PLOS Computational Biology

RESEARCH ARTICLE

# From noise to models to numbers: Evaluating negative binomial models and parameter estimations in single-cell RNA-seq

**Yiling Wang**[1☯], **Zhanpeng Shu**[2☯], **Zhixing Cao**[1,3*], **Ramon Grima**[4*]

**1** State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China, **2** School of Electrical Engineering, Shanghai Dianji University, Shanghai, China, **3** Department of Chemical Engineering, Queen's University, Kingston, Canada, **4** School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

☯These authors contributed equally to this work.
* z.cao@queensu.ca (ZC); ramon.grima@ed.ac.uk (RG)



OPEN ACCESS

**Citation:** Wang Y, Shu Z, Cao Z, Grima R (2026) From noise to models to numbers: Evaluating negative binomial models and parameter estimations in single-cell RNA-seq. PLoS Comput Biol 22(3): e1014014. https://doi.org/10.1371/journal.pcbi.1014014

**Editor:** Ilya Ioshikhes, Ottawa, CANADA

**Received:** January 23, 2026

**Accepted:** February 12, 2026

**Published:** March 16, 2026



**Data availability statement:** The code for this paper is available at https://github.com/quark0211/aeBIC.

**Funding:** This work was supported by Shanghai Action Plan for Technological Innovation Grant (23S41900500 to ZC), the Natural Science and Engineering Research Council of Canada Discovery Grant (RGPIN-2024-06015 to ZC),

## Abstract

The Negative Binomial (NB) distribution is widely used to approximate transcript count distributions in single-cell RNA sequencing (scRNA-seq) data, yet the reason for its ubiquity is not fully understood. Here, we employ a computationally efficient model selection technique to map the relationship between the best-fit models – Beta-Poisson (Telegraph), NB, and Poisson – and the kinetic parameters that govern gene expression stochasticity. Our findings reveal that the NB distribution closely approximates simulated data (incorporating both biological and technical noise) within an intermediate range of the sum of the gene activation and inactivation rates normalized by the mRNA degradation rate. This range expands with decreasing mean expression, increasing technical noise, and larger sample sizes. The results imply that: (i) good NB fits occur in diverse parameter regimes without exclusively indicating transcriptional bursting; (ii) for small sample sizes, biological noise predominantly shapes the NB profile even when technical noise is present; (iii) under steady-state conditions, gene-specific parameters (burst size and frequency) estimated in regions where the NB model fits well, typically show large relative errors, even after corrections for technical noise, and (iv) gene ranking by burst frequency remains reliably accurate, suggesting that burst parameters are most informative in a relative sense. Finally, applying technical-noise–corrected model fitting to scRNA-seq data confirms that a substantial fraction of mammalian genes fall within these NB-fitting regimes, despite lacking transcriptional bursting.

## Author summary

Single-cell RNA sequencing (scRNA-seq) measures mRNA molecule counts in individual cells. For most genes, these counts are well fit by a negative binomial

and the Leverhulme Trust (RPG-2024-082 to RG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

(NB) distribution, and NB fits are often interpreted as evidence for transcriptional bursting. We asked when an NB model is expected to arise from a mechanistic gene-expression process, and what biological meaning can be safely assigned to its parameters. We combine the standard two-state telegraph model of promoter switching with a binomial model of transcript capture, and introduce the approximate expected Bayesian information criterion (aeBIC). aeBIC predicts which distribution—telegraph, NB, or Poisson—would be chosen by likelihood/BIC model selection. We show that NB fits are optimal in an intermediate regime of promoter switching relative to mRNA decay, and that this regime expands for low mean expression, larger sample sizes, and increased cell-to-cell variability in capture probability. Consequently, excellent NB fits can occur well outside the classical bursting limit. In these regimes, estimating burst size and burst frequency from NB parameters can incur large absolute errors, although relative comparisons are more robust: ranking genes by inferred burst frequency is usually preserved. Our results provide practical guidance for model choice and for interpreting fitted burst parameters in single-cell genomics.

## 1 Introduction

Single-cell RNA sequencing (scRNA-seq) facilitates the quantitative characterization of cellular heterogeneity by providing genome-wide transcriptomic profiles at single-cell resolution across hundreds or thousands of individual cells [1–4]. The use of unique molecular identifiers (UMIs) has substantially improved the reliability of sequencing data by enabling the distinction between original RNA molecules and PCR duplicates. By tagging each molecule with a unique barcode prior to amplification, UMIs mitigate amplification biases and sequencing artifacts, resulting in molecular counts that more accurately reflect true biological abundance [5]. The transcript count distribution of most genes is unimodal and well approximated by a negative binomial (NB) distribution, a two-parameter model often interpreted as a Gamma–Poisson mixture, where one parameter reflects the number of successes and the other the success probability [6–11]. The significance of the NB distribution is highlighted by its extensive application in various computational tools for scRNA-seq analysis [10,12–19]. The fitting of the NB distribution to data yields two parameters for each gene, providing a convenient numerical signature of its observed stochasticity. However, the origin of the universality of the NB distribution in single-cell sequencing data remains unclear. In fact, this distribution is also commonly found using single-molecule fluorescence in situ hybridization (smFISH), suggesting that its ubiquity cannot be simply dismissed as stemming from the significant technical noise of sequencing technologies [20].

Over the past two decades, using fluorescence-based single-cell transcriptomics, it has been shown that biological noise for each gene is well described by the two-state telegraph model [21–27] or by a three or more-state extension of this model [28–30]. Here, we focus on the telegraph model because it has been shown to very well

approximate the steady-state distribution of mRNA counts predicted by models with a higher number of states [31–33]. The telegraph model describes a gene that switches between an active state ($G$) and an inactive state ($G^*$). Transcription occurs only from the active state $G$, and subsequently mRNA decay occurs with first-order kinetics. For an illustration, see Fig 1a. The solution of the chemical master equation that describes the telegraph model in steady-state conditions leads to a Beta-Poisson distribution of mRNA counts [34–36]. Typically, scRNA-seq measurements detect a significantly smaller fraction of transcripts than smFISH and hence extending the telegraph model to also account for technical noise is important. It has been shown that even after this modification, the steady-state distribution of transcript counts in a single cell follows a Beta-Poisson distribution [37,38] — technical noise simply leads to a rescaling of the effective transcription rate by the probability of transcript capture in the cell, but does not change the type of distribution. Depending on the parameter values, the Beta-Poisson distribution can be unimodal with a peak at zero, unimodal with a peak at a non-zero value, or bimodal with peaks at zero and non-zero values [39]. All of these distribution shapes have been measured using smFISH [20,40,41]. Hence, clearly, the classical model of stochastic gene expression, even when extended to account for technical noise, does not generally predict an NB distribution. There are two known conditions under which the Beta-Poisson distribution reduces to an NB distribution.

Case (i). Say that a gene is always in the active state, i.e., there is no gene state switching. The telegraph model predicts that, in each cell, the transcript counts are sampled from a Poisson distribution with a parameter given by the effective transcription rate divided by the degradation rate [42]. If the effective transcription rate, which is equal to the product of the transcription rate and the probability of transcript capture, is the same in each cell, then the distribution of transcript counts across all cells is

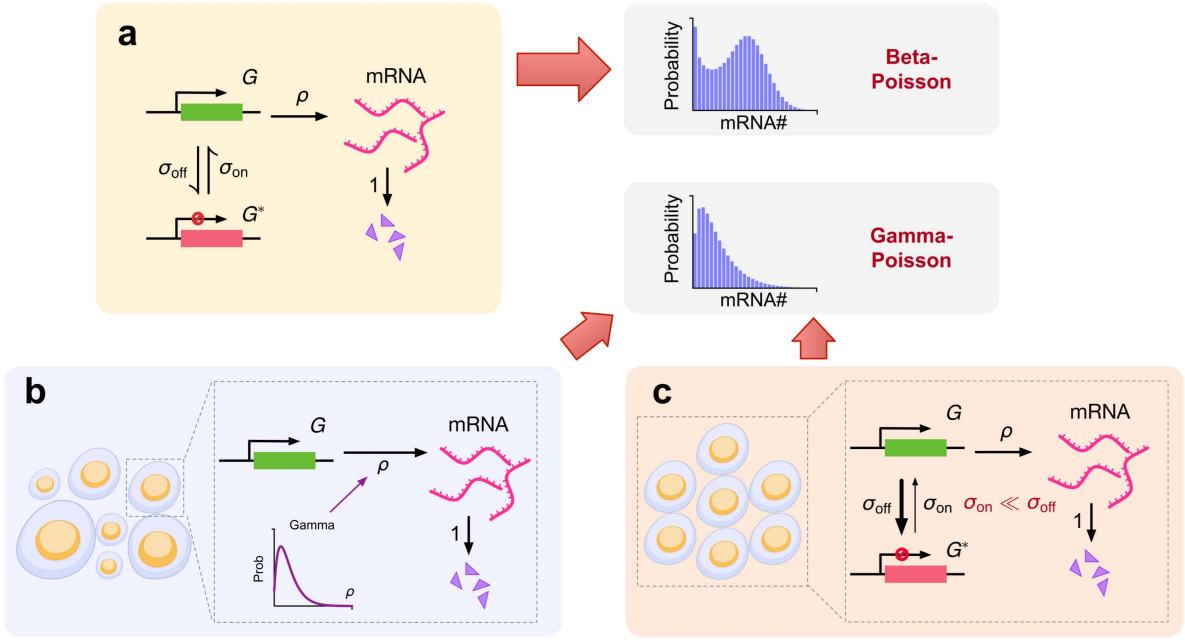


**Fig 1. Schematic comparison of the telegraph model of gene expression and two of its limiting cases, illustrating how distinct mechanisms can converge to the same negative binomial mRNA distribution.** (a) Schematic illustrating the telegraph model of gene expression. A gene switches between active (green) and inactive states (red) with rates $\sigma_{on}$ and $\sigma_{off}$. Synthesis of transcripts occurs from the active state with rate $\rho$. The transcripts are subsequently degraded with rate *1*. The rates are all normalized by the degradation rate. The steady-state distribution of transcript numbers is a Beta-Poisson compound distribution; (b) Schematic showing the special case where the gene is always in the active state and the transcription rate $\rho$ varies from cell to cell according to a Gamma distribution. In this case, the mRNA distribution predicted by the telegraph model reduces to a Gamma-Poisson compound distribution (an NB distribution); (c) Schematic showing the special case where the gene spends most of its time in the inactive state ($\sigma_{on} \ll \sigma_{off}$) which leads to transcription occurring in short-lived bursts that are well separated from each other. All cells are identical, i.e., the rate constants do not vary from cell to cell. In this case, the mRNA distribution predicted by the telegraph model also reduces to an NB distribution.

https://doi.org/10.1371/journal.pcbi.1014014.g001

necessarily Poisson. However, if the effective transcription rate varies from cell to cell according to a Gamma distribution, then it immediately follows that the observed distribution is a Gamma-Poisson compound distribution, which is an NB distribution. Hence, in this case, the NB character of the distribution stems from extrinsic (cell-to-cell variation in the transcription rate) or technical noise (cell-to-cell variation in the transcript capture probability), and not from the intrinsic dynamics of a gene. For an illustration, see Fig 1b. This case is unlikely to be the main reason for the universality of the NB in scRNA-seq data because the visualization of transcription in living cells using live-cell imaging conclusively shows alternating periods of gene activity and inactivity [43]. This is due to many factors such as reversible binding of transcription factors, enhancer-promoter interactions, the clustering dynamics of Pol II and the opening and closing of chromatin [44–48].

Case (ii). Now consider the case where the gene spends most of its time in the inactive state and synthesizes mRNA in a burst, i.e., in the short time that the gene is active. This occurs when the gene inactivation rate is much larger than the gene activation rate. This phenomenon is often referred to as transcriptional bursting [36]; for an illustration, see Fig 1c. If the effective transcription rate is the same in each cell, then the Beta-Poisson distribution of the telegraph model has been shown to reduce to an NB distribution [49]. This case has sometimes been used to explain how the NB distribution of mRNA counts in sequencing data can arise from a physical model of gene expression [11]. However, this is also unlikely to explain the universality of the NB distribution in scRNA-seq data because the constraint that the activation rate is much lower than the inactivation rate represents a very small part of the available parameter space. Also because the assumption that the effective transcription rate is the same in all cells implies negligible variation in the probability of mRNA capture from one cell to another — which is difficult to reconcile with the wide distribution of total UMI counts (from all genes) per cell in typical scRNA-seq datasets [17,50,51].

In summary, there is no strong theoretical reason behind the universality of the NB distribution in scRNA-seq data. The two existing results in the literature on stochastic gene expression can only explain the NB distribution by invoking strong assumptions that are not realistic. Furthermore, it remains unclear what biological or biophysical meaning is to be imparted to the two parameters of the NB distribution. Currently, this meaning is only clear when gene expression is bursty (as in Case (ii)), in which case after correcting for technical noise, the two parameters can be used to extract the burst frequency (the rate at which mRNA bursts are transcribed) and the burst size (the mean number of mRNA produced when the gene is active) [52].

In this paper, we overcome these limitations by deriving new parametric conditions under which the observed distribution of transcript counts is well approximated by an NB distribution. We find that these conditions include, but are not limited to, transcriptional bursting. We also show that while the absolute values of the burst frequency and the burst size cannot generally be reliably extracted from the two parameters of the NB distribution, nevertheless these still contain useful information about differential gene expression.

## 2 Results

### 2.1 The NB distribution can provide a good approximation to the mRNA distribution of the telegraph model even when transcriptional bursting is absent

Consider an idealized scenario where each cell is identical to each other, i.e., for a given gene, there is no cell-to-cell variation in the parameters of the telegraph model (Fig 1a). This implies that all cells are of the same type. Furthermore, assume that all transcripts from each cell can be detected. In this case, the distribution of mRNA counts is given by the steady-state solution of the chemical master equation of the telegraph model:

$$P(n) = \frac{\rho^n}{n!}\frac{(\sigma_{on})_n}{(\sigma_{on}+\sigma_{off})_n}{}_1F_1(\sigma_{on}+n, \sigma_{on}+\sigma_{off}+n, -\rho), \tag{1}$$

where $n$ represents the mRNA count, $(x)_n = \prod_{k=0}^{n-1}(x-k)$ denotes the Pochhammer symbol, and ${}_1F_1(a,b,z) = \sum_{n=0}^{\infty}\frac{(a)_n z^n}{(b)_n n!}$ is the Kummer confluent hypergeometric function [34]. Note that $\sigma_{on}$ is the gene activation rate, $\sigma_{off}$ is the gene deactivation

rate, $\rho$ is the transcription rate; these rates are non-dimensional because they are normalized by the degradation rate (this convention will be used throughout the paper). Eq (1) is equivalent to the compound Beta-Poisson distribution because the transcript numbers are distributed as follows:

$$n \sim \text{Poisson}(\rho x), \quad x \sim \text{Beta}(\sigma_{on}, \sigma_{off}). \tag{2}$$

As mentioned previously, it is well known that if gene expression is bursty then Eq (1) is well approximated by an NB distribution, or its continuous counterpart, the Gamma distribution [36,49]; related derivations for the distribution of protein numbers can be found in Refs. [53,54]. Specifically, transcriptional bursting occurs in the limit $\sigma_{off} \to \infty$ taken such that $\rho/\sigma_{off}$ remains constant [49,55]. A simple way to see how this limit leads to an NB distribution makes use of the Beta-Poisson formulation of the telegraph model. From Eq (2), it follows that

$$\lim_{\sigma_{off}\to\infty} \text{Poisson}\left(\frac{\rho}{\sigma_{off}}\sigma_{off}\text{Beta}(\sigma_{on}, \sigma_{off})\right) = \text{Poisson}\left(\frac{\rho}{\sigma_{off}}\text{Gamma}(\sigma_{on}, 1)\right) = \text{NB}\left(\sigma_{on}, \frac{1}{1+\rho/\sigma_{off}}\right), \tag{3}$$

where in the second step we used the standard statistical result: $\lim_{\sigma_{off}\to\infty} \sigma_{off}\text{Beta}(\sigma_{on}, \sigma_{off}) = \text{Gamma}(\sigma_{on}, 1)$.
*It is presently unclear if the converse is true: Does an excellent NB fit to the mRNA count distribution of the telegraph model imply transcriptional bursting?*
To answer this question, we proceed as follows. It is straightforward to show using Eq (1) that the mean and variance of mRNA counts for the telegraph model are given by

$$\langle n\rangle_{tele} = \frac{\rho\sigma_{on}}{\sigma_{on}+\sigma_{off}}, \qquad \text{Var}_{tele}(n) = \langle n^2\rangle_{tele} - \langle n\rangle^2_{tele} = \frac{\rho\sigma_{on}[(\sigma_{on}+\sigma_{off})(\sigma_{on}+\sigma_{off}+1)+\rho\sigma_{off}]}{(\sigma_{on}+\sigma_{off})^2(\sigma_{on}+\sigma_{off}+1)}, \tag{4}$$

where $\langle\bullet\rangle$ is the averaging operator. In contrast, the mean and variance of an NB distribution $\text{NB}(r, p)$ are

$$\langle n\rangle_{NB} = \frac{r(1-p)}{p}, \qquad \text{Var}_{NB}(n) = \frac{r(1-p)}{p^2}. \tag{5}$$

By equating $\langle n\rangle_{tele} = \langle n\rangle_{NB}$ and $\text{Var}_{tele}(n) = \text{Var}_{NB}(n)$ in Eqs (4) and (5) and solving for $r$ and $p$, we can construct an NB distribution that has the same first and second moments as the mRNA count distribution of the telegraph model:

$$\text{NB}\left(\frac{\sigma_{on}(\sigma_{on}+\sigma_{off}+1)}{\sigma_{off}}, \frac{(\sigma_{on}+\sigma_{off})(\sigma_{on}+\sigma_{off}+1)}{(\sigma_{on}+\sigma_{off})(\sigma_{on}+\sigma_{off}+1)+\rho\sigma_{off}}\right). \tag{6}$$

This distribution is hereafter referred to as the effective NB distribution denoted as $\text{NB}(r_e, p_e)$. Of course, this distribution and the telegraph model potentially can differ in their third and higher moments, and hence there is no guarantee that the shapes of the two distributions will generally be similar. We also note that while maximum likelihood estimation or other methods can be used to obtain an effective negative binomial distribution, they often lack closed-form solutions for parameter estimates, which can hinder further quantitative analysis. In contrast, moment matching provides analytical expressions, facilitating deeper insight.

In Fig 2, we compare the distributions given by Eq (1) and Eq (6) for four distinct parameter sets. In each case, the effective NB distribution is practically perfectly aligned with the corresponding telegraph distribution. However, note that only one of these four cases (Fig 2a) corresponds to the classical case of $\sigma_{off} \gg \sigma_{on}$. The three counterexamples (Fig 2b–2d) clearly show that excellent fitting of the telegraph model to an NB distribution does NOT necessarily indicate transcriptional bursting.

