Transient power-law behaviour following induction distinguishes between competing models of stochastic gene expression

Andrew G. Nicoll[1], Juraj Szavits-Nossan[1], Martin R. Evans[2], Ramon Grima[1,*]
[1] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom.
[2] School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom.
* Correspondence: ramon.grima@ed.ac.uk

What features of transcription can be learnt by fitting mathematical models of gene expression to mRNA count data? Given a suite of models, fitting to data selects an optimal one, thus identifying a probable transcriptional mechanism. Whilst attractive, the utility of this methodology remains unclear. Here, we sample steady-state, single-cell mRNA count distributions from parameters in the physiological range, and show they cannot be used to confidently estimate the number of inactive gene states, i.e. the number of rate-limiting steps in transcriptional initiation. Distributions from over 99% of the parameter space generated using models with 2, 3, or 4 inactive states can be well fit by one with a single inactive state. However, we show that if the mRNA lifetime is hours long, then for many minutes following induction, the increase in the mean mRNA count obeys a power law whose exponent equals the sum of the number of states visited from the initial inactive state to the active state and the number of rate-limiting post-transcriptional processing steps. Our study shows that non-linear regression estimation of the exponent from eukaryotic data is sufficient to estimate the total number of regulatory steps in transcription initiation, splicing, and nuclear export.

I. INTRODUCTION

A broad molecular-level picture of transcription in eukaryotes is well established [1–3]. This picture involves the opening of chromatin which allows RNA polymerase II (RNAP) to access the promoter region, the sequential binding of several transcriptional factors and of RNAP to the promoter leading to the closed preinitiation complex (PIC), the unwinding of promoter DNA to create an open PIC that starts productive elongation of nascent RNA and finally the detachment of the RNAP leading to a mature RNA molecule. Despite this detailed molecular knowledge, it has thus far proved impossible to quantitatively predict the kinetics of gene expression at the single-cell level. In particular, it is unclear how molecular mechanisms can explain the large degree of heterogeneity in gene expression observed in a population of isogenic cells in a common environment [4–6].

Mathematical modelling provides a potential approach to understand how the observed statistics of gene expression emerges from the molecular description of transcription [7]. While in principle, models can be constructed that encapsulate each and every biochemical reaction step that is relevant to transcription, in practice this strategy is futile, because given the huge parameter space, its numerical exploration via simulations would require prohibitive amounts of computational power. A practical way forward is through the use of much simpler models composed of a (relatively) small set of effective reactions, where each reaction captures the overall effect of a large number of elementary reactions [8]. These reactions often correspond to the rate-limiting steps (the slowest steps of the mechanism) since these are principal determinants of a model's output.

A large variety of such minimal models of stochastic gene expression have been developed [9, 10]. Once a model is formulated, its output can be fit to experimental data, thus leading to the estimation of the rate constants [11–13]. Moreover, by assessing the relative goodness of fit between a suite of models, one can potentially arrive at a best model that captures salient features of the data [14–16]. This model selection holds great promise since in principle it can be used to sift between competing mechanisms that at first sight appear equally likely to explain a set of measurements.

The majority of the published minimal stochastic models of transcription are either special cases of, or can be reduced to, the following N-state model:

$$\begin{aligned}
&G_1 \xrightarrow{k_1} G_2 \xrightarrow{k_2} G_3 \xrightarrow{k_3} \ldots \xrightarrow{k_{N-2}} G_{\text{N-1}} \xrightarrow{k_{N-1}} G_{\text{N}} \xrightarrow{k_N} G_1, \\
&G_N \xrightarrow{\rho} G_N + M, \\
&M \xrightarrow{d} \emptyset.
\end{aligned} \tag{1}$$

Here G_1 to G_{N-1} (for $N > 1$) are inactive states from which mRNA (M) synthesis is not possible and G_N is an active state from which it is. The rate (constant) parameters governing the transition from gene state i to $i+1$, mRNA synthesis in gene state N and mRNA degradation are k_i, ρ, and d, respectively. Each of the various reactions represents an important rate-limiting step in the transcriptional process (the rates of each reaction are shown above the reaction arrows). We note that while an extension of this model to include reversible reactions between gene

states or transcription from multiple gene states is likely more realistic [17–19], nevertheless it has many more rate parameters than the irreversible model and hence the latter is typically preferred. It can be shown that this model predicts the Fano factor (variance divided by the mean) of mRNA counts to be greater than or equal to 1, which is the case for most genes; however this means that it cannot describe the sub-Poissonian expression (Fano factor less than 1) measured for a few eukaryotic and prokaryotic genes [16, 20]. The case $N = 1$, i.e. no inactive states, is that of constitutive (non-regulated) expression. Bursty expression, i.e. where transcription occurs only in short bursts, separated by intervals of inactivity implies the presence of at least one inactive state, i.e. $N \geq 2$. For example, the commonly used telegraph model, also called the 2-state model, is a special case of this model with $N = 2$ (one inactive state and one active state) [11–13, 21–24]. Many genes seemingly exhibit bursty expression [25]. For these there is evidence that generally the number of active states is just one (since the distribution of the total time spent in the active states is measured to be exponential [26, 27]) but the number of inactive states can be greater than one (since the distribution of the total time spent in the inactive states can be non-monotonic with a peak at a non-zero value [26–28]). Most experimental evidence suggests that the number of inactive states is just one for prokaryotic genes [15]. On the other hand, for eukaryotic genes, some studies make the case that it is larger than one [14, 15, 27–31] while others are consistent with one [12, 23, 26]. In particular the telegraph model, with one active and one inactive state, well fits the statistics of mRNA counts for thousands of sequenced mammalian genes [23], seemingly contradicting studies that make a case for models with more inactive states. We emphasize that because the number of inactive states can be interpreted as the number of rate-limiting steps in initiation, their determination is of biological importance since these are often the likely points of gene regulation.

Unfortunately, commonly used methods to determine the number of rate-limiting steps present several challenges. These methods can be divided into two types according to the type of data used: (i) direct or indirect measurements of the distribution of the total time spent in the inactive states [26, 27, 31–33]; (ii) measurements of the time-dependent distribution of mRNA counts following a perturbation [14, 15]. Methods of type (i) are the most reliable because if the distribution of time in the inactive state is measured to be non-exponential then the number of active states is greater than one. However, direct measurements of this type are challenging because live-cell measurements are low-throughput and require genome-editing [34, 35]. Methods of type (ii) while indirect are often the preferred choice since mRNA count data per cell can nowadays be easily obtained for several genes using single molecule fluorescence *in situ* hybridization (smFISH) [36] or single-cell sequencing (scRNA-Seq) [23]. These require measurements of the distribution of mRNA counts at several time points, each of which is calculated from an independent sub-population of cells. Inference methods using this type of data maximize the likelihood [14, 37] or utilise a Bayesian framework [15] to fit stochastic models of gene expression with varying number of inactive states to temporal count data, in the process selecting an optimal, best-fit model. It is currently unclear how far away from steady state does one need to be for these methods to perform robustly – if the shape or another property of the steady-state distribution of mRNA counts encapsulates information about the number of inactive states then the latter could be estimated. There is some evidence from comparison of the telegraph and the N-state model with $N = 3$ that it is not easy to estimate the number of gene states from steady-state count data [38, 39] but a systematic analysis, especially one for larger values of N, remains missing.

These methods also suffer from some notable disadvantages: they are not easy to use because of their technical sophistication and the quality of inference is highly dependent on how well one can correct the data for static extrinsic noise, i.e. cell-to-cell variability in the rate parameters. The latter is a particularly significant issue because static extrinsic noise tends to increase the variance of the distribution of mRNA counts (relative to its mean) and hence can lead to the incorrect estimation of rate parameters [40]; in addition, it can lead to wrong deduction of one or more inactive states when actually there are none. For example, in Ref. [41] it was shown that for 150 mammalian genes after correcting for cellular phenotypic states such as size and cell cycle stage, the variance of fluctuations was reduced to that of a Poisson distribution which is obtained from a model with no inactive states.

Summarising, we have identified two open questions: (i) is steady-state count data sufficient to estimate the number of inactive states? (ii) can one devise an estimation method that is very simple to use, does not need large volumes of temporal data and which is robust to static extrinsic noise? A schematic summarising these questions and the whole model deduction process is illustrated in Fig. 1.

In this paper we provide answers to these two questions. We consider the model (1) with one active state and a general number of inactive states, i.e. $N \geq 2$. We first conclusively show that steady-state mRNA count data is not typically sufficient to distinguish between the $N = 2$ model [21] and more complex models with $N = 3$, 4 and 5 gene states. In particular we identify a mapping of rate parameters between the telegraph model and the other types of models that leads to distributions of mRNA counts that in the vast majority of cases are practically model-independent. Next we devise a simple method to overcome the difficulties of the inference method based on steady-state data and other methods in the literature that utilise the time-dependent distribution of mRNA counts. We analytically show that following induction, the initial increase of the mean mRNA count with time is described by a power law whose exponent is solely dependent on the number of gene states between the initial inactive state

and the active state. We confirm the accuracy of the theoretical result using simulations and then apply it to data collected for various cell types to provide a robust estimate of a lower bound for the total number of inactive gene states. Interestingly, our results show that this bound: is not affected by static extrinsic noise due to cell-to-cell variability; *is* valid for a large class of networks of gene states (of which (1) is a special case); and *is* sufficient to accurately infer the number of gene states using bulk data, rather than single-cell data.

