'(0)$ and hence is given by

$$\alpha = \frac{\sigma_{\text{on}} + \sigma_{\text{off}}}{\sigma_{\text{on}} \rho}.$$

By taking the inverse Laplace transform of Eq. (7), we obtain the probability generating function of nuclear mRNA

$$G_N(z_1) = \frac{\nu_+ + x_1 d\sigma_{\text{on}}/(\sigma_{\text{off}} + \sigma_{\text{on}})}{d\theta} \mathrm{e}^{-\nu_- \tau} - \frac{\nu_- + x_1 d\sigma_{\text{on}}/(\sigma_{\text{off}} + \sigma_{\text{on}})}{d\theta} \mathrm{e}^{-\nu_+ \tau}, \tag{S6}$$

with $\nu_\pm = d(r \pm \theta - 1)/2$, $\theta = [(x_1 - \sigma_{\text{off}}/d - \sigma_{\text{on}}/d)^2 + 4\sigma_{\text{on}} x_1/d]^{\frac{1}{2}}$, $x_1 = \rho u_1/d$ and $u_1 = z_1 - 1$.

It follows from Eqs. (4) and (S5) that

$$C_r = \frac{(\rho/d)^r (1+\sigma_{\rm on}/d)_r}{r!(1+\sigma_{\rm off}/d+\sigma_{\rm on}/d)_r},$$

which together with Eq. (6) leads to the probability generating function of cytoplasmic mRNA

$$G_M(z_2) = M(\sigma_{\rm on}/d, \sigma_{\rm off}/d+\sigma_{\rm on}/d, \rho u_2/d), \tag{S7}$$

where $u_2 = z_2 - 1$ and M is Kummer's confluent hypergeometric function.

To calculate $G(z_1, z_2)$ using Eq. (5), we need to calculate $\Psi(z_1, z_2)$. Using

$$\mathcal{L}_x\{\phi(rd, x)\} = \frac{\rho[dr(s+\sigma_{\rm on}) + \sigma_{\rm on}(s+\sigma_{\rm on}+\sigma_{\rm off})]}{[dr(dr+\rho+\sigma_{\rm off}) + (dr+\rho)\sigma_{\rm on}][s(s+\rho+\sigma_{\rm off}) + (s+\rho)\sigma_{\rm on}]}, \tag{S8}$$

we find that the Laplace transform of $\Psi(z_1, z_2)$ (as given by Eq. (8)) is given by

$$\begin{aligned}\mathcal{L}_\tau\{\Psi(z_1,z_2)\}(s) =& \frac{\rho\sigma_{\rm on}(z_1-1)}{s(\sigma_{\rm off}+\sigma_{\rm on})(d+\sigma_{\rm off}+\sigma_{\rm on})(s(s+\rho-z_1\rho+\sigma_{\rm off}) + (s+\rho-z_1\rho)\sigma_{\rm on})} \\ &\times \Big[(d+\sigma_{\rm on}+\sigma_{\rm off})(s+\sigma_{\rm on}+\sigma_{\rm off})(M(\sigma_{\rm on}/d, 1+\sigma_{\rm on}/d+\sigma_{\rm off}/d, \rho u_2/d) - 1) \\ &+ \rho(s+\sigma_{\rm on})(z_2-1)M(1+\sigma_{\rm on}/d, 2+\sigma_{\rm on}/d+\sigma_{\rm off}/d, \rho u_2/d)\Big].\end{aligned} \tag{S9}$$

Substituting the inverse Laplace transform of Eq. (S9), and Eqs. (S6) and (S7) in Eq. (5), we find after simplification using `Mathematica` [63] that the solution for $G(z_1, z_2)$ is given by Eq. (10).

III. Derivation of Eq. (10) using the delay master equation

A. Derivation of the delay master equations

Consider the system of reactions given by Eq. (9). To derive the delay master equations describing the stochastic dynamics, we want to obtain the probability $P(\varphi, n, m, t+\Delta t)$, i.e., the probability of observing n nuclear mRNAs and m cytoplasmic mRNAs at time $t+\Delta t$ given that the gene state is φ. Note that the inactive and active gene states are denoted by $\varphi = 0$ and $\varphi = 1$, respectively. There are four sets of events which contribute to this probability in the time interval $(t, t+\Delta t)$:

(A) a nuclear mRNA transcript is produced while the gene is active;

(B) either the gene state changes or a cytoplasmic mRNA is degraded;

(C) a nuclear mRNA transcript is exported to the cytoplasm and becomes a cytoplasmic mRNA;

(D) no reaction occurs.

In the infinitesimal limit $\Delta t \to 0$, the probability of event (A) simply follows the conventional propensity associated with a firing of the transcription reaction:

$$P(1, n, m, t+\Delta t; 1, n-1, m, t) = \rho\Delta t P(1, n-1, m, t), \tag{S10}$$

where $P(\varphi, n, m, t; \varphi', n', m', t')$ is the joint probability of having gene state φ, n nuclear mRNAs and m cytoplasmic mRNAs at time t and gene state φ', n' nuclear mRNAs and m' cytoplasmic mRNAs at time t'.

Similarly, the joint probabilities of the reactions associated with event (B) are

$$\begin{aligned}P(0, n, m, t+\Delta t; 1, n, m, t) =& \sigma_{\rm off}\Delta t P(1, n, m, t), \\ P(1, n, m, t+\Delta t; 0, n, m, t) =& \sigma_{\rm on}\Delta t P(0, n, m, t), \\ P(\varphi, n, m, t+\Delta t; \varphi, n, m+1, t) =& d(m+1)P(\varphi, n, m+1, t).\end{aligned} \tag{S11}$$

The probability of event (C) occurring requires a careful consideration of the history of the process: (i) $(1, n', m', t) \to (1, n'+1, m', t-\tau+\Delta t)$; (ii) $(1, n'+1, m', t-\tau+\Delta t) \to (\varphi, n+1, m-1, t)$; (iii) $(\varphi, n+1, m-1, t) \to (\varphi, n, m, t+\Delta t)$, where n' and m' are the number of nuclear and cytoplasmic RNAs at time $t-\tau$. The elementary probability of event (C) can hence be decomposed into

$$\begin{aligned}
&P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) \\
=&P(1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part I)} \\
&\times P(\varphi, n+1, m-1, t | 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part II)} \\
&\times P(\varphi, n, m, t+\Delta t | \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part III)}
\end{aligned} \tag{S12}$$

The probability of Part III is 1 since every production event is followed by an export event a time τ later, i.e.,

$$P(\text{Part III}) = 1. \tag{S13}$$

According to Eq. (S10), the probability of Part I is

$$P(\text{Part I}) = \rho \Delta t P(1, n', m', t-\tau). \tag{S14}$$

The two "+1"s of nuclear mRNAs in the probability of Part II represent a nuclear mRNA was born in time $(t-\tau, t-\tau+\Delta t)$ and stays in the system till time t, and notably does not participate in any reaction in time $(t-\tau+\Delta t, t)$. Thus, we have

$$P(\varphi, n+1, m-1, t | 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) = P(\varphi, n, m-1, t | 1, n', m', t-\tau+\Delta t). \tag{S15}$$

Since all nuclear mRNAs born before time $t-\tau$ (of which the number is n') are exported from nucleus to cytoplasm, the number of nuclear mRNAs (n) at time t is independent of n' and m', but only depends on the gene state at time $t-\tau$. Besides, the number of cytoplasmic mRNAs ($m-1$) at time t is determined by the number of nuclear mRNAs out of n' exported to cytoplasm and the number of cytoplasm RNA out of m' degraded in the time interval $(t-\tau+\Delta t, t)$. As such, we have

$$P(\varphi, n, m-1, t | 1, n', m', t-\tau+\Delta t) = P(\varphi, n, \bullet, t | 1, 0, 0, t-\tau+\Delta t) P(\bullet, \bullet, m-1, t | 1, n', m', t-\tau+\Delta t), \tag{S16}$$

where the symbol $\bullet$ stands for the marginalization over the variable in the same position. For instance,

$$P(\varphi, n, \bullet, t | 1, 0, 0, t-\tau+\Delta t) = \sum_m P(\varphi, n, m, t | 1, 0, 0, t-\tau+\Delta t). \tag{S17}$$

It follows that by means Eqs. (S12)-(S17) that the total probability of event (C) is given by

$$\begin{aligned}
&P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t) \\
=&\sum_{n', m'} P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) \\
=&\rho \Delta t P(\varphi, n, \bullet, t | 1, 0, 0, t-\tau+\Delta t) \sum_{n', m'} P(\bullet, \bullet, m-1, t | 1, n', m', t-\tau+\Delta t) P(1, n', m', t-\tau).
\end{aligned} \tag{S18}$$

Note that here we used the law of total probability. See Fig. S3 for an illustration of the history of event (C).

The probability that event (D) occurs is obtained by subtracting all the probabilities of one reaction occurring from $P(\varphi, n, m, t+\Delta t)$. Specifically, we have

$$\begin{aligned}
P(\varphi, n, m, t+\Delta t; \varphi, n, m, t) =& P(\varphi, n, m, t) - \delta_\varphi P(\varphi, n+1, m, t+\Delta t; \varphi, n, m, t) - P(1-\varphi, n, m, t+\Delta t; \varphi, n, m, t) \\
&- P(\varphi, n-1, m+1, t+\Delta t; \varphi, n, m, t) - P(\varphi, n, m+1, t+\Delta t; \varphi, n, m, t),
\end{aligned} \tag{S19}$$

where δ_φ is Kronecker delta, and $\delta_\varphi = 1$ if and only if $\varphi = 1$.

By means of Eqs. (S10), (S11), (S18) and (S19), it can be shown that

$$\begin{cases} P(0,n,m,t+\Delta t) =\sigma_{\text{off}}\Delta t P(1,n,m,t) + d(m+1)\Delta t P(0,n,m+1,t) \\ \qquad + \rho\Delta t P(0,n,\bullet,t|1,0,0,t-\tau+\Delta t)\sum_{n',m'} P(\bullet,\bullet,m-1,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau) \\ \qquad + [P(0,n,m,t) - \sigma_{\text{on}}\Delta t P(0,n,m,t) - dm\Delta t P(0,n,m,t) \\ \qquad - \rho\Delta t P(0,n-1,\bullet,t|1,0,0,t-\tau+\Delta t)\sum_{n',m'} P(\bullet,\bullet,m,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau)], \\ P(1,n,m,t+\Delta t) =\sigma_{\text{on}}\Delta t P(0,n,m,t) + \rho\Delta t P(1,n-1,m,t) + d(m+1)\Delta t P(1,n,m+1,t) \\ \qquad + \rho\Delta t P(1,n,\bullet,t|1,0,0,t-\tau+\Delta t)\sum_{n',m'} P(\bullet,\bullet,m-1,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau) \\ \qquad + [P(1,n,m,t) - \sigma_{\text{off}}\Delta t P(1,n,m,t) - \rho\Delta t P(1,n,m,t) - dm\Delta t P(1,n,m,t) \\ \qquad - \rho\Delta t P(1,n-1,\bullet,t|1,0,0,t-\tau+\Delta t)\sum_{n',m'} P(\bullet,\bullet,m,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau)]. \end{cases} \tag{S20}$$

Dividing both sides of Eq. (S20) by Δt and letting $\Delta t \to 0$, the delay master equation of the delayed RNA export model is obtained:

$$\begin{cases} \dfrac{\mathrm{d}P(0,n,m,t)}{\mathrm{d}t} =\sigma_{\text{off}}P(1,n,m,t) - \sigma_{\text{on}}P(0,n,m,t) + d(m+1)P(0,n,m+1,t) - dmP(0,n,m,t) \\ \qquad + \rho[P(0,n,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau) - P(0,n-1,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m,t;1,\bullet,\bullet,t-\tau)], \\ \dfrac{\mathrm{d}P(1,n,m,t)}{\mathrm{d}t} =\sigma_{\text{on}}P(0,n,m,t) - \sigma_{\text{off}}P(1,n,m,t) + d(m+1)P(1,n,m+1,t) - dmP(1,n,m,t) + \rho P(1,n-1,m,t) - \rho P(1,n,m,t) \\ \qquad + \rho[P(1,n,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau) - P(1,n-1,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m,t;1,\bullet,\bullet,t-\tau)], \end{cases} \tag{S21}$$

in which the following simplification step is used

$$\sum_{n',m'} P(\bullet,\bullet,m-1,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau) = P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau).$$

Next we seek solutions to the two probabilities, i.e., $P(\varphi,n,\bullet,t|1,0,0,t-\tau)$ and $P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau)$. The n nuclear mRNAs of the two conditional probabilities $P(\varphi,n,\bullet,t|1,0,0,t-\tau)$ for $\varphi = 0,1$ in Eq. (S21) are produced during time $(t-\tau,t)$, and the pertinent dynamics are that of a reaction system composed of the first three reactions in Eq. (9), i.e., $G \rightleftharpoons G^*$ and $G \to G + N$. Specifically, we have

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\bar{P}_0(n,t)}{\mathrm{d}t} = -\sigma_{\text{on}}\bar{P}_0(n,t) + \sigma_{\text{off}}\bar{P}_1(n,t), \\ \dfrac{\mathrm{d}\bar{P}_1(n,t)}{\mathrm{d}t} = \sigma_{\text{on}}\bar{P}_0(n,t) - \sigma_{\text{off}}\bar{P}_1(n,t) + \rho[\bar{P}_1(n-1,t) - \bar{P}_1(n,t)], \end{cases}} \tag{S22}$$

with initial conditions $\bar{P}_0(n,0) = 0$ for any n, $\bar{P}_1(n,0) = 1$ for $n = 0$ and equal to 0 otherwise, as well as $P(\varphi,n,\bullet,t|1,0,0,t-\tau) = \bar{P}_\varphi(n,\tau)$ for $\varphi = 0,1$.