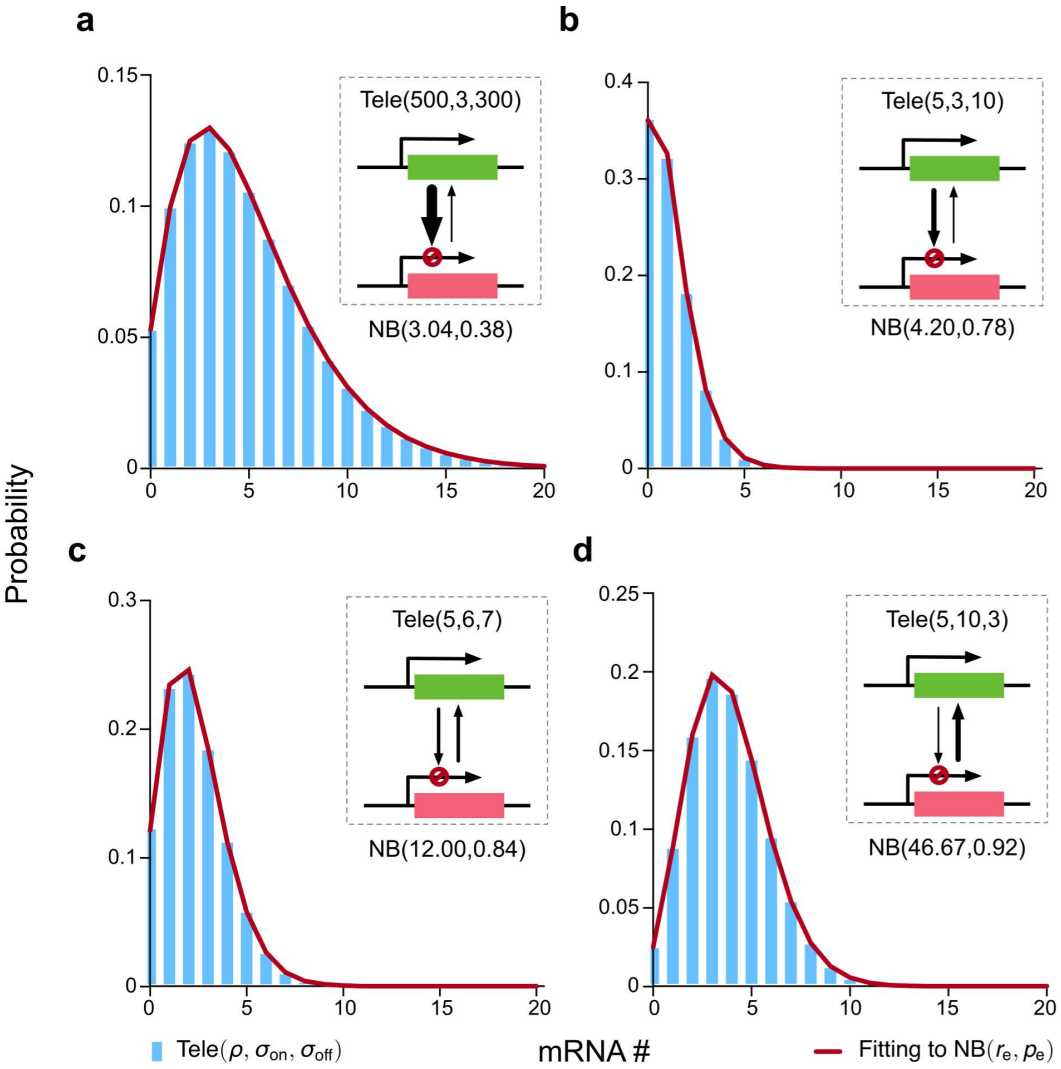


**Fig 2. Comparison of the steady-state mRNA distribution of the telegraph model (Eq (1) denoted as $\text{Tele}(\rho, \sigma_{on}, \sigma_{off})$) with the effective NB distribution (Eq (6) denoted as $\text{NB}(r_e, p_e)$) for the case of perfectly identical cells (parameters do not vary from cell-to-cell).** The two parameters of the effective NB distribution are chosen so that its first and second moments of mRNA counts exactly agree with those of the telegraph model (the values of the two parameters, $r_e$ and $p_e$, are stated to 2 decimal places in the figure). In all four parameter cases, the effective NB distribution exceptionally well fits the corresponding telegraph distribution, and yet only in case (a) we have $\sigma_{on} \ll \sigma_{off}$ (the classical case of transcriptional bursting). This demonstrates that a good fit of an NB distribution to the telegraph model distribution does not imply the presence of transcriptional bursting.

https://doi.org/10.1371/journal.pcbi.1014014.g002

### 2.2 An alternative set of conditions under which the NB distribution provides a good approximation to the telegraph model distribution

We start by reparameterizing the gene switching rates as

$$\sigma_{on} = f_{on} N_\sigma, \qquad \sigma_{off} = (1 - f_{on}) N_\sigma, \tag{7}$$

where $N_\sigma = \sigma_{on} + \sigma_{off}$ represents the timescale of the transition rates relative to the rate of mRNA degradation and $f_{on}$ is the fraction of time spent in the active state, which is given by

$$f_{\text{on}} = \frac{\sigma_{\text{on}}}{\sigma_{\text{on}} + \sigma_{\text{off}}}. \tag{8}$$

Furthermore, for convenience, we denote the telegraph model distribution, Eq (1), by $\text{Tele}(\rho, \sigma_{\text{on}}, \sigma_{\text{off}})$. Given these definitions, we state the following theorem:

*Theorem 1:* In the limit $N_\sigma \to \infty$, the telegraph model distribution $\text{Tele}(\rho, f_{\text{on}}N_\sigma, (1-f_{\text{on}})N_\sigma)$ and the effective negative binomial distribution $\text{NB}(r_{\text{e}}, p_{\text{e}})$ (Eq (6)) converge to the same distribution, both with a convergence rate of $O(1/\sqrt{N_\sigma})$.

The key idea behind proving this theorem is to show that the Beta and Gamma distributions converge to the same limiting distribution – specifically, a normal distribution – as $N_\sigma \to \infty$. This then implies a convergence of the Beta-Poisson (telegraph model) distribution to a Gamma-Poisson (NB) distribution. This convergence is not immediately obvious because the Beta distribution is defined on $(0, 1)$ while the Gamma distribution is defined on $(0, \infty)$. Details of the proof can be found in Methods Sect 4.1.

This proof makes more precise the statement in the caption of SI Fig 4 of Ref [56] which states that the mRNA count distribution of the telegraph model is bimodal (and therefore not similar to an NB distribution) when $\sigma_{\text{on}} < 1$ and $\sigma_{\text{off}} < 1$; it is unimodal and resembles an NB distribution when $\sigma_{\text{on}} > 1$ and $\sigma_{\text{off}} > 1$. In particular, Theorem 1 shows that $\sigma_{\text{on}} > 1$ and $\sigma_{\text{off}} > 1$ are not sufficient by itself to guarantee an excellent fit of the NB distribution to that of the telegraph model. We also note that the conditions identified by the theorem are different from those given by the classical result that requires $\sigma_{\text{on}} \ll \sigma_{\text{off}}$. In fact, while the classical results cannot explain the excellent fits shown in Fig 2, these can be explained by Theorem 1, because in each of these cases $N_\sigma$ is sufficiently large.

In the top left panel of Fig 3, we use the Kullback–Leibler (KL) divergence, defined as $D_{\text{KL}}(P \,\|\, Q) = \sum_n P(n) \ln \left[\frac{P(n)}{Q(n)}\right]$ to quantify the proximity between the effective negative binomial distribution ($P(n)$) and the telegraph model distribution ($Q(n)$). For fixed values of $\rho$ and $f_{\text{on}}$, increasing $N_\sigma$ results in a rapid decrease in the KL divergence, which agrees with Theorem 1. In the same figure, we also show the KL divergence between the Poisson distribution (with its parameter set to the mean of the telegraph model distribution) and the telegraph model distribution. It can be deduced that the fit of the Poisson distribution to the telegraph model distribution also becomes more accurate as $N_\sigma$ increases, although it is always a poorer fit than the effective NB distribution. Of course, in the limit of large $N_\sigma$, the KL divergence is very small for both effective NB and Poisson distributions and hence, practically speaking, in this limit both will provide an excellent approximation to the telegraph model distribution. These observations are further supported by comparing in Fig 3 the mRNA count distributions predicted by the telegraph model with the effective NB distribution and the Poisson distribution for 5 different values of $N_\sigma$ (these correspond to the points A-E in the plot of the KL divergence versus $N_\sigma$ in the top left corner of Fig 3).

### 2.3 The NB distribution provides an optimal approximation to the telegraph model distribution for an intermediate range of the sum of the gene switching rates

The distribution comparison in Fig 3 suggests that the effective NB distribution is an optimal fit to the telegraph model distribution in the intermediate range of $N_\sigma$. This is because: (i) for small $N_\sigma$, Theorem 1 does not hold. As well, in this case, it is already known that small $\sigma_{\text{on}}$ and $\sigma_{\text{off}}$ imply bimodal distributions of the telegraph model, which clearly cannot be well fitted by the (unimodal) effective NB distribution [56,57]; (ii) for large $N_\sigma$, the Poisson distribution practically provides an equally good fit as the effective NB distribution, but is preferable by the principle of Occam's razor because it has one less parameter than the effective NB approximation.

To make this observation more rigorous, we will use a model selection approach. An often used method to select which of two models provides a best fit to the data involves (i) maximizing the likelihood function of each model on the sample data; (ii) evaluating the Bayesian Information Criterion (BIC) for each model which is a function of both the maximum

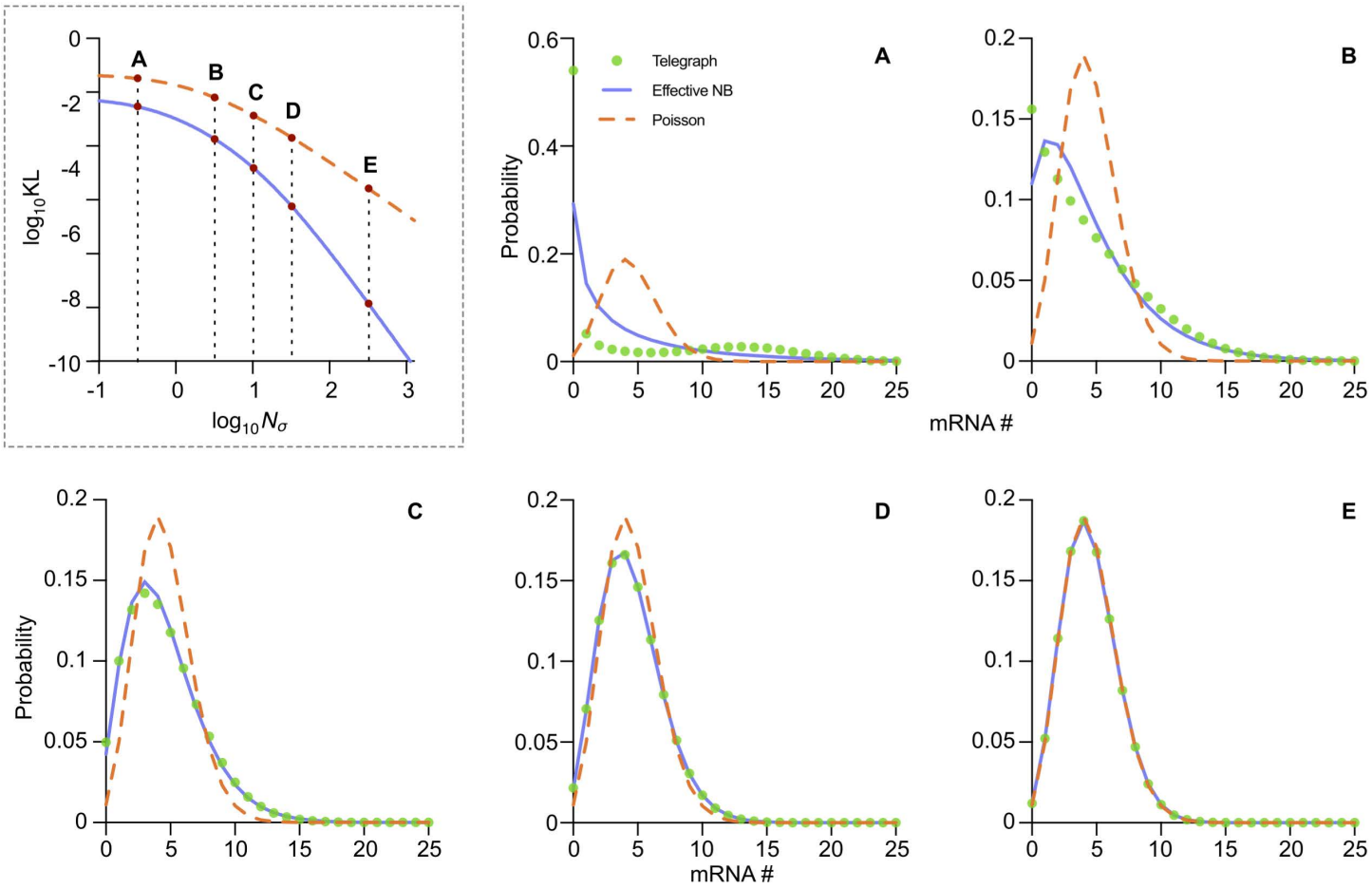


**Fig 3. The mRNA distribution of the telegraph model converges to that of effective NB distribution as the sum of the gene-state switching rates relative to the mRNA degradation rate ($N_\sigma$) increases.** In the dashed box, we show that the effective NB distribution (blue solid line) exhibits a lower KL divergence to the telegraph model distribution compared to the Poisson approximation (orange dashed line) as $N_\sigma$ grows. The distributions of the telegraph model (green dots), effective NB (blue solid lines), and Poisson (orange dashed lines) are shown for Points A-E, as indicated in the KL divergence plot. The other parameters are fixed at $\rho = 15$ and $f_{\text{on}} = 0.3$. Point A: $N_\sigma = 0.3$, Point B: $N_\sigma = 3$, Point C: $N_\sigma = 10$, Point D: $N_\sigma = 30$, Point E: $N_\sigma = 300$.

https://doi.org/10.1371/journal.pcbi.1014014.g003

value of the likelihood and the number of parameters; (iii) the optimal model is selected to be the one with the smallest BIC score.

In particular, the BIC is defined as

$$\text{BIC}(\mathcal{M}, \mathcal{X}) = |\mathcal{M}| \ln |\mathcal{X}| - 2 \sum_{i=1}^{|\mathcal{X}|} \ln P_{\mathcal{M}}(n_i|\theta_{\text{MLE}}), \tag{9}$$

where $\mathcal{X}$ represents the set of count data $\{n_i\}$ for $i = 1, \cdots, n_c$, with $|\mathcal{X}| = n_c$ being the size of the dataset (number of cells). The proposed model is denoted by $\mathcal{M}$, its number of parameters by $|\mathcal{M}|$ and the distribution that defines the model (its likelihood function) by $P_{\mathcal{M}}(n_i|\theta)$ where $\theta$ represents the model parameters. The parameter $\theta_{\text{MLE}}$ is the maximum likelihood estimate (MLE) obtained by maximizing the log-likelihood function

$$\mathcal{L}_\theta = \sum_{i=1}^{|\mathcal{X}|} \ln P_{\mathcal{M}}(n_i|\theta).$$

In our case, given a parameter set $\{\sigma_{on}, \sigma_{off}, \rho\}$, the count data is simulated by randomly sampling the telegraph model distribution $n_c$ times. The difficulty with this approach is that for each parameter set, there is an infinite number of datasets that can be simulated, and in principle, the algorithm can select a different best model for each dataset. Rather than relying on individual samples, our goal is to associate a unique best model with each set of underlying parameters. The most straightforward way to achieve this is by averaging the BIC over an infinite number of samples, i.e., $\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\text{BIC}(\mathcal{M}, \mathcal{X})]$. However, this approach is computationally intensive, as model parameters must be re-estimated for each sample.

To address this challenge, we introduce a new measure—the approximate expected Bayesian information criterion (aeBIC)—as an efficient estimator of the expected BIC across repeated samples. The aeBIC is defined as

$$\text{aeBIC}(\mathcal{M}, n_c) = |\mathcal{M}| \ln n_c + 2n_c\varepsilon_c(\mathcal{M}_{\text{MLE}}, \mathcal{G}). \tag{10}$$

Here, $\varepsilon_c$ represents the cross-entropy between two distributions (representing the model and the ground truth), defined as

$$\varepsilon_c(\mathcal{M}, \mathcal{G}) = -\sum_{n=0}^{\infty} P_{\mathcal{G}}(n) \ln P_{\mathcal{M}}(n|\theta), \tag{11}$$

where $P_{\mathcal{G}}(n)$ and $P_{\mathcal{M}}(n|\theta)$ are the probabilities of observing $n$ molecules in distributions $\mathcal{G}$ (the ground-truth distribution which for us is the telegraph model distribution) and $\mathcal{M}$ (the proposed model which can be telegraph, NB or Poisson with parameters $\theta$), respectively. Note that the subscript MLE denotes that the kinetic parameters of the proposed distribution are estimated via maximum likelihood. Specifically, $\mathcal{M}_{\text{MLE}}$ represents the model whose parameters $\theta$ minimize Eq (11). If $\mathcal{M} = \mathcal{G}$, the cross entropy $\varepsilon_c(\mathcal{M}, \mathcal{G})$ reduces to the information entropy $\varepsilon(\mathcal{G})$. Additionally, the cross entropy can be decomposed as [58]

$$\varepsilon_c(\mathcal{M}, \mathcal{G}) = \varepsilon(\mathcal{G}) + \text{KL}(\mathcal{M}||\mathcal{G}), \tag{12}$$

where $\text{KL}(\mathcal{M}||\mathcal{G})$ is the KL divergence between $\mathcal{M}$ and $\mathcal{G}$. This decomposition implies that the closer the proposed distribution $\mathcal{M}$ is to the ground-truth distribution $\mathcal{G}$, the smaller the KL divergence becomes, and consequently, the cross entropy decreases as well.

In S1 Text, Sect 4.2 we show that the expectation of the BIC over an infinite number of independent samples (each of size $n_c$) is related to the aeBIC by:

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\text{BIC}(\mathcal{M}, \mathcal{X})] \leq \text{aeBIC}(\mathcal{M}, n_c). \tag{13}$$

This shows that aeBIC serves as an upper bound for the expectation of the BIC. It can also be shown that the equality in Eq (13) is achieved as $n_c \to \infty$.

However, in practice, we find that the difference between $\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\text{BIC}(\mathcal{M}, \mathcal{X})]$ and $\text{aeBIC}(\mathcal{M}, n_c)$ is negligible. To demonstrate this, for a fixed set of parameters $\{\rho, \sigma_{on}, \sigma_{off}\}$, we randomly sampled $n_c$ counts from the telegraph model distribution $\text{Tele}(\rho, \sigma_{on}, \sigma_{off})$ (using its Beta-Poisson formulation in Eq (2)) and repeated this step $n_{\text{trial}} = 10^3$ times. For each trial, we fitted the sampled counts to both Poisson and NB distributions using MLE and computed the BIC for both distributions. The BIC values were then averaged across all $n_{\text{trial}}$ datasets to estimate the expectation of the BIC. Additionally, we computed the aeBIC using Eq (10) for both distributions for each parameter set of the telegraph model distribution. The relative error of aeBIC compared to the expectation of BIC was calculated for each set of parameters. This comparison process is

illustrated by a cartoon in Fig 4a and the results for numerous parameter sets that cover the range $N_\sigma \in [0.1, 1000]$, $f_{on} \in [0.1, 0.9]$ and $\rho \in [1, 15]$ are shown in Fig 4b and Fig Aa,b in S1 Text (these parameter sets correspond to a mean number of transcripts that ranges between *0* and *15*). Note that, independent of the type of the fitted distribution, the relative error decreases with increasing sample size $n_c$. Notably, even for small sample sizes, the magnitude of the relative error is very small, and thereby for all intents and purposes, we can equate the aeBIC with the average of the BIC computed over many independent samples. This is very convenient because it is far faster to compute the aeBIC than the BIC since, for the former, the maximum likelihood estimation only needs to be done once using Eq. (11).