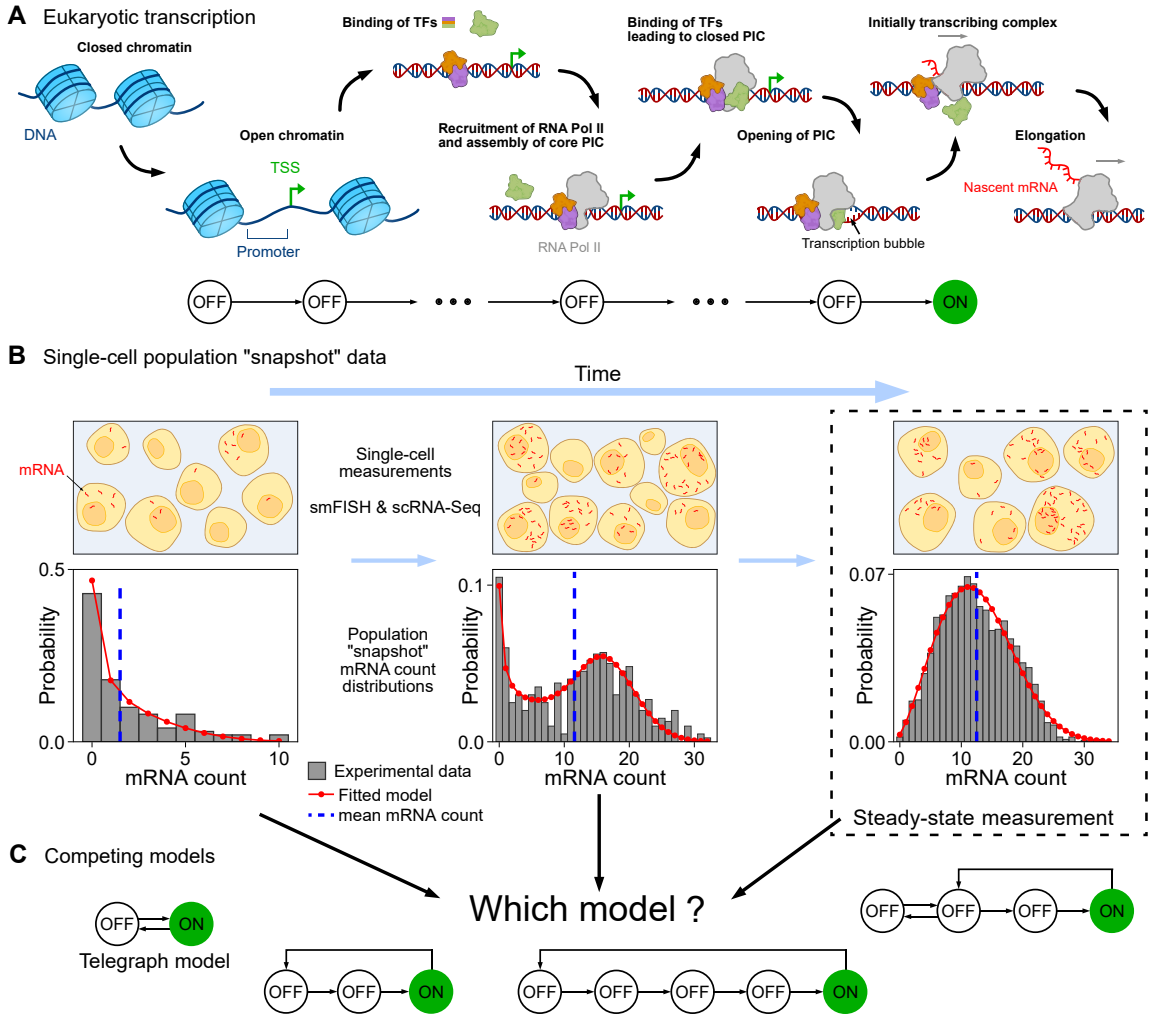

FIG. 1. Inference of the number of rate-limiting steps in transcription from mRNA count data using mathematical models of gene expression. (A) Schematic of the key stages in eukaryotic transcription. This includes the opening of closed chromatin, exposing the transcription start site (TSS); the binding of various transcription factors (TF) to the promoter; recruitment of RNAP and assembly of the PIC; and successful elongation. From a detailed biological description, it is unclear what are the actual number of rate-limiting steps. (B) Schematic of single-cell population snapshot measurements of the mRNA count in time (including at steady-state) and the corresponding mRNA count distributions. Mathematical models (red lines) can be fitted to experimental data (histograms), obtained using smFISH and scRNA-Seq, with the aim of inferring probable transcriptional mechanisms. (C) It is currently unclear if by fitting steady-state distributions only, one can identify the number of rate-limiting steps and the network topology of the true transcriptional mechanism. For example, can we tell apart the telegraph model from various 3, 4, and 5-state models? Here OFF represents a transcriptionally inactive state and ON an active state. Equally unclear is the success of inference using time-dependent distributions due to static extrinsic noise stemming from cell-to-cell variability. In this article, we address these questions and devise an efficient, simple-to-use method that utilises only the mean mRNA count (dashed blue line) to overcome the limitations of current methods.

RESULTS

Theoretical limitations of inference from steady-state mRNA count data

We first remark that there are two fundamental limitations to what can be inferred from the steady-state distribution of the number of mRNA counts of the N-state model defined by reaction scheme (1). The steady-state distribution of mRNA counts, $P(m)$, is dependent only on the values of k_i/d and ρ/d, i.e. the rate parameters normalised by the mRNA degradation rate parameter. Hence, if by means of some fitting algorithm we had to achieve a perfect match between the theoretical and an experimentally measured distribution of mRNA counts, it would not be possible to infer the absolute values of k_i and ρ unless the degradation rate d was also estimated by means of a separate experiment. A second limitation is that the steady-state distribution $P(m)$ is invariant to permutations of the set of rate constants $k_1, k_2, ..., k_{N-1}$, i.e. the rate constants of the reactions from one inactive state to another or to the active state are interchangeable. For example, this means that by matching of the steady-state mRNA distributions measured experimentally to that of an N-state model, we cannot distinguish between a model where there is fast transcription factor binding to the promoter followed by slow RNAP binding from a model where the speeds of these two processes are reversed. See Appendix A for a queuing theory based derivation of the steady-state distribution of mRNA counts of the N-state model (see also [42]) and a proof of the two aforementioned results.

A physically meaningful mapping between the rate parameters of the telegraph and N-state ($N > 2$) models

We note that the steady-state distribution of reaction scheme (1) is not invariant with respect to N and hence in principle it might be possible to distinguish between models with different total number of states. Of course, there is always the possibility that two models with different N have very similar steady-state distributions of mRNA counts for particular choices of rate parameter values, in which case they would be practically indistinguishable. In particular since the telegraph model is the commonest model in the literature, it is of interest to understand whether it can accurately match the steady-state mRNA distributions of gene expression models with more than 2 states, i.e. with multiple inactive states.

The natural way to investigate this question would be to (i) generate a large number of rate parameters sets of the N-state model; (ii) for each of these sets, use an optimization algorithm to find rate parameters of the telegraph model which minimize the distance between its steady-state mRNA distributions and that of a given N-state model. If this procedure identifies rate parameters for which the minimal distribution distance is very small then it means that the telegraph and the N-state model are essentially indistinguishable. While desirable, this procedure is computationally expensive and would be difficult to employ to explore large swathes of parameter space.

To avoid these difficulties we devise a different method. We define the distribution $f_N(\tau)$ of the waiting time τ between two successive transcription events of the N-state model, its mean denoted $\langle\tau\rangle_N$ and variance $\sigma^2_{\tau,N}$. The procedure then consists of the following steps: (i) generate thousands of rate parameters sets of the N-state model with $N = 3, 4, 5$; (ii) for each parameter set, calculate the rate parameters of an effective telegraph model by matching its waiting time statistics to that of the N-state model. In particular we enforce $f_N(0) = f_2(0)$, $\langle\tau\rangle_N = \langle\tau\rangle_2$ and $\sigma^2_{\tau,N} = \sigma^2_{\tau,2}$. This means that the rate parameters of the effective telegraph model (denoted by a star superscript) are given in terms of the rate parameters of the N-state model by:

$$\rho^* = \rho, \quad k_1^* = \frac{1}{\langle\tau_{1\to N}\rangle} \cdot \frac{2}{\mathrm{CV}^2_{1\to N} + 1}, \quad k_2^* = k_N \cdot \frac{2}{\mathrm{CV}^2_{1\to N} + 1}, \tag{2}$$

where $\tau_{1\to N}$ is the random time it takes to go from state 1 to state N in the N-state model and $\mathrm{CV}_{1\to N}$ is its coefficient of variation (ratio of the standard deviation to the mean). For a derivation of Eq. (2) see Appendix A and for their interpretation see the Discussion section. Finally, step (iii), using the rate parameters calculated in the previous step, calculate the Wasserstein distance (WD) between the cumulative mRNA count distribution of the effective telegraph model (P^c) and that of the N-state model (Q^c) defined as $\mathrm{WD} = \sum_i |P^c_i - Q^c_i|$ where i is the number of mRNA counts. If the WD is small over the whole scanned parameter space then we know that generally there exists an effective telegraph model that captures, to a good degree, the salient features of the mRNA count distribution of the N-state model. We note that our choice of the WD as a measure of similarity is motivated by the fact that it has a simple interpretation (it considers how much and how far probability mass has to be moved to reshape a histogram into another [43]) and that it does not suffer from well-known issues of other common metrics [44].