By summing over all possible n on both sides of Eq. (S21), we obtain

$$\begin{cases} \dfrac{\mathrm{d}P(0,m,t)}{\mathrm{d}t} =\sigma_{\text{off}}P(1,m,t) - \sigma_{\text{on}}P(0,m,t) + d(m+1)P(0,m+1,t) - dmP(0,m,t) \\ \qquad + \rho[P(m-1,t;1,\bullet,\bullet,t-\tau) - P(m,t;1,\bullet,\bullet,t-\tau)] \times P(0,\bullet,\bullet,t|1,0,0,t-\tau), \\ \dfrac{\mathrm{d}P(1,m,t)}{\mathrm{d}t} =\sigma_{\text{on}}P(0,m,t) - \sigma_{\text{off}}P(1,m,t) + d(m+1)P(1,m+1,t) - dmP(1,m,t) \\ \qquad + \rho[P(m-1,t;1,\bullet,\bullet,t-\tau) - P(m,t;1,\bullet,\bullet,t-\tau)] \times P(1,\bullet,\bullet,t|1,0,0,t-\tau). \end{cases} \tag{S23}$$

Summing up the two equations in Eq. (S23) leads to

$$\frac{\mathrm{d}P(m,t)}{\mathrm{d}t} = d(m+1)P(m+1,t) - dmP(m,t) + \rho[P(m-1,t;1,\bullet,\bullet,t-\tau) - P(m,t;1,\bullet,\bullet,t-\tau)], \tag{S24}$$

where $P(m,t) = \sum_\varphi P_\varphi(m,t)$. Given the nature of the delayed export, the dynamics of cytoplasmic mRNAs is equivalent to that of a conventional telegraph model a τ time earlier. The telegraph model encompasses the following set of reactions

$$G \underset{\sigma_{\text{on}}}{\overset{\sigma_{\text{off}}}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{\rho} G + M, \qquad M \xrightarrow{d} \varnothing,$$

of which the governing master equations are

$$\boxed{\begin{cases} \dfrac{\mathrm{d}Q_0(m,t)}{\mathrm{d}t} = \sigma_{\text{off}} Q_1(m,t) - \sigma_{\text{on}} Q_0(m,t) + d(m+1)Q_0(m+1,t) - dmQ_0(m,t), \\ \dfrac{\mathrm{d}Q_1(m,t)}{\mathrm{d}t} = -\sigma_{\text{off}} Q_1(m,t) + \sigma_{\text{on}} Q_0(m,t) + d(m+1)Q_0(m+1,t) - dmQ_0(m,t) + \rho[Q_1(m-1,t) - Q_1(m,t)]. \end{cases}} \tag{S25}$$

Also by summing up the two equations in Eq. (S25), we have that

$$\frac{\mathrm{d}Q(m,t)}{\mathrm{d}t} = d(m+1)Q(m+1,t) - dmQ(m,t) + \rho[Q_1(m-1,t) - Q_1(m,t)], \tag{S26}$$

where $Q(m,t) = \sum_\varphi Q_\varphi(m,t)$. Noted that $P(m,t) = Q(m,t-\tau)$, and by comparing Eqs. (S24) and (S26), we conclude that for $t > \tau$

$$P(m,t;1,\bullet,\bullet,t-\tau) = Q_1(m,t-\tau).$$

By replacing the two conditional probabilities in Eq. (S21) with $\bar{P}_\varphi(n,\tau)$ and $Q_1(m,t-\tau)$, Eq. (S21) becomes

$$\begin{cases} \dfrac{\mathrm{d}P(0,n,m,t)}{\mathrm{d}t} = \sigma_{\text{off}} P(1,n,m,t) - \sigma_{\text{on}} P(0,n,m,t) + d(m+1)P(0,n,m+1,t) - dmP(0,n,m,t) \\ \qquad + \rho \mathbf{1}_{t-\tau}[\bar{P}_0(n,\tau)Q_1(m-1,t-\tau) - \bar{P}_0(n-1,\tau)Q_1(m,t-\tau)], \\ \dfrac{\mathrm{d}P(1,n,m,t)}{\mathrm{d}t} = \sigma_{\text{on}} P(0,n,m,t) - \sigma_{\text{off}} P(1,n,m,t) + d(m+1)P(1,n,m+1,t) - dmP(1,n,m,t) + \rho P(1,n-1,m,t) - \rho P(1,n,m,t) \\ \qquad + \rho \mathbf{1}_{t-\tau}[\bar{P}_1(n,\tau)Q_1(m-1,t-\tau) - \bar{P}_1(n-1,\tau)Q_1(m,t-\tau)], \end{cases} \tag{S27}$$

which together with Eqs. (S22) and (S25) fully characterizes the stochastic dynamics of the system of reactions in Eq. (9). Note that in Eq. (S27), $\mathbf{1}_{t-\tau} = 1$ if and only if $t > \tau$. Note that the delay master equation [64, 65] is not the standard master equation [66], because the presence of a deterministic export time means that the model is non-Markovian.

B. Solution conditional on the gene state

In this subsection, we solve Eq. (S27) in steady state conditions by means of the generating function method. By defining generating functions $G_\varphi(u_1,u_2,t) = \sum_{n,m}(u_1+1)^n(u_2+1)^m P(\varphi,n,m,t)$, $G_{N\varphi}(u_1,t) = \sum_n (u_1+1)^n \bar{P}_\varphi(n,t)$ and $G_{M\varphi}(u_2,t) = \sum_m (u_2+1)^m Q_\varphi(m,t)$, Eqs. (S22), (S25) and (S27) are transformed into

$$\begin{cases} \partial_t G_0 = \sigma_{\text{off}} G_1 - \sigma_{\text{on}} G_0 - du_2 \partial_{u_2} G_0 + \rho(u_2 - u_1) G_{N0}^\tau G_{M1}, \\ \partial_t G_1 = -\sigma_{\text{off}} G_1 + \sigma_{\text{on}} G_0 - du_2 \partial_{u_2} G_1 + \rho u_1 G_1 + \rho(u_2 - u_1) G_{N1}^\tau G_{M1}, \end{cases} \tag{S28}$$

$$\begin{cases} \partial_t G_{N0} = \sigma_{\text{off}} G_{N1} - \sigma_{\text{on}} G_{N0}, \\ \partial_t G_{N1} = -\sigma_{\text{off}} G_{N1} + \sigma_{\text{on}} G_{N0} + \rho u_1 G_{N1}, \end{cases} \tag{S29}$$

and

$$\begin{cases} \partial_t G_{M0} = \sigma_{\text{off}} G_{M1} - \sigma_{\text{on}} G_{M0} + du_2 \partial_{u_2} G_{M0}, \\ \partial_t G_{M1} = -\sigma_{\text{off}} G_{M1} + \sigma_{\text{on}} G_{M0} + du_2 \partial_{u_2} G_{M1} + \rho u_2 G_{M1}, \end{cases} \tag{S30}$$

where the arguments u_1, u_2 and t are suppressed for brevity, and G^{τ}_{N0} is short for $G_{N0}(u_1,\tau)$. Prior to solving Eq. (S28), we need to find the solutions to Eqs. (S29) and (S30).

The exact solution to Eq. (S30) has been reported in Ref. [67]. At steady state, Eq. (S30) admits the solution

$$G_{M0} = \frac{\sigma_{\text{off}}}{\sigma_{\text{off}}+\sigma_{\text{on}}}\mathbf{M}_{-1,0}, \qquad G_{M1} = \frac{\sigma_{\text{on}}}{\sigma_{\text{off}}+\sigma_{\text{on}}}\mathbf{M}_{0,0}, \tag{S31}$$

where $x_2 = \rho u_2/d$ and the Kummer function

$$\mathbf{M}_{i,j} := M\left(1+\frac{\sigma_{\text{on}}}{d}+i, 1+\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}+j, x_2\right).$$

Note that Eq. (S29) consists of two ordinary differential equations (ODEs). We solve G_{N1} as a function of G_{N0} from the first equation in Eq. (S29) and insert the solution into the second one to obtain a second-order ODE. Using initial conditions $G_{N0}=0$ and $G_{N1}=1$ at time $t=0$ we then solve the second-order ODE at time τ to obtain

$$G^{\tau}_{N0} = \frac{\sigma_{\text{off}}}{d\theta}\exp\left[-\tau d\left(\frac{r-\theta-1}{2}\right)\right] - \frac{\sigma_{\text{off}}}{d\theta}\exp\left[-\tau d\left(\frac{r+\theta-1}{2}\right)\right], \tag{S32}$$

with $x_1 = \rho u_1/d$, $\alpha = \sigma_{\text{on}}/(\sigma_{\text{off}}+\sigma_{\text{on}})$, $r = 1+(\sigma_{\text{off}}+\sigma_{\text{on}})/d - x_1$ and

$$\theta = \sqrt{\left(\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}-x_1\right)^2 + \frac{4\sigma_{\text{on}}x_1}{d}}.$$

The probability associated with the active gene state immediately follows

$$G^{\tau}_{N1} = \left(\frac{r+\theta-1}{2\theta} - \frac{\sigma_{\text{on}}}{d\theta}\right)\exp\left[-\tau d\left(\frac{r+\theta-1}{2}\right)\right] - \left(\frac{r-\theta-1}{2\theta} - \frac{\sigma_{\text{on}}}{d\theta}\right)\exp\left[-\tau d\left(\frac{r-\theta-1}{2}\right)\right]. \tag{S33}$$

Letting $\partial_t G_0 = \partial_t G_1 = 0$, we solve G_1 as a function of G_0 from the first equation in Eq. (S28), and have that

$$G_1 = [\sigma_{\text{on}}G_0 + du_2\partial_{u_2}G_0 - \rho(u_2-u_1)G^{\tau}_{N0}G_{m1}]/\sigma_{\text{off}}. \tag{S34}$$

Summing up the two equations in Eq. (S28) leads to

$$du_2\partial_{u_2}(G_0+G_1) = \rho u_1 G_1 + \rho(u_2-u_1)(G^{\tau}_{N0}+G^{\tau}_{N1})G_{M1}. \tag{S35}$$

Inserting Eqs. (S31) and (S34) into Eq. (S35), we then get an equation that only depends on G_0:

$$\begin{aligned} &u_2^2\partial^2_{u_2}G_0 + \left(1+\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}-x_1\right)\partial_{u_2}G_0 - \frac{\sigma_{\text{on}}}{d}x_1G_0 \\ =&\underbrace{(u_2-u_1)x_2G^{\tau}_{N0}\partial_{u_2}G_{M1} + x_2G^{\tau}_{N0}G_{M1} + (x_2-x_1)(\sigma_{\text{off}}/d-x_1)G^{\tau}_{N0}G_{M1} + (x_2-x_1)\sigma_{\text{off}}G^{\tau}_{N1}G_{M1}/d}_{\mathcal{A}}. \end{aligned} \tag{S36}$$

By means of Eqs. (13.3.3), (13.3.4) and (13.3.15) in Ref [68], the term $\mathcal{A}$ simplifies to

$$\begin{aligned} \mathcal{A} =& \frac{\alpha(1+\sigma_{\text{on}}/d)(x_2-x_1)x_2}{1+(\sigma_{\text{off}}+\sigma_{\text{on}})/d}G^{\tau}_{N0}\mathbf{M}_{1,1} + \alpha x_2G^{\tau}_{N0}\mathbf{M}_{0,0} + \alpha(x_2-x_1)(\sigma_{\text{off}}/d-x_1)G^{\tau}_{N0}\mathbf{M}_{0,0} + \alpha(x_2-x_1)\sigma_{\text{off}}G^{\tau}_{N1}\mathbf{M}_{0,0}/d \\ =&\alpha(x_2-x_1)(1+\sigma_{\text{on}}/d)G^{\tau}_{N0}\mathbf{M}_{1,0} + \alpha(x_2-x_1)\left[\left(\frac{\sigma_{\text{off}}-\sigma_{\text{on}}}{d}-x_1\right)G^{\tau}_{N0} + \frac{\sigma_{\text{off}}}{d}G^{\tau}_{N1}\right]\mathbf{M}_{0,0} + \alpha x_1G^{\tau}_{N0}\mathbf{M}_{0,0}. \end{aligned} \tag{S37}$$