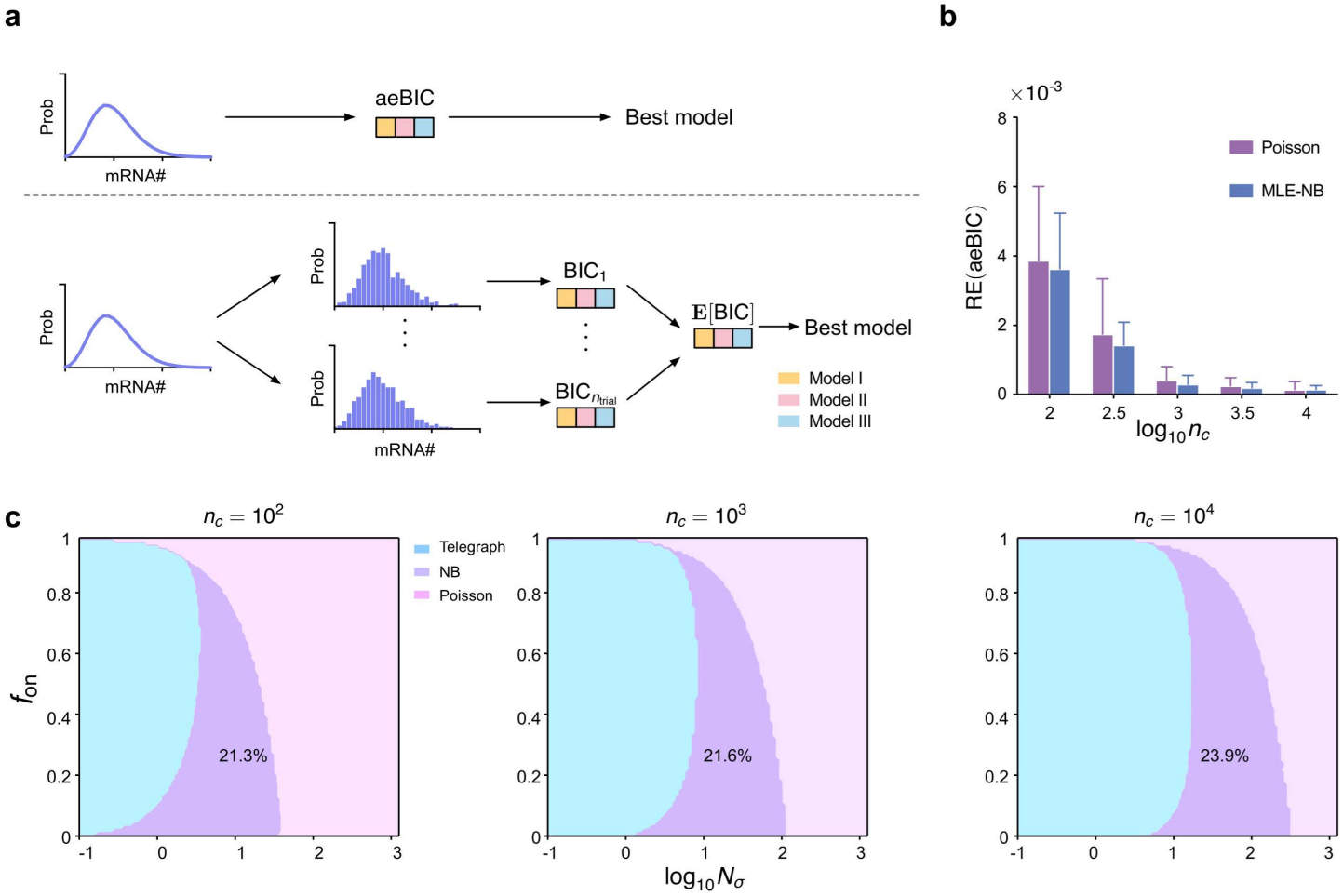


**Fig 4. Benchmarking aeBIC against. E**[BIC] shows that a single-sample-based criterion can reliably recover the expected model-selection landscape across sample sizes and telegraph-model regimes. (a) Cartoon illustrating a computational approach to compare the aeBIC (top) with E[BIC] — the expected value of the BIC (bottom). The aeBIC utilizes a single score to select the best model distribution (telegraph, NB or Poisson) given that the ground-truth mRNA distribution is that of the telegraph model. The BIC method assigns a score to each different sample of simulated data from the telegraph model and then all these scores are averaged leading to **E**[BIC]. (b) The relative error (RE) of aeBIC compared to **E**[BIC] for two distributions (Poisson and NB) as a function of sample size $n_c$ for 10 parameter sets (see Table A in S1 Text for the values of $N_\sigma$ and $f_{on}$; $\rho$ is fixed to 15). Error bars show the standard error of the mean. (c) Phase diagram showing the regions of parameter space where the telegraph, NB and Poisson distributions are selected as optimal by the aeBIC, given that the ground-truth mRNA distribution is that of the telegraph model. Here $n_c$ is the sample size, $N_\sigma$ is the sum of gene-state switching rates normalised by the degradation rate of mRNA, and $f_{on}$ is the fraction time spent in the active state. The fraction of the total parameter space occupied by the region where the NB distribution is optimally selected is shown on the plots. Note that the transcription rate is fixed to $\rho$ = 15 which implies that the maximum mean number of transcripts in the phase plots is 15.

https://doi.org/10.1371/journal.pcbi.1014014.g004

Next, we use Eqs (10)–(11) to compute the aeBIC and identify the best-fitted model (the one with the smallest aeBIC) among Poisson, NB and telegraph model distributions across the $N_\sigma - f_{\text{on}}$ parameter space (while keeping the transcription rate $\rho$ constant), given that the ground-truth data is generated by the telegraph model. Note that this model selection can be performed for different sample sizes because the aeBIC is a function of $n_c$. Note also that the mean mRNA is proportional to $f_{\text{on}}$ since the mean mRNA of the telegraph model is $\rho f_{\text{on}}$ (Eq (4)) and $\rho$ is fixed in our analysis. The results are shown in Fig 4c — the model associated with the smallest aeBIC score corresponds to the model most frequently selected by the BIC calculated over many independent samples of the same size, thus providing a further test of the validity of the aeBIC approach (Table B in S1 Text). In particular, aeBIC preferentially selects the telegraph model distribution for small $N_\sigma$, while the Poisson distribution is preferred for large $N_\sigma$. This agrees with the distribution comparison in Fig 3 which suggested that the effective NB distribution is an optimal fit to the telegraph model distribution in an intermediate range of $N_\sigma$. The NB distribution is more likely to be selected as $f_{\text{on}}$ decreases, which is consistent with the fact that the telegraph model distribution converges to an NB distribution when the gene spends most of its time in the inactive state ($f_{\text{on}}$ is small). Interestingly, the shape and area of the crescent-shaped region where the NB distribution is an optimal model remains largely unchanged as the sample size increases, but its position shifts horizontally. In particular, the fraction of space where the NB distribution is the optimal model increases monotonically from 21.3% to 30% as the sample size increases over four orders of magnitude, from $10^2$ to $10^6$ cells (Fig G in S1 Text).

Note that an alternative (and faster) way to obtain the phase space plots in Fig 4c is to calculate the aeBIC with the value of the cross-entropy computed using moment matching between the proposed and the ground-truth distribution (instead of MLE). Specifically, the aeBIC using moment-matching is defined as

$$\text{aeBIC}(\mathcal{M}, n_c) = |\mathcal{M}| \ln n_c + 2n_c\varepsilon_c(\mathcal{M}_{\text{MOM}}, \mathcal{G}), \tag{14}$$

where $\varepsilon_c(\mathcal{M}_{\text{MOM}}, \mathcal{G})$ is determined by evaluating Eq (11) with parameters $\theta$ determined by moment-matching between the proposed distribution ($P_{\mathcal{M}}(n|\theta)$) and the ground-truth distribution ($P_{\mathcal{G}}(n)$). In Fig B in S1 Text top left corner, we demonstrate that these two methods lead to very similar minimum values for the cross entropy and hence there is no appreciable difference on the model selection phase plots in Fig 4c. On a MacBook Air with an Apple M2 chip and 16 GB memory, generating a single phase space plot with this method ($2.06 \times 10^4$ pairs of $N_\sigma - f_{\text{on}}$) takes about 68 seconds (CPU time) and uses only 3.6 MB of memory, demonstrating its computational efficiency.

### 2.4 Cell-to-cell variability in transcript capture probability significantly affects the region of parameter space where the NB distribution is optimally selected

Thus far, we have assumed that scRNA-seq data can be directly explained by the telegraph model. In practice, however, not all transcripts are captured. During the reverse transcription stage, each transcript is captured with probability $p_1$. UMIs are attached to molecules before PCR, enabling PCR duplicates to be collapsed during downstream processing and thereby correcting amplification bias. Finally, during sequencing, the sequencing depth determines the detection probability ($p_2$) of each UMI-labeled transcript. The observed UMI counts for a gene therefore represent the number of transcripts successfully captured, labeled, and detected. Collectively, the overall probability linking the true transcript count to the observed count is denoted as $p_{\text{cap}}$ (including the effect of $p_1$ and $p_2$). Therefore, it is necessary to account for technical noise in the data. A simple way to do this is to assume that the number of zeros in the data is artificially high and to generate simulated scRNA-seq data using a zero-inflated telegraph model (a distribution that is a sum of a Dirac-delta function and the telegraph model distribution). While this approach or variants of it are widespread [14,18,19,59] and attempts have been made to explain zero-inflation using mechanistic models [60], it is not ideal because (i) the downsampling of transcripts due to imperfect capture results in an increase of not only zeros but also 1's, 2's, etc; (ii) there is evidence that scRNA-seq data with UMIs barely suffer from zero-inflation [6,61]. A more principled and increasingly used model to

explain the effect of imperfect capture is the binomial capture model [10,38,62]. This model assumes that each transcript is captured and observed with a probability $p_{cap}$ (Fig 5a). For current standard droplet-based sequencing technologies, the capture probability typically ranges between 0.05 and 0.3, depending on the sequencing technique used [11,63–65]. Extrinsic noise, such as variability in transcription rates, is a major contributor to cell-to-cell heterogeneity in gene expression [66,67]. However, in most cases, variability in capture probability and transcription rate cannot be disentangled mathematically (see S1 Text, Sect 4.3), unless additional experimental controls, such as spike-ins, are employed [8]. Therefore, in the following, we focus on variability in transcript capture probability.

For simplicity, first we consider the case where the capture probability does not vary from cell to cell and from gene to gene, i.e., the distribution of the capture probability is a Dirac-delta function, Dirac($p_{cap}$). We investigated the impact of $p_{cap}$ on model selection outcomes (using aeBIC) and the results are shown in Fig 5b. Note that in this case $\mathcal{G}$ in Eq (10) is the telegraph model distribution Eq (1) with $\rho$ rescaled to $\rho p_{cap}$ [37,38]. Comparing Fig 4c with Fig 5b, we see that for a fixed sample size, the fraction of parameter space where the NB distribution is preferentially selected is almost constant for capture probabilities in the range $p_{cap} = 0.2 - 1.0$. The accuracy of scRNA-seq techniques is constantly improving, with commonly used platforms such as 10x having a capture probability that exceeds 0.2 [65]. Hence, we can conclude that if the capture probability is roughly the same from one cell to another, then the universality of the NB distribution in scRNA-seq datasets has little to do with the low capture efficiency.

Of course, for some scRNA-seq methods the capture probability may be very low and then in that case the technical noise will be so large that there will be a significant impact on model selection. For example, in the case of $p_{cap} = 0.05$ and sample size $n_c = 10^2$ in Fig 5b, the Poisson distribution emerges as the optimal model across most of the parameter space, with the NB distribution being selected only in a limited region. The telegraph model, however, is never identified as the optimal choice under these conditions. This is expected because the bimodal character of the telegraph model distribution at small $N_\sigma$ becomes less obvious at lower capture probabilities — downsampling causes the two peaks of the distribution to become closer together or even to merge and hence the best-fit distributions are exclusively unimodal (Poisson or NB distributions).

The wide distribution of the total counts (from all genes) per cell in typical scRNA-seq datasets (see for, e.g., Fig 1A of [17]) strongly suggests that the capture probability varies considerably across cells (Fig 6a). To address this, we simulate scRNA-seq data by sampling from the underlying mRNA distribution given by Eq (2) with $\rho$ rescaled to $\rho p_{cap}$, and $p_{cap}$ that varies between cells according to three different distributions (Dirac-delta function and two different Beta distributions), all of which have mean $\langle p_{cap} \rangle = 0.3$ but with different coefficients of variation (CV = *0* for the Dirac-delta function, and 0.11 and 0.21 for the two different Beta distributions). See Fig 6b for a plot of the three distributions. Note that the Beta distribution was chosen because it samples numbers in the range $(0, 1)$, a necessity given $p_{cap}$ is a probability. The aeBIC was used to select between the standard telegraph model distribution (no assumption of variability of rates or capture probability between cells), NB and Poisson models with the ground-truth (measured) mRNA distribution $\mathcal{G}$ given by that of the telegraph model with $p_{cap}$ distributed according to a Beta distribution (S1 Text, Sect 4.4). Note that the aeBIC remains an accurate predictor of the expectation of the BIC computed from MLE and hence suitable for model selection even when the transcript numbers are heavily influenced by technical noise (Fig Ac in S1 Text). The phase plots generated by aeBIC-based model selection are shown in Fig 6c. We find that for fixed sample size $n_c$, an increase in the CV of the Beta distribution, leads to very small changes in the fraction of parameter space where the telegraph model distribution is selected, but the fraction of parameter space where the NB or Poisson distributions are selected changes significantly. In particular, an increase in the cell-to-cell variability in the capture probability tends to favor the selection of the NB distribution over the Poisson distribution. This is because variability in the capture probability increases the Fano factor (variance divided by the mean) of mRNA distributions (S1 Text, Sect 4.5) — Fano factors larger than 1 can be captured by an NB distribution but the Poisson distribution has a fixed Fano factor of 1. For example, for a sample size of $10^3$ cells, as the CV of the Beta distribution of $p_{cap}$ is increased from *0* to 0.21 (keeping the mean constant at 0.3), the fraction of space where

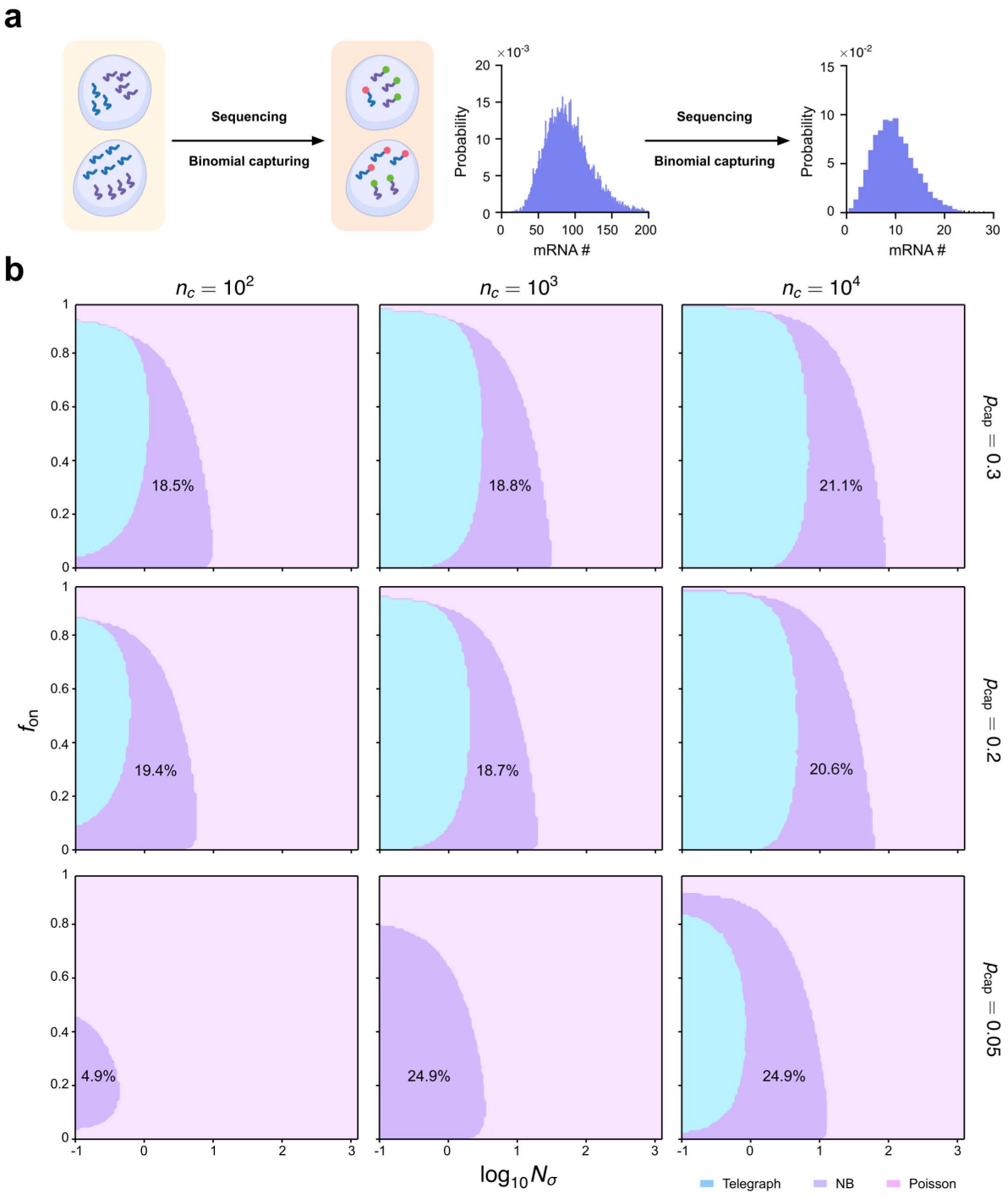


**Fig 5. The binomial capture model for scRNA-seq reveals how incomplete transcript capture systematically shifts the effective model-selection landscape across telegraph-model parameter regimes.** (a) Schematic illustrating the binomial capture model for scRNA-seq. Transcripts in each cell are captured with some probability $p_{cap}$. This causes a downsampling of the distribution of mRNA counts. (b) Phasediagram showing the regions of parameter space where the telegraph, NB and Poisson distributions are selected as the optimal ones by the aeBIC, given that the ground-truth mRNA distribution is that of the telegraph model. The phase diagrams are shown for three different values of $p_{cap}$ (values stated next to the plots). Here $n_c$ is the sample size, $N_\sigma$ is the sum of gene-state switching rates normalised by the degradation rate of mRNA, and $f_{on}$ is the fraction time spent in the active state. The fraction of the total parameter space occupied by the region where the NB distribution is optimally selected is shown on the plots. Note that the transcription rate is fixed to $\rho = 15$ which implies that the maximum mean number of transcripts in the phase plots is $15p_{cap} = 4.5, 3$ and $0.75$ for the phase plots in rows 1, 2 and 3, respectively. These phase plots do not appreciably change if aeBIC is determined using moment-matching instead of MLE (Fig B Text).