The theoretical justification of step (ii) is as follows. Using queuing theory, in Appendix A we show that the steady-state mRNA count distribution of the N-state model is completely determined by the Laplace transform of $f_N(\tau)$, which we shall denote by $f^*_N(s)$, where s is the variable in Laplace space. Now $f_N(0) = \lim_{s\to\infty}[s f^*_N(s)]$,

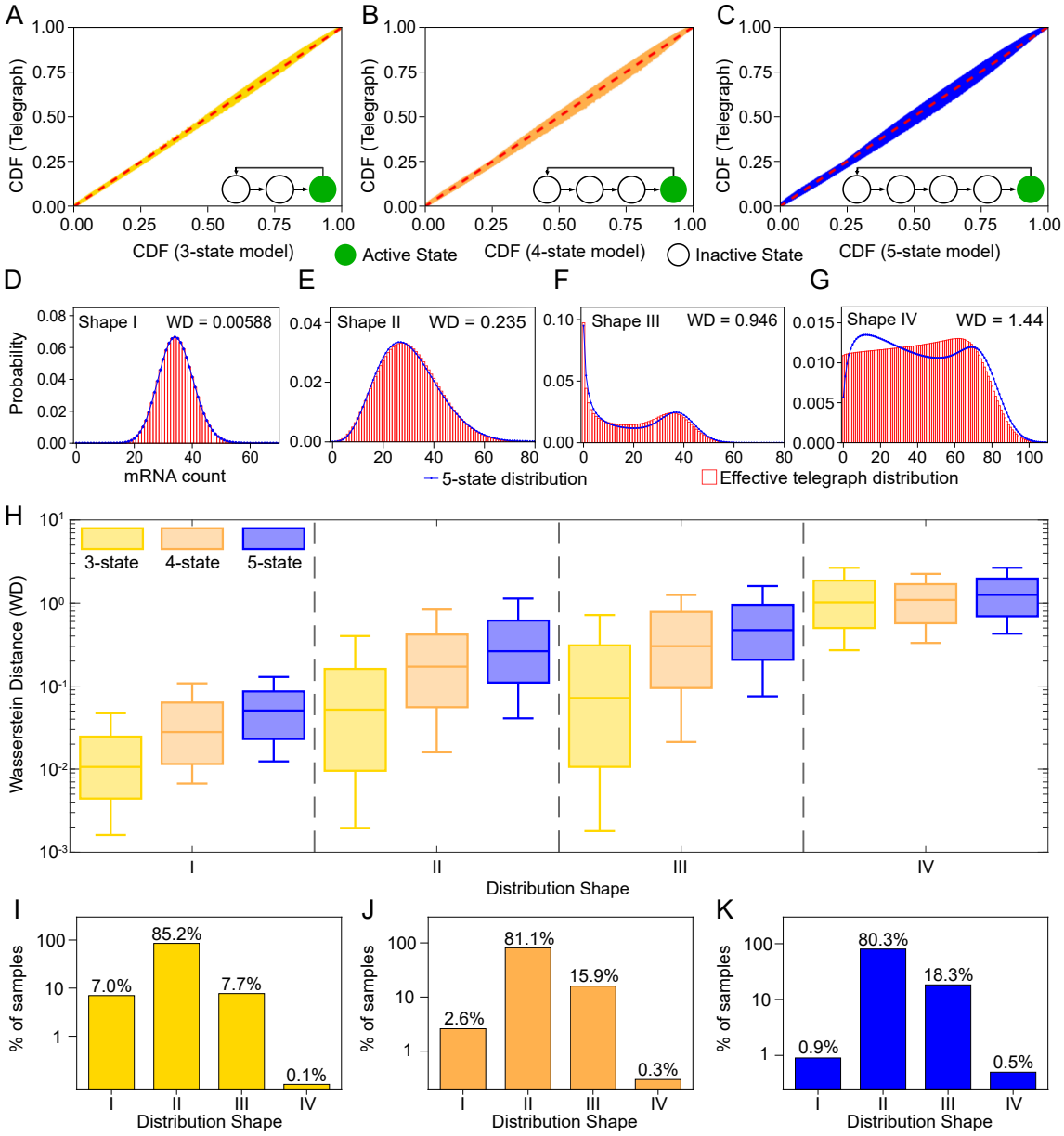

FIG. 2. The steady-state mRNA count distribution of the N-state model can often be well fit by an effective telegraph model. Thousands of rate parameter sets of the N-state model ($N = 3, 4, 5$) were sampled in the range relevant for eukaryotic gene expression. For each parameter set we use the mapping giving by Eq. (2) to obtain the rate parameters of an effective telegraph model. (A-C) P-P plots comparing the cumulative distribution function of the N-state and effective telegraph models. These are constructed from the 500 rate parameter sets with the largest Wasserstein distance (WD) values between the distributions of the two models. Deviations from the line $y = x$ (red) indicate differences between the distributions. (D-G) Typical steady-state mRNA count distributions of the 5-state (solid blue lines) and effective telegraph models (histograms) for four classes of distributions (Shape I: Unimodal with Fano factor < 2; Shape II: Unimodal with Fano factor ≥ 2; Shape III: Bimodal with one peak at zero and another at a non-zero value; Shape IV: Bimodal with peaks at non-zero values. (H) Boxplots displaying summary statistics of the WD between the steady-state distribution of mRNA counts of the N-state and effective telegraph model. The horizontal line and the box show the median and the interquartile range, respectively; whiskers extend to the 10th and 90th percentiles. (I-K) Bar charts displaying the percentage of each distribution shape sampled across the parameter space. The agreement between the N-state and its effective telegraph model becomes worse as N increases and is worst for Shape IV because this shape is qualitatively impossible to capture by the telegraph model. For discussion see main text and for implementation details see Appendix E.

which implies that the condition $f_N(0) = f_2(0)$ forces $f_N^*(s)$ and $f_2^*(s)$ to agree in the limit of large s. In addition, by matching the first two moments of the waiting distribution, $\langle\tau\rangle_N = \langle\tau\rangle_2$ and $\sigma^2_{\tau,N} = \sigma^2_{\tau,2}$, we effectively are matching $f_N^*(s)$ and $f_2^*(s)$ for small s. This is because by the definition of the Laplace transform the rth moment of $f_N(\tau)$ is determined by the rth derivative of $f_N^*(s)$ with respect to s, hence the series expansion (and Pade approximation) of $f_N^*(s)$ and $f_2^*(s)$ must agree for small s. Thus by step (ii), we are guaranteed that $f_N^*(s)$ and $f_2^*(s)$ agree for both small and large s which in turn implies that the steady-state mRNA count distribution of the N-state and the telegraph models are also matched to a good degree. Further justifications for step (ii) are that it is very quick to compute since it requires only the evaluation of analytical formulae and that the procedure guarantees that the rate constants of the effective telegraph model are positive and hence physically meaningful (see Appendix A for a proof). Whilst this is not the only matching procedure that will accomplish this, common alternative matching procedures cannot guarantee this; for example, matching of the first three moments of the waiting time distributions or the first three moments of the mRNA count distributions of the N-state and telegraph models can lead to negative rate parameter values.

The number of inactive gene states of the N-state model is difficult to estimate from steady-state data

In Fig. 2 the steady-state mRNA count distributions of thousands of rate parameter sets for the $N = 3, 4, 5$ state models are compared with those of their effective telegraph models (computed using the three-step procedure detailed in the previous section). Note that to ensure biological relevance, the rate parameter ranges for each model were chosen to comfortably cover the bounds of the telegraph model rate parameters estimated using a range of eukaryotic data in various published studies (see Table 1 of [45] for these estimates and see Appendix E for further details on generating the rate parameter sets).

Fig. 2A-C shows the probability-probability (P-P) plots comparing cumulative distribution functions of the effective telegraph and the N-state models for the 500 rate parameter sets with the largest WD of those sampled. Distributions which are very similar fall very close to the dashed red line, which represent identical distributions, whereas greater deviations from this line represent distributions that significantly disagree. For each of the $N = 3, 4, 5$ state models we see minor deviations from this line, indicating broad agreement between the compared distributions. As the number of states increases, we see greater, but not significantly so, deviations as the bulge around this line broadens. Overall these P-P plots show that even for the most statistically distinct distributions, the difference between these distributions is not very large. This suggests that in many cases the telegraph model provides an excellent approximation to the N-state models with $N > 2$.

To obtain a deeper insight into the deviations, we categorised the $N = 3, 4, 5$ state rate parameter sets according to the shape and statistical properties of their mRNA count distributions. All sampled distribution shapes fell into the following four groups: shape I - unimodal with Fano factor < 2; shape II - unimodal with Fano factor ≥ 2; shape III - bimodal with one zero and one non-zero mode; shape IV - bimodal with two non-zero modes. Note that the single mode of both shapes I and II can be either at zero or at a non-zero value. A typical distribution of each shape is seen in Fig. 2D-G, along with the WD between the 5-state and the effective telegraph distribution. A comparison of the WD of these four shapes suggests that the quality of the effective telegraph model fits decrease with increasing shape number, being best for shape I and worst for shape IV.

In Fig. 2H we show by means of box plots that this is indeed the picture that emerges when considering all rate parameter sets sampled: comparing box plots of the same color, i.e. same number of states, we see that shape I distributions are associated with the lowest median WD, followed by shapes II and III distributions which have similar median WDs that are approximately 5-10 times larger than those of shape I, followed further by shape IV distributions which have median WDs that are 3-20 times larger than of shapes II and III. These trends become less clear when considering not just the medians but also the whiskers of the box plots which shows that there is still considerable variation of the WD within each shape type (Fig. S1 and Appendix F). In particular for the 3-state model (the commonest N-state model with $N > 2$ in the literature) the lower whisker of the box plots extends nearly to 10^{-3} for shapes I-III; this suggests that a substantial portion of the parameter space of this model is very well approximated by the effective telegraph model, since WDs of this magnitude are essentially indistinguishable by eye (Fig. 2D). Within each of shapes I, II and III, the median WD increases with the number of states of the N-state model (compare the three differently colored box plots for each category). Hence, as expected, the telegraph model fit worsens as the number of gene states increases, though the worst fits (Shape III for the $N = 5$ model), when visually inspected, still manage to capture the main features of the N-state distribution (the two peaks and their positions; Fig. 2F). Note that while in principle this suggests that as we increase N, the effective telegraph model will eventually become glaringly distinguishable from the N-state model, in reality this is unlikely because until now, the maximum number of states estimated from experimental data is 6 and this was an isolated case [29].

Therefore, for mRNA count distributions of shapes I-III, our results strongly suggest that the actual number of

gene states cannot be reliably inferred from steady-state mRNA count distributions. The similarity of the N-state and effective telegraph model means that the maximum likelihood of each distribution computed using mRNA count data from each cell would also be similar and hence using common model selection criteria such as the Akaike Information Criterion or the Bayesian Information Criterion (which penalise more strongly models with a larger number of parameters), the telegraph model would be selected as the optimal one. Note that our comparison of distributions was done assuming an infinite number of cells; noise in the shape of the distribution introduced by the finite sample sizes of experimental data will make it even more difficult to distinguish models than suggested by our study.