It can be ascertained that Eq. (S36) is a linear inhomogeneous second-order ODE. Therefore, the solution to Eq. (S36) is a combination of solutions to the general ODEs of the following form

$$x^2f'' + Axf' + Bf = Hx^iM(a,b,cx),$$

where A, B, H, a, b, c are constants and $i=0,1$. The general solution is given by

$$f = c_1x^{\frac{1}{2}(1-A-\lambda)} + c_2x^{\frac{1}{2}(1-A+\lambda)} + \frac{Hx^i}{(i+\ell_-)\lambda}{}_2F_2\left(a, i+\ell_-; b, 1+i+\ell_-; cx\right) - \frac{Hx^i}{(i+\ell_+)\lambda}{}_2F_2\left(a, i+\ell_+; b, 1+i+\ell_+; cx\right)$$

with $\lambda = \sqrt{1-2A+A^2-4B}$, $\ell_- = (A-1-\lambda)/2$ and $\ell_+ = (A-1+\lambda)/2$. By defining the hypergeometric function

$$\mathbf{F}^{\pm}_{i,j,k,l} := {}_2F_2\left(1+i+\frac{\sigma_{\text{on}}}{d}, \frac{1+r\pm\theta}{2}+j; 1+k+\frac{\sigma_{\text{on}}+\sigma_{\text{off}}}{d}, \frac{3+r\pm\theta}{2}+l; x_2\right),$$

the general solution to Eq. (S36) reads

$$\begin{aligned} G_0 =& k_A x_2\left(\frac{2}{1+r-\theta}\mathbf{F}^-_{1,0,0,0} - \frac{2}{1+r+\theta}\mathbf{F}^+_{1,0,0,0}\right) + k_B x_2\left(\frac{2}{1+r-\theta}\mathbf{F}^-_{0,0,0,0} - \frac{2}{1+r+\theta}\mathbf{F}^+_{0,0,0,0}\right) \\ &- k_A x_1\left(\frac{2}{r-\theta-1}\mathbf{F}^-_{1,-1,0,-1} - \frac{2}{r+\theta-1}\mathbf{F}^+_{1,-1,0,-1}\right) - k_C x_1\left(\frac{2}{r-\theta-1}\mathbf{F}^-_{0,-1,0,-1} - \frac{2}{r+\theta-1}\mathbf{F}^+_{0,-1,0,-1}\right) + c_1 u_2^{\frac{1-r-\theta}{2}} + c_2 u_2^{\frac{1-r+\theta}{2}} \end{aligned} \tag{S38}$$

where

$$k_A = \frac{\alpha(d+\sigma_{\text{on}})G^\tau_{N0}}{d\theta}, \qquad k_B = \frac{\alpha[(\sigma_{\text{off}}-\sigma_{\text{on}}-dx_1)G^\tau_{N0}+\sigma_{\text{off}}G^\tau_{N1}]}{d\theta}, \qquad k_C = k_B - \alpha G^\tau_{N0}/\theta,$$

and constants c_1 and c_2 remain to be determined.

Since $1-r-\theta<0$ and $u_2\in[-1,0]$, the term $u_2^{\frac{1-r-\theta}{2}}$ becomes ill-defined when u_2 approaches 0 from the negative real axis, thereby indicating $c_1=0$. Additionally, $(1-r+\theta)/2$ can hardly be an integer, and the generating function is a polynomial of only integer powers, thereby suggesting $c_2=0$.

By plugging Eq. (S38) into Eq. (S34), we obtain an equation for G_1 of the form

$$\begin{aligned} G_1 =& k_A k_1^- x_2 \mathbf{F}^-_{1,0,0,0} + k_A k_{23}^- x_2^2 \mathbf{F}^-_{2,1,1,1} - k_A k_1^+ x_2 \mathbf{F}^+_{1,0,0,0} - k_A k_{23}^+ x_2^2 \mathbf{F}^+_{2,1,1,1} + k_B k_1^- x_2 \mathbf{F}^-_{0,0,0,0} + k_B k_{13}^- x_2^2 \mathbf{F}^-_{1,1,1,1} - k_B k_1^+ x_2 \mathbf{F}^+_{0,0,0,0} \\ &- k_B k_{13}^+ x_2^2 \mathbf{F}^+_{1,1,1,1} + k_A k_2^- \mathbf{F}^-_{1,-1,0,-1} - k_A k_{21}^- x_1 x_2 \mathbf{F}^-_{2,0,1,0} + k_A k_2^+ \mathbf{F}^+_{1,-1,0,-1} + k_A k_{21}^+ x_2 \mathbf{F}^+_{2,0,1,0} + k_C k_2^- \mathbf{F}^-_{0,-1,0,-1} - k_C k_{11}^- x_2 \mathbf{F}^-_{1,0,1,0} \\ &+ k_C k_2^+ \mathbf{F}^+_{0,-1,0.-1} + k_C k_{11}^+ x_1 x_2 \mathbf{F}^+_{1,0,1,0} - \alpha d(x_2-x_1) G^\tau_{N0} \mathbf{M}_{0,0}/\sigma_{\text{off}}, \end{aligned} \tag{S39}$$

with

$$k_1^\pm = \frac{2(d+\sigma_{\text{on}})}{\theta\sigma_{\text{off}}(1+r\pm\theta)}, \qquad k_2^\pm = \frac{2\sigma_{\text{on}}x_1}{\sigma_{\text{off}}(r\pm\theta-1)}, \qquad k_{ij}^\pm = \frac{2d(id+\sigma_{\text{on}})}{\theta\sigma_{\text{off}}(d+\sigma_{\text{on}}+\sigma_{\text{off}})(j+r\pm\theta)}.$$

C. Solution independent of the gene state

Since the solutions Eqs. (S38) and (S39) are rather complicated, we derive much simpler solutions that are independent of the gene state. Note that the solutions Eqs. (S32) and (S33) can be compactly written as

$$G^\tau_{N0} = \frac{\sigma_{\text{off}}}{d\theta}\left(e^{-\nu_-\tau} - e^{-\nu_+\tau}\right), \tag{S40a}$$

$$G^\tau_{N1} = \frac{\nu_+ - \sigma_{\text{on}}}{d\theta}e^{-\nu_+\tau} - \frac{\nu_- - \sigma_{\text{on}}}{d\theta}e^{-\nu_-\tau}, \tag{S40b}$$

using the definition $\nu_\pm = d(r\pm\theta-1)/2$.

Additionally, by combining the two equations in Eq. (S28) in the steady-state limit, we obtain the following ODE for the joint probability generating function $G = G_0 + G_1$,

$$d^2u_2^2\partial_{u_2}^2 G + du_2(\sigma_{\text{on}}+\sigma_{\text{off}}-\rho u_1+d)\partial_{u_2}G - \sigma_{\text{on}}\rho u_1 G = -\rho u_1 I_0 + (\sigma_{\text{on}}+\sigma_{\text{off}})I + du_2\partial_{u_2}I, \tag{S41}$$

where $I_0 = \rho(u_2-u_1)G^\tau_{N0}G_{M1}$, $I_1 = \rho(u_2-u_1)G^\tau_{N1}G_{M1}$ and $I = I_0+I_1$. The solution to this equation takes a simple form in terms of two Kummer functions, which precisely agrees with Eq. (10), a result that has been previously derived using queueing theory.

IV. Stochastic description of reaction system Eq. (12) using queueing theory

Without loss of any generality, here we present the solution for $d=1$. The probability density of the time between two successive nuclear mRNA production events for the reaction system Eq. (9) is a phase-type distribution,

$$f(t) = \kappa \mathrm{e}^{-D_0 t}(-D_0\mathbf{1}^\top)$$

with $\kappa = [0,\ 0,\ 1]^\top$, $\mathbf{1} = [1,\ 1,\ 1]^\top$ and

$$D_0 = \begin{bmatrix} -\sigma_\mathrm{u} & \sigma_\mathrm{u} & 0 \\ 0 & -\lambda & \lambda \\ \sigma_\mathrm{b} & 0 & -\sigma_\mathrm{b} - \rho \end{bmatrix}.$$

The corresponding Laplace transform of $f(t)$ is

$$\phi(s) = \kappa(sI - D_0)^{-1}(-D_0\mathbf{1}^\tau) = \frac{\rho(s+\lambda)(s+\sigma_\mathrm{u})}{-\lambda\sigma_\mathrm{b}\sigma_\mathrm{u} + (s+\lambda)(s+\rho+\sigma_\mathrm{b})(s+\sigma_\mathrm{u})}. \tag{S42}$$

The mean interarrival time α follows from $\alpha = -\phi'(0)$ and is given by

$$\alpha = \frac{\sigma_\mathrm{b}\sigma_\mathrm{u} + \lambda\sigma_\mathrm{b} + \lambda\sigma_\mathrm{u}}{\lambda\rho\sigma_\mathrm{u}}. \tag{S43}$$

Using Eqs. (7), (S42) and (S43), the Laplace transform of $G_N(z_1)$ with respect to the nuclear processing time τ is

$$\mathcal{L}\left\{G_N(z_1)\right\}(s) = \frac{1}{s} + \frac{(z_1-1)[1-\phi(s)]}{\alpha s^2[1-z_1\phi(s)]} = \frac{1}{s} + \frac{\lambda x_1\sigma_\mathrm{u}\left[(s+\lambda)(s+\sigma_\mathrm{b}) + (s+\lambda+\sigma_\mathrm{b})\sigma_\mathrm{u}\right]}{\mathcal{S}\mu_0},$$

where $u_1 = z_1 - 1$, $x_1 = \rho u_1$, $\mu_0 = \sigma_\mathrm{b}\sigma_\mathrm{u} + \lambda\sigma_\mathrm{b} + \lambda\sigma_\mathrm{u}$ and

$$\mathcal{S} = s[s(s+\lambda)(s+\sigma_\mathrm{b}-\rho u_1) + (s+\lambda)(s-\rho u_1)\sigma_\mathrm{u} + s\sigma_\mathrm{b}\sigma_\mathrm{u}].$$

It follows from Eq. (4) that C_j for $i = 1, 2, \cdots$ are given by

$$C_j = \prod_{i=1}^{j} \frac{\phi(i)}{1-\phi(i)} = \frac{\rho^i(1+\lambda)_i(1+\sigma_u)_i}{\Gamma(i)(1+v^-)_i(1+v^+)_i}, \tag{S44}$$

where

$$v^\pm = \frac{\lambda+\sigma_\mathrm{b}+\sigma_\mathrm{u}}{2} \pm \frac{\sqrt{(\lambda-\sigma_\mathrm{b})^2 - 2(\lambda+\sigma_\mathrm{b})\sigma_\mathrm{u} + \sigma_\mathrm{u}^2}}{2}.$$