https://doi.org/10.1371/journal.pcbi.1014014.g005

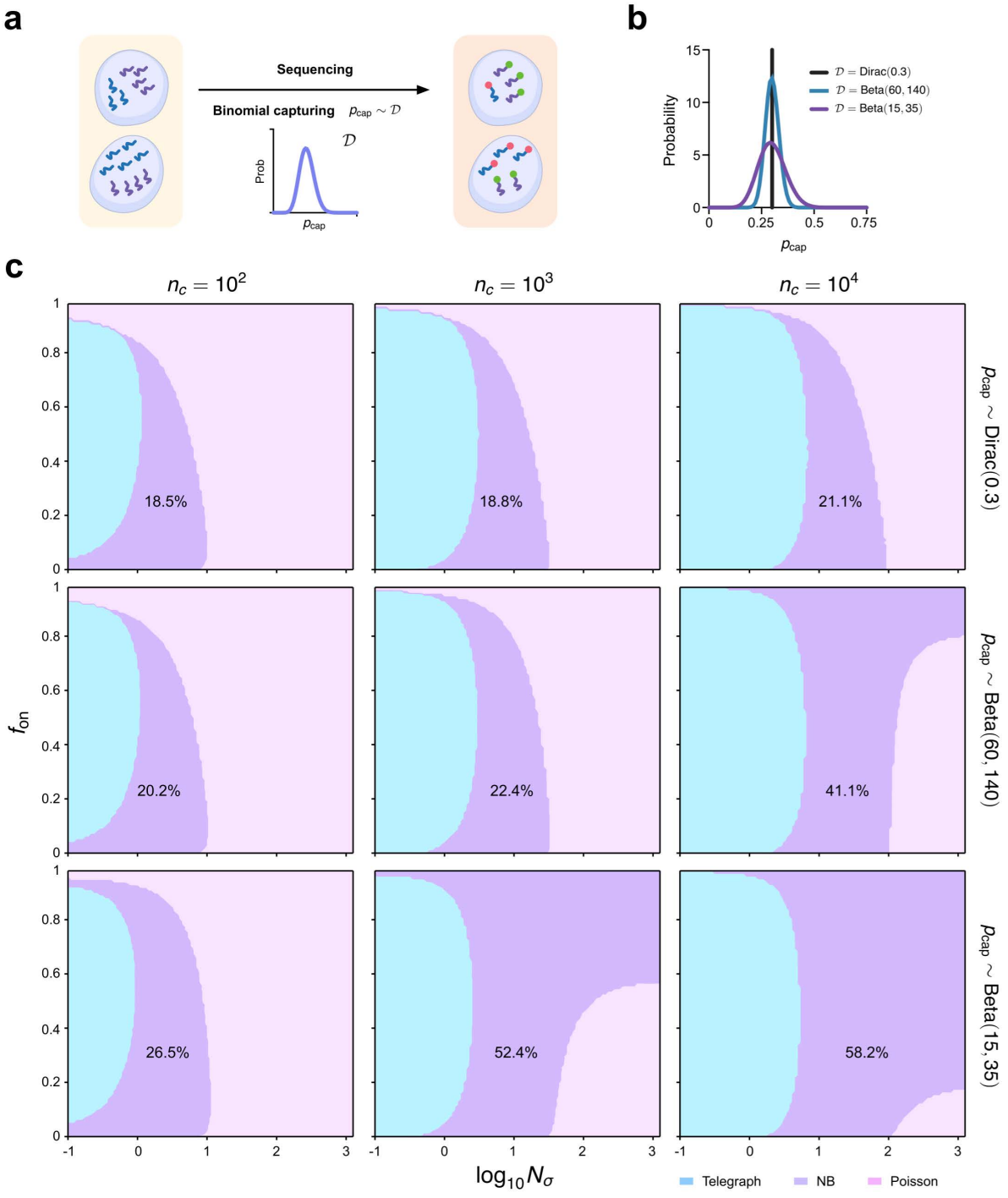


**Fig 6. Heterogeneity in scRNA-seq capture efficiency across cells systematically alters the effective model-selection landscape, shifting the regions in which telegraph, NB and Poisson distributions are favoured.** (a) Schematic illustrating the binomial capture model for scRNA-seq with a probability of mRNA capture, $p_{cap}$, that varies between cells according to some distribution. (b) We consider three different distributions all with mean $\langle p_{cap} \rangle = 0.3$ but with varying coefficient of variation (CV): (i) Dirac(0.3) with CV = 0; (ii) Beta(60, 140) with CV = 0.11; (iii) Beta(15, 35) with CV = 0.21. (c) Phase diagram showing the regions of parameter space where the telegraph, NB and Poisson distributions are selected as the optimal ones by the aeBIC, given that the ground-truth mRNA distribution is that of the telegraph model with effective transcription rate $\rho p_{cap}$ where $p_{cap}$ is sampled from the 3 distributions mentioned above. Here $n_c$ is the sample size, $N_\sigma$ is the sum of gene-state switching rates normalised by the degradation rate of mRNA, and $f_{on}$ is the fraction time spent in the active state. The fraction of the total parameter space occupied by the region where the NB distribution is optimally selected is shown on the plots. Note that the transcription rate is fixed to $\rho = 15$ which implies that the maximum mean number of transcripts in the phase plots is $15\langle p_{cap} \rangle = 4.5$.

https://doi.org/10.1371/journal.pcbi.1014014.g006

the telegraph model distribution is optimal decreases slightly from 32.1% to 29.1%, where the NB distribution is optimal increases from 18.8% to 52.4%, and where the Poisson distribution is optimal decreases from 49.1% to 18.5%.

### 2.5 In NB-optimal parameter space, burst parameter estimation accuracy is typically low, but ranking genes by burst frequency remains reliable

We also seek to understand if any biologically relevant interpretation can be imparted to the two parameters of the NB distribution fitted to the scRNA-seq data. Consider the ideal scenario in which we can perfectly correct for technical noise. In regions of parameter space where the NB distribution is optimally selected, can we obtain accurate estimates of the burst size and burst frequency from the two parameters of the best fitted NB distribution?

The aeBIC Eq (10) used for model selection depends on the choice of the ground-truth model describing the distribution of the data ($\mathcal{G}$) and the model(s) ($\mathcal{M}$) to be fitted to this data. In our case, $\mathcal{G}$ is the telegraph model with effective transcription rate $\rho p_{\text{cap}}$, and $p_{\text{cap}}$ distributed according to a Beta distribution, $\text{Beta}(a, b)$. Previously, we assumed that the models $\mathcal{M}$ are given by the standard telegraph, NB and Poisson distributions. These models do not have any information on the distribution of $p_{\text{cap}}$ and hence this is the case where model selection using aeBIC is done without correcting for technical noise. In contrast, a perfect correction for technical noise is possible if instead the models $\mathcal{M}$ are corrected for technical noise, i.e., the mRNA distributions of the telegraph model, NB and Poisson models are integrated over the $\text{Beta}(a, b)$ distribution (this of course assumes perfect knowledge of the parameters $a$ and $b$ characterizing the technical noise). For a derivation of these distributions, see Sect 4.4. The differences between the conventional models and those with $p_{\text{cap}}$ sampled from a distribution are illustrated in Fig 7a.

In Fig 7b we contrast model selection using conventional and technical-noise-corrected models for the case where $p_{\text{cap}}$ is distributed according to a $\text{Beta}(60, 140)$ distribution which has a mean of $\langle p_{\text{cap}} \rangle = 0.3$ and CV$= 0.11$. We find that the regions of parameter space where the conventional and corrected telegraph models are selected are very similar. However, the region where the corrected NB distribution is selected (rainbow-colored region) is smaller than the region where the conventional NB distribution is selected — in fact, there is a region (shown in grey) where the conventional NB distribution and the corrected Poisson distribution are selected, i.e., in this region the apparent NB character of the mRNA count distribution is purely due to technical noise. The discrepancies between the two approaches are negligible for small sample sizes ($10^2$) but significant for large samples $10^4$.

Next, in the parameter space region where the corrected NB distribution is selected as the optimal model, we estimate the two parameters of the corrected NB distribution by the method of moments. We equate the first two moments of the latter distribution with those of the model simulating the data, i.e., the corrected telegraph model. This mimics the fitting of the corrected NB distribution to the data using the method of maximum likelihood in the limit of large sample sizes. In Sect 4.5 we show that this procedure leads to $\text{NB}(r, p)$ where $r = r_{\text{e}}$ and $p = p_{\text{e}}$ are exactly as given by the effective NB distribution in Eq (6). Clearly, because $r_{\text{e}}$ and $p_{\text{e}}$ do not depend solely on the true burst frequency $\sigma_{\text{on}}$ and burst size $\rho/\sigma_{\text{off}}$ of the telegraph model [68], the latter two parameters cannot be directly estimated by moment matching.

This issue can be circumvented by interpreting the best-fitting corrected NB distribution $\text{NB}(r_{\text{e}}, p_{\text{e}})$ as equal to the steady-state NB distribution solution $\text{NB}\left(\hat{\beta}_f, 1/(1 + \hat{\beta}_s)\right)$ of the bursty gene expression circuit:

$$G \xrightarrow{\hat{\beta}_f} G + sM, \quad M \xrightarrow{1} \varnothing, \tag{15}$$

where $s$ is a geometrically distributed random burst size with mean $\hat{\beta}_s$ and the burst frequency is $\hat{\beta}_f$ [69]. This model has experimental support [26] and its steady-state distribution is exactly the same as that of the telegraph model in the limit of transcriptional bursting, i.e., when the inactivation rate is much larger than the activation rate.

Hence the described inference procedure leads to estimated burst frequency and size that are given by

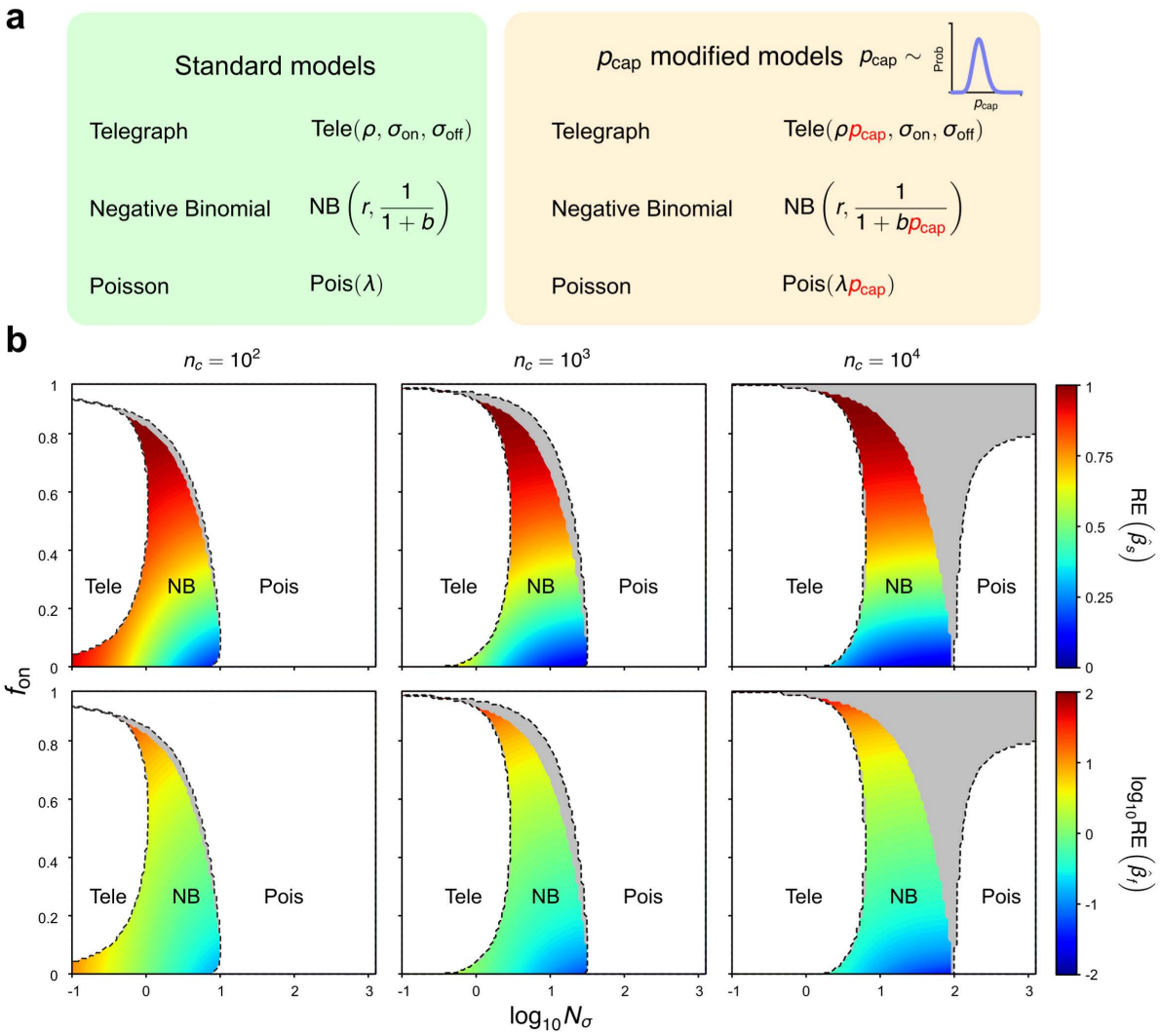


**Fig 7. Technical-noise correction for heterogeneous capture efficiency reshapes the aeBIC model-selection landscape and reveals its impact on the accuracy of inferred bursting kinetics.** (a) Illustration of the differences between the standard and technical-noise-corrected models; (b) Phase diagrams produced by aeBIC model selection based on standard or corrected models. For both, the ground-truth model for observed data is the telegraph model with $p_{cap}$ distributed according to the Beta(60, 140) distribution which has mean $\langle p_{cap} \rangle = 0.3$ and CV=0.11. The transcription rate $\rho$ is fixed to 15; the maximum mean number of transcripts is 4.5. The labels "Tele", "NB" and "Pois" denote the regions selected using the aeBIC procedure with corrected models. The dashed lines demarcate the same regions but using the aeBIC procedure with standard models. The "Pois" area is divided into a white part (where both aeBIC procedures select the Poisson distribution) and a grey part (where the aeBIC with standard models selects the NB distribution while the aeBIC with corrected models selects the Poisson distribution). The heatmap shows the magnitude of the relative errors in the estimated burst frequency ($\hat{\beta}_f$) and burst size ($\hat{\beta}_s$) in the NB-optimal region (using the aeBIC with corrected models). The errors are computed using Eq (17) — note that this approach assumes full knowledge of the distribution of probability capture, an ideal case. In the plots, $n_c$ denotes sample size, $N_\sigma$ is the sum of gene-state switching rates normalised by the degradation rate of mRNA, and $f_{on}$ is the fraction time spent in the active state.

https://doi.org/10.1371/journal.pcbi.1014014.g007

$$\hat{\beta}_f = r_e = \frac{f_{on}(1+N_\sigma)}{1-f_{on}}, \quad \hat{\beta}_s = \frac{1}{p_e} - 1 = \frac{\rho(1-f_{on})}{1+N_\sigma}. \tag{16}$$

Since we know that the true burst frequency of the telegraph model is $\sigma_{on}$ and the true burst size is $\rho/\sigma_{off}$, it follows that the relative errors are given by

$$\mathrm{RE}(\hat{\beta}_s) = \left|\frac{(\hat{\beta}_s - \rho/\sigma_{\mathrm{off}})}{\rho/\sigma_{\mathrm{off}}}\right| = 1 - \frac{(1-f_{\mathrm{on}})^2 N_\sigma}{1+N_\sigma}, \quad \mathrm{RE}(\hat{\beta}_f) = \left|\frac{(\hat{\beta}_f - \sigma_{\mathrm{on}})}{\sigma_{\mathrm{on}}}\right| = \frac{1+f_{\mathrm{on}}N_\sigma}{(1-f_{\mathrm{on}})N_\sigma}. \tag{17}$$

In Fig 7b we show a heatmap of the relative errors in the NB-optimal region (using the aeBIC with technical-noise-corrected models). The magnitude of these errors vary significantly and are smallest when $f_{\mathrm{on}}$ is small, i.e., when $\sigma_{\mathrm{off}} \gg \sigma_{\mathrm{on}}$ and $\sigma_{\mathrm{off}} \gg 1$. Hence, this implies that even though the NB distribution may be optimally selected, nevertheless this does not guarantee the accuracy of the estimated burst parameters. Indeed, the lack of accuracy is not very surprising given that the NB distribution of the reaction scheme in Eq (15) is only a good approximation to the telegraph model distribution when $\sigma_{\mathrm{on}} \ll \sigma_{\mathrm{off}}$ [49] and that we have previously established that it is possible to have excellent NB fits even when this inequality does not hold (Fig 2). Note that the burst size can generally be estimated more reliably than the burst frequency — this is also apparent from Eq (17) which implies $\mathrm{RE}(\hat{\beta}_s) < 1$ and $\mathrm{RE}(\hat{\beta}_f) < \infty$.

Next, we sought to understand whether reliable information of some type can be extracted from the burst parameters, even though their absolute values are generally inaccurate. In particular, we seek to understand whether ranking of genes by their burst parameter values can be accurate since this is based on relative rather than absolute information. To answer this question, we used the following protocol:

1. Randomly sample two pairs of parameter sets $(f_{\mathrm{on}}^{(1)}, N_\sigma^{(1)})$ and $(f_{\mathrm{on}}^{(2)}, N_\sigma^{(2)})$, each associated with a different gene, from the parameter space region where the NB distribution (corrected for technical noise) is selected as the optimal model for $p_{\mathrm{cap}}$ distributed according to the Dirac $(0, 3)$, $\mathrm{Beta}(60, 140)$ and $\mathrm{Beta}(15, 35)$ distributions. For a sample size of $n_c = 10^4$, these regions are shown in purple in Fig C in S1 Text. Note that the transcription rate $\rho$ is fixed to 15 in all cases.
2. Calculate the true burst frequency and size for each gene. For the first gene, we denote the true burst frequency and burst size as $\beta_f^{(1)} = \sigma_{\mathrm{on}}^{(1)} = f_{\mathrm{on}}^{(1)} N_\sigma^{(1)}$ and $\beta_s^{(1)} = \rho/\sigma_{\mathrm{off}}^{(1)} = \rho/(N_\sigma^{(1)}(1-f_{\mathrm{on}}^{(1)}))$, respectively; for the second gene, $\beta_f^{(2)} = \sigma_{\mathrm{on}}^{(2)} = f_{\mathrm{on}}^{(2)} N_\sigma^{(2)}$ and $\beta_s^{(2)} = \rho/\sigma_{\mathrm{off}}^{(2)} = \rho/(N_\sigma^{(2)}(1-f_{\mathrm{on}}^{(2)}))$, respectively. These relationships follow from Eq (7).
3. For each gene, use the method of moments to fit an NB distribution integrated over the distribution of $p_{\mathrm{cap}}$ to the distribution of the data, i.e., telegraph model distribution integrated over the distribution of $p_{\mathrm{cap}}$. This leads to the estimated burst frequency-size pairs for each gene: $(\hat{\beta}_f^{(1)}, \hat{\beta}_s^{(1)})$ and $(\hat{\beta}_f^{(2)}, \hat{\beta}_s^{(2)})$ which are given by Eq (16).
4. Compute the ground-truth ratios of burst sizes for the pair of genes as $r_{\mathrm{true}} = \min(\beta_s^{(1)}, \beta_s^{(2)})/\max(\beta_s^{(1)}, \beta_s^{(2)})$. Note that by placing the smaller burst size in the numerator, we guarantee $r_{\mathrm{true}} < 1$. The ground-truth ratio for the burst frequencies can be computed similarly.
5. Compute the estimated ratio of burst sizes for the pair of genes $r_{\mathrm{estimate}}$, preserving the original pair order used for the ground-truth ratio of burst sizes. That is, if $r_{\mathrm{true}} = \beta_s^{(1)}/\beta_s^{(2)}$ then $r_{\mathrm{estimate}} = \hat{\beta}_s^{(1)}/\hat{\beta}_s^{(2)}$; if $r_{\mathrm{true}} = \beta_s^{(2)}/\beta_s^{(1)}$ then $r_{\mathrm{estimate}} = \hat{\beta}_s^{(2)}/\hat{\beta}_s^{(1)}$. The estimated ratio for the burst frequencies can be computed similarly.