We now discuss in detail the comparison of distributions of the N-state model of shape IV and those of the effective telegraph model. This shape presents an interesting case as it is the only shape out of the four considered for which the telegraph model cannot even qualitatively capture the shape of the distribution of the N-state model. As we see for the example in Fig. 2G, the number and position of the modes differs between the 5-state model (blue) and the effective telegraph model (red) - while the former has two modes both at non-zero values, the effective telegraph model has a single, non-zero mode. Now it has been proved that when the steady-state distribution of the telegraph model is bimodal then one of its peaks must lie at zero [46]. Hence clearly for shapes of type IV, it is impossible for the telegraph model to provide an adequate approximation of the models with $N > 2$. It is also interesting to consider that while slow promoter switching compared to the mRNA degradation rate is a necessary condition for the telegraph model ($k_1/d < 1,\ k_2/d < 1$) to display bimodal distributions [47], this is not the case for the emergence of bimodal distributions of shape IV in the models with $N > 2$; in particular we found this distribution shape can be obtained when $k_N/d < 1$ and $k_i/d \geq 1$ for $i = 1, \ldots, N-1$. Hence we conclude that shape IV distributions are associated with an intermediate switching regime.

Of course, the relevance of our results depend on the actual frequency of the four distribution shapes in the biologically relevant parameter space for eukaryotic transcription. In Fig. 2I-K we see how many of each distribution shape were sampled in the parameter sweep for each $N = 3, 4, 5$ state model. Despite the aforementioned distinction of distributions of shape IV, we actually see that for each $N = 3, 4, 5$ the percentage of shape IV distributions sampled was incredibly low, at 0.1%, 0.3%, and 0.5% respectively. Therefore, even though shape IV distributions are distinguishable from their effective telegraph model distributions in steady-state, since they appear so infrequently within the relevant parameter space, we do not expect them to appear commonly from experimental data. Hence our overall conclusion is that the number of gene states is difficult, if not impossible, to reliably estimate from steady-state mRNA count distributions.

The number of gene states can be estimated from short-time mean count data following induction

Now that we have established the difficulty of telling apart the telegraph model from the N-state model with $N = 3, 4, 5$ using steady-state distributions, we look at how this can instead be achieved by considering time-dependent statistics. Inference of the gene state number N from distributions of the mRNA counts measured at several points in time has been previously shown using the method of maximum likelihood [14] and also using a Bayesian approach [15]. In what follows we will describe a much simpler method which requires only the application of regression on the mean mRNA count computed for a short interval of time following gene induction.

Consider an experimental setup whereby some gene of interest is inactive before induction by an inducer molecule. From application of the inducer to its binding to DNA, there is some time lag due to nuclear translocation and other processes. Hence we shall consider $t = 0$ to be the point at which induction starts in the nucleus. We seek to understand what information about the number of gene states N can be deduced from the the moments of the mRNA counts measured a short time after induction.

Specifically consider the case where gene expression is described by reaction scheme (1), the gene is initially in the inactive state j where $j \in 1, .., N-1$ and the number of mRNA is zero. Note that generally j is unknown. In Appendix B starting from the chemical master equation (CME) for the N-state model, we use perturbation theory to show that in the limit of short times, the mean $\langle m(t) \rangle$ of mRNA counts at time t is given by:

$$\langle m(t) \rangle = \rho \frac{\prod_{i=j}^{N-1} k_i}{(N-j+1)!} t^{N-j+1} + O(t^{N-j+2}). \tag{3}$$

Note that in this equation both time and the rate parameters are non-dimensional since they are rescaled by the mRNA degradation rate parameter, i.e. $t \mapsto td$, $k_i \mapsto k_i/d$ and $\rho \mapsto \rho/d$.

The expression for the mean suggests a simple method to extract information about the number of inactive states. For short times after induction, a log-log plot of the mean mRNA count versus time leads to a straight line with slope equal to $N - j + 1$. Since $j = 1, .., N-1$, the slope provides a lower bound on the total number of gene states N.

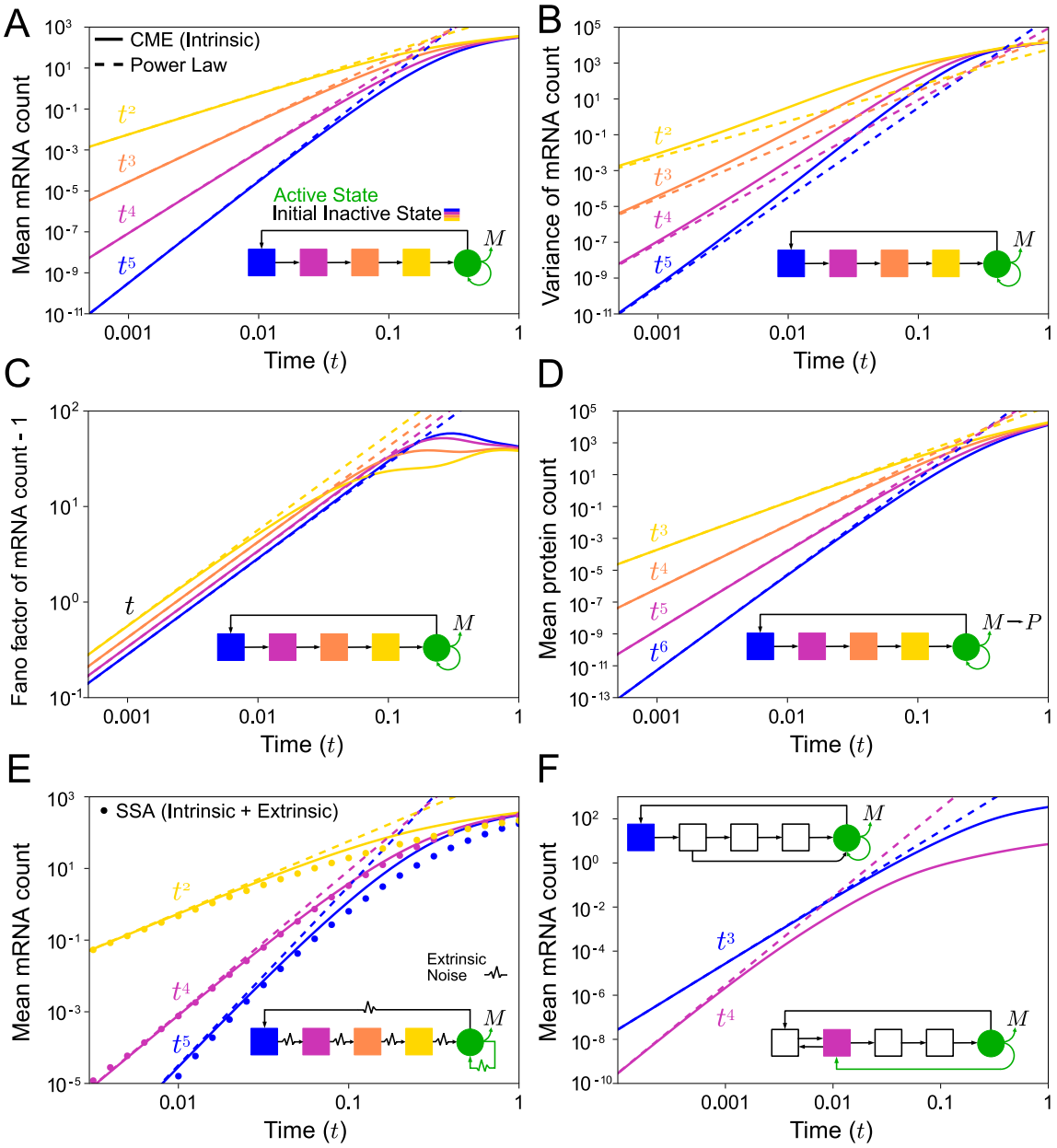

FIG. 3. Simulations confirm the theoretical predictions for the time-dependence of the statistics of mRNA counts following gene induction. (A)-(C) Mean of mRNA counts, variance of mRNA counts and Fano factor of mRNA counts for the 5-state model, i.e. the system (1) with $N = 5$. (D) Mean of mRNA counts for the 5-state model with one extra step modeling the translation of mRNA (M) to protein (P). (E) Mean of mRNA counts for the 5-state model with cell-to-cell variation of rate parameters (static extrinsic noise); each rate parameter is drawn from a gamma distribution. (F) Mean mRNA count from extensions of the 5-state model that incorporate reversible reactions between gene states and state change upon mRNA synthesis; these are special cases of the system (5). The theory is shown by the dashed lines and simulations by solid lines (obtained by direct numerical integration of the moment equations of the CME). In (E) the simulations are done in two ways: solid lines are obtained by direct numerical integration of the moment equations of the CME with rate parameters fixed to the average of the gamma distributions and dots are from the SSA with rate parameters randomly drawn from these distributions. In all cases there is excellent agreement between theory and simulations for short times. All panels use a log-log scale hence a power-law relationship is represented by a straight line with gradient corresponding to the exponent of the power law. The colours: blue, pink, orange, and yellow, represent the initial inactive gene state after induction: $j = 1, 2, 3, 4$ respectively and are illustrated by the cartoon model diagram in each panel. The outward pointing green arrow represents mRNA synthesis and the other one shows the state after synthesis has occurred. For discussion see main text and for simulation details see Appendix E.

Specifically the slope estimates the "distance" between the initial inactive state j and the active state N. This result also means that it is difficult to distinguish N-state models for which $j = N - 1$ from the 2-state telegraph model because in both cases the measured power law exponent is 2. In Fig. 3A we confirm the derived power law by direct numerical integration of the time-evolution equations for the first moment of mRNA numbers of a 5-state model of gene expression – since the system (1) is composed of purely first-order reactions, these equations whilst derived from the CME, in this case they are none other than the deterministic rate equations.