Using Eqs. (6) and (S44), we obtain the probability generating function for cytoplasmic mRNA

$$G_M(z_2) = 1 + \sum_{m=1}^{\infty} \frac{C_{m-1}}{\alpha m}(z_2-1)^m = 1 + \mathbf{F}_{0,0,0,0}, \tag{S45}$$

where $u_2 = z_2 - 1$, $x_2 = \rho u_2$ and $\mathbf{F}_{i,j,k,l} = {}_2F_2\left(\lambda+i, \sigma_\mathrm{u}+j; v^- + k, v^+ + l; x_2\right)$. Next we compute the Laplace transform $\mathcal{L}\left\{\phi(r,x)\right\}$ as

$$\begin{aligned}\mathcal{L}\left\{\phi(r,x)\right\}(s) =& \kappa(sI-D_0)^{-1}(rI-D_0)^{-1}(-D_0\mathbf{1}^\tau) \\ =& \frac{\rho rs(r+\lambda)(s+\lambda)}{h(r)h(s)} + \frac{\rho[(r+s)(r+\lambda)(s+\lambda) + \lambda(r+s+\lambda)\sigma_\mathrm{b}]\sigma_\mathrm{u}}{h(r)h(s)} + \frac{\rho[(r+\lambda)(s+\lambda) + \lambda\sigma_\mathrm{b}]\sigma_\mathrm{u}^2}{h(r)h(s)},\end{aligned} \tag{S46}$$

where $h(x) = x(x+\lambda)(x+\rho+\sigma_\mathrm{b}) + (x+\lambda)(x+\rho)\sigma_\mathrm{u} + x\sigma_\mathrm{b}\sigma_\mathrm{u}$. According to Eq. (S46), the summation term in Eq. (8) becomes

$$\begin{aligned}\sum_{r=1}^{\infty} C_r \frac{\mathcal{L}_x\left\{\phi(r,x)\right\}(s)}{\phi(r)}(z_2-1)^r =& -\frac{(s+\lambda)(s+\sigma_\mathrm{b}) + (s+\lambda+\sigma_\mathrm{b})\sigma_\mathrm{u}}{h(s)} + \frac{(s+\lambda)(s+\sigma_\mathrm{b}) + (s+\lambda+\sigma_\mathrm{b})\sigma_\mathrm{u}}{h(s)}\mathbf{F}_{0,0,1,1} \\ &+ \frac{x_2[s^2(1+\lambda+\sigma_\mathrm{u}) + s(\lambda+\sigma_\mathrm{u})(1+\lambda+\sigma_\mathrm{u}) + \lambda\sigma_u(1+\lambda+\sigma_\mathrm{u}+\sigma_\mathrm{b})]}{h(s)\mu_1}\mathbf{F}_{1,1,2,2} + \frac{{x_2}^2(1+\lambda)(s+\lambda)(1+\sigma_\mathrm{u})(s+\sigma_\mathrm{u})}{h(s)\mu_1\mu_2}\mathbf{F}_{2,2,3,3},\end{aligned} \tag{S47}$$

where $\mu_1 = (1+\sigma_\mathrm{b})(1+\sigma_\mathrm{u}) + \lambda(1+\sigma_\mathrm{b}+\sigma_\mathrm{u})$, $\mu_2 = (2+\sigma_\mathrm{b})(2+\sigma_\mathrm{u}) + \lambda(2+\sigma_\mathrm{b}+\sigma_\mathrm{u})$. Given Eq. (S47), the Laplace transform of $\Psi(z_1, z_2)$ becomes

$$\begin{aligned}
&\mathcal{L}_T\left\{\Psi(z_1,z_2)\right\}(s)\\
=&-\mathcal{L}\left\{G_N(z_1)\right\}(s) + \frac{\lambda x_1\sigma_\mathrm{u}\left[(s+\lambda)(s+\sigma_\mathrm{b}) + (s+\lambda+\sigma_\mathrm{b})\sigma_\mathrm{u}\right]}{\mathcal{S}\mu_0}\mathbf{F}_{0,0,1,1}\\
&+\underbrace{\frac{\lambda x_1\sigma_\mathrm{u}x_2[s^2(1+\lambda+\sigma_\mathrm{u}) + s(\lambda+\sigma_\mathrm{u})(1+\lambda+\sigma_\mathrm{u}) + \lambda\sigma_u(1+\lambda+\sigma_\mathrm{b}+\sigma_\mathrm{u})]}{\mathcal{S}\mu_0\mu_1}\mathbf{F}_{1,1,2,2}}_{\text{Term B}} + \underbrace{\frac{\lambda x_1\sigma_u {x_2}^2(1+\lambda)(s+\lambda)(1+\sigma_\mathrm{u})(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0\mu_1\mu_2}\mathbf{F}_{2,2,3,3}}_{\text{Term A}}.
\end{aligned} \tag{S48}$$

Next we simplify Terms A and B using the relation $bz_2F_2(a+1,b+1;c+1,d+1;z) = cd_2F_2(a+1,b;c,d;z) - cd_2F_2(a,b;c,d;z)$ [69]. This leads to

$$\begin{aligned}
\text{Term A} =& \frac{\lambda x_1\sigma_\mathrm{u}x_2(1+\lambda)(s+\lambda)(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0\mu_1}\mathbf{F}_{2,1,2,2} - \frac{\lambda x_1\sigma_\mathrm{u}x_2(1+\lambda)(s+\lambda)(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0\mu_1}\mathbf{F}_{1,1,2,2}\\
=& \underbrace{\frac{\lambda x_1(1+\lambda)(s+\lambda)(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0}\mathbf{F}_{2,0,1,1}}_{\text{Term C}} - \frac{\lambda x_1(1+\lambda)(s+\lambda)(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0}(2\mathbf{F}_{1,0,1,1} - \mathbf{F}_{0,0,1,1}),
\end{aligned} \tag{S49}$$

$$\text{Term B} = \frac{\lambda x_1[s^2(1+\lambda+\sigma_\mathrm{u}) + s(\lambda+\sigma_\mathrm{u})(1+\lambda+\sigma_\mathrm{u}) + \lambda\sigma_u(1+\lambda+\sigma_\mathrm{b}+\sigma_\mathrm{u})]}{\mathcal{S}\mu_0}(\mathbf{F}_{1,0,1,1} - \mathbf{F}_{0,0,1,1}). \tag{S50}$$

Given another relation, $a_2F_2(a+1,b;c,d;z) = b_2F_2(a,b+1;c,d;z) + (a-b)_2F_2(a,b;c,d;z)$ [69], we simplify Term C to

$$\text{Term C} = \frac{\lambda x_1(s+\lambda)(s+\sigma_\mathrm{u})\sigma_\mathrm{u}}{\mathcal{S}\mu_0}\mathbf{F}_{1,1,1,1} + \frac{\lambda x_1(s+\lambda)(s+\sigma_\mathrm{u})(1+\lambda-\sigma_\mathrm{u})}{\mathcal{S}\mu_0}\mathbf{F}_{1,0,1,1}. \tag{S51}$$

Using Eqs. (S49), (S50) and (S51), Eq. (S48) can be simplified to

$$\mathcal{L}_T\left\{\Psi(z_1,z_2)\right\}(s) = -\mathcal{L}\left\{G_N(z_1)\right\}(s) + \frac{\lambda x_1\sigma_\mathrm{u}\sigma_\mathrm{b}(s+\sigma_u)}{\mathcal{S}\mu_0}\mathbf{F}_{0,0,1,1} + \frac{\lambda x_1\sigma_\mathrm{u}\lambda\sigma_\mathrm{b}}{\mathcal{S}\mu_0}\mathbf{F}_{1,0,1,1} + \frac{\lambda x_1\sigma_\mathrm{u}(s+\lambda)(s+\sigma_\mathrm{u})}{\mathcal{S}\mu_0}\mathbf{F}_{1,1,1,1}. \tag{S52}$$

Subsequently, we need to calculate the inverse Laplace transform of $\mathcal{L}_T\left\{\Psi(z_1,z_2)\right\}(s)$. Note that the last three terms in Eq. (S52) share the same denominator, of which the poles are denoted here by r_1, r_2, r_3; these are the roots of the equation $\mathcal{S} = 0$. By using the residue method, we derive the following results:

$$\begin{aligned}
\mathcal{L}^{-1}\left\{\frac{s+\sigma_\mathrm{u}}{s(s-r_1)(s-r_2)(s-r_3)}\right\} &= \frac{\mathrm{e}^{r_1\tau}(r_1+\sigma_\mathrm{u})}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}(r_2+\sigma_\mathrm{u})}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}(r_3+\sigma_\mathrm{u})}{(r_3-r_1)(r_3-r_2)r_3} - \frac{\sigma_\mathrm{u}}{r_1r_2r_3},\\
\mathcal{L}^{-1}\left\{\frac{1}{s(s-r_1)(s-r_2)(s-r_3)}\right\} &= \frac{\mathrm{e}^{r_1\tau}}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}}{(r_3-r_1)(r_3-r_2)r_3} - \frac{1}{r_1r_2r_3},\\
\mathcal{L}^{-1}\left\{\frac{(s+\lambda)(s+\sigma_\mathrm{u})}{s(s-r_1)(s-r_2)(s-r_3)}\right\} &= \frac{\mathrm{e}^{r_1\tau}(r_1+\lambda)(r_1+\sigma_\mathrm{u})}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}(r_2+\lambda)(r_2+\sigma_\mathrm{u})}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}(r_3+\lambda)(r_3+\sigma_\mathrm{u})}{(r_3-r_1)(r_3-r_2)r_3} - \frac{\lambda\sigma_\mathrm{u}}{r_1r_2r_3}.
\end{aligned}$$

Using these equations, the function $\Psi(z_1, z_2)$ is found to be

$$\begin{aligned}
\Psi(z_1,z_2) =& -G_N(z_1)\\
&+\frac{\lambda x_1\sigma_\mathrm{u}\sigma_\mathrm{b}}{\mu_0}\left[\frac{\mathrm{e}^{r_1\tau}(r_1+\sigma_\mathrm{u})}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}(r_2+\sigma_\mathrm{u})}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}(r_3+\sigma_\mathrm{u})}{(r_3-r_1)(r_3-r_2)r_3} - \frac{\sigma_\mathrm{u}}{r_1r_2r_3}\right]\mathbf{F}_{0,0,1,1}\\
&+\frac{\lambda^2 x_1\sigma_\mathrm{u}\sigma_\mathrm{b}}{\mu_0}\left[\frac{\mathrm{e}^{r_1\tau}}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}}{(r_3-r_1)(r_3-r_2)r_3} - \frac{1}{r_1r_2r_3}\right]\mathbf{F}_{1,0,1,1}\\
&+\frac{\lambda x_1\sigma_\mathrm{u}}{\mu_0}\left[\frac{e^{r_1\tau}(r_1+\lambda)(r_1+\sigma_\mathrm{u})}{r_1(r_1-r_2)(r_1-r_3)} + \frac{\mathrm{e}^{r_2\tau}(r_2+\lambda)(r_2+\sigma_\mathrm{u})}{(r_2-r_1)r_2(r_2-r_3)} + \frac{\mathrm{e}^{r_3\tau}(r_3+\lambda)(r_3+\sigma_\mathrm{u})}{(r_3-r_1)(r_3-r_2)r_3} - \frac{\lambda\sigma_\mathrm{u}}{r_1r_2r_3}\right]\mathbf{F}_{1,1,1,1}.
\end{aligned} \tag{S53}$$

Solving the three roots r_1, r_2 and r_3 of the equation $\mathcal{S}=0$ we find

$$r_1=\frac{k}{2^{\frac{1}{3}}3}-\frac{m}{3}-\frac{2^{\frac{1}{3}}(-m^2+3n)}{3k},\qquad r_2=a+b\mathrm{i},\qquad r_3=a-b\mathrm{i}, \tag{S54}$$

where

$$a=-\frac{k}{2^{\frac{1}{3}}6}-\frac{m}{3}+\frac{-m^2+3n}{2^{\frac{2}{3}}3k},\qquad b=\frac{k}{2^{\frac{1}{3}}2\sqrt{3}}+\frac{-m^2+3n}{2^{\frac{2}{3}}\sqrt{3}k},$$

$$m=\lambda+\sigma_{\mathrm{b}}+\sigma_{\mathrm{u}}-x_1,\qquad n=\lambda\sigma_{\mathrm{u}}+\lambda\sigma_{\mathrm{b}}+\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}-(\lambda+\sigma_{\mathrm{u}})x_1,\qquad p=-\lambda\sigma_{\mathrm{u}}x_1,$$

and

$$k=\left(-2m^3+9mn-27p+3\sqrt{3}\sqrt{-m^2n^2+4n^3+4m^3p-18mnp+27p^2}\right)^{\frac{1}{3}}.$$

Finally using Eqs. (S53), (S54) and (5), the generating function solution is found to be given by

$$G_{N,M}(z_1,z_2)=\mathbf{F}_{0,0,0,0}+k_1\mathbf{F}_{0,0,1,1}+k_2\mathbf{F}_{1,0,1,1}+k_3\mathbf{F}_{1,1,1,1} \tag{S55}$$

together with

$$\begin{aligned}
k_1=&-\frac{\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}}{\mu_0}+\frac{\mathrm{e}^{r_1\tau}x_1\lambda\sigma_{\mathrm{b}}\sigma_u(r_1+\sigma_{\mathrm{u}})}{r_1\mathcal{B}\mu_0}-\frac{\mathrm{e}^{a\tau}x_1\lambda\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}\left[\mathcal{J}+(2a-r_1)\sigma_{\mathrm{u}}\right]\cos(b\tau)}{\mathcal{B}\mathcal{J}\mu_0}\\
&+\frac{\mathrm{e}^{a\tau}x_1\lambda\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}\left[\mathcal{J}(a-r_1)+(a^2-b^2-ar_1)\sigma_{\mathrm{u}}\right]\sin(b\tau)}{b\mathcal{B}\mathcal{J}\mu_0},\\
k_2=&-\frac{\lambda\sigma_{\mathrm{b}}}{\mu_0}+\frac{\mathrm{e}^{r_1\tau}x_1\lambda^2\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}}{r_1\mathcal{B}\mu_0}-\frac{\mathrm{e}^{a\tau}x_1\lambda^2\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}(2a-r_1)\cos(b\tau)}{\mathcal{B}\mathcal{J}\mu_0}+\frac{\mathrm{e}^{a\tau}x_1\lambda^2\sigma_{\mathrm{b}}\sigma_{\mathrm{u}}(a^2-ar_1-b^2)\sin(bt)}{b\mathcal{B}\mathcal{J}\mu_0},\\
k_3=&-\frac{\lambda\sigma_{\mathrm{u}}}{\mu_0}+\frac{\mathrm{e}^{r_1\tau}x_1\lambda\sigma_{\mathrm{u}}(r_1+\lambda)(r_1+\sigma_{\mathrm{u}})}{r_1\mathcal{B}\mu_0}-\frac{\mathrm{e}^{a\tau}x_1\lambda\sigma_{\mathrm{u}}\left[\mathcal{J}(r_1+\lambda+\sigma_{\mathrm{u}})+(2a-r_1)\lambda\sigma_{\mathrm{u}}\right]\cos(b\tau)}{\mathcal{B}\mathcal{J}\mu_0}\\
&+\frac{\mathrm{e}^{a\tau}x_1\lambda\sigma_{\mathrm{u}}\left[\mathcal{J}(b^2+(a-r_1)(a+\lambda))+\mathcal{J}(a-r_1)\sigma_{\mathrm{u}}+(a^2-b^2-ar_1)\lambda\sigma_{\mathrm{u}}\right]\sin(b\tau)}{b\mathcal{B}\mathcal{J}\mu_0},
\end{aligned}$$

and $\mathcal{J}=a^2+b^2$, $\mathcal{B}=(a-r_1)^2+b^2$.