In Fig 8a we show a plot of $r_{\mathrm{true}}$ versus $r_{\mathrm{estimate}}$. Each point in this scatter plot corresponds to a randomly selected pair of genes. Points are classified by their color: (i) blue if $r_{\mathrm{estimate}} > 1$ indicates a reversal of pair order which implies that burst parameter inference fails to preserve the relative magnitude of the burst parameters; (ii) orange if $r_{\mathrm{estimate}} < 1$ and $r_{\mathrm{estimate}} > r_{\mathrm{true}}$ indicates that the order of the estimated burst parameters is the same as that of the ground truth parameters but that the estimation leads to an amplification of the difference between the burst parameters of the two genes; (iii) green if $r_{\mathrm{estimate}} < 1$ and $r_{\mathrm{estimate}} < r_{\mathrm{true}}$ indicates that the order of the estimated burst parameters is the same as that of the ground truth parameters but that the estimation leads to a reduction of the difference between the parameters of the two genes.

The results show that in many instances the ranking of a pair of genes by the size of their burst frequency is correctly estimated — these are the orange and green points in Fig 8a which imply gene ranking accuracy by burst frequency

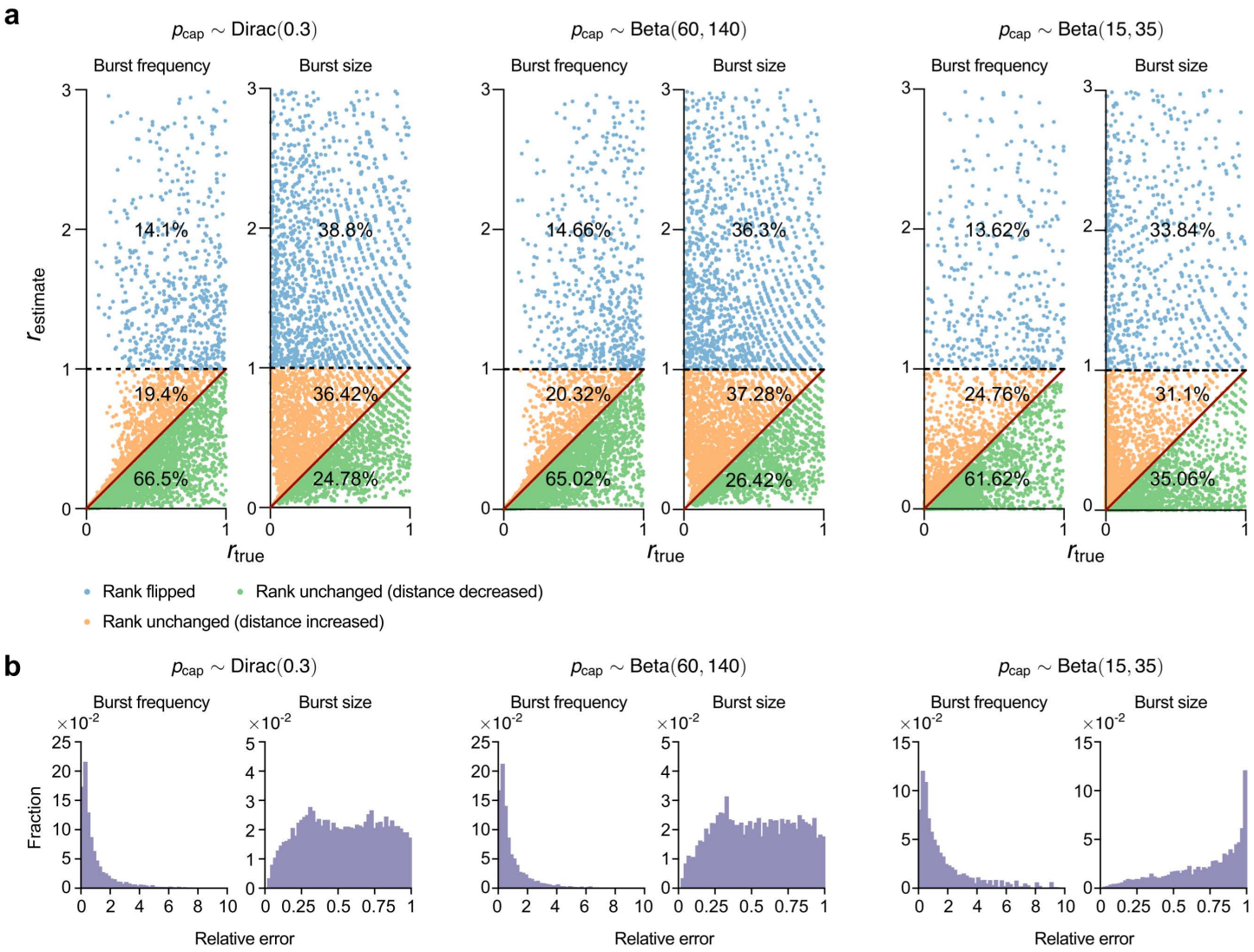


**Fig 8. Technical-noise-corrected inference in the NB-optimal regime reveals that relative ordering of gene bursting parameters can be robustly recovered.** (a) Scatter plot of the estimated ratio of burst parameters $r_{estimate}$ of a pair of genes and of the ground-truth ratio of the burst parameters $r_{true}$ of the same pair of genes. All parameter sets are sampled from the region of parameter space where the technical-noise-corrected NB distribution is optimally selected using the aeBIC approach (purple regions in Fig C in S1 Text). Points marked as blue are those pairs of genes for which the estimation led to an incorrect ordering of genes by the size of the burst parameter. The order was correctly inferred for gene pairs corresponding to orange and green points. Perfect ratio estimation is shown by the solid red line; gene pairs corresponding to orange (green) points overestimate (underestimate) the distance between the burst parameters of the gene pairs. (b) Distributions of the relative errors in burst frequency and burst size for those pairs of genes for which the order was correctly inferred (orange and green points) in (a). Note that the inference and model selection approach here assumes full knowledge of the distribution of probability capture, an ideal case.

https://doi.org/10.1371/journal.pcbi.1014014.g008

in 86 – 87% of cases. However, the ranking of genes by burst size is significantly less accurate: it is correct in merely 58 – 69% of all cases, though it is better than expected by chance. In Fig 8b we show the relative errors of the burst parameters for the pairs of genes that were correctly ranked (orange and green points in (a)). This verifies that generally relative errors and ranking accuracy are not related to each other since the ranking can be accurate and yet the relative errors can be large.

The main limitations of our approach are that we assume: (i) perfect knowledge of the distribution of capture probability; (ii) the information in the first two moments of the transcript count distribution is enough to accurately infer the burst parameters; (iii) we exclusively perform parameter inference using the technical-noise corrected NB model. To overcome these limitations, in Sect 4.6, we describe an approach which simulates the workflow of an actual scRNA-seq experiment: (i) synthetic count data is generated for thousands of genes in $n_c$ cells, which includes noise from both biological and technical sources; technical noise is modeled by a Beta distributed transcript capture probability; (ii) a maximum likelihood-based approach is used to infer the burst frequency and size, and to select between the technical-noise corrected telegraph, NB and Poisson models derived in Sect 4.4. Note that this approach relies on using the measured total counts from all genes in a cell as a proxy for the (unknown) capture probability for that cell; the approach also does not simply use the first two moments of the observed distribution of counts, but the full distribution information and hence overcomes some of the known limitations of moment-matching approaches [70]; (iii) for those genes for which the technical-noise corrected NB model is selected as optimal by the BIC, the burst frequency and size are estimated separately from the technical-noise corrected telegraph and NB models. These are compared to the ground-truth values to calculate the relative errors. In addition, the percentage of pairs of genes that are correctly ranked according to the magnitude of their burst frequency and size, are calculated for each model. The whole methodology is illustrated in Fig Da in S1 Text. We considered three variations of this procedure, according to the generation of synthetic data in step (i):(a) $n_c = 10^3$ cells, $N_\sigma = 10$ and the distribution of $p_{\text{cap}}$ is fixed to $\text{Beta}(15, 35)$; (b) $n_c = 10^4$ cells, $N_\sigma = 10$ and the distribution of $p_{\text{cap}}$ is fixed to $\text{Beta}(60, 140)$; (c) $n_c = 10^4$ cells, $N_\sigma = 20$ and the distribution of $p_{\text{cap}}$ is fixed to $\text{Beta}(60, 140)$.

The results for (a) are shown in Fig D in S1 Text and for (b) and (c) are shown in Fig E in S1 Text. We find that for each fixed value of $N_\sigma$, the relative errors of the burst frequency and size increase with the fraction of time that the gene spends in the active state $f_{\text{on}}$. This agrees with the results shown in Fig 7. Interestingly, even though the genes analyzed are those for which the technical-noise corrected NB model is selected as the optimal model by BIC, the parameter estimates are more accurate using the technical-noise corrected telegraph model (compare blue and green bars in Fig Db, Fig Ea and Fig Ec in S1 Text). In addition, a high percentage of pairs of genes are correctly ranked according to the magnitude of their burst frequency, both by the technical noise-corrected NB and telegraph models; however, ranking by the burst size was much less reliable (see the piecharts in Fig Dc, Fig Eb and Fig Ed in S1 Text). The rankings also agree with the results shown in Fig 8. Hence, overall, the results obtained from the MLE-based inference and BIC verify the accuracy of those previously obtained using moment-matching and aeBIC.

### 2.6 Analysis of experimental data reveals a substantial fraction of genes best fitted by NB but not transcriptionally bursty

Finally, we sought to determine how many genes that are best fitted by the negative binomial (NB) distribution are truly transcriptionally bursty, and whether this fraction is significant. Prior review papers, such as [71], provide ranges of switching rates ($\sigma_{\text{on}}$ and $\sigma_{\text{off}}$) that partially address this question. To answer it more comprehensively, we analyzed scRNA-seq data from mouse fibroblasts [72], preprocessing the dataset to ensure reliability (see S1 Text, Sect 4.7). After preprocessing, 670 cells and 21,684 genes were retained.

We fitted the three $p_{\text{cap}}$-modified models in Fig 7a to this dataset using MLE. The distribution of $p_{\text{cap}}$ was estimated via kernel density estimation, and the computation of count distributions over $p_{\text{cap}}$ followed the procedure described in Sect 4.8, consistent with the approach outlined in Supplementary Information Section VI of Ref [73]. After parameter inference, we applied BIC to select the best-fitting model for each gene and found that ~80% of genes were best fitted by the NB model (Fig 9a).

For these NB-fitted genes, we refined the inference using Bayesian estimation with the Turing.jl package in Julia, fitting each gene's count distribution with the $p_{\text{cap}}$-modified telegraph model. We then quantified the confidence interval (CI) for $f_{\text{on}}$. If the CI range (2.5%–97.5%) divided by the median was less than 0.4 (i.e., ±20% around the median), we considered

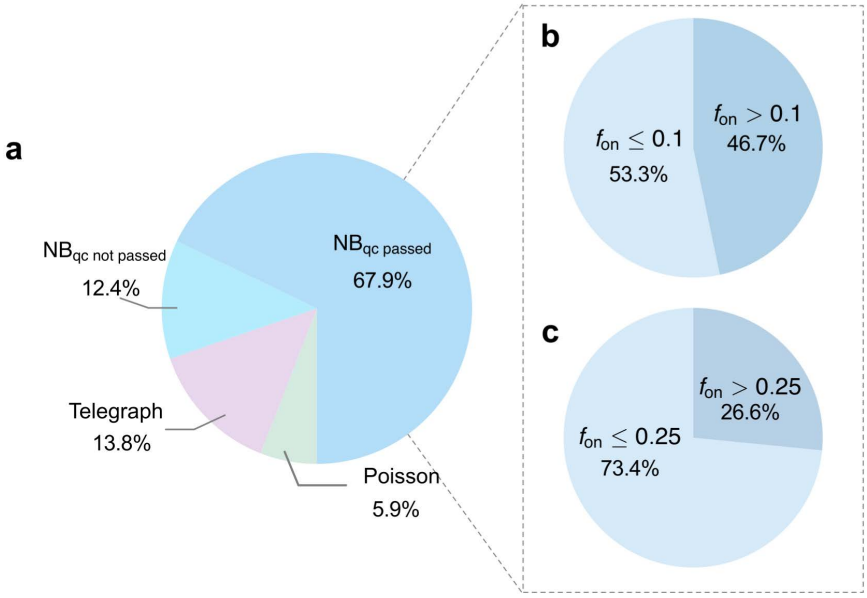


**Fig 9. Analysis of mouse fibroblast scRNA-seq data reveals that many genes best fitted by the NB distribution are not transcriptionally bursty. (a)** Model selection using BIC on 21,684 genes (670 cells) after preprocessing shows that ~80% of genes are best fitted by the NB model. **(b,c)** For these NB-fitted genes, Bayesian inference of $f_{on}$ was performed using the $p_{cap}$-modified telegraph model, with reliability assessed by the confidence interval criterion (CI range/median <0.4). The fraction of NB-fitted genes classified as transcriptionally bursty depends on the threshold chosen for $f_{on}$: 0.1 in **(b)** and 0.25 in **(c)**. In both cases, a substantial fraction of genes are best fitted by the NB model yet are not transcriptionally bursty.

https://doi.org/10.1371/journal.pcbi.1014014.g009

the estimate of $f_{on}$ reliable. This yielded 14,722 genes, of which the inferred parameters have been summarized and deposited on GitHub (see Data availability statement).

We next computed the fraction of genes classified as transcriptionally bursty under different thresholds for $f_{on}$ (0.1 in Fig 9b and 0.25 in Fig 9c). In both cases, a considerable fraction of genes were best fitted by the NB model yet not transcriptionally bursty. This result highlights the importance of our theoretical framework for correctly interpreting NB fits in scRNA-seq data.

## 3 Discussion

In this study, we have investigated why the NB distribution effectively approximates the transcript count distribution observed in scRNA-seq experiments and evaluated the biological implications of its two parameters. Presently, the universality of the NB distribution in UMI count data is puzzling given that a decade of analytical studies have unearthed only one stringent condition under which the distribution of the telegraph model of stochastic gene expression is well approximated by an NB distribution: genes must be predominantly inactive in a population of identical cells — conditions that are unlikely met in scRNA-seq experiments.

We first showed using theory and a new model selection criterion (aeBIC) that in the ideal case where the noise in the data is purely of biological origin, the NB distribution provides an accurate and optimal approximation of the transcript count distribution of a gene in a crescent-shaped region of its $f_{on}$ vs $N_\sigma$ parameter space where $f_{on}$ is the fraction of time spent in the active state, while $N_\sigma$ is the sum of the activation and deactivation rates (normalized by the degradation rate). This region occupied about 20% of the scanned parameter space, showing little dependence with the sample size (number of cells). In particular, the NB distribution is an excellent approximation in an intermediate range of $N_\sigma$ and the size of this range increases with decreasing $f_{on}$. The classical parameter regime where the NB distribution is thought to be a good

approximation to the telegraph model, i.e., when $f_{on}$ is small (genes mostly inactive) and transcriptional bursting is apparent [49], is a subset of the parameter regime that we have identified. In fact, this new region encompasses genes whose $f_{on}$ ranges from small to large, thus showing that a good fit of the NB to the measured distribution of transcript counts is generally not indicative of transcriptional bursting.

Next, we investigated the more realistic case where the noise, i.e., the variability of mRNA numbers across cells, is due to both biological and technical noise. We simulated observed scRNA-seq data by downsampling the counts of the telegraph model of gene expression, assuming that the transcript capture probability for each cell is either constant or beta distributed. The selection of models through aeBIC allowed us to quantify how the proportion $F$ of the parameter space where the NB distribution optimally fits the data varies with sample size and the shape of the capture probability distribution. We found that for a sample size of *100* cells where all noise is biological noise, $F \approx 21\%$; where technical noise is added assuming that each cell has a transcript capture probability of 0.3, $F \approx 19\%$; where technical noise is added assuming that the transcript capture probability is sampled from a Beta distribution with mean 0.3 and CV= 0.21, $F \approx 27\%$ . In contrast, for a sample size of $10^4$ cells, the parameter space fractions for the aforementioned three cases were $F \approx 24\%$, $F \approx 21\%$ and $F \approx 58\%$. Note that a mean capture probability of 0.3 is typical of 10x Genomics Chromium Single Cell 3' version 3 reagent chemistry [65]. From these comparisons, we can deduce the following. If the capture probability distribution is: (i) narrow, the results are similar to those obtained under the assumption that all noise is biological; (ii) wide but the sample size is small (order 100 cells) then the results are also similar to those obtained under the assumption that all noise is biological; (iii) wide but for sample sizes of the order of thousands of cells, we find a significant enhancement of the region of parameter space where the NB distribution is the optimal model compared to the case where all noise is biological. Specifically, the proportion of the parameter space favoring the telegraph model remains largely constant, while the proportion favoring the NB distribution increases at the expense of the Poisson distribution. Hence, given that the capture probability distribution is wide in many cases (reflected by wide distributions of the total UMI counts per cell [17,50,51]), our results suggest that for small sample sizes, biological noise from those genes with an intermediate-sized $N_\sigma$ parameter is the main determinant of the NB character of the count distribution, while for large sample sizes, technical noise leads to NB distributions in regions of parameter space where biological noise cannot.

A previous study by Tang et al. [37] generated a simulated data set consisting of 500 cells and 7000 genes by sampling from the steady-state distribution of the telegraph model (Eq (1)). For each gene, the Akaike Information Criterion (AIC) scores were calculated for the telegraph, NB, and Poisson models, and the model with the lowest AIC was selected as the optimal fit. The set of genes for which the Poisson and NB models were favored tended to have a lower mean expression; however, the distributions of the means of the genes in the three groups (see Fig 3a of [37]) overlapped significantly, indicating that the mean alone is not enough to explain the model that best fits the expression of each gene. In the phase diagrams in Fig 4c we show the regions where the three models are selected as a function of $f_{on}$ and $N_\sigma$ — clearly $N_\sigma$ , not $f_{on}$, is the primary determinant of the chosen model. Since the mean mRNA is proportional to $f_{on}$ (the mean mRNA of the telegraph model is $\rho f_{on}$ and $\rho$ is fixed in these phase plots), it follows that the selection of the optimal model is only weakly determined by the mean RNA. This explains why Tang et al. [37] did not observe a strong relationship between the selected model and the mean mRNA count. Another study [74] has also shown that the NB distribution is often complex enough to describe the simulated scRNA-seq data generated from the telegraph model but did not map the relationship between the choice of the best-fit model and the regions of parameter space, or account for technical noise, as we have done here. The complete mapping that we have reported was only possible due to the use of the novel model selection criterion, aeBIC, which is a computationally efficient and accurate estimator of the model most frequently selected by the standard maximum likelihood procedure followed by BIC.

We also investigated the reliability of the burst frequency and burst size estimated from the two parameters of the best-fitting NB distribution in the parameter space region where this model is selected as the optimal one. These are commonly estimated by interpreting the NB distribution as that arising from a simpler mechanistic model than the telegraph

model, namely the bursty gene expression model (with reaction scheme Eq (15)), which was first studied in [69] and is now commonly used as the basis for more sophisticated stochastic models of gene expression [75–80] including those used to fit scRNA-seq data [6,52,81–84]. This model can be derived from the telegraph model under the assumption that the gene inactivation rate is much larger than the gene activation rates [49], i.e., the transcriptional bursting regime. However, our analysis shows that when the parameter $N_\sigma$ is large, the steady-state distribution of the telegraph model is well approximated by the negative binomial (NB) distribution, a scenario that applies beyond the bursting regime. This means that estimating the burst frequency and size of a gene by fitting the NB distribution solution of the bursty model Eq (15) to scRNA-seq measurements of its expression (after correcting for technical noise) can lead to potentially large errors, even if the data is excellently fit by a NB distribution. Our analysis supports this notion — we find that in the region of parameter space where the distribution of the telegraph model is best fit by an NB distribution, the errors in the estimated burst frequency and size are generally large, easily exceeding 50% for the burst frequency and at most 100% for the burst size. However, interestingly, we observe that the relative magnitudes of the burst parameters still hold some validity (as also suggested by a recent study of the cell-cycle dependence of bursty gene expression [6]). In particular, we find that the ranking of two genes by their burst frequency estimate is correct in about $86-87\%$ of the cases; for the burst size, the ranking is correct in $58-69\%$ of the cases, showing less reliability. We also found that in the parameter space region where the NB model is preferentially selected, the limitations discussed above remain largely the same if instead one estimates the burst size and burst frequency from the three parameters obtained by fitting the distribution of the telegraph model (after correcting for technical noise) to the data using the method of maximum likelihood [35,37,85–88].