As expected, the power law behavior is valid only for short times (less than 0.1 in non-dimensional time units) and hence one may question its usefulness in practice. In Appendix B we derive a condition for how short the observation time should be so that the t^{N-j+1} scaling law for the mean is valid. As a rule of thumb, the power law for the mean can be considered to be valid over time intervals that are much smaller than the mRNA degradation timescale ($t = 1$ in non-dimensional time units). For mammalian cells, the median half-life of mRNA molecules is many hours, e.g. it is about 9 hours for NIH3T3 mouse fibroblasts cells [48], and hence the power law should in this case be observable for many minutes. For prokarytic cells and simple eukaryotes, the half-lives of mRNA are much less (typically ranging from a few minutes to about an hour and a half for *E. coli* and budding yeast [49, 50]) and hence the power law may be more difficult to observe for these organisms. Note that if the initial mean mRNA (m_0) is not zero then there would be an additional term, $m_0 e^{-t}$, present in Eq. (3). Because of the length of the mRNA lifetime in mammalian cells, this term can be considered constant for short times and hence $\langle m(t)\rangle - m_0$ will typically display the power law on the right hand side of Eq. (3).

In Appendix C we extend the analytical results to the variance σ_m^2 and Fano factor FF (variance divided by the mean) of mRNA counts:

$$\sigma_m^2(t) = \rho \frac{\prod_{i=j}^{N-1} k_i}{(N-j+1)!} t^{N-j+1} + O(t^{N-j+2}),\ FF_m(t) = \frac{\sigma_m^2(t)}{\langle m(t)\rangle} = 1 + \frac{2\rho}{N-j+2} t + O(t^2). \tag{4}$$

Note that the short-time law for the variance is precisely the same as that for the mean. In Figs. 3B and C we confirm the accuracy of these expressions for short times by comparison with the results obtained by direct numerical integration of the time-evolution equations for the first and second moment of mRNA numbers obtained from the CME. This confirms that a log-log plot of the variance of mRNA counts versus time does not provide more information than the mean does and that from the exponent of the Fano factor minus 1 scaling law it is not possible to obtain any information on the number of states. By comparing Figs. 3A and B, we also note that the true time-evolution of the variance of mRNA counts deviates quicker from the short-time power law than that for the mean of mRNA counts – in Appendix C we provide a general theoretical proof of this property. In fact this is a consequence of a general property of any Markovian system, i.e. the approach to the steady-state form of the variance occurs asymptotically at twice the rate of approach of the mean [51]. The implication of this is that practically speaking, the method for extracting the number of states from short-time data is likely only useful when the mean counts are used.

We note that our results also point to a simple method by which some of the rate parameters of the N-state model can be estimated. In particular, from simultaneous short-time measurements of the mean and the Fano factor it is possible to estimate the transcription rate ρ and the product of rate constants $\prod_{i=j}^{N-1} k_i$. The procedure is to first estimate $N - j + 1$ from the slope of a log-log plot of the mean versus time (or from non-linear regression fit of a power law to the data). Using this value and the slope of the FF versus time graph, we obtain ρ. Finally given the latter and the earlier estimate of $N - j + 1$, the product of rate constants is obtained from the intercept of the log-log plot of the mean versus time.

The short time results are valid for more general models that include post-transcriptional mRNA processing, cell-to-cell parameter variability and complex networks of gene state interactions

Post-transcriptional processing of mRNA may affect the power law exponent. It is simple to show that if splicing and nuclear export can each effectively be described by a single first-order reaction then the short-time exponent deduced from unspliced nuclear mRNA is one less and two less than the exponents deduced from spliced nuclear mRNA data and cytoplasmic mRNA, respectively. Similarly if we interpret the mRNA species (M) in the N-state model as the cytoplasmic mRNA, a first-order processing step to describe its translation to protein would imply that the short-time power law exponent for the latter is one more than that for the cytoplasmic mRNA (Appendix B). By comparison of Fig. 3A and D, it is clear that these results are verified by simulations (direct numerical integration of the time-evolution equations for the means of the mRNA and of the species following from it due to processing). On the other hand, if splicing and export can be described by fixed-delay reactions then there is no difference in the exponents deduced from different types of mRNA. In practice the effect of mRNA post-processing on the exponent might not be highly relevant since often only the total mRNA counts are measured. But if count data is available for

mRNA at different stages of its lifecycle then a comparison of the exponents of the short-time power laws for each stage can reveal information about the effective reactions describing processing.

Realistically, the rate parameters of the N-state model will generally vary from one cell to another one, a phenomenon called static extrinsic noise [40, 52]. This means that the mean mRNA measured in an experiment is in reality described by $\langle m(t)\rangle_e = \int \langle m(t)\rangle P(k_1,..,k_n,\rho)dk_1..dk_N d\rho$ where $P(k_1,..,k_n,\rho)$ is the joint distribution of all rate parameters in the N-state model. Given the solution Eq. (3) it immediately follows that while static extrinsic noise changes the prefactor in front of the power law it does not change the exponent and hence it follows that the lower bound of N estimated from a log-log plot of the mean versus time is insensitive to static extrinsic noise. Note that while in principle j (the integer labelling the initial inactive state) can be variable between cells, in practice this is unlikely because the inducer used for gene induction is associated with a particular stage of the transcriptional initiation process, hence implying a fixed value of j for all cells in the population. In Fig. 3E we confirm using stochastic simulations the robustness of the short-time power law to static extrinsic noise. Here we use the stochastic simulation algorithm (SSA) to simulate mRNA count data in 10^6 cells, assuming that the time-independent rate parameters for each cell were drawn from independent gamma distributions. This distribution is useful because it is defined on the positive real line, has a simple functional form, and its mean and variance can be tuned independently of each other. Note that while static extrinsic noise has no impact on the short-time dynamics, it does change the long-time behaviour (compare points obtained using the SSA with static extrinsic noise vs solid lines obtained by numerical integration of the time-evolution equations for the mean mRNA counts, with rate parameter values equal to the mean of the gamma distributions used for the SSA).

Thus far we have derived results for short-time behaviour assuming the N-state model. Whilst fairly general, one could argue that this model can be extended further to account for complex initiation mechanisms. Thus we consider a model of the type:

$$G_i \underset{k_{j,i}}{\overset{k_{i,j}}{\rightleftharpoons}} G_j,\ i,j = 1,\ldots,N, \quad G_N \xrightarrow{\rho} G_K + M, \quad M \xrightarrow{d} \emptyset. \tag{5}$$

Here we have N gene states of which only one is active (state N). Reversible reactions are allowed between any of pair of states and when transcription occurs the state changes from N to K where the latter can be any integer between 1 and N. This state change models reinitiation where some of the essential transcription factors (but not all) remain bound to the promoter [19]; this can also model promoter proximal-pausing [53]. The N-state model (1) is a special case of this model where all gene state reactions are irreversible and no state changes are allowed upon synthesis of a transcript. In Appendix D we show using queuing theory that for short times after induction, the model also predicts a mean mRNA count that increases with time as a power law. In this case, the exponent is equal to the minimal number of states that are visited between the initial inactive state and the state from which the synthesis of mRNA occurs – hence independent of the connectivity of the reaction pathway describing initiation, the exponent is a lower bound on the number of gene states N. In Fig. 3F we show that the theoretical results are in good agreement with numerical integration of the time-evolution equations for the mean mRNA counts for two special cases of reaction scheme (5), one where there is an extra reaction (compared to the N-state model) between states 2 and 5 and another one with reversible reactions between states 1 and 2 and a gene state change from 5 to 2 upon transcription.

Synthetic data generated using biologically relevant parameter values validates the method's practical use for eukaryotic cells, even for time regimes where the power law breaks down

To test more directly the practical utility of the theory, next we mimic computation of the mean mRNA count from experimental data, by using realistic conditions for mammalian cells. Since the median number of mRNA molecules in mammalian cells in steady-state growth is about 17 [48], we first randomly generated thousands of rate parameter sets for which the mean mRNA count of the N-state model was at least greater than 10 molecules when $t = 0.1$ (for more details on the selection procedure see Appendix E). Note that one time unit equals the inverse of the mRNA degradation rate whose median is about 9 hours for mouse fibroblasts [48] and hence $t = 0.1$ equates to doing measurements of the mRNA count over about an hour, following induction. This is achievable as has been demonstrated by various studies using eukaryotic cells [14, 54, 55]. We then collected data at five time points ($t = 0.02, 0.04, 0.06, 0.08, 0.1$). For each of the rate parameter sets, we integrated the time-evolution equations for the mean mRNA counts of the N-state model with an initial condition of zero mRNA and starting from inactive state j to calculate the mean mRNA at each of the five time points. Note that by the use of these equations, here we made the implicit assumption that the number of cells is infinite (in practice this means the number of cells is larger than a few thousand such that the standard error of the mean is very small). Finally we applied linear regression to estimate the slope of a log-log plot of mean vs time which estimates the exponent of the power law. In all cases, the Pearson correlation coefficient of the best linear fit through the data points was larger than 0.99 thus firmly establishing the

existence of a power law – however as clear from the example shown in Fig. 4A, because of the length of the time interval used, the fitted power law was different than the theoretical one and hence it is of interest to understand the relationship between the two.