V. The first and second-order moments of Models I and III converge under fast export conditions

By taking the first-order derivatives of Eq. (10) with respect to z_1 and z_2 and letting $z_1=1$ and $z_2=1$, we obtain the first-order moments of Model I:

$$\begin{aligned}
\langle n_{\mathrm{I}}\rangle&=\frac{\rho\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}},\\
\langle m_{\mathrm{I}}\rangle&=\frac{\rho\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})}.
\end{aligned} \tag{S56}$$

From the master equation of Model III (or the rate equations) we can obtain the first-order moments, which are

$$\begin{aligned}
\langle n_{\mathrm{III}}\rangle&=\frac{\rho\sigma_{\mathrm{on}}}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\gamma},\\
\langle m_{\mathrm{III}}\rangle&=\frac{\rho\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})}.
\end{aligned}$$

Since $\gamma=1/\tau$, it immediately follows that that $\langle n_{\mathrm{I}}\rangle=\langle n_{\mathrm{III}}\rangle$ and $\langle m_{\mathrm{I}}\rangle=\langle m_{\mathrm{III}}\rangle$.

Similarly, by taking the second-order derivatives of Eq. (10) with respect to z_1 and z_2 and letting $z_1=1$ and $z_2=1$, all the second-order moments of Model I can be obtained

$$\begin{aligned}
\mathrm{var}(n_{\mathrm{I}})&=\frac{\rho\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}}+\frac{2\rho^2\sigma_{\mathrm{on}}\sigma_{\mathrm{off}}\left[-1+\mathrm{e}^{-(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau}+(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau\right]}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^4},\\
\mathrm{var}(m_{\mathrm{I}})&=\frac{\rho\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})}+\frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})},\\
\mathrm{cov}(n_{\mathrm{I}},m_{\mathrm{I}})&=\frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}\left[1-\mathrm{e}^{-(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau}\right]}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^3(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})}.
\end{aligned} \tag{S57}$$

Using the master equation of Model III we can also derive the exact second-order moments, which are

$$\begin{aligned}
\mathrm{var}(n_{\mathrm{III}}) &= \frac{\rho^2\sigma_{\mathrm{on}}\sigma_{\mathrm{off}}\tau^2}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(1+\sigma_{\mathrm{off}}\tau+\sigma_{\mathrm{on}}\tau)} + \frac{\rho\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}},\\
\mathrm{var}(m_{\mathrm{III}}) &= \frac{\rho\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})} + \frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})(1+\tau d)(1+\sigma_{\mathrm{off}}\tau+\sigma_{\mathrm{on}}\tau)}\\
&\quad + \frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}\tau}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(1+\tau d)(1+\sigma_{\mathrm{off}}\tau+\sigma_{\mathrm{on}}\tau)},\\
\mathrm{cov}(n_{\mathrm{III}},m_{\mathrm{III}}) &= \frac{\rho^2\sigma_{\mathrm{on}}(1+\sigma_{\mathrm{on}}\tau)\tau}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})(1+\sigma_{\mathrm{off}}\tau+\sigma_{\mathrm{on}}\tau)(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}+d)(1+\tau d)} + \frac{\rho^2\sigma_{\mathrm{on}}^2\tau}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}+d)(1+\tau d)}\\
&\quad + \frac{\rho^2\sigma_{\mathrm{on}}(1+\sigma_{\mathrm{on}}\tau)\tau^2}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})(1+\sigma_{\mathrm{off}}\tau+\sigma_{\mathrm{on}}\tau)(1+\tau d)} - \frac{\rho^2\sigma_{\mathrm{on}}^2\tau}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2}.
\end{aligned}$$

By expanding all the second-order moments of Models I and III as a series in powers of τ about $\tau = 0$, we find that

$$\begin{aligned}
\mathrm{var}(n_{\mathrm{I}}) = \mathrm{var}(n_{\mathrm{III}}) &= \frac{\rho\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}} + \mathcal{O}(\tau^2),\\
\mathrm{var}(m_{\mathrm{I}}) = \mathrm{var}(m_{\mathrm{III}}) &= \frac{\rho\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})} + \frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})} + \mathcal{O}(\tau),\\
\mathrm{cov}(n_{\mathrm{I}},m_{\mathrm{I}}) = \mathrm{cov}(n_{\mathrm{III}},m_{\mathrm{III}}) &= \frac{\rho^2\sigma_{\mathrm{off}}\sigma_{\mathrm{on}}\tau}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})} + \mathcal{O}(\tau^2).
\end{aligned}$$

Hence given only information about the first and second moments of mRNA counts, the two models become indistinguishable in the limit of fast nuclear export.

VI. Estimation of the ratio of export to degradation times from experimental data

The steady-state means of nuclear and cytoplasmic mRNA counts according to Model I are given by Eq. (S56). This implies that $\langle n\rangle/\langle m\rangle = \tau d$. To estimate τd, we extracted the average numbers of nuclear and cytoplasmic mRNAs in mouse MIN6 cells and Liver cells from Table S2 of Ref. [32]. We found that approximately 80% of genes for which the mean number of nuclear and cytoplasmic mRNAs is not zero, are characterised by $\tau d < 1$ (Fig. S2).

VII. Parameter inference using the generating function solution for synthetic data without extrinsic noise and for one gene copy

Consider a sample of n_c cells in steady-state conditions. Specifically, in each cell i, we measure its nuclear mRNA counts (n_i) and cytoplasmic mRNA counts (m_i). The empirical joint distribution for the cell population is

$$\hat{P}(n,m) = \frac{1}{n_c}\sum_{i=1}^{n_c}\mathbf{1}_{n=n_i,m=m_i},$$

where $\mathbf{1}_{n=n_i,m=m_i} = 1$ if and only if $n = n_i$ and $m = m_i$. The empirical generating function for the cell population is then computed as

$$\bar{G}(z_1,z_2) = \sum_{n,m}\hat{P}(n,m)z_1^n z_2^m = \sum_{i=1}^{n_c}\frac{1}{n_c}z_1^{n_i}z_2^{m_i}. \tag{S58}$$

To infer the kinetic parameters ϕ, we aim to minimize the loss function in the form of the L^2 norm

$$J_{\mathrm{PGF}}(\phi) = \int_0^1\int_0^1 \left|\mathcal{G}(z_1,z_2,\phi) - \bar{G}(z_1,z_2)\right|^2 \mathrm{d}z_1\mathrm{d}z_2, \tag{S59}$$

where $\mathcal{G}(z_1,z_2,\phi)$ is the generating function of the model. For the Model I (Eq. (9)), $\mathcal{G}(z_1,z_2,\phi)$ is the generating function solution given by Eq. (10), and $\phi = \{\rho,\sigma_{\mathrm{on}},\sigma_{\mathrm{off}},\tau\}$; for the Poisson model Eq. (16),

$$\mathcal{G}(z_1,z_2,\phi) = \exp\left[\rho\tau(z_1-1) + \rho(z_2-1)/d\right], \tag{S60}$$

and $\phi = \{\rho, \tau\}$. Note that since we are only considering the steady state limit, only the kinetic parameters normalised by d can be estimated. Note also that Eq. (S60) can be derived from the general queueing theory result, Eq. (5).

To improve computational efficiency, we utilize the Gauss quadrature method and approximate the integral by

$$J_{\mathrm{PGF}}(\phi) \approx \frac{1}{4}\sum_{j=1}^{N_y}\sum_{k=1}^{N_y} w_j w_k \left|\mathcal{G}\left(\frac{y_{1,j}+1}{2}, \frac{y_{2,k}+1}{2}, \phi\right) - \bar{G}\left(\frac{y_{1,j}+1}{2}, \frac{y_{2,k}+1}{2}\right)\right|^2, \tag{S61}$$

where $y_{1,j} \in [-1, 1]$ and $y_{2,k} \in [-1, 1]$ are the j-th and k-th integral points of the N_y-th order Gauss quadrature approximation, and w_j, w_k are the corresponding integral weights given by the command `gausslegendre` in Julia.

Finally, we employ the Nelder-Mead algorithm to minimize $J_{\mathrm{PGF}}(\phi)$. We use this algorithm because: (i) it is more robust against the selection of initial conditions, particularly compared to gradient-based optimization algorithms; (ii) the computation of the gradients $\partial_\phi J_{\mathrm{PGF}}(\phi)$ is rather time demanding. The optimization is implemented in the package `Optim.jl` in Julia.

The numerical recipe of the generating-function-based inference method is summarised in Algorithm 1. In this study, the default threshold used in the Julia command `optimize` is gradient tolerance `g_tol`$= 10^{-20}$ (the absolute value of the gradient of the loss function $J_{\mathrm{PGF}}(\phi)$ less than 10^{-20}) and maximal iterations `iterations`$= 1000$.

Algorithm 1 Generating-function-based parameter inference for nuclear-cytoplasmic mRNA count data

Input: Number of cells (n_c), the counts of nuclear and cytoplasmic mRNAs (n_i, m_i) for cell i
Output: Kinetic parameters ϕ

```
Generate Gauss quadrature points and weights (y_{1,j}, y_{2,k}, w_j, w_k) for z_1, z_2 by the command gausslegendre
Compute the empirical generating function by using Eq. (S58)
Initialize the inferred parameters φ
while Threshold reached do
    Compute the generating function for the inferred parameters φ using Eq. (10) or Eq. (S60)
    Compute the loss function J_PGF(φ) using Eq. (S61)
    Update the inferred parameters φ by the Nelder-Mead optimization algorithm
end while
return Kinetic parameters φ
```

VIII. Parameter inference for data with extrinsic noise and multiple gene copies

Consider a sample of n_c cells in steady-state conditions. Specifically, in each cell i, we measure its nuclear mRNA counts (n_i), cytoplasmic mRNA counts (m_i) and cellular volume (V_i). The empirical distribution for the cell population is calculated using Eq. (S58). The normalized volume for cell i is given by

$$\beta_i = \frac{V_i}{\sum_i V_i / n_c}. \tag{S62}$$

Because of extrinsic noise due to the coupling of the transcription rate to cell volume, a common property of many eukaryotic cells [46, 47], the transcription step in Model I (Eq. (9)) and the Poisson model (Eq. (16)) is modified as

$$G \xrightarrow{\rho} G + N \qquad \Rightarrow \qquad G \xrightarrow{\bar{\rho}\beta} G + N$$

where β is a random variable sampled from an unknown distribution $p(\beta)$, representing the transcriptional extrinsic noise. The unknown distribution is estimated by using the kernel density estimation (KDE) method [50] and the estimate is denoted as $\hat{p}(\beta)$. For Model I (Eq. (9)), the kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\mathrm{off}}, \sigma_{\mathrm{on}}, \tau\}$ are the ones which we will infer; for the Poisson model (Eq. (16)), $\phi = \{\bar{\rho}, \tau\}$.