Our study also has some limitations. We have accounted for cell-to-cell variability in transcript counts due to intrinsic noise and differences in the effective transcription rate from one cell to another due to a varying transcript capture probability. Any variation in the effective transcription rate due to variability in the transcription rate (extrinsic noise on the transcription rate) between cells is indistinguishable from variability in the transcript capture probability and hence is automatically accounted for in our present method (see S1 Text, Sect 4.3). However, we have assumed that there is no cell-to-cell variation in the activation and inactivation rates, which is, of course, a simplification of biological reality. Within our phase diagram methodology, these effects can be properly incorporated using a ground-truth distribution $\mathcal{G}$ in Eq (10) that is derived via an integration of the distribution of the telegraph model over the distributions of the activation and inactivation rates — these are typically unknown but could be approximated by gamma or lognormal distributions as in [52]. The issues we have identified regarding the inaccuracy of the absolute values of the burst parameters are possibly because we have exclusively used the steady-state distributions of stochastic gene expression models. These are most commonly used for two reasons: (i) scRNA-seq data often come from a single snapshot measurement; (ii) it is much easier to solve the chemical master equation of gene expression models in steady-state conditions than in time [89]. However, the median mRNA half-life in mammalian cells is a sizeable fraction of the cell-cycle duration (for, e.g., $\approx 7$ hours [90] compared to $\approx 13$ hours [91] in mouse embryonic stem cells), and hence it is unlikely that a steady state assumption generally holds in this case. To overcome this issue, one can first use methods like DeepCycle [92] or VeloCycle [93] to assign a cell age $\theta$ (which varies between $0$ at the beginning of the cell-cycle and $1$ at cell division) to each cell and then fit an age-dependent model of gene expression to the age-resolved scRNA-seq data [6]. Alternative ways to possibly circumvent the issues we have reported using steady-state models is to instead fit models that account for different cell states due to differentiation [94] or models that use a variety of single-cell data (transcriptomic, proteomic and epigenomic) [95]. These approaches are computationally challenging and are actively under investigation.

Additionally, the eigenvalues of the Hessian matrix of the log-likelihood quantify the curvature of the likelihood surface and reflect the degree of parameter degeneracy. As the telegraph model converges to the NB and further to the Poisson model with increasing $N_\sigma$, the number of effective parameters decreases from three to two and finally to one. This reduction leads to broader confidence intervals for inferred parameters, with eigenvalue ratios diverging to infinity. However, as shown in Fig F in S1 Text, these ratios increase only gradually with $N_\sigma$ and do not provide clear thresholds for model

selection. In contrast, the aeBIC method yields well-defined selection boundaries, demonstrating that Hessian eigenvalues are not reliable indicators for model selection.

In conclusion, our findings establish that the NB distribution is a robust approximation of scRNA-seq transcript count distributions across diverse noise conditions and parameter regimes. Notably, its broad applicability challenges the conventional theoretical link to bursty gene expression, since both biological and technical noise can induce NB-like behavior even outside of classical conditions. Our analysis cautions against direct burst parameter estimation from NB or telegraph model fits and calls for nuanced model selection and parameter estimation approaches in single-cell genomics.

## 4 Methods

### 4.1 Proof of Theorem 1

It is well known that the probability distribution of the Beta-Poisson process

$$x \sim \text{Beta}(\sigma_{\text{on}}, \sigma_{\text{off}}), \qquad n \sim \text{Pois}(\rho x),$$

is identical to the steady-state distribution of the telegraph model given by Eq (1) [35]. Another standard result is that the probability distribution of the Gamma-Poisson process

$$y \sim \text{Gamma}\left(r_{\text{e}}, \frac{p_{\text{e}}}{1-p_{\text{e}}}\right), \qquad n \sim \text{Pois}(y)$$

is identical to that of a negative binomial distribution $\text{NB}(r_{\text{e}},\ p_{\text{e}})$. In what follows we shall fix these two parameters to be

$$r_{\text{e}} = \frac{\sigma_{\text{on}}(\sigma_{\text{on}} + \sigma_{\text{off}} + 1)}{\sigma_{\text{off}}}, \quad p_{\text{e}} = \frac{(\sigma_{\text{on}} + \sigma_{\text{off}})(\sigma_{\text{on}} + \sigma_{\text{off}} + 1)}{(\sigma_{\text{on}} + \sigma_{\text{off}})(\sigma_{\text{on}} + \sigma_{\text{off}} + 1) + \rho\sigma_{\text{off}}}. \tag{18}$$

Note also that the definition of the Gamma distribution, $\text{Gamma}(\alpha, \beta)$, that we use is $f(y) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y},\ y > 0$.

Proposition 7 in Ref. [96] states that for two mixed Poisson distributions, denoted as $\text{MP}(g_1)$ and $\text{MP}(g_2)$, the convergence $\text{MP}(g_1) \to \text{MP}(g_2)$ holds if and only if $g_1 \to g_2$ where $\to$ denotes convergence in distribution. Consequently, to prove Theorem 1, it suffices to demonstrate that the distribution of the random variable $x$ converges to that of the random variable $y/\rho$ as $N_\sigma = \sigma_{\text{on}} + \sigma_{\text{off}} \to \infty$.

To proceed, we first determine the density function of the transformed random variable $\bar{y} = y/\rho$. This is computed as

$$\pi(\bar{y}) = \pi(\rho\bar{y}) \times \frac{\text{d}(\rho\bar{y})}{\text{d}\bar{y}} = \frac{\rho}{\Gamma(r_{\text{e}})}\left(\frac{p_{\text{e}}}{1-p_{\text{e}}}\right)^{r_{\text{e}}} (\rho\bar{y})^{r_{\text{e}}-1}\, \text{e}^{-\frac{p_{\text{e}}\rho\bar{y}}{1-p_{\text{e}}}} = \frac{1}{\Gamma(r_{\text{e}})}\left(\frac{\rho p_{\text{e}}}{1-p_{\text{e}}}\right)^{r_{\text{e}}} (\bar{y})^{r_{\text{e}}-1}\, \text{e}^{-\frac{p_{\text{e}}\rho\bar{y}}{1-p_{\text{e}}}},$$

which precisely corresponds to the density function of $\text{Gamma}(r_{\text{e}}, \frac{\rho p_{\text{e}}}{1-p_{\text{e}}})$.

To show $x \to \bar{y}$, we need to establish the convergence

$$\text{Beta}(\sigma_{\text{on}}, \sigma_{\text{off}}) \to \text{Gamma}\left(r_{\text{e}}, \frac{\rho p_{\text{e}}}{1-p_{\text{e}}}\right), \tag{19}$$

as $N_\sigma = \sigma_{\text{on}} + \sigma_{\text{off}} \to \infty$. Rather than directly proving Eq (19), we instead demonstrate that both distributions converge to a common limiting distribution, denoted as Dist, i.e.,

$$\text{Gamma}\left(r_{\text{e}}, \frac{\rho p_{\text{e}}}{1-p_{\text{e}}}\right) \to \text{Dist}, \qquad \text{Beta}(\sigma_{\text{on}}, \sigma_{\text{off}}) \to \text{Dist}. \tag{20}$$

It turns out that Dist is a normal distribution.

Next, we show that the Gamma distribution in Eq (20) converges to a normal distribution as $N_\sigma \to \infty$. To achieve this, we introduce the moment generating function (MGF) for the random variable $\bar{y}$,

$$M_{\bar{y}}(t) = \langle e^{t\bar{y}} \rangle.$$

Consider the linear transformation

$$\tilde{y} = \frac{\bar{y} - \mu}{\sigma},$$

where

$$\mu = f_{\text{on}}, \qquad \sigma = \sqrt{\frac{f_{\text{on}}(1 - f_{\text{on}})}{N_\sigma + 1}},$$

and $\sqrt{r_e} = \mu/\sigma$. The MGF of the transformed random variable $\tilde{y}$ is

$$M_{\tilde{y}}(t) = \langle e^{t\tilde{y}} \rangle = \langle e^{t(\frac{\bar{y}}{\sigma} - \sqrt{r_e})} \rangle = e^{-\sqrt{r_e}t} \langle e^{\frac{t\bar{y}}{\sigma}} \rangle = e^{-\sqrt{r_e}t} M_{\bar{y}}\left(\frac{t}{\sigma}\right).$$

Using the MGF of the Gamma distribution, this simplifies to

$$M_{\tilde{y}}(t) = e^{-\sqrt{r_e}t} \left(1 - \frac{t}{\sqrt{r_e}}\right)^{-r_e}.$$

As $N_\sigma \to \infty$, $r_e \to \infty$, the expression further reduces to

$$\lim_{N_\sigma \to \infty} M_{\tilde{y}}(t) = \exp\left[-\sqrt{r_e}t - r_e \ln\left(1 - \frac{t}{\sqrt{r_e}}\right)\right] = \exp\left[\frac{t^2}{2} + O\left(\frac{1}{\sqrt{r_e}}\right) t^3\right] = e^{\frac{t^2}{2}} + O\left(\frac{1}{\sqrt{N_\sigma}}\right),$$

which is the MGF of the standard normal distribution $\mathcal{N}(0, 1)$. Therefore, we conclude that

$$\lim_{N_\sigma \to \infty} \text{Gamma}\left(r_e, \frac{\rho p_e}{1 - p_e}\right) = \mathcal{N}(\mu, \sigma),$$

with convergence rate $O(1/\sqrt{N_\sigma})$.

Finally, we show that $\text{Beta}(\sigma_{\text{on}}, \sigma_{\text{off}}) = \text{Beta}(f_{\text{on}} N_\sigma, (1 - f_{\text{on}}) N_\sigma) \to \mathcal{N}(\mu, \sigma)$ as $N_\sigma \to \infty$. Note that we use the definition of the Beta distribution, $\text{Beta}(\alpha, \beta)$, given by

$$f(x) = \frac{x^{\alpha-1}(1 - x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1, \tag{21}$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$. The density function of this Beta distribution at the point $x = \mu + \sigma t$ is given by

$$f(\mu + \sigma t) = \underbrace{\frac{\Gamma(N_\sigma)}{\Gamma(f_{\text{on}} N_\sigma)\Gamma((1 - f_{\text{on}})N_\sigma)} \mu^{f_{\text{on}} N_\sigma - 1}(1 - \mu)^{(1 - f_{\text{on}})N_\sigma - 1}}_{\mathcal{A}} \underbrace{\left(1 + \frac{\sigma t}{\mu}\right)^{f_{\text{on}} N_\sigma - 1} \left(1 - \frac{\sigma t}{1 - \mu}\right)^{(1 - f_{\text{on}})N_\sigma - 1}}_{\mathcal{B}}. \tag{22}$$

Using Stirling's formula, $\Gamma(x) \to \sqrt{2\pi/x}(x/e)^x[1 + O(1/x)]$ as $x \to \infty$, the term $\mathcal{A}$ in Eq (22) reduces to

$$\begin{aligned}\lim_{N_\sigma \to \infty} \mathcal{A} &= \frac{\sqrt{\frac{2\pi}{N_\sigma}}(N_\sigma/e)^{N_\sigma} f_{\text{on}}{}^{f_{\text{on}}N_\sigma - 1}(1-f_{\text{on}})^{N_\sigma(1-f_{\text{on}})-1}[1 + O(1/N_\sigma)]}{\sqrt{\frac{2\pi}{f_{\text{on}}N_\sigma}}(f_{\text{on}}N_\sigma/e)^{f_{\text{on}}N_\sigma}\sqrt{\frac{2\pi}{N_\sigma - f_{\text{on}}N_\sigma}}[(N_\sigma - f_{\text{on}}N_\sigma)/e]^{N_\sigma - f_{\text{on}}N_\sigma}[1 + O(1/(f_{\text{on}}N_\sigma))][1 + O(1/(N_\sigma - f_{\text{on}}N_\sigma))]} \\ &= \sqrt{\frac{N_\sigma}{2\pi f_{\text{on}}(1-f_{\text{on}})}}\left[1 + O\left(\frac{1}{N_\sigma}\right)\right] \\ &\approx \frac{1}{\sqrt{2\pi}\sigma} + O\left(\frac{1}{N_\sigma}\right).\end{aligned} \tag{23}$$

Next, we rewrite the term $\mathcal{B}$ in Eq (22) as

$$\mathcal{B} = \exp\left[(f_{\text{on}}N_\sigma - 1)\ln\left(1 + \frac{\sigma t}{\mu}\right) + (N_\sigma - f_{\text{on}}N_\sigma - 1)\ln\left(1 - \frac{\sigma t}{1-\mu}\right)\right].$$

Noting that $\sigma/\mu \to 0$ and $\sigma/(1-\mu) \to 0$ as $N_\sigma \to \infty$, we expand $\mathcal{B}$ around $t = 0$ using a Taylor series expansion, yielding

$$\begin{aligned}\mathcal{B} = \exp&\left[\underbrace{\sigma t\left(\frac{f_{\text{on}}N_\sigma - 1}{\mu} - \frac{N_\sigma(1-f_{\text{on}})-1}{1-\mu}\right)}_{\mathcal{B}_1} - \underbrace{\frac{\sigma^2 t^2}{2}\left(\frac{f_{\text{on}}N_\sigma - 1}{\mu^2} + \frac{N_\sigma(1-f_{\text{on}})-1}{(1-\mu)^2}\right)}_{\mathcal{B}_2}\right] \\ \times \exp&\left[\underbrace{O\left(\frac{(f_{\text{on}}N_\sigma - 1)\sigma^3}{\mu^3} - \frac{(N_\sigma(1-f_{\text{on}})-1)\sigma^3}{(1-\mu)^3}\right)t^3}_{\mathcal{B}_3}\right],\end{aligned} \tag{24}$$

where

$$\mathcal{B}_1 = \frac{t}{\sqrt{N_\sigma + 1}}\left(\sqrt{\frac{\sigma_{\text{on}}}{\sigma_{\text{off}}}} - \sqrt{\frac{\sigma_{\text{off}}}{\sigma_{\text{on}}}}\right),$$

$$\mathcal{B}_2 = \frac{t^2}{2(N_\sigma + 1)}\left(N_\sigma - \frac{1-f_{\text{on}}}{f_{\text{on}}} - \frac{f_{\text{on}}}{1-f_{\text{on}}}\right)$$

$$= \frac{t^2}{2} + O\left(\frac{1}{N_\sigma}\right)t^2,$$

and

$$\mathcal{B}_3 = O\left(\frac{1}{\sqrt{N_\sigma}}\right)t^3.$$

Using Eqs (22)–(24), we obtain

$$f(x) = f(\mu + \sigma t) \approx \frac{e^{-t^2/2}}{\sqrt{2\pi}\,\sigma} = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma},$$

which corresponds to the probability density function of the normal distribution $\mathcal{N}(\mu, \sigma)$, with a convergence rate of $O(1/\sqrt{N_\sigma})$. This concludes the proof of Theorem 1.

### 4.2 Mathematical relationship between aeBIC and BIC

Since the data in set $\mathcal{X}$ (the set of count data $\{n_i\}$ for $i = 1, \cdots, n_c$) are randomly sampled from the ground-truth distribution $\mathcal{G}$, the expectation of the BIC (Eq (9)) is given by

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathrm{BIC}(\mathcal{M}, \mathcal{X})] = |\mathcal{M}| \ln n_c - 2\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}\left[\max_\theta \mathcal{L}_\theta\right] \leq |\mathcal{M}| \ln n_c - 2 \max_\theta \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathcal{L}_\theta], \tag{25}$$

where the inequality follows from the fact that the expectation of the maximum is greater than or equal to the maximum of the expectation. A proof of this inequality is as follows.

For any fixed $\theta$, it is true that

$$\mathcal{L}_\theta \leq \max_{\theta'} \mathcal{L}_{\theta'}.$$

So taking expectations on both sides yields

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathcal{L}_\theta] \leq \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\max_{\theta'} \mathcal{L}_{\theta'}].$$

Because this is true for any $\theta$, it follows that

$$\max_\theta \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathcal{L}_\theta] \leq \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\max_\theta \mathcal{L}_\theta].$$

Furthermore, we note that

$$\max_\theta \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathcal{L}_\theta] = \max_\theta \left[\sum_{n=0}^{\infty}\sum_{i=1}^{n_c} \mathbf{1}_{n_i=n} \ln P_{\mathcal{M}}(n|\theta)\right] = n_c \max_\theta \left[\sum_{n=0}^{\infty} P_{\mathcal{G}}(n) \ln P_{\mathcal{M}}(n|\theta)\right] = -n_c \varepsilon_{\mathrm{c}}(\mathcal{M}_{\mathrm{MLE}}, \mathcal{G}), \tag{26}$$

according to Eq (11). Note that the first step in Eq (26) is a reformulation of the log-likelihood function: the term $\mathbf{1}_{n_i=n}$ is an indicator function equal to 1 if and only if $n_i = n$, and hence $P_{\mathcal{G}}(n) = \sum_{i=1}^{n_c} \mathbf{1}_{n_i=n}$. From Eqs (10), (25), and (26), it follows that

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}[\mathrm{BIC}(\mathcal{M}, \mathcal{X})] \leq \mathrm{aeBIC}(\mathcal{M}, n_c), \tag{27}$$

indicating that aeBIC serves as an upper bound for the expectation of the BIC. For a fixed distribution $\mathcal{M}$ with fixed parameters, $\mathcal{L}_\theta \to -n_c\varepsilon_{\mathrm{c}}(\mathcal{M}, \mathcal{G})$ as $n_c \to \infty$, which further suggests

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}\left[\max_\theta \mathcal{L}_\theta\right] \to \mathbf{E}_{\mathcal{X}\sim\mathcal{G}}\left[-\max_\theta n_c\varepsilon_{\mathrm{c}}(\mathcal{M}, \mathcal{G})\right].$$

By noting that

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}\left[-\max_{\theta} n_c\varepsilon_c(\mathcal{M},\mathcal{G})\right] = -\max_{\theta} n_c\varepsilon_c(\mathcal{M},\mathcal{G}),$$

one can conclude that

$$\mathbf{E}_{\mathcal{X}\sim\mathcal{G}}\left[\max_{\theta}\mathcal{L}_\theta\right] \to n_c\varepsilon_c(\mathcal{M}_{\mathrm{MLE}},\mathcal{G})$$

as $n_c \to \infty$.

### 4.3 Breakdown of variability involved in transcription

Variability in scRNA-seq data arises from three major sources: intrinsic noise, extrinsic noise, and technical noise. The first two correspond to biological variability, whereas the latter is introduced by sequencing procedures.

**Intrinsic noise** mainly reflects stochastic promoter and chromatin activity, which is well captured by the gene-state switching dynamics of the telegraph model, as supported experimentally in [67]. The probability generating function (PGF) of the steady-state mRNA distribution under this model is

$$G_{\mathrm{tele}}(z) = \sum_n z^n P(n) = {}_1F_1(\sigma_{\mathrm{on}}, \sigma_{\mathrm{on}}+\sigma_{\mathrm{off}}, \rho(z-1)),$$

where $P(n)$ is given in Eq (1).

**Extrinsic noise** is primarily attributed to cell-to-cell variation in transcription rates [66,67], manifested as heterogeneity in $\rho$.