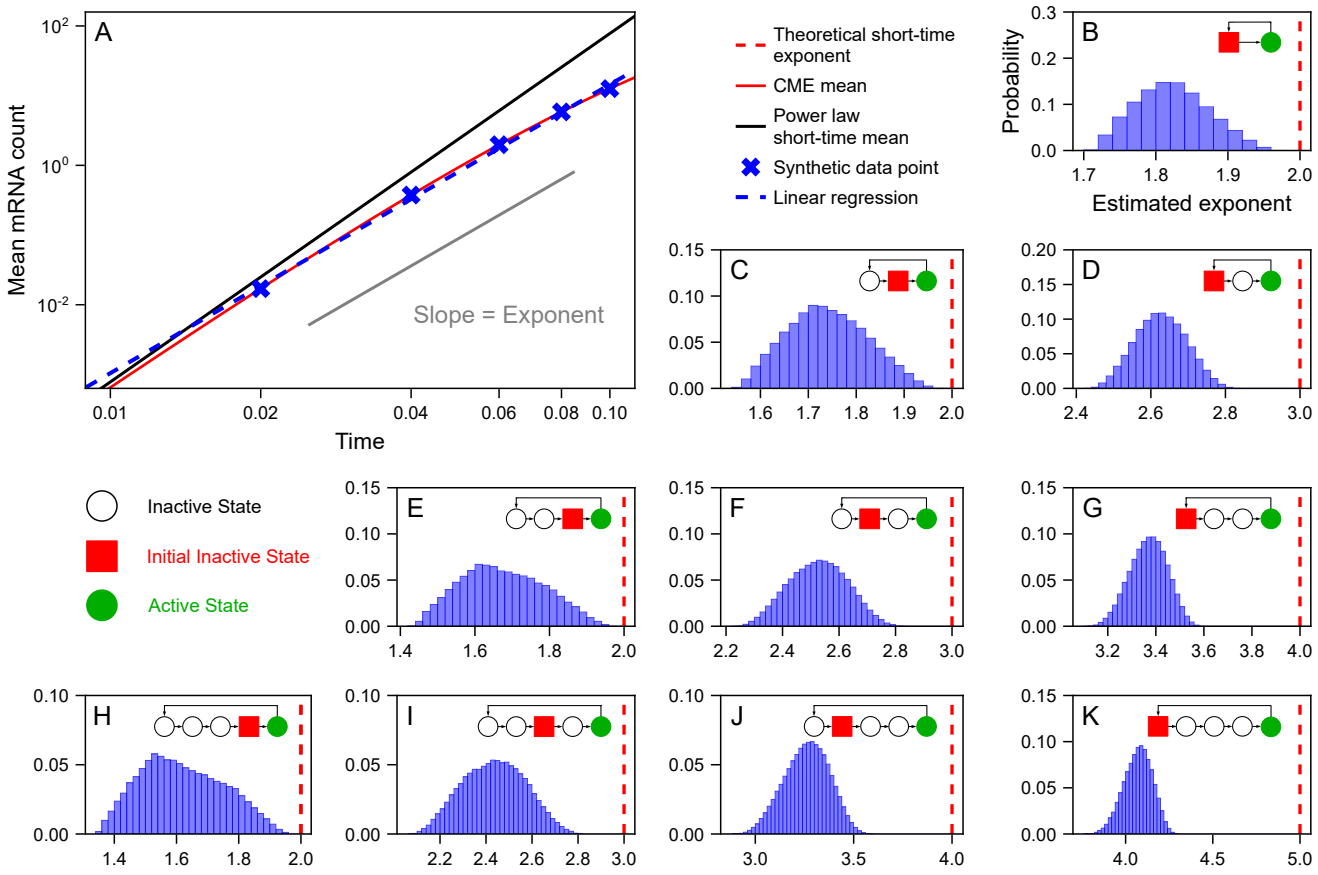

FIG. 4. Synthetic data generated using biologically relevant rate parameter values for mammalian cells validates the practical use of the short-time mean mRNA count data to estimate a lower bound for the number of gene states N. By numerical integration of the time-evolution equations for the mean mRNA of the N-state model, we obtained curves of the mean mRNA versus time for 5 timepoints $t = 0.02, 0.04, 0.06, 0.08, 0.1$. Note that time is non-dimensional since it is multiplied by the mRNA degradation rate: hence $t = 0.1$ corresponds to about one hour for mammalian cells. These curves were generated for a large number of parameter sets, in each of which the mean mRNA count reached at least 10 by $t = 0.1$ (this is about 1/2 of the median mRNA count in steady-state growth conditions for mammalian cells [48]). (A) An example curve for one rate parameter set showing deviations of the true power law (dashed blue) from the theoretical power law (solid black) for biologically relevant times. (B)-(K) For each parameter set, we used linear regression to extract the true power law exponent from the slope of the log-log plot of the mean versus time and then plotted the distributions of the exponent (blue histograms). The model diagram in each figure illustrates the active state (solid green circle), inactive states (open circles) and initial inactive state (solid red square) for each system. Note that N (the total number of gene states) varies between 2 and 5 while the initial inactive state j varies between 1 and $N-1$. The exponent $N-j+1$ estimated by our theory (Eq. (3)) is shown by a vertical red dashed line – in all cases the maximum value of the numerical exponent is below the theoretical value which itself is a lower bound for the number of gene states N. For discussion see main text and for simulation details see Appendix E.

In Fig. 4B-K we show histograms of the exponent (computed over all rate parameter sets) for N-state models with $N = 2-5$ and $j = 1, .., N-1$. The distributions of exponent values are wide but the maximum value of the exponents is always less than the theoretical exponent $N-j+1$ (shown by a vertical red dashed line). This is because the next-to-leading order term in the series expansion of the mean in powers of time has a negative sign (Appendix B) and hence the leading-order term proportional to t^{N-j+1} always overestimates the true mean mRNA count. Similar histograms of the exponent values are found if we compute the exponents from data collected over a longer period of time – 15 data points collected over the time interval $(0.02, 0.3)$, with the same constraint that the mean mRNA count is at least greater than 10 molecules when $t = 0.1$. Hence we have shown using biologically realistic rate parameter constraints that independent of N, j and the time interval over which the data is collected, the slope calculated from

log-log plots of the mean versus time is a reliable means to estimate the lower bound of the number of gene states N.

Confirmation of the power law in eukaryotic data and its use to estimate the number of rate-limiting steps for initiation, splicing and export

Next we sought to show that the predicted power law indeed can be seen in experimental data. We considered two different independent experimental data sets from Refs. [54] and [55] which we next briefly describe.

The first data set consists of simultaneous measurements of nuclear and cytoplasmic mRNA following the induction of two genes, CTT1 and STL1, involved in the osmotic stress response of *Saccharomyces cerevisiae* [54]. The experiment used 0.2mol and 0.4mol NaCl to induce this response, which we refer to as experiments 1 and 2, respectively. smFISH was used to quantify the number of mRNA molecules for each gene, in the nucleus and the cytoplasm of each imaged cell, at several time points over 60 minutes. This led to a total of eight data sets (2 experiments $\times$ 2 genes $\times$ 2 RNA species).

The second data set consists of simultaneous measurements of (unspliced) pre-mRNA and (spliced) mRNA following tumor necrosis factor (TNF) induced expression of several inflammatory genes in mouse fibroblasts and bone marrow-derived macrophages (BMDM). Using quantitative PCR (qPCR) to detect a distinction between pre-mRNA and mRNA, the induction curves for several inflammatory genes were measured (see Fig. 2 and Fig. S2 of [55]), of which we chose six induction curves (for the genes Tnfaip3, Icam1, Cxcl1, and IL-1b) whose initial RNA count was approximately 0 and for which the mRNA curves lagged behind the pre-mRNA curves by several minutes.

For these two datasets we used non-linear regression to fit the power law, $\langle m(t)\rangle = c_1 t^{c_2}$, to the various mean mRNA count curves following induction. The estimated exponents (based on at least 5 data points) can be seen with the corresponding data and regression curves in Fig. 5A-D for the yeast dataset and in Fig. 5E-H for the mammalian dataset. The exponents together with their standard errors are also summarised in Table I. The excellent fits (Pearson correlation coefficient > 0.99) confirm the theoretical prediction that the initial transients for all genes following induction are described by power laws and verify that these manifest over a sufficiently long period of time that they can be measured. Next we extract information from the exponents of the power laws to make various deductions of biological relevance.

First we consider the yeast dataset in Fig. 5A-D. For the gene CTT1, in both experiments 1 and 2, the estimated exponents from the mean nuclear mRNA count are very similar, 2.60 and 2.77; for the gene STL1, the estimated exponents from the same type of data are 3.89 and 4.54. These values are consistent with an N-state model where the minimum number of gene states is $N = 3$ (two inactive states) for CTT1 and $N = 4$ or 5 (three or four inactive states) for STL1. The difference between the nuclear and cytoplasmic exponents for CTT1 was computed to be 1.53 and 1.29 from the two experiments, and for STL1 it was 1.87 and 0.857. Accounting for the standard error in these estimates gives a difference of at least one across both experiments (Table I). This is consistent with at least one first-order reaction between the synthesis of nuclear RNA and export to the cytoplasm, but could indicate there are more, potentially two first-order rate-limiting steps. We note that a delayed export reaction is not suggested because in that case there would be no difference between the nuclear and cytoplasmic exponents.

We note that in Ref. [14] a multi-state stochastic model with multiple active gene states (connected by reversible reactions) was successfully fitted to data very similar to the yeast data used in our analysis, except that the mRNA was not separated into nuclear and cytoplasmic components. By means of a maximum likelihood estimation procedure with cross-validation, they found that two- and three-state models are too simple, while a five-state model structure was prone to over-fitting; hence they settled on an optimal model with four states. In a later study [37] the model (with four states) was refined by performing inference using the measured joint distribution of nuclear and cytoplasmic mRNA counts – the values for the inferred rate parameters can be found in Tables S3 and S4 in Ref. [37]. An inspection of the inferred transcription rates from each gene state for both CTT1 and STL1 shows that the rate from one gene state is approximately 5-1000 times larger than the rate from the three other gene states. Hence the four-state model with four active states is well approximated by a simpler model in which there are three inactive states and one active state. This is broadly consistent with our estimation from power law fitting of at least two inactive states for CTT1 and three inactive states for STL1.