A. Generating-function-based parameter inference

Let the joint distribution of nuclear and cytoplasmic mRNA counts (n, m) for cell i according to a model be $P(n, m, \phi, \beta_i)$, and let the corresponding generating function for a single copy gene in cell i be $G(z_1, z_2, \phi, \beta_i)$. Assuming

each cell has two independent copies of each gene, the generating function used in Eq. (S59) for the cell population reads

$$\mathcal{G}(z_1, z_2, \phi) = \int_0^\infty G^2(z_1, z_2, \phi, \beta)\hat{p}(\beta)\mathrm{d}\beta. \tag{S63}$$

As before, to increase computational efficiency we utilize the Gauss quadrature method and transform the integral variable, thereby Eq. (S63) becomes

$$\mathcal{G}(z_1, z_2, \phi) = \frac{\beta_{\max} - \beta_{\min}}{2} \int_{-1}^{1} G^2(z_1, z_2, \phi, \beta_x)p(\beta_x)\mathrm{d}x, \tag{S64}$$

where

$$\beta_x = \frac{\beta_{\max} - \beta_{\min}}{2}x + \frac{\beta_{\max} + \beta_{\min}}{2}.$$

Here $\beta_{\max} = \max_i \beta_i$, $\beta_{\min} = \min_i \beta_i$ and x is the new integral variable. The numerical approximation to Eq. (S64) is then

$$\mathcal{G}(z_1, z_2, \phi) \approx \frac{\beta_{\max} - \beta_{\min}}{2} \sum_{j=1}^{N_\beta} w_j G^2(z_1, z_2, \phi, \beta_{x_j})p(\beta_{x_j}), \tag{S65}$$

where x_j is the j-th integral point of the N_β-th order Gauss quadrature approximation, and w_j is the corresponding integral weight, which is calculated based on Legendre polynomials and automatically computed by the Julia command `gausslegendre`. In Algorithm 2 we summarise the numerical recipe.

Algorithm 2 Generating-function-based parameter inference accounting for extrinsic noise and multiple gene copies

Input: Number of cells (n_c), the counts of nuclear and cytoplasmic mRNAs (n_i, m_i) and cellular size (V_i) for cell i
Output: Kinetic parameters ϕ.

```
Calculate the normalized cellular volume β_i by using Eq. (S62)
Use KDE to estimate p̂(β) according to the data {β_i}
Generate Gauss quadrature points and weights (x_j, w_j) for β by the command gausslegendre
Generate Gauss quadrature points and weights (y_{1,j}, y_{2,k}, w_j, w_k) for z_1, z_2 by the command gausslegendre
Compute the empirical generating function using Eq. (S58)
Initialize the inferred parameters φ
while Threshold reached do
    Compute the generating function for the inferred parameters φ by using Eq. (S65) where G(z_1, z_2, φ, β_{x_j}) is calculated
    from Eq. (10) or Eq. (S60)
    Compute the loss function J_PGF(φ) by using Eq. (S61)
    Update the inferred parameters φ by the Nelder-Mead optimization algorithm
end while
return Kinetic parameters φ
```

B. Finite-state-projection-based parameter inference

Given the n_c tuples of counts $\mathcal{D} = \{(n_i, m_i)\}$, the total likelihood of all observations is the product over every tuple, i.e.,

$$\mathcal{L}(\mathcal{D}|\phi) = \prod_{i=1}^{n_c} P(n_i, m_i|\phi).$$

The inference of the kinetic parameters proceeds by minimizing the negative log likelihood, that is

$$J_{\mathrm{FSP}}(\phi) = -\sum_{i=1}^{n_c} \log P(n_i, m_i|\phi). \tag{S66}$$

The optimization details of minimizing Eq. (S66) are the same as those of the inference method using the generating function. For Model I (Eq. (9)), we use finite state projection (FSP) to solve the joint probability for a single gene

copy and for a given β_{x_j} ($P(n, m, \beta_{x_j}|\phi)$) from Eqs. (S22), (S25) and (S27) in the steady state limit. The stationary distribution for the cell population is computed using

$$P(n, m|\phi) \approx \frac{\beta_{\max} - \beta_{\min}}{2} \sum_{j=1}^{N_\beta} w_j P^{*2}(n, m, \beta_{x_j}|\phi) p(\beta_{x_j}), \tag{S67}$$

for the same tuple (β_{x_j}, w_j) in Eq. (S65). The probability $P^{*2}(n, m, \beta_{x_j}|\phi)$ is the convolution of the probability $P(n, m, \beta_{x_j}|\phi)$. By plugging $P(n, m|\phi)$ into Eq. (S66), the negative log likelihood can be computed.

The numerical recipe for inference using FSP is summarised in Algorithm 3, and the optimization details are the same as those of the inference method using the generating function.

Algorithm 3 FSP-based parameter inference for Model I that accounts for extrinsic noise and multiple gene copies

Input: Number of cells (n_c), the counts of nuclear and cytoplasmic RNAs (n_i, m_i) and cellular length (V_i) for cell i
Output: Kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\text{off}}, \sigma_{\text{on}}, \tau\}$.

```
Calculate the normalized cellular volume β_i by using Eq. (S62)
Use KDE to estimate p̂(β) according to the data {β_i}
Generate Gauss quadrature points and weights (x_j, w_j) for β by the command gausslegendre
Initialize the inferred parameters φ
while Threshold reached do
    Compute the single-copy probability P(n, m, β_{x_j}|φ) for the inferred parameters φ and β_{x_j}
    Compute the single-cell probability P^{*2}(n, m, β_{x_j}|φ) by convolutions
    Compute the probability P(n, m, |φ) for the cell population by using Eq. (S67)
    Compute the loss function J_FSP(φ) by using Eq. (S66)
    Update the inferred parameters φ by the Nelder-Mead optimization algorithm
end while
return Kinetic parameters φ = {ρ̄, σ_off, σ_on, τ}
```

C. Method-of-moments-based parameter inference

We begin by deriving the first- and second-order moments for nuclear and cytoplasmic mRNA counts transcribed from two identical and independent gene copies in the presence of extrinsic noise. We here assume expression from each gene is described by Model I (9).

Let n_i and m_i denote the nuclear and cytoplasmic mRNA counts of gene copy i respectively. According to Eq. (S56), the corresponding means are

$$\begin{aligned} \langle n \rangle &= \langle n_1 + n_2 \rangle = \langle n_1 \rangle + \langle n_2 \rangle = 2\langle n_1 \rangle = 2\langle n_2 \rangle = \frac{2\bar{\rho}\langle \beta \rangle \sigma_{\text{on}} \tau}{\sigma_{\text{off}} + \sigma_{\text{on}}}, \\ \langle m \rangle &= \langle m_1 + m_2 \rangle = \langle m_1 \rangle + \langle n_2 \rangle = 2\langle n_1 \rangle = 2\langle n_2 \rangle = \frac{2\bar{\rho}\langle \beta \rangle \sigma_{\text{on}}}{d(\sigma_{\text{off}} + \sigma_{\text{on}})}. \end{aligned} \tag{S68}$$

Applying the law of total variance $\text{var}(Y) = \langle \text{var}(Y|Z) \rangle + \text{var}(\langle Y|Z \rangle)$, and the law of total covariance $\text{cov}(X, Y) = \langle \text{cov}(X, Y|Z) \rangle + \text{cov}(\langle X|Z \rangle, \langle Y|Z \rangle)$, we find that the second-order moments are given by

$$\begin{aligned} &\text{var}(n_1 + n_2) = \langle \text{var}(n_1 + n_2|\beta) \rangle + \text{var}(\langle n_1 + n_2|\beta \rangle), \\ &\text{var}(m_1 + m_2) = \langle \text{var}(m_1 + m_2|\beta) \rangle + \text{var}(\langle m_1 + m_2|\beta \rangle), \\ &\text{cov}(n_1 + n_2, m_1 + m_2) = \langle \text{cov}(n_1 + n_2, m_1 + m_2|\beta) \rangle + \text{cov}(\langle n_1 + n_2|\beta \rangle, \langle m_1 + m_2|\beta \rangle). \end{aligned} \tag{S69}$$

Since n_1 and n_2 are independent given a certain β, it follows that

$$\langle \text{var}(n_1 + n_2|\beta) \rangle = 2\langle \text{var}(n_1|\beta) \rangle. \tag{S70}$$

By similar arguments, we obtain that

$$\langle \text{var}(m_1 + m_2|\beta) \rangle = 2\langle \text{var}(m_1|\beta) \rangle, \qquad \langle \text{cov}(n_1 + n_2, m_1 + m_2|\beta) \rangle = 2\langle \text{cov}(n_1, m_1|\beta) \rangle. \tag{S71}$$

Next, using the linearity of the expectation operator, we find that

$$\begin{aligned}
&\mathrm{var}(\langle n_1+n_2|\beta\rangle) = \mathrm{var}(\langle n_1|\beta\rangle + \langle n_2|\beta\rangle) = \mathrm{var}(2\langle n_1|\beta\rangle) = 4\mathrm{var}(\langle n_1|\beta\rangle),\\
&\mathrm{var}(\langle m_1+m_2|\beta\rangle) = \mathrm{var}(\langle m_1|\beta\rangle + \langle m_2|\beta\rangle) = \mathrm{var}(2\langle m_1|\beta\rangle) = 4\mathrm{var}(\langle m_1|\beta\rangle),\\
&\mathrm{cov}(\langle n_1+n_2|\beta\rangle, \langle m_1+m_2|\beta\rangle) = \mathrm{cov}(2\langle n_1|\beta\rangle, 2\langle m_1|\beta\rangle) = 4\mathrm{cov}(\langle n_1|\beta\rangle, \langle m_1|\beta\rangle).
\end{aligned} \tag{S72}$$

Combining Eqs. (S69)-(S72), we obtain

$$\begin{aligned}
&\mathrm{var}(n) = \mathrm{var}(n_1+n_2) = 2\langle \mathrm{var}(n_1|\beta)\rangle + 4\mathrm{var}(\langle n_1|\beta\rangle),\\
&\mathrm{var}(m) = \mathrm{var}(m_1+m_2) = 2\langle \mathrm{var}(m_1|\beta)\rangle + 4\mathrm{var}(\langle m_1|\beta\rangle),\\
&\mathrm{cov}(n,m) = \mathrm{cov}(n_1+n_2, m_1+m_2) = 2\langle \mathrm{cov}(n_1,m_1|\beta)\rangle + 4\mathrm{cov}(\langle n_1|\beta\rangle, \langle m_1|\beta\rangle).
\end{aligned} \tag{S73}$$

The terms on the RHS of Eq. (S73) are modified from those in Eqs. (S56) and (S57). These are given by

$$\begin{aligned}
&\langle \mathrm{var}(n_1|\beta)\rangle = \frac{\bar{\rho}\langle\beta\rangle\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}} + \frac{2\bar{\rho}^2\langle\beta^2\rangle\sigma_{\mathrm{on}}\sigma_{\mathrm{off}}\left[-1+\mathrm{e}^{-(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau}+(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau\right]}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^4},\\
&\mathrm{var}(\langle n_1|\beta\rangle) = \left(\frac{\bar{\rho}\sigma_{\mathrm{on}}\tau}{\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}}}\right)^2 \mathrm{var}(\beta),\\
&\langle \mathrm{var}(m_1|\beta)\rangle = \frac{\bar{\rho}\langle\beta\rangle\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})} + \frac{\bar{\rho}^2\langle\beta^2\rangle\sigma_{\mathrm{on}}\sigma_{\mathrm{off}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})},\\
&\mathrm{var}(\langle m_1|\beta\rangle) = \left[\frac{\bar{\rho}\sigma_{\mathrm{on}}}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})}\right]^2 \mathrm{var}(\beta),\\
&\langle \mathrm{cov}(n_1,m_1|\beta)\rangle = \frac{\bar{\rho}^2\langle\beta^2\rangle\sigma_{\mathrm{on}}\sigma_{\mathrm{off}}\left[1-\mathrm{e}^{-(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})\tau}\right]}{(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^3(d+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})},\\
&\mathrm{cov}(\langle n_1|\beta\rangle, \langle m_1|\beta\rangle) = \frac{\bar{\rho}^2\sigma_{\mathrm{on}}^2\tau}{d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})^2}\mathrm{var}(\beta).
\end{aligned} \tag{S74}$$

Parameter inference is performed by optimizing the loss function, which is given by

$$J_{\mathrm{MOM}}(\phi) = \left(\frac{\langle n\rangle - \langle n\rangle_D}{\langle n\rangle_D}\right)^2 + \left(\frac{\langle m\rangle - \langle m\rangle_D}{\langle m\rangle_D}\right)^2 + \left(\frac{\mathrm{var}(n)-\mathrm{var}_D(n)}{\mathrm{var}_D(m)}\right)^2 + \left(\frac{\mathrm{var}(m)-\mathrm{var}_D(m)}{\mathrm{var}_D(m)}\right)^2 + \left(\frac{\mathrm{cov}(n,m)-\mathrm{cov}_D(n,m)}{\mathrm{cov}_D(n,m)}\right)^2,$$

where the model's moments are given by Eq. (S68), Eq. (S73) and Eq. (S74), while the moments with subscript D are those directly computed from data. This optimization problem is also solved by the Nelder-Mead algorithm.