**Technical noise** arises during sequencing, most notably from variability in transcript capture efficiency. Assuming each transcript is independently captured with probability $p_{\mathrm{cap}}$ [10,62], the observed counts in each cell follow the PGF

$$G_{\mathrm{obs}}(z) = {}_1F_1(\sigma_{\mathrm{on}}, \sigma_{\mathrm{on}}+\sigma_{\mathrm{off}}, p_{\mathrm{cap}}\rho(z-1)). \tag{28}$$

This result follows from the general property that binomial downsampling of a distribution with PGF $G(z)$ produces a new PGF $G(1-p_{\mathrm{cap}}(1-z))$ (see [49], SI p. 9 for the special case $p_{\mathrm{cap}} = 1/2$; for a more general discussion see [97]).

Importantly, Eq (28) shows that the cell-specific transcription rate $\rho$ and capture probability $p_{\mathrm{cap}}$ always appear as a product, $\lambda = \rho p_{\mathrm{cap}}$, whose distribution can in principle be calculated from knowledge of the joint distribution $p(\rho, p_{\mathrm{cap}})$. This implies that, at the level of transcript count distributions, variability in transcription rates and variability in capture probability are mathematically indistinguishable. For this reason, in the following we focus our discussion on variability in transcript capture probability.

### 4.4 Derivation of the observed count distributions under the assumption of Beta-distributed capture rates

**4.4.1 Telegraph model.** If the capture rate $p_{\mathrm{cap}}$ follows a Beta distribution, $p_{\mathrm{cap}} \sim \mathrm{Beta}(a, b)$, defined by

$$P(p_{\mathrm{cap}}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_{\mathrm{cap}}^{a-1}(1-p_{\mathrm{cap}})^{b-1},$$

the PGF of the observed counts is given by integrating Eq (28) over the distribution of $p_{\mathrm{cap}}$, which yields a generalized hypergeometric function

$$G_{\text{obs}}(z) = \int_0^1 {}_1F_1(\sigma_{\text{on}}, \sigma_{\text{on}} + \sigma_{\text{off}}, p_{\text{cap}}\rho(z-1))P(p_{\text{cap}})dp_{\text{cap}} = {}_2F_2(a, \sigma_{\text{on}}; a+b, \sigma_{\text{on}} + \sigma_{\text{off}}; \rho(z-1)). \tag{29}$$

Finally, we obtain the closed-form probability of observing $n$ counts

$$P(n) = \frac{1}{n!}\frac{d^n}{dz^n}G_{\text{obs}}(z)\Big|_{z=0} = \frac{\rho^n}{n!}\frac{(\sigma_{\text{on}})_n a_n}{(\sigma_{\text{on}} + \sigma_{\text{off}})_n (a+b)_n} {}_2F_2(a+n, \sigma_{\text{on}} + n; a+b+n, \sigma_{\text{on}} + \sigma_{\text{off}} + n; -\rho). \tag{30}$$

Note that this method of accounting for technical noise, i.e., via compound distributions, is essentially the same as that used to account for static extrinsic noise [98,99].

**4.4.2 Negative binomial model.** The PGF of negative binomial distribution, $\text{NB}(r, p)$, is given by

$$G_{\text{NB}}(z) = (1 - (1-p)(z-1)/p)^{-r}.$$

If $p_{\text{cap}}$ is the same in all cells, the PGF of the observed counts becomes

$$G_{\text{obs}}(z) = (1 - p_{\text{cap}}(1-p)(z-1)/p)^{-r}.$$

If we have a $\text{Beta}(a, b)$ distribution of $p_{\text{cap}}$ then we integrate the above expression over the Beta distribution, yielding

$$G_{\text{obs}}(z) = \int_0^1 (1 - p_{\text{cap}}(1-p)(z-1)/p)^{-r} P(p_{\text{cap}})dp_{\text{cap}} = {}_2F_1(a, r; a+b; (1-p)(z-1)/p). \tag{31}$$

Finally, we obtain the probability distribution of the observed counts

$$P(n) = \frac{1}{n!}\frac{d^n}{dz^n}G_{\text{obs}}(z)\Big|_{z=0} = \frac{\rho^n}{n!}\frac{(r)_n (a)_n}{(a+b)_n} {}_2F_1(a+n, r+n; a+b+n; -(1-p)/p).$$

**4.4.3 Poisson model.** The PGF of the Poisson distribution, $\text{Pois}(\lambda)$, is given by

$$G_{\text{pois}}(z) = \exp(\lambda(z-1)).$$

If $p_{\text{cap}}$ is the same in all cells, the PGF of the observed counts becomes

$$G_{\text{obs}}(z) = \exp(p_{\text{cap}}\lambda(z-1)).$$

If we have a $\text{Beta}(a, b)$ distribution of $p_{\text{cap}}$ then we integrate the above expression over the Beta distribution, yielding

$$G_{\text{obs}}(z) = {}_1F_1(a; a+b; \lambda(z-1)).$$

Finally, we obtain the probability distribution of the observed counts

$$P(n) = \frac{1}{n!}\frac{d^n}{dz^n}G_{\text{obs}}(z)\Big|_{z=0} = \frac{\lambda^n}{n!}\frac{(a)_n}{(a+b)_n} {}_1F_1(a+n, a+b+n, -\lambda). \tag{32}$$

### 4.5 Inference of burst parameters assuming ideal conditions

We consider ideal conditions: (i) infinite sample size so that the measured moments of mRNA numbers do not suffer from finite sample size effects and thus the method of moments can be used for reliable inference; (ii) knowledge of the distribution describing cell-to-cell variability of the capture probability; this means that we can completely correct for technical noise.

**4.5.1 Moments of the measured mRNA counts in scRNA-seq data.** We assume that the distribution of the measured counts from a gene of interest is given by the telegraph model with capture probability sampled from a Beta$(a, b)$ distribution, i.e., Eq (30).

Given a PGF of observed counts $G_{\text{observed}}(z)$, the first two orders of moments can be derived as

$$\langle n\rangle = \frac{d}{dz}G_{\text{observed}}(z)\Big|_{z=1}, \qquad \text{var}(n) = \frac{d^2}{dz^2}G_{\text{observed}}(z)\Big|_{z=1} + \langle n\rangle - \langle n\rangle^2. \tag{33}$$

Substituting Eq (29)into Eq (33), we obtain expressions for the mean and the variance of the measured count distribution

$$\langle n\rangle_{\text{tele,observed}} = \frac{a\rho\sigma_{\text{on}}}{(a+b)(\sigma_{\text{on}}+\sigma_{\text{off}})}, \tag{34}$$

and

$$\text{Var}_{\text{tele,observed}}(n) = \frac{a(a+1)\rho^2\sigma_{\text{on}}(1+\sigma_{\text{on}})}{(a+b)(a+b+1)(\sigma_{\text{off}}+\sigma_{\text{on}})(1+\sigma_{\text{off}}+\sigma_{\text{on}})} + \frac{a\rho\sigma_{\text{on}}}{(a+b)(\sigma_{\text{off}}+\sigma_{\text{on}})} - \left(\frac{a\rho\sigma_{\text{on}}}{(a+b)(\sigma_{\text{off}}+\sigma_{\text{on}})}\right)^2. \tag{35}$$

It follows that the Fano factor of mRNA counts (variance divided by the mean) is given by

$$\text{FF}_{\text{observed}} = 1 + \langle p_{\text{cap}}\rangle \frac{\rho\sigma_{\text{off}}}{(\sigma_{\text{on}}+\sigma_{\text{off}})(1+\sigma_{\text{on}}+\sigma_{\text{off}})} + \text{CV}^2_{p_{\text{cap}}}\langle p_{\text{cap}}\rangle \frac{\rho(1+\sigma_{\text{on}})}{1+\sigma_{\text{on}}+\sigma_{\text{off}}}, \tag{36}$$

where $\langle p_{\text{cap}}\rangle$ and $\text{CV}_{p_{\text{cap}}}$ are the mean and the coefficient of variation (standard deviation divided by the mean) of the distribution of the capture probability, Beta$(a, b)$.

**4.5.2 Moments of the best fit model.** We assume that we are in a region of parameter space where aeBIC selects the NB distribution with Beta-distributed capture probability as the optimal model. Substituting Eq (31) into Eq (33), we find expressions for the mean and variance of this model

$$\langle n\rangle_{\text{NB,observed}} = \frac{ar(1-p)}{(a+b)p}, \tag{37}$$

and

$$\text{Var}_{\text{NB,observed}}(n) = \frac{a(a+1)(1-p)^2r(r+1)}{(a+b)(a+b+1)p^2} + \frac{ar(1-p)}{(a+b)p} - \left(\frac{ar(1-p)}{(a+b)p}\right)^2. \tag{38}$$

**4.5.3 Fitting the model to the data and parameter inference.** Since we are assuming that the parameters of the Beta distribution, i.e., $a$ and $b$, are known, the only parameters to be inferred are $r$ and $p$. These two unknown parameters can be determined exactly by matching the mean and variance of the optimal model Eqs (37)–(38) with the mean and variance of the data Eqs (34)–(35). This leads to two simultaneous equations for $r$ and $p$ which when solved lead to expressions for the inferred two parameters of the NB distribution

$$r = \frac{\sigma_{on}(\sigma_{on} + \sigma_{off} + 1)}{\sigma_{off}}, \qquad p = \frac{(\sigma_{on} + \sigma_{off})(\sigma_{on} + \sigma_{off} + 1)}{(\sigma_{on} + \sigma_{off})(\sigma_{on} + \sigma_{off} + 1) + \rho\sigma_{off}}.$$

Note that these expressions do not have any dependence on the parameters of the technical noise ($a$ and $b$), which indeed show that the inference method has perfectly corrected for technical noise. In fact, $r$ and $p$ determined in this way are none other than those parameterizing the effective NB distribution Eq (6) that best fits the conventional telegraph model.

### 4.6 Parameter inference without exact knowledge of the capture rate distribution

Here, we describe an inference approach that simulates the workflow of an actual scRNA-seq experiment including the subsequent analysis of the data. Note that in the inference and model selection parts of the procedure, we do not assume that the distribution of the capture probability is known — rather we use the distribution of the measured total counts per cell as a proxy.

**4.6.1 Simulating genomic data.** We generated synthetic count data for a genome using the following procedure:

- For each cell $j = 1, \ldots, n_c$ ($n_c = 1000$), sample the capture rate $p_{cap,j}$ from a $\mathrm{Beta}(15, 35)$ distribution which has a mean of $\langle p_{cap} \rangle = 0.3$ and a CV of 0.21. This accounts for cell-to-cell variability in the transcript capture probability.
- Choose the fraction of time spent by genes in the active state from the set $f_{on} \in \{0.1, 0.2, \ldots, 0.7\}$, and fix $N_\sigma = 10$ and $\rho = 15$. Use the telegraph model as the ground-truth generative model for all genes.
- For each value of $f_{on}$ and each cell $j$, sample the transcript counts of gene $i$, $x_{ij} \sim \mathrm{Tele}(\rho, f_{on}N_\sigma, (1 - f_{on})N_\sigma)$, independently for 400 genes. This results in transcript numbers per cell for a total of $n_g = 2800$ genes.
- Simulate observed counts in each cell by downsampling the counts according to the Binomial distribution: $y_{ij} \sim \mathrm{Binomial}(x_{ij}, p_{cap,j})$.

**4.6.2 Parameter inference and model selection.**

- Compute the total observed count for each cell $j$ as

$$t_j = \sum_{i=1}^{n_g} y_{ij}.$$

- Estimate the normalization factor $\beta_j$ using the known mean capture rate $\langle p_{cap} \rangle$ (e.g., obtained via spike-in controls) [37]:

$$\beta_j = \frac{\langle p_{cap} \rangle t_j}{\langle t_j \rangle}.$$

- Use kernel density estimation (KDE) to estimate the distribution $p(\beta)$ of the normalization factors.
- For each gene $i$, estimate model parameters by minimizing the negative log likelihood across cells:

$$J(\theta_i) = -\sum_{j=1}^{n_c} \log P(y_{ij} \mid \theta_i), \tag{39}$$

where the likelihood $P(y_{ij} \mid \theta_i)$ is computed by marginalizing over $\beta$:

$$P(y_{ij} \mid \theta_i) = \int_0^\infty p(y_{ij} \mid \theta_i, \beta)\, p(\beta)\, \mathrm{d}\beta. \tag{40}$$

where $p(y_{ij} \mid \theta_i, \beta)$ for each candidate model is given by

- Telegraph: Tele($\rho\beta, \sigma_{on}, \sigma_{off}$), with $\theta_i = \{\rho, \sigma_{on}, \sigma_{off}\}$;
- Negative Binomial: NB($r$, 1/(1 + $b\beta$)), with $\theta_i = \{r, b\}$;
- Poisson: Pois($\lambda\beta$), with $\theta_i = \{\lambda\}$.

- Evaluate the integral in Eq (40) using Gauss quadrature for computational efficiency. Minimize the negative log likelihood using the Nelder-Mead algorithm.
- Compute the Bayesian Information Criterion (BIC) for model selection based on the log-likelihood in Eq (39).
- If the NB model is selected, estimate the burst size and the burst frequency from the parameter estimates of the telegraph and negative binomial models. For the telegraph model, the burst size is $\rho/\sigma_{off}$ and the burst frequency is $\sigma_{on}$. For the NB model, the estimates are given by $\hat{\beta}_f = r$ for the burst frequency and $\hat{\beta}_s = b$ for the burst size.
- Calculate the relative errors in the burst parameter estimates given the actual values used for simulating the genomic data. Randomly select $10^3$ sets of two pairs of burst size and burst frequency estimates. In each set, rank genes by the magnitude of the burst frequency and separately by the burst size. Calculate the percentage of rankings that are flipped compared to the true ranking.

### 4.7 Preprocessing of scRNA-seq dataset

We obtained the raw UMI count matrix for mouse fibroblasts from the file ss3_n682_fibs_umiCounts.rds, available at https://github.com/sandberg-lab/lncRNAs_bursting/tree/main/data. To ensure data quality, we applied standard quality control (QC) procedures consisting of two filtering steps. (i) Cell filtering: cells were removed if fewer than 30% of genes had nonzero counts, thereby excluding cells with insufficient transcript coverage. (ii) Gene filtering: genes were removed if they were expressed (nonzero counts) in fewer than 1% of all cells, thereby excluding genes with highly sparse expression. The resulting dataset was then used for downstream statistical modeling and analysis.

### 4.8 MLE-based parameter inference

Given $n_c$ observed mRNA counts $\mathcal{D} = \{n_i\}_{i=1}^{n_c}$, the likelihood of the dataset under kinetic parameters $\varphi$ is

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

Parameter inference is performed by minimizing the negative log-likelihood,

$$J(\phi) = -\sum_{i=1}^{n_c} \log P(n_i \mid \phi). \tag{41}$$

We optimize $J(\phi)$ using the Nelder–Mead algorithm implemented in the Optim.jl package in Julia. This derivative-free approach avoids costly gradient evaluations of $\partial_\phi J(\phi)$ while maintaining robustness to initialization. Unless otherwise noted, we adopt a gradient tolerance of g_tol = $10^{-20}$ and a maximum of iterations = 1000, following the default settings of the optimize command.

To account for variability in capture probability $p_{\text{cap}}$ across cells, we define the normalized capture probability

$$\beta_i = \frac{p_{\text{cap},i}}{\langle p_{\text{cap}} \rangle} = \frac{V_i}{\sum_i V_i / n_c}, \tag{42}$$

where $V_i$ is the total transcript count of cell $i$. The distribution of $\beta$, denoted $p(\beta)$, is not known *a priori* and is approximated by KDE from the observed values $\{\beta_i\}$, yielding an estimate $\hat{p}(\beta)$.

Efficient computation of $P(n_i \mid \phi)$ requires integrating over the latent variability in $\beta$. Instead of computing this integral directly, we approximate it using Gauss–Legendre quadrature:

$$P(n_i \mid \phi) \approx \frac{\beta_{\max} - \beta_{\min}}{2} \sum_{j=1}^{N_\beta} w_j \, P(n_i \mid \beta_{x_j}, \phi) \, p(\beta_{x_j}), \tag{43}$$

where

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

Here $\beta_{\max} = \max_i \beta_i$ and $\beta_{\min} = \min_i \beta_i$, and $x_j, w_j$ are the quadrature nodes and weights generated using gausslegendre. The probability $P(n_i \mid \beta_{x_j}, \phi)$ is evaluated using the three $p_{\text{cap}}$-modified models shown in Fig 7, with $p_{\text{cap}}$ replaced by $\beta$. Under this formulation, the inferred parameters – $\rho$ for the telegraph model, $b$ for the NB model, and $\lambda$ for the Poisson model – are effectively scaled by the mean capture probability $\langle p_{\text{cap}} \rangle$, which is typically unknown.

Algorithm 1 summarizes the numerical procedure for MLE under variable capture probability.

**Algorithm 1. MLE-based parameter inference accounting for $p_{\text{cap}}$ variability.**

**Input:** Number of cells $n_c$, mRNA counts $\{n_i\}$, and total mRNA counts $V_i$ for each cell $i$
**Output:** Inferred kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\text{off}}, \sigma_{\text{on}}\}$
Compute normalized capture probabilities $\beta_i = V_i / (\sum_i V_i / n_c)$
Estimate the density $\hat{p}(\beta)$ via KDE using the samples $\{\beta_i\}$
Generate Gauss-Legendre quadrature nodes and weights $(x_j, w_j)$ using gausslegendre
Initialize parameter values $\phi$
**While** convergence criterion not met **do**
  Evaluate $P(n_i \mid \beta_{x_j}, \phi)$ for each cell $i$ and quadrature node $\beta_{x_j}$
  Compute $P(n_i \mid \phi)$ using Eq (43)
  Evaluate the loss $J(\phi)$ via Eq (41)
  Update $\phi$ using the Nelder–Mead algorithm
**End While**
**Return** $\phi = \{\bar{\rho}, \sigma_{\text{off}}, \sigma_{\text{on}}\}$

## Supporting information

**S1 Appendix. Supplemental Figures, Supplemental Tables, and References.** Supplemental figures include the relative error of aeBIC compared with the expectation of BIC for Poisson and negative binomial (NB) distributions across sample sizes and selected parameter sets (Fig A), the relative error between the minimum cross-entropy determined by MLE and

moment matching for the NB approximation under different transcript capture probabilities (Fig B), phase diagrams showing aeBIC-based model selection among technical-noise corrected telegraph/NB/Poisson models for different capture-probability distributions (Fig C), a simulated scRNA-seq workflow with technical noise and downstream comparisons of burst-frequency/burst-size estimation error and ranking consistency (Fig D), recalculated versions of the statistics in Fig D under alternative capture distributions and switching-rate regimes ($N_\sigma$ = 10 and *20*) (Fig E), an evaluation showing that Hessian eigenvalue ratios are not reliable indicators for model selection compared with aeBIC boundaries (Fig F), and an aeBIC phase diagram for telegraph/NB/Poisson model selection at large sample size ($n_c = 10^6$) (Fig G). Supplemental tables include the parameter sets used in Fig 4b and Fig A(a,b) (Table A), and a comparison of model selection by aeBIC and BIC for the simulations in Fig 4c with $n_c$ = 100 cells (Table B). References are provided at the end of the appendix.
(PDF)

## Acknowledgments

The authors are grateful to Augustinas Sukys for his valuable feedback on previous drafts of this work.