Next we consider the mammalian dataset in Fig. 5E-H. Considering the power laws fitted to unspliced mean count data (and their standard errors), we deduce that the minimum number of inactive states is 3 for the genes Tnfaip3 and IL-1b ($N = 4$ model) and at least one for Cxcl1 ($N = 2$ model). The results for the latter gene reinforce the idea that not all eukaryotic genes need to be described by a model more complex than the telegraph model. Comparing the exponents from the induction curves of Cxcl1 in macrophages we get 2.04 and 3.53 for the pre-mRNA and mRNA curves respectively, with a difference of 1.49 and a standard error of 0.41. This therefore suggests there is at least one first-order reaction during the splicing process of pre-mRNA to mRNA. Similarly comparing the exponents from the pre-mRNA and mRNA induction curves of IL-1b in macrophages (4.33 and 5.70), the difference between the two

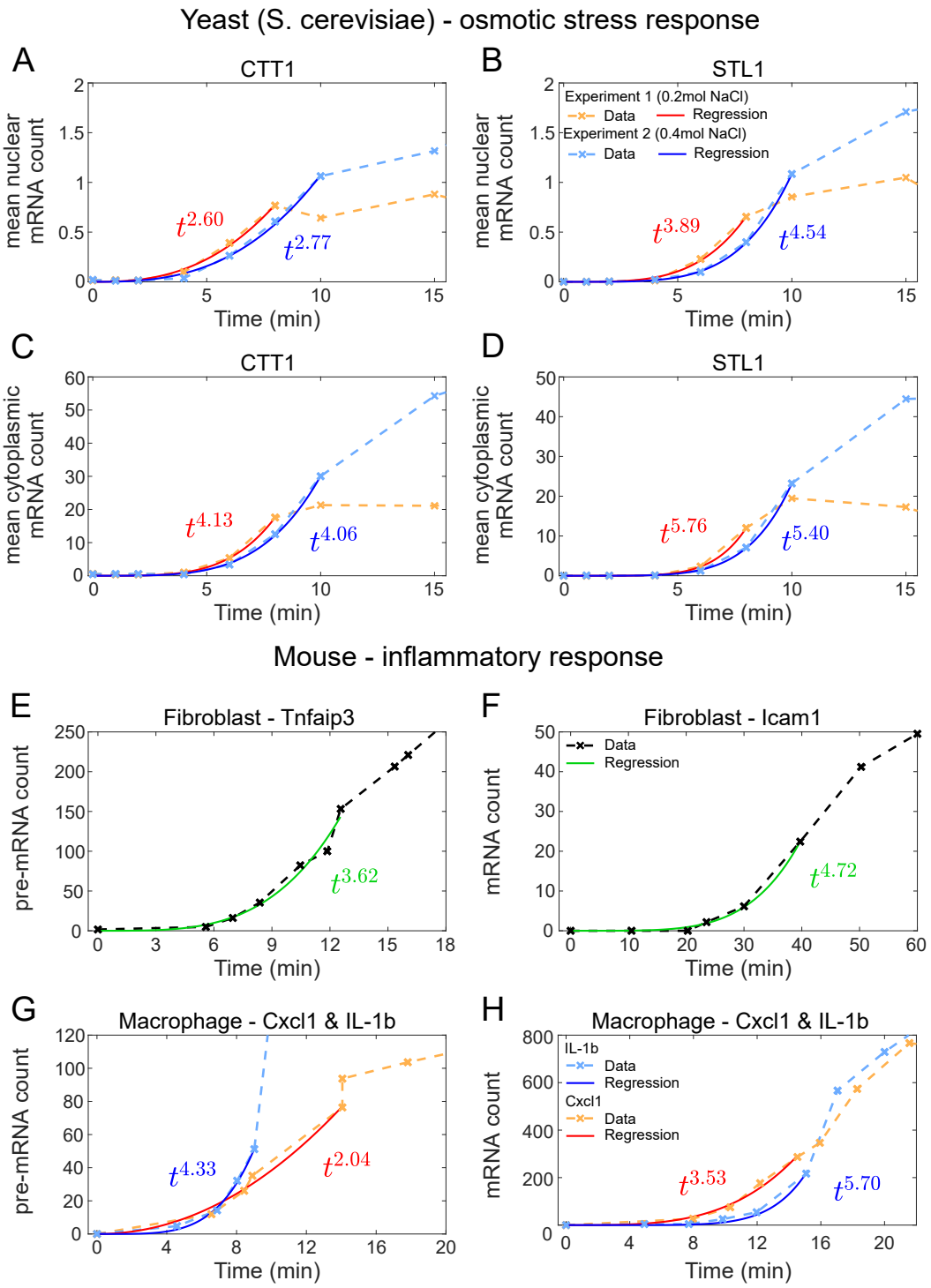

FIG. 5. Non-linear regression fitting of RNA induction curves confirms the theoretically predicted short-time power law behaviour in eukaryotic cells. The estimated exponent is indicated next to each curve. (A-D) Induction curves of CTT1 and STL1 genes following osmotic stress response in *S. cerevisiae*. (A-B) Mean nuclear mRNA counts of CTT1 and STL1, respectively. (C-D) Mean cytoplasmic mRNA counts of CTT1 and STL1, respectively. Experiment 1 (0.2mol NaCl) data and regression curves are shown in orange and red, respectively and for experiment 2 (0.4mol NaCl) are shown in light and dark blue, respectively. (E-H) Induction curves of Tnfaip3, Icam1, Cxcl1, and IL-1b genes following TNF-induced inflammatory response in mouse fibroblasts and macrophages. (E-F) Pre-mRNA induction curve of Tnfaip3 and mRNA induction curve of Icam1 in mouse fibroblasts, respectively. Data is shown in black and regression curve in green. (G-H) Pre-mRNA and mRNA induction curves of Cxcl1 and IL-1b in mouse macrophages, respectively. Cxcl1 data and regression curves are shown in orange and red, respectively and for IL-1b are shown in light and dark blue, respectively. Further details on the experimental data and regression analysis are given in Appendix E.

(1.37 with a standard error of 0.41) also supports a one step process with first-order kinetics for splicing of these transcripts. As for nuclear export, the difference in the exponents of the power laws fitted to spliced and unspliced data excludes the possibility of a step with fixed time delay to describe splicing.

Experiment	Gene	RNA species	Estimated exponent ± std error	Exponent difference ± std error
1 (0.2mol NaCl)	CTT1	Nuclear	2.60 ± 0.170	1.53 ± 0.290
		Cytoplasmic	4.13 ± 0.0235	
	STL1	Nuclear	3.89 ± 0.256	1.87 ± 0.284
		Cytoplasmic	5.76 ± 0.123	
2 (0.4mol NaCl)	CTT1	Nuclear	2.77 ± 0.151	1.29 ± 0.215
		Cytoplasmic	4.06 ± 0.154	
	STL1	Nuclear	4.54 ± 0.0475	0.857 ± 0.0868
		Cytoplasmic	5.40 ± 0.0726	

Cell type	Gene	RNA species	Estimated exponent ± std error	Exponent difference ± std error
Fibroblast	Tnfaip3	pre-mRNA	3.62 ± 0.48	-
	Icam1	mRNA	4.72 ± 0.25	-
Macrophage	Cxcl1	pre-mRNA	2.04 ± 0.19	1.49 ± 0.41
		mRNA	3.53 ± 0.36	
	IL-1b	pre-mRNA	4.33 ± 0.32	1.37 ± 0.41
		mRNA	5.70 ± 0.26	

TABLE I. Exponents and their standard error estimated by non-linear regression fitting of power laws to the short-time mean mRNA count curves for yeast and mammalian genes following induction. Top panel: Induction of CTT1 and STL1 genes in *S. cerevisiae* during osmotic stress response following addition of two different concentrations of NaCl (see Fig. 5A-D). Both the mean nuclear and cytoplasmic mRNA was measured and used in the analysis. Bottom Panel: Induction of Tnfaip3 and Icam1 genes in mouse fibroblasts, and Cxcl1 and IL-1b genes in mouse macrophages, after the addition of TNF which triggers an inflammatory gene response (see Fig. 5E-H). Both (unspliced) pre-mRNA and (spliced) mRNA were measured and used in the analysis. In the last column, we show the difference between the exponents (and the associated standard error) estimated from nuclear and cytoplasmic data for *S. cerevisiae* genes, and from pre-mRNA and mRNA for mammalian genes.

DISCUSSION

In this paper we have shown that (i) in most cases, steady-state mRNA count data is not sufficient to estimate the number of inactive states in transcription; (ii) when the steady-state mRNA distribution of the telegraph model distribution well fits the distribution of the N-state model ($N > 2$), there exists a simple relationship between the rate parameters of the two models; (iii) for a short time after induction, the increase in the mean mRNA counts follows a universal power law whose exponent provides a lower bound on the number of gene states – this is valid for a large family of models of gene expression, and it is independent of the values of the rate parameters of the gene expression model and the size of static extrinsic noise. We confirm this law by analyzing published yeast and mammalian data. In agreement with estimation using other methods [14, 27], our results suggest that the number of eukaryotic inactive gene states can be larger than one, however they also show that the common practice of using spliced mRNA counts (nuclear, cytoplasmic or total mature mRNA counts) for inference can lead to a significant overestimation of this number.

We note that while (i) was previously suspected, until now there has been no conclusive evidence for it. The success of the telegraph model to fit mRNA count distribution measured for a large number of genes [11–13, 22–24] could be due to two distinct reasons. Either, in reality, the number of rate-limiting steps for most genes is very small, i.e. one for gene inactivation and one for gene activation or else the number of rate-limiting steps is larger but the telegraph model is the simplest of the N-state models that can describe the shape of the steady-state distribution of mRNA counts that are commonly measured. In support of the latter reasoning, Ref. [38] showed that the maximum Hellinger distance (HD) computed between the steady-state count distribution of the 3-state model and an effective telegraph model (which spent the same mean time in the active and inactive states as the 3-state model) was quite small (less

than 0.06; note that the HD can vary between 0 and 1). In further support of this apparent flexibility of the telegraph model, in Ref. [56] a mapping was elucidated between the rate parameters of this model and those of more complex models with the same number of gene states but involving feedback loops which led to good agreement between the time-dependent distributions of both models. As well, more recently it was found that the telegraph model well fits the steady-state distributions of the same model but with static extrinsic noise [40, 57]. Our results add further evidence suggesting a significant difficulty in distinguishing between the telegraph model and more complex models. We found that for 3 of the 4 possible types of distribution shapes of the N-state models with $N > 2$, i.e. unimodal with Fano factor < 2 (shape I), unimodal with Fano factor ≥ 2 (shape II) and bimodal distributions with zero and non-zero peaks (shape III), the telegraph model provides a good fit. One may hypothesize that the underlying theoretical reasons for this are that perhaps for shapes I-III, the distributions of the N-state model look like simple distributions which are limiting cases of the telegraph model, e.g. the Poisson distribution with Fano factor $= 1$ (zero inactivation rate), the negative binomial distribution with Fano factor > 1 (the inactivation rate is much larger than the activation rate) and the sum of a Dirac delta function and a Poisson distribution with a non-zero mode (small activation and inactivation rates, combined with large enough transcription rate). While these distributions may explain the agreement between the N-state and the telegraph model in some cases, generally this explanation is not sufficient because we have shown that within each shape type there is quite a variation in distribution type (Fig. S1 and Appendix F). What is perhaps more intriguing is that we found the existence of a distribution shape of the N-state models with $N > 2$ (shape IV) which to our knowledge has not been described before and which cannot be fit by any rate parameter choice of the telegraph model – this distribution has two peaks, both of them at a non-zero value. We note that while an extension of the telegraph model to include transcription from both the active and inactive states (the so-called leaky telegraph model [58]) would predict a qualitatively similar distribution shape (in the limit of slow switching between active and inactive states), there still remains an interpretation issue. This is because the notion of a leaky off-state implies low expression and hence a peak which is very close to zero – but as is clear from Fig. 2G, both peaks can be far away from zero – hence in this case both the telegraph model and its leaky variant can be safely excluded.