IX. Processing of MERFISH data

We sourced total mRNA counts and nuclear mRNA counts from the files `pnas.1912459116.sd12.csv` and `pnas.1912459116.sd14.csv` of Ref. [14]. There are 10,050 genes across a population of $n_c = 1368$ human osteosarcoma (U-2 OS) cells reported in the dataset. By excluding genes for which the total mRNA counts < 1, only 6724 genes remained. Afterwards, we processed the data from each gene into the tuples of nuclear-cytoplasmic mRNA counts (n_i, m_i) where i is the cell index. The number of tuples varied from gene to gene because for each gene we removed those cells whose nuclear or cytoplasmic counts were outliers; specifically from the data for gene j, we removed those cells whose counts for gene j were outside of the range $(\mathrm{mean} - 5\ \mathrm{std}, \mathrm{mean} + 5\ \mathrm{std})$ where the mean and standard deviation of counts for gene j are computed over the whole cell population.

X. Generating-function-based model selection using 10-fold cross validation and ontology analysis

The purpose of model selection in this context is to identify the type of gene activity. Since the generating-function-based inference is not based on the conventional maximum likelihood method, it is difficult to see how any generating-function-based model selection method can be based on information criterion (such as the Akaike or Bayesian Information Criteria).

To avoid this difficulty, we propose to use 10-fold cross validation. We start by arranging candidate models in order of increasing complexity (number of parameters) and dividing the cell data into 10 equal-sized subsamples. Inference is repeated 10 times, where for each time, nine of the subsamples are used to infer parameters for candidate models, while the tenth is used to compute a performance score $J(\hat{\phi}_i)$ using Eq. (14). This process generates a score vector $\mathcal{J}_{\text{model}_j} = [J(\hat{\phi}_1), \cdots, J(\hat{\phi}_{10})]^\top$ for each model j (this index increases with model complexity as described earlier). We denote the index of the proposed best model, i.e. that with the smallest mean score, by $j = \mathcal{I}$. The mean and standard deviation of this model's score vector ($\mathcal{J}_{\text{best}} = \mathcal{J}_{\text{model}_\mathcal{I}}$) are denoted $\bar{\mathcal{J}}_{\text{best}}$ and σ_{best}, respectively. To balance model complexity with fitting accuracy, we need to compare the proposed best model with simpler models having a smaller number of parameters. Defining $\bar{\mathcal{J}}_{\text{model}_j}$ as the mean of the score vector for model j, for models with $j < \mathcal{I}$, we apply the inequality

$$\bar{\mathcal{J}}_{\text{model}_j} \leq \bar{\mathcal{J}}_{\text{best}} + \sigma_{\text{best}} \sqrt{1 - \text{cor}(\mathcal{J}_{\text{best}}, \mathcal{J}_{\text{model}_j})}, \tag{S75}$$

selecting the model with the smallest j that satisfies Eq. (S75) as the final best model. Note that $\text{cor}(x, y)$ stands for the correlation coefficient of two score vectors x and y.

We summarise the entire model selection procedure in the following algorithm.

Algorithm 4 Generating-function-based model selection method using 10-fold cross validation

Input: Number of cells (n_c), the equal sized counts data $\{\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_{10}\}$, the set of candidate models $\{\text{model}_1, \cdots, \text{model}_n\}$ ordered by the model complexity (the number of kinetic parameters)
Output: Best-fitting model

```
for each candidate model model_i do
    for each fold j do
        Use Algorithm 2 to infer kinetic parameters φ̂_j based on the data {D_1, ···, D_j, D_{j+1}, ···, D_10}
        Compute the performance score J_PGF(φ̂_j) on the validation dataset D_j
    end for
    Collect all the performance scores for model_i J_model_i = [J_PGF(φ̂_1), ···, J_PGF(φ̂_10)]^⊤
    Compute the corresponding mean (J̄_model_i) and standard deviation σ_model_i
end for
Find the minimal performance score J̄_best and its index I
for each candidate model model_i and i ≤ I do
    Calculate the correlation coefficient cor(J_best, J_model_i) of the performance score vectors of the best model and model_i
    Calculate the threshold of performance score thresh_model_i using the RHS of Eq. (S75)
    if J̄_model_i ≤ thresh_model_i then
        Accept model_i as the best-fitting model and relabel the index to ℓ
        Break
    end if
end for
return Best-fitting model_ℓ
```

For the MERFISH data, we apply two candidate models — Poisson (Eq. (16)) and Model I (Eq. (9)) — to identify gene activity. The PGFs for a single gene copy without extrinsic noise are provided in Eqs. (S60), and (10), respectively. Using Eq. (S65) and these PGFs, we compute the PGF accounting for extrinsic noise and multiple gene copies, enabling the application of Algorithm 4. Notably, we exclude the refractory model (Model II) as a candidate model due to its low accuracy in being correctly selected when using sample sizes of $\sim 10^3 - 10^4$ cells (see Fig. S4).

Given the genes classified according to one of 2 best fitting models, subsequently we utilized the DAVID software [57] for gene ontology (GO) analysis of each of the 2 gene categories. The analysis was contextualized against a background of filtered 6724 genes. GO terms were extracted from the `GOTERM_BP_FAT` category. Our criteria for inclusion were a Benjamini p-value (p_{adj}) of less than 0.01 and a minimum of 15 enriched genes per term. These terms were then ranked based on Gene Ratio. The comprehensive results are detailed in Excel files stored in Github.

XI. Code and data availability

All codes and data have been deposited at https://github.com/edwardcao3026/pgf-inf.

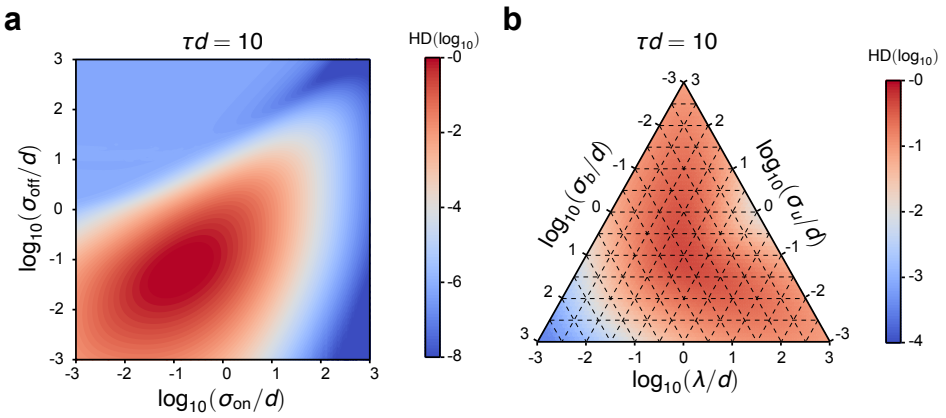

FIG. S1. Hellinger distance (HD) between the joint distributions of non-Markovian models (Models I and II) and Markovian Models (Model III and IV) as a function of the normalised gene state switching rates σ_{on}/d and σ_{off}/d (or σ_{u}/d, σ_{b}/d and λ/d), and the normalised export time $\tau d = 10$ (ρ is fixed to 5). (a) Comparison of Model I vs Model III; (b) Comparison of Model II vs Model IV.

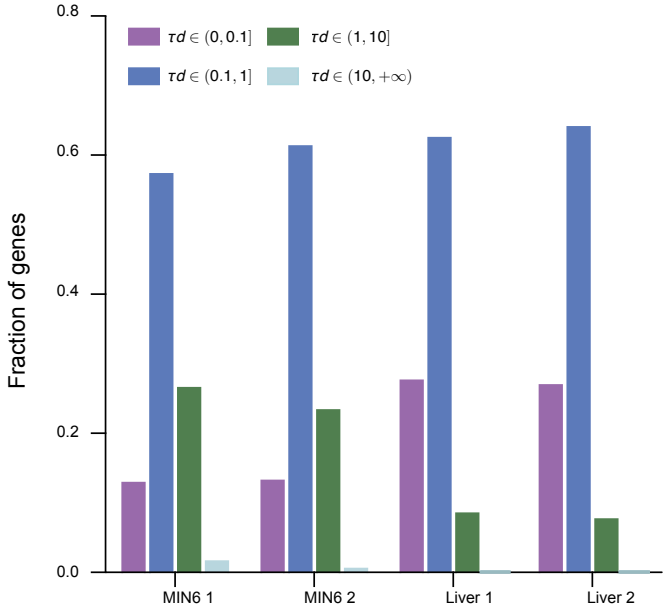

FIG. S2. Distributions of τd estimated from mRNA data measured in two samples of MIN6 cells and liver cells.

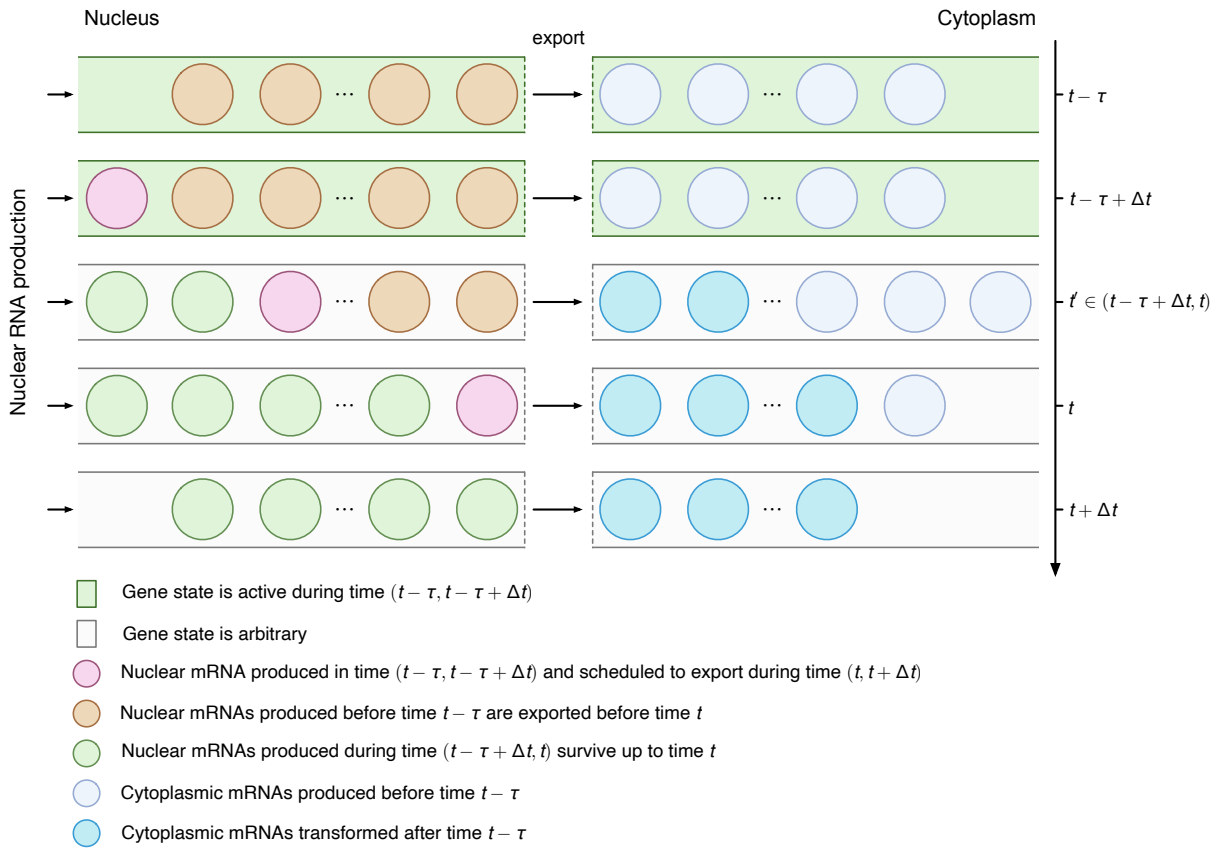

FIG. S3. Illustration of the chain of events culminating in the export of a nuclear mRNA in Model I (refer to event (C)). Let's assume there are n' nuclear mRNAs (brown balls) and m' cytoplasmic mRNAs (purple balls) produced before time $t-\tau$. In the next time step $t-\tau+\Delta t$, there is a newborn nuclear mRNA (red ball). During $(t-\tau+\Delta t, t)$, the first n' brown mRNAs are progressively exported and turned into cytoplasmic mRNAs (blue balls) and in the meanwhile n new nuclear mRNAs (green balls) are produced. Since each nuclear RNA stays for a fixed time τ, the red RNA definitely leaves in the time interval $(t, t+\Delta t)$. Note that during the time interval $(t-\tau+\Delta t, t)$, the red RNA does not participate in any dynamics of the system which explains Eq. (S15) and the dynamics of the green mRNAs follows Eq. (S22) which is a simple telegraph model without decay from $(0, \tau)$.