## Author contributions

**Conceptualization:** Zhixing Cao, Ramon Grima.

**Formal analysis:** Yiling Wang, Zhanpeng Shu.

**Investigation:** Yiling Wang.

**Methodology:** Zhanpeng Shu.

**Project administration:** Zhixing Cao, Ramon Grima.

**Software:** Yiling Wang, Zhanpeng Shu.

**Supervision:** Zhixing Cao, Ramon Grima.

**Visualization:** Yiling Wang, Zhanpeng Shu.

**Writing – original draft:** Zhixing Cao, Ramon Grima.

**Writing – review & editing:** Zhixing Cao, Ramon Grima.

## References

1. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N. mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods. 2009;6(5):377–82.
2. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21(7):1160–7. https://doi.org/10.1101/gr.110882.110 PMID: 21543516
3. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777–82. https://doi.org/10.1038/nbt.2282 PMID: 22820318
4. Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks G-J, Larsson AJM, et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol. 2020;38(6):708–14. https://doi.org/10.1038/s41587-020-0497-0 PMID: 32518404
5. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9(1):72–4. https://doi.org/10.1038/nmeth.1778 PMID: 22101854
6. Sukys A, Grima R. Cell-cycle dependence of bursty gene expression: insights from fitting mechanistic models to single-cell RNA-seq data. Nucleic Acids Res. 2025;53(7):gkaf295. https://doi.org/10.1093/nar/gkaf295 PMID: 40240003
7. Mangiola S, Thomas EA, Modrák M, Vehtari A, Papenfuss AT. Probabilistic outlier identification for RNA sequencing generalized linear models. NAR Genom Bioinform. 2021;3(1):lqab005. https://doi.org/10.1093/nargab/lqab005 PMID: 33709073
8. Svensson V. Droplet scRNA-seq is not zero-inflated. Nat Biotechnol. 2020;38(2):147–50. https://doi.org/10.1038/s41587-019-0379-5 PMID: 31937974
9. Vieth B, Ziegenhain C, Parekh S, Enard W, Hellmann I. powsimR: power analysis for bulk and single cell RNA-seq experiments. Bioinformatics. 2017;33(21):3486–8. https://doi.org/10.1093/bioinformatics/btx435 PMID: 29036287

10. Tang W, Bertaux F, Thomas P, Stefanelli C, Saint M, Marguerat S, et al. bayNorm: Bayesian gene expression recovery, imputation and normalization for single-cell RNA-sequencing data. Bioinformatics. 2020;36(4):1174–81. https://doi.org/10.1093/bioinformatics/btz726 PMID: 31584606
11. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods. 2014;11(6):637–40. https://doi.org/10.1038/nmeth.2930 PMID: 24747814
12. Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian analysis of single-cell sequencing data. PLoS Comput Biol. 2015;11(6):e1004333. https://doi.org/10.1371/journal.pcbi.1004333 PMID: 26107944
13. Qiu X, Hill A, Packer J, Lin D, Ma Y-A, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nat Methods. 2017;14(3):309–15. https://doi.org/10.1038/nmeth.4150 PMID: 28114287
14. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun. 2019;10(1):390. https://doi.org/10.1038/s41467-018-07931-2 PMID: 30674886
15. Andrews TS, Hemberg M. M3Drop: dropout-based feature selection for scRNASeq. Bioinformatics. 2019;35(16):2865–7. https://doi.org/10.1093/bioinformatics/bty1044 PMID: 30590489
16. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018;15(7):539–42. https://doi.org/10.1038/s41592-018-0033-z PMID: 29941873
17. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296. https://doi.org/10.1186/s13059-019-1874-1 PMID: 31870423
18. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053–8. https://doi.org/10.1038/s41592-018-0229-2 PMID: 30504886
19. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods. 2014;11(7):740–2. https://doi.org/10.1038/nmeth.2967 PMID: 24836921
20. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell. 2014;55(2):319–31. https://doi.org/10.1016/j.molcel.2014.06.029 PMID: 25038413
21. Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. Single-cell analysis of transcription kinetics across the cell cycle. Elife. 2016;5:e12175. https://doi.org/10.7554/eLife.12175 PMID: 26824388
22. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-Fos transcriptional bursts. Cell Rep. 2014;8(1):75–83. https://doi.org/10.1016/j.celrep.2014.05.053 PMID: 24981864
23. Fu X, Patel HP, Coppola S, Xu L, Cao Z, Lenstra TL, et al. Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions. Elife. 2022;11:e82493. https://doi.org/10.7554/eLife.82493 PMID: 36250630
24. Ochiai H, Sugawara T, Sakuma T, Yamamoto T. Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci Rep. 2014;4:7125. https://doi.org/10.1038/srep07125 PMID: 25410303
25. Bahar Halpern K, Tanami S, Landen S, Chapal M, Szlak L, Hutzler A, et al. Bursty gene expression in the intact mammalian liver. Mol Cell. 2015;58(1):147–56. https://doi.org/10.1016/j.molcel.2015.01.027 PMID: 25728770
26. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123(6):1025–36. https://doi.org/10.1016/j.cell.2005.09.031 PMID: 16360033
27. Tunnacliffe E, Chubb JR. What is a transcriptional burst? Trends in Genetics. 2020;36(4):288–97.
28. Zoller B, Nicolas D, Molina N, Naef F. Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol Syst Biol. 2015;11(7):823. https://doi.org/10.15252/msb.20156257 PMID: 26215071
29. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011;332(6028):472–4. https://doi.org/10.1126/science.1198817 PMID: 21415320
30. Zhou T, Zhang J. Analytical results for a multistate gene model. SIAM J Appl Math. 2012;72(3):789–818. https://doi.org/10.1137/110852887
31. Cao Z, Filatova T, Oyarzún DA, Grima R. A stochastic model of gene expression with polymerase recruitment and pause release. Biophys J. 2020;119(5):1002–14. https://doi.org/10.1016/j.bpj.2020.07.020 PMID: 32814062
32. Jiao F, Li J, Liu T, Zhu Y, Che W, Bleris L. What can we learn when fitting a simple telegraph model to a complex gene expression model?. PLOS Computational Biology. 2024;20(5):e1012118.
33. Nicoll AG, Szavits-Nossan J, Evans MR, Grima R. Transient power-law behaviour following induction distinguishes between competing models of stochastic gene expression. Nat Commun. 2025;16(1):2833. https://doi.org/10.1038/s41467-025-58127-4 PMID: 40121209
34. Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. Theoretical Population Biology. 1995;48(2):222–34. https://doi.org/10.1006/tpbi.1995.1027
35. Kim JK, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. Genome Biol. 2013;14(1):R7. https://doi.org/10.1186/gb-2013-14-1-r7 PMID: 23360624
36. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biology. 2006;4(10):e309.
37. Tang W, Jørgensen ACS, Marguerat S, Thomas P, Shahrezaei V. Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics. Bioinformatics. 2023;39(7):btad395. https://doi.org/10.1093/bioinformatics/btad395 PMID: 37354494

38. Trzaskoma P, Jung S, Pękowska A, Bohrer CH, Wang X, Naz F, et al. 3D chromatin architecture, BRD4, and Mediator have distinct roles in regulating genome-wide transcriptional bursting and gene network. Sci Adv. 2024;10(32):eadl4893. https://doi.org/10.1126/sciadv.adl4893 PMID: 39121214
39. Jiao F, Sun Q, Tang M, Yu J, Zheng B. Distribution modes and their corresponding parameter regions in stochastic gene transcription. SIAM J Appl Math. 2015;75(6):2396–420. https://doi.org/10.1137/151005567
40. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008;15(12):1263–71. https://doi.org/10.1038/nsmb.1514 PMID: 19011635
41. Wang M, Zhang J, Xu H, Golding I. Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality. Nat Microbiol. 2019;4(12):2118–27. https://doi.org/10.1038/s41564-019-0553-z PMID: 31527794
42. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. Science. 2012;336(6078):183–7. https://doi.org/10.1126/science.1216379 PMID: 22499939
43. Brouwer I, Lenstra TL. Visualizing transcription: key to understanding gene expression dynamics. Curr Opin Chem Biol. 2019;51:122–9. https://doi.org/10.1016/j.cbpa.2019.05.031 PMID: 31284216
44. Liu Z, Tjian R. Visualizing transcription factor dynamics in living cells. J Cell Biol. 2018;217(4):1181–91. https://doi.org/10.1083/jcb.201710038 PMID: 29378780
45. Bartman CR, Hsu SC, Hsiung CC-S, Raj A, Blobel GA. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. Mol Cell. 2016;62(2):237–47. https://doi.org/10.1016/j.molcel.2016.03.007 PMID: 27067601
46. Viñuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, et al. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. BMC Biol. 2013;11:15. https://doi.org/10.1186/1741-7007-11-15 PMID: 23442824
47. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer-promoter topology and gene activity. Nat Genet. 2018;50(9):1296–303. https://doi.org/10.1038/s41588-018-0175-z PMID: 30038397
48. Cho W-K, Jayanth N, English BP, Inoue T, Andrews JO, Conway W, et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife. 2016;5:e13617. https://doi.org/10.7554/eLife.13617 PMID: 27138339
49. Cao Z, Grima R. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proc Natl Acad Sci U S A. 2020;117(9):4682–92. https://doi.org/10.1073/pnas.1910888117 PMID: 32071224
50. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049. https://doi.org/10.1038/ncomms14049 PMID: 28091601
51. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;161(5):1202–14. https://doi.org/10.1016/j.cell.2015.05.002 PMID: 26000488
52. Grima R, Esmenjaud P-M. Quantifying and correcting bias in transcriptional parameter inference from single-cell data. Biophys J. 2024;123(1):4–30. https://doi.org/10.1016/j.bpj.2023.10.021 PMID: 37885177
53. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A. 2008;105(45):17256–61. https://doi.org/10.1073/pnas.0803850105 PMID: 18988743
54. Friedman N, Cai L, Xie XS. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett. 2006;97(16):168302. https://doi.org/10.1103/PhysRevLett.97.168302 PMID: 17155441
55. Jia C. Simplification of Markov chains with infinite state space and the mathematical theory of random gene expression bursts. Phys Rev E. 2017;96(3–1):032402. https://doi.org/10.1103/PhysRevE.96.032402 PMID: 29346865
56. Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, Sexton DW, et al. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. Nat Biotechnol. 2013;31(8):748–52. https://doi.org/10.1038/nbt.2642 PMID: 23873083
57. Thomas P, Popović N, Grima R. Phenotypic switching in gene regulatory networks. Proc Natl Acad Sci U S A. 2014;111(19):6994–9. https://doi.org/10.1073/pnas.1400049111 PMID: 24782538
58. Öcal K, Sanguinetti G, Grima R. Model reduction for the Chemical Master Equation: An information-theoretic approach. J Chem Phys. 2023;158(11):114113. https://doi.org/10.1063/5.0131445 PMID: 36948813
59. Jiang R, Sun T, Song D, Li JJ. Statistics or biology: the zero-inflation controversy about scRNA-seq data. Genome Biol. 2022;23(1):31. https://doi.org/10.1186/s13059-022-02601-5 PMID: 35063006
60. Jia C. Kinetic Foundation of the Zero-Inflated Negative Binomial Model for Single-Cell RNA Sequencing Data. SIAM J Appl Math. 2020;80(3):1336–55. https://doi.org/10.1137/19m1253198
61. Cao Y, Kitanovski S, Küppers R, Hoffmann D. UMI or not UMI, that is the question for scRNA-seq zero-inflation. Nat Biotechnol. 2021;39(2):158–9. https://doi.org/10.1038/s41587-020-00810-6 PMID: 33526946
62. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187–201. https://doi.org/10.1016/j.cell.2015.04.044 PMID: 26000487
63. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017;65(4):631-643.e4. https://doi.org/10.1016/j.molcel.2017.01.023 PMID: 28212749

64. Salomon R, Kaczorowski D, Valdes-Mora F, Nordon RE, Neild A, Farbehi N, et al. Droplet-based single cell RNAseq tools: a practical guide. Lab Chip. 2019;19(10):1706–27. https://doi.org/10.1039/c8lc01239c PMID: 30997473
65. 10x Genomics. What fraction of mRNA transcripts are captured per cell?. 10x Genomics Knowledge Base. 2025. [cited 2025 April 2]. https://kb.10xgenomics.com/hc/en-us/articles/360001539051-What-fraction-of-mRNA-transcripts-are-captured-per-cell
66. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;135(2):216–26. https://doi.org/10.1016/j.cell.2008.09.050 PMID: 18957198
67. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. Science. 2004;304(5678):1811–4. https://doi.org/10.1126/science.1098641 PMID: 15166317
68. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, et al. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A. 2012;109(43):17454–9. https://doi.org/10.1073/pnas.1213530109 PMID: 23064634
69. Paulsson J, Berg OG, Ehrenberg M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci U S A. 2000;97(13):7148–53. https://doi.org/10.1073/pnas.110057697 PMID: 10852944
70. Munsky B, Li G, Fox ZR, Shepherd DP, Neuert G. Distribution shapes govern the discovery of predictive models for gene regulation. Proc Natl Acad Sci U S A. 2018;115(29):7533–8. https://doi.org/10.1073/pnas.1804060115 PMID: 29959206
71. Meeussen JVW, Lenstra TL. Time will tell: comparing timescales to gain insight into transcriptional bursting. Trends Genet. 2024;40(2):160–74. https://doi.org/10.1016/j.tig.2023.11.003 PMID: 38216391
72. Johnsson P, Ziegenhain C, Hartmanis L, Hendriks G-J, Hagemann-Jensen M, Reinius B, et al. Transcriptional kinetics and molecular functions of long noncoding RNAs. Nat Genet. 2022;54(3):306–17. https://doi.org/10.1038/s41588-022-01014-1 PMID: 35241826
73. Wang Y, Szavits-Nossan J, Cao Z, Grima R. Joint Distribution of Nuclear and Cytoplasmic mRNA Levels in Stochastic Models of Gene Expression: Analytical Results and Parameter Inference. Phys Rev Lett. 2025;135(6):068401. https://doi.org/10.1103/q5sd-tpms PMID: 40864937
74. Amrhein L, Harsha K, Fuchs C. A mechanistic model for the negative binomial distribution of single-cell mRNA counts. bioRxiv. 2019. https://doi.org/10.1101/657619
75. Singh A, Bokes P. Consequences of mRNA transport on stochastic variability in protein levels. Biophys J. 2012;103(5):1087–96. https://doi.org/10.1016/j.bpj.2012.07.015 PMID: 23009859
76. Kumar N, Platini T, Kulkarni RV. Exact distributions for stochastic gene expression models with bursting and feedback. Phys Rev Lett. 2014;113(26):268105. https://doi.org/10.1103/PhysRevLett.113.268105 PMID: 25615392
77. Kumar N, Singh A, Kulkarni RV. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. PLoS Comput Biol. 2015;11(10):e1004292. https://doi.org/10.1371/journal.pcbi.1004292 PMID: 26474290
78. Jia C, Grima R. Small protein number effects in stochastic models of autoregulated bursty gene expression. J Chem Phys. 2020;152(8):084115. https://doi.org/10.1063/1.5144578 PMID: 32113345
79. Gorin G, Pachter L. Special function methods for bursty models of transcription. Phys Rev E. 2020;102(2–1):022409. https://doi.org/10.1103/PhysRevE.102.022409 PMID: 32942485
80. Jia C, Grima R. Frequency Domain Analysis of Fluctuations of mRNA and Protein Copy Numbers within a Cell Lineage: Theory and Experimental Validation. Phys Rev X. 2021;11(2). https://doi.org/10.1103/physrevx.11.021032
81. Gorin G, Pachter L. Modeling bursty transcription and splicing with the chemical master equation. Biophys J. 2022;121(6):1056–69. https://doi.org/10.1016/j.bpj.2022.02.004 PMID: 35143775
82. Wu B, Holehouse J, Grima R, Jia C. Solving the time-dependent protein distributions for autoregulated bursty gene expression using spectral decomposition. J Chem Phys. 2024;160(7):074105. https://doi.org/10.1063/5.0188455 PMID: 38364008
83. Gorin G, Yoshida S, Pachter L. Assessing Markovian and Delay Models for Single-Nucleus RNA Sequencing. Bull Math Biol. 2023;85(11):114. https://doi.org/10.1007/s11538-023-01213-9 PMID: 37828255
84. Chari T, Gorin G, Pachter L. Biophysically interpretable inference of cell types from multimodal sequencing data. Nat Comput Sci. 2024;4(9):677–89. https://doi.org/10.1038/s43588-024-00689-2 PMID: 39317762
85. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, et al. Genomic encoding of transcriptional burst kinetics. Nature. 2019;565(7738):251–4. https://doi.org/10.1038/s41586-018-0836-1 PMID: 30602787
86. Luo S, Zhang Z, Wang Z, Yang X, Chen X, Zhou T, et al. Inferring transcriptional bursting kinetics from single-cell snapshot data using a generalized telegraph model. R Soc Open Sci. 2023;10(4):221057. https://doi.org/10.1098/rsos.221057 PMID: 37035293
87. Luo S, Wang Z, Zhang Z, Zhou T, Zhang J. Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics. Nucleic Acids Res. 2023;51(1):68–83. https://doi.org/10.1093/nar/gkac1204 PMID: 36583343
88. Ramsköld D, Hendriks G-J, Larsson AJM, Mayr JV, Ziegenhain C, Hagemann-Jensen M, et al. Single-cell new RNA sequencing reveals principles of transcription at the resolution of individual bursts. Nat Cell Biol. 2024;26(10):1725–33. https://doi.org/10.1038/s41556-024-01486-9 PMID: 39198695
89. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. J Phys A: Math Theor. 2017;50(9):093001. https://doi.org/10.1088/1751-8121/aa54d9

90. Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MSH. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. 2009;16(1):45–58. https://doi.org/10.1093/dnares/dsn030 PMID: 19001483

91. Waisman A, Sevlever F, Elías Costa M, Cosentino MS, Miriuka SG, Ventura AC, et al. Cell cycle dynamics of mouse embryonic stem cells in the ground state and during transition to formative pluripotency. Sci Rep. 2019;9(1):8051. https://doi.org/10.1038/s41598-019-44537-0 PMID: 31142785

92. Riba A, Oravecz A, Durik M, Jiménez S, Alunni V, Cerciat M, et al. Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. Nat Commun. 2022;13(1):2865. https://doi.org/10.1038/s41467-022-30545-8 PMID: 35606383

93. Lederer AR, Leonardi M, Talamanca L, Bobrovskiy DM, Herrera A, Droin C, et al. Statistical inference with a manifold-constrained RNA velocity model uncovers cell cycle speed modulations. Nat Methods. 2024;21(12):2271–86. https://doi.org/10.1038/s41592-024-02471-8 PMID: 39482463

94. Fang M, Gorin G, Pachter L. Trajectory inference from single-cell genomics data with a process time model. PLoS Comput Biol. 2025;21(1):e1012752. https://doi.org/10.1371/journal.pcbi.1012752 PMID: 39836699

95. Peidli S, Green TD, Shen C, Gross T, Min J, Garda S, et al. scPerturb: harmonized single-cell perturbation data. Nat Methods. 2024;21(3):531–40. https://doi.org/10.1038/s41592-023-02144-y PMID: 38279009

96. Karlis D, Xekalaki E. Mixed poisson distributions. International Statistical Review/Revue Internationale de Statistique. 2005;73(1):35–58. https://doi.org/10.1111/j.1751-5823.2005.tb00250.x

97. Zabaikina I, Grima R. Imperfect molecular detection renormalizes apparent kinetic rates in stochastic gene regulatory networks. arXiv preprint. 2025. https://arxiv.org/abs/251202908

98. Ham L, Brackston RD, Stumpf MPH. Extrinsic Noise and Heavy-Tailed Laws in Gene Expression. Phys Rev Lett. 2020;124(10):108101. https://doi.org/10.1103/PhysRevLett.124.108101 PMID: 32216388

99. Ham L, Jackson M, Stumpf MP. Pathway dynamics can delineate the sources of transcriptional noise in gene expression. Elife. 2021;10:e69324. https://doi.org/10.7554/eLife.69324 PMID: 34636320