We also elucidated a novel relationship between the rate parameters of the 2-state and N-state models with $N > 2$ when the steady-state mRNA count distributions of the two models are similar. In particular Eq. (2) provides an interesting insight: the inverse of the activation rate of the telegraph model underestimates the total time spent in the inactive states $1, .., N-1$ of the N-state model, i.e. $\langle \tau_{1\to N} \rangle$. Similarly, the inverse of the inactivation rate of the telegraph model underestimates the time spent in the active state N of the N-state model, i.e. $1/k_N$. The underestimation is because of the factor $2/(\mathrm{CV}^2_{1\to N}+1)$ which is bounded between 1 and 2 because $\mathrm{CV}^2_{1\to N} \geq 1$ and $\lim_{N\to\infty} \mathrm{CV}_{1\to N} = 0$. This also implies that the relative error in the estimates of the time spent in the inactive and active states is at most 50%. We note that the factor $2/(\mathrm{CV}^2_{1\to N}+1)$ naturally emerges from requiring the variance of the interarrival time between successive mRNA synthesis events of the 2 and N-state models to be matched, which as we argued earlier is a necessary condition to ensure matching of the distribution of mRNA counts. Our results hence suggest that fitting the telegraph model to steady-state mRNA count data from a gene with a larger number of states leads to a systematic bias in the estimated speed of promoter switching. We note that an alternative way of thinking about the N-state model is as a simple birth-death model with fixed degradation rate but an mRNA synthesis rate that randomly alternates in time between 0 and a positive constant ρ, i.e. a type of dynamic extrinsic noise [59] whereby a rate parameter varies randomly not only from cell to cell but also in time for the same cell. Hence from this perspective, our results point to a systematic overestimation of certain rate parameters in the presence of dynamic extrinsic noise, which is interesting considering that static extrinsic noise has also recently been shown to lead to systematic over- or underestimation of rate parameters inferred from steady-state count data [40]. Overall these studies imply that generally reliable rate parameter estimation from steady-state count data is difficult unless extrinsic noise sources are identified and properly taken into account.

Clearly, the best way to overcome the limitations of steady-state data is to use temporal data. Previously, a few methods have utilised the information contained in the distributions of mRNA counts measured at various time points [14, 15, 37] to simultaneously estimate the rate parameters and the number of gene states. Motivated by the fact that this data is not commonly available and the inference methods used are rather intricate, we here developed a much simpler method which utilises only the mean mRNA count measured for a short time after gene induction. In regard to whether dynamical low-order moment data is sufficient to distinguish between models with varying number of states, we are not aware of any study which has studied this question before; an exhaustive study of moment-based and distribution-based inference approaches compared their accuracy to estimate rate parameters at a fixed number of gene states [37]. The main advantages of our method are the ease with which the input data (the mean mRNA counts) can be measured (qPCR, smFISH, single-molecule and bulk RNA sequencing) and with which the number of gene states is estimated (requires only non-linear regression applied to a plot of the mean mRNA counts vs time). The method can only estimate a lower bound for the number of gene states, not the true number of states – this limitation is common to all inference methods since it is only possible to identify the simplest N-state model that is consistent with the data. Perhaps the main limitation of the inference method is that a sufficient number of measurements need

to be done in the short period of time following the start of gene expression after induction – this period is of the order of minutes if the mRNA half life is many minutes or hours and hence the method is seemingly only practical for eukaryotic genes. Assessing the period of time over which it is optimal to fit the power law can be challenging when the number of measurements is small; in this case the exponent deduced can be fairly sensitive to the addition or removal of a data point and hence there is less confidence in the results. For the results presented in this paper, at least 5 data points were used to fit power laws with high correlation coefficient (over 0.97) to cytoplasmic, nuclear and unspliced mRNA count data for two yeast and four mammalian genes – this provided clear evidence of the predicted power law behaviour. The exponents from nuclear and cytoplasmic mean count data suggest three to four gene states (two to three inactive states) for two yeast genes (in agreement with inference by a far more complex inference method [14]), and one or two rate-limiting steps for nuclear export (in agreement with observations using single molecule microscopy [60, 61]). The exponents from spliced and unspliced mRNA count data for two mammalian genes suggest two to four gene states (one to three inactive states) and one rate-limiting step for splicing. We found no evidence that nuclear export or splicing can be described by an effective reaction step with a fixed time delay – rather these seem to be described by one rate-limiting step with first-order kinetics. These results are broadly in agreement with a recent model-based analysis of single-cell and single-nucleus RNA sequencing data [62]. Importantly, our analysis shows that failing to account for post-transcriptional processing will lead to an overestimation of the number of gene states.

Concluding, we have shown that important details of transcription such as the number of rate-limiting steps in initiation have signatures in the mean mRNA data collected a short time after gene induction. The method's simplicity, particularly its avoidance of the need for the distribution of mRNA counts and of sophisticated inference techniques, shows that simple mathematical tools can also be used to gain insight from transcriptomic data.

DATA AVAILABILITY

Data is available online at https://github.com/agnflame/PowerLaw.

ACKNOWLEDGMENTS

This work was supported by a BBSRC EASTBIO PhD studentship to AGN and a Leverhulme Trust research award to RG. (RPG-2020-327).

[1] S. Sainsbury, C. Bernecky, and P. Cramer, Structural basis of transcription initiation by rna polymerase ii, Nature reviews Molecular cell biology **16**, 129 (2015).
[2] R. G. Roeder, 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms, Nature structural & molecular biology **26**, 783 (2019).
[3] I. Jonkers and J. T. Lis, Getting up to speed with transcription elongation by rna polymerase ii, Nature reviews Molecular cell biology **16**, 167 (2015).
[4] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, Proceedings of the National Academy of Sciences **99**, 12795 (2002).
[5] D. Nicolas, N. E. Phillips, and F. Naef, What shapes eukaryotic transcriptional bursting?, Molecular BioSystems **13**, 1280 (2017).
[6] E. Tunnacliffe and J. R. Chubb, What is a transcriptional burst?, Trends in Genetics **36**, 288 (2020).
[7] R. E. Kingston and M. R. Green, Modeling eukaryotic transcriptional activation, Current Biology **4**, 325 (1994).
[8] B. P. Ingalls, *Mathematical modeling in systems biology: an introduction* (MIT press, 2013).
[9] J. Paulsson, Models of stochastic gene expression, Physics of life reviews **2**, 157 (2005).
[10] B. Munsky, G. Neuert, and A. Van Oudenaarden, Using gene expression noise to understand gene regulation, Science **336**, 183 (2012).
[11] K. B. Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz, Bursty gene expression in the intact mammalian liver, Molecular cell **58**, 147 (2015).
[12] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, Single-cell analysis of transcription kinetics across the cell cycle, Elife **5**, e12175 (2016).
[13] X. Fu, H. P. Patel, S. Coppola, L. Xu, Z. Cao, T. L. Lenstra, and R. Grima, Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mrna distributions, Elife **11**, e82493 (2022).
[14] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. Van Oudenaarden, Systematic identification of signal-activated stochastic gene regulation, Science **339**, 584 (2013).

[15] Z. Kilic, M. Schweiger, C. Moyer, D. Shepherd, and S. Pressé, Gene expression model inference from snapshot rna data using bayesian non-parametrics, Nature Computational Science **3**, 174 (2023).

[16] D. E. Weidemann, J. Holehouse, A. Singh, R. Grima, and S. Hauf, The minimal intrinsic stochasticity of constitutively expressed eukaryotic genes is sub-poissonian, Science Advances **9**, eadh5138 (2023).

[17] J. Zhang, L. Chen, and T. Zhou, Analytical distribution and tunability of noise in a model of promoter progress, Biophysical Journal **102**, 1247 (2012).

[18] C. Jia and Y. Li, Analytical time-dependent distributions for gene expression models with complex promoter switching mechanisms, SIAM Journal on Applied Mathematics **83**, 1572 (2023).

[19] J. Szavits-Nossan and R. Grima, Steady-state distributions of nascent rna for general initiation mechanisms, Physical Review Research **5**, 013064 (2023).

[20] A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro, Dynamics of transcription driven by the teta promoter, one event at a time, in live escherichia coli cells, Nucleic acids research **40**, 8472 (2012).

[21] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis, Theoretical population biology **48**, 222 (1995).

[22] D. Zenklusen, D. R. Larson, and R. H. Singer, Single-rna counting reveals alternative modes of gene expression in yeast, Nature structural & molecular biology **15**, 1263 (2008).

[23] A. J. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Genomic encoding of transcriptional burst kinetics, Nature **565**, 251 (2019).

[24] H. Ochiai, T. Sugawara, T. Sakuma, and T. Yamamoto, Stochastic promoter activation affects nanog expression variability in mouse embryonic stem cells, Scientific reports **4**, 7125 (2014).

[25] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Transcriptional burst frequency and burst size are equally modulated across the human genome, Proceedings of the National Academy of Sciences **