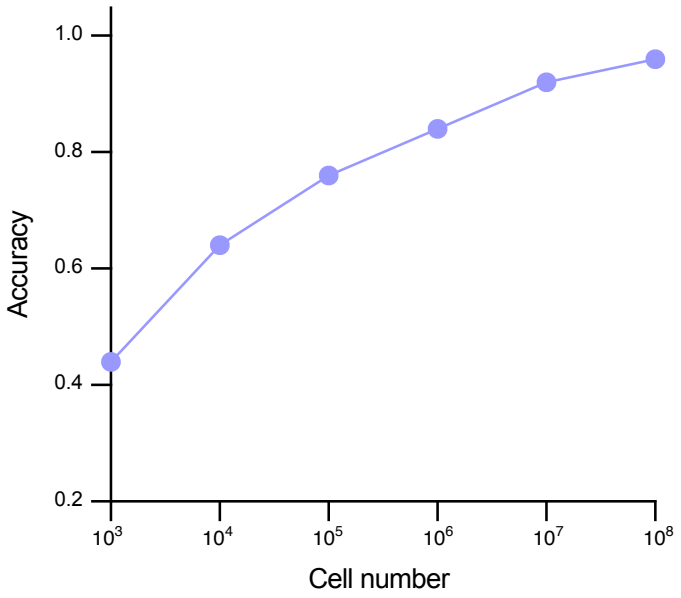

FIG. S4. Accuracy of identifying the true model (Model II, Eq. (12)) as a function of the number of cells in the sample (n_c). The counts data are generated by simulating Model II using the delay SSA. The kinetic parameters are shown in Table S4. The model selection procedure in SM X is applied to the data to select between two candidate models: Models I and II. The accuracy of identifying the correct model increases with the number of cells. However, in a typical sample size $n_c = 10^3 - 10^4$, the two competing models are difficult to discern. Note that for simplicity, here we assumed that all noise is intrinsic and there is a single gene copy; a similar outcome is expected in the presence of extrinsic noise and for multiple gene copies.

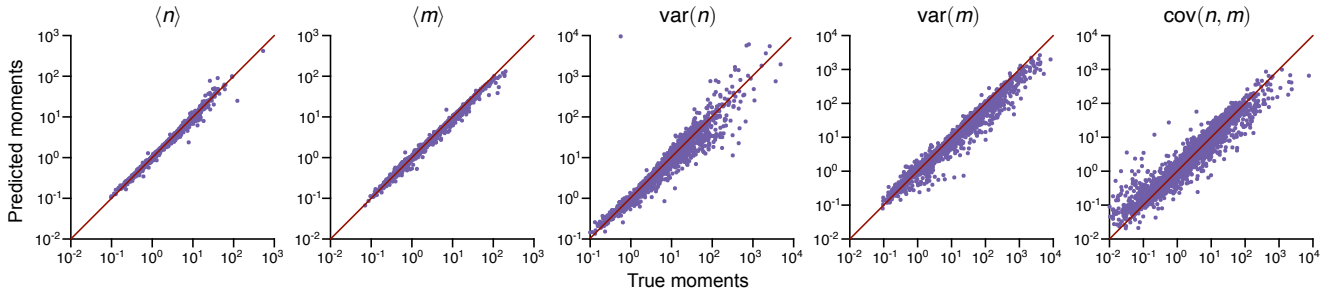

FIG. S5. Accuracy of the first- and second-order moments predicted by fitting Model I to the mRNA count data from 2,722 non-Poissonian genes in the MERFISH dataset. The moments are calculated using Eqs. (S68), (S73), and (S74) and compared with the true moments from experimental data. 2,184 genes meet the criterion that the average relative error across the five moments is less than 35%.

TABLE S1. Kinetic parameters used in the P-P plots of Fig. 2a. The degradation rate d is set to 1.

	ρ/d	σ_{off}/d	σ_{on}/d	σ_{u}/d	λ/d	σ_{b}/d	τd
Set 1	120	100	2	/	/	/	0.8
Set 2	5	0.041	0.2	/	/	/	0.8
Set 3	20	0.1	4.5	/	/	/	0.8
Set 4	5	/	/	6	0.3	0.05	3
Set 5	10	/	/	0.9	5	0.5	3
Set 6	15	/	/	1	1	0.4	2

TABLE S2. Inferred and true kinetic parameters from simulated data – data summarised in Fig. 3b. The true parameters are uniformly sampled from $\bar{\rho}/d \in [0, 80]$, $\sigma_{\text{on}}/d \in [0, 2]$ and $\sigma_{\text{off}}/d \in [0, 40]$, — these are roughly in the range estimated for various types of eukaryotic cells in Table I of Ref. [58]. The parameter $d\tau$ is sampled from $d\tau \in [0, 1]$ since about 80% of genes fall in this range according to Fig. S2.

	PGF				FSP				True			
	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd
Set 1	57.46	24.31	2.01	0.14	69.08	28.07	1.91	0.15	60.97	24.71	1.94	0.15
Set 2	29.19	20.53	0.49	0.66	24.75	19.33	0.53	0.63	27.14	19.48	0.52	0.61
Set 3	36.53	9.51	0.1	0.28	36.64	7.05	0.08	0.34	40.62	8.84	0.09	0.32
Set 4	74.48	37.82	0.92	0.26	52.06	22.88	0.85	0.27	73.26	30.41	0.79	0.28
Set 5	24.29	13.17	0.25	0.57	18.36	8.69	0.23	0.58	25.01	11.87	0.22	0.57
Set 6	3.90	0.11	0.38	0.60	3.79	0.08	0.32	0.59	4.06	0.16	0.42	0.61
Set 7	39.73	22.60	1.41	0.20	58.11	40.48	1.61	0.18	44.07	28.13	1.43	0.18
Set 8	19.28	32.81	1.2	0.8	9.35	14.02	1.12	0.8	18.53	31.42	1.18	0.82
Set 9	51.12	19.68	1.42	0.5	49.9	18.51	1.38	0.51	51.13	19.59	1.42	0.51
Set 10	35.36	4.56	1.12	0.26	41.34	5.41	1.45	0.28	35.34	4.59	1.08	0.26
Set 11	23.23	9.37	0.56	0.64	23.94	11.58	0.62	0.62	26.67	12.13	0.61	0.58
Set 12	4.52	0.25	0.19	0.58	4.26	0.22	0.18	0.57	4.44	0.23	0.17	0.57
Set 13	32.28	16.63	0.40	0.73	25.91	14.09	0.42	0.67	27.14	15.04	0.43	0.61
Set 14	26.11	7.50	0.35	0.19	32.86	9.29	0.35	0.20	39.14	10.04	0.31	0.17
Set 15	49.8	21.16	1.55	0.38	46.11	19.66	1.56	0.38	47.25	20.94	1.61	0.38

TABLE S3. Inferred and true kinetic parameters and extrinsic nose from simulated data – data summarised in Fig. 3c. The true parameters are uniformly sampled from $\bar{\rho}/d \in [0, 25]$, $\sigma_{\text{on}}/d \in [0, 10]$, $\sigma_{\text{off}}/d \in [0, 10]$, $\tau d \in [0, 1]$. The extrinsic noise is sampled as $\beta \in \text{Gamma}(40, 0.025)$.

	Joint				Moments				True			
	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd
Set 1	7.132	3.838	2.041	0.745	6.644	2.922	1.743	0.746	7.1	3.67	1.98	0.711
Set 2	18.211	1.195	6.462	0.28	22.445	3.96	16.384	0.277	17.915	1.677	9.08	0.282
Set 3	14.508	1.256	4.792	0.342	13.46	0.539	2.859	0.359	14.225	1.06	4.739	0.344
Set 4	14.095	6.335	3.533	0.461	12.864	4.631	2.919	0.478	15.531	6.961	3.29	0.48
Set 5	25.567	2.773	2.825	0.986	26.083	3.441	3.303	0.987	24.335	2.45	2.812	0.956
Set 6	23.553	5.498	1.056	0.506	21.522	5.004	1.057	0.503	22.538	5.162	1.046	0.494
Set 7	18.991	1.334	6.041	0.276	17.535	0.486	3.243	0.29	19.29	1.651	6.965	0.275
Set 8	4.792	10.121	2.1	0.495	4.949	9.467	1.874	0.514	4.472	9.755	2.244	0.502
Set 9	11.499	4.763	2.836	0.553	10.97	4.835	3.067	0.552	11.128	4.812	2.983	0.558
Set 10	4.207	7.81	4.273	0.202	2.988	3.328	3.199	0.201	3.895	8.666	5.451	0.186
Set 11	8.45	1.37	0.9	0.679	8.498	1.394	0.885	0.685	8.469	1.4	0.898	0.656
Set 12	8.174	4.189	5.198	0.383	7.597	2.713	3.785	0.416	7.904	3.399	4.411	0.391
Set 13	7.88	4.614	1.101	0.529	8.743	5.835	1.254	0.504	7.555	4.562	1.128	0.527
Set 14	6.257	0.187	1.606	0.421	6.423	0.377	2.402	0.429	6.494	0.389	2.457	0.417
Set 15	1.273	1.624	0.737	0.559	1.344	1.618	0.733	0.517	1.531	2.253	0.723	0.559

TABLE S4. The sets of kinetic parameters used for generating counts data for model selection in Fig. S4. These parameters are sampled from $\rho/d \in [0, 30]$, $\sigma_{\mathrm{u}}/d \in [0, 10]$, $\lambda/d \in [0, 10]$, $\sigma_{\mathrm{b}}/d \in [0, 10]$, $\tau d \in [0, 5]$ and $d = 1$.

	Parameters				
	ρ/d	σ_{u}/d	λ/d	σ_{b}/d	τd
Set 1	7.218	5.533	7.089	7.148	0.286
Set 2	6.575	2.567	4.518	8.341	4.598
Set 3	8.329	4.982	4.374	5.063	4.564
Set 4	9.487	6.055	6.067	6.366	3.178
Set 5	13.404	5.804	1.592	8.433	3.625
Set 6	15.858	1.573	7.57	9.463	1.914
Set 7	10.974	4.815	5.871	3.92	2.05
Set 8	27.07	3.895	4.48	6.472	2.668
Set 9	2.453	3.566	8.616	5.964	3.893
Set 10	14.149	5.464	1.975	9.687	4.062
Set 11	0.675	9.511	2.835	5.225	3.558
Set 12	28.847	7.508	6.949	2.481	4.648
Set 13	3.022	1.967	9.764	1.52	0.777
Set 14	1.463	0.597	9.872	1.859	0.245
Set 15	9.632	8.053	7.964	2.949	0.092
Set 16	12.865	5.673	5.804	4.397	2.641
Set 17	8.296	2.601	0.968	7.805	4.486
Set 18	19.2	9.526	5.221	0.445	2.083
Set 19	5.62	5.143	1.069	9.366	1.774
Set 20	2.545	9.837	3.254	3.468	2.365
Set 21	14.819	1.984	0.524	7.203	0.11
Set 22	16.171	3.753	2.294	2.984	0.889
Set 23	16.479	2.499	9.822	4.521	0.796
Set 24	24.963	8.934	7.695	0.67	0.69
Set 25	19.399	8.837	9.969	5.244	0.724

TABLE S5. Dependence of the results of the model selection and parameter inference procedure on various assumptions and the type of input single-cell data. The first column shows whether the model accounts for both intrinsic and extrinsic noise (tick) or only intrinsic noise (cross). The second column is the assumed number of gene copies; for eukaryotic genes, this can be 2 or 4 depending on the position in the cell-cycle (information that is not available for the MERFISH dataset). The third column shows the type of data used as input to the inference algorithm: total counts (sum of nuclear and cytoplasmic counts) or joint data. The fourth, fifth and sixth columns show the medians of the estimated f_{on}, burst size and burst frequency for Model I. The number of genes that are best fit by the Poisson model and Model I are shown in columns 7 and 8, respectively. The number of non-Poissonian genes that pass the quality criteria (similar to Fig. S5) are shown in the last column; the parameters for Model I are only estimated for these genes. When performing inference on total count data, we set $z_1 = 1$ in Eq. (10); this reduces the generating function solution of Model I to that of the conventional telegraph model [12]. This choice conforms with the common practice of interpreting the mRNA count prediction of the telegraph model as the total mRNA per cell. For experimental sets where extrinsic noise was not considered, we fixed $\beta = 1$ in Algorithm 2. In experimental sets involving four gene copies, we set the exponent of G in Eq. (S65) to 4.

	Extrinsic noise	**Gene copy numbers**	**Data type**	f_{on}	**Burst size**	**Burst frequency**	**# of Poissonian genes**	**# of non-Poissonian genes**	**# of non-Poissonian genes passing fitting quality check**
Set 1	×	2	Joint	0.66	7.84	0.59	617	6107	4811
Set 2	×	2	Total	0.19	3.05	1.59	626	6098	6056
Set 3	√	2	Joint	0.64	6.83	0.69	4022	2722	2184
Set 4	√	2	Total	0.40	5.41	1.14	4172	2552	2523
Set 5	×	4	Joint	0.50	6.52	0.29	624	6100	5389
Set 6	×	4	Total	0.15	3.10	0.76	622	6102	6069
Set 7	√	4	Joint	0.44	3.88	0.39	3987	2737	2241
Set 8	√	4	Total	0.34	6.35	0.48	4169	2555	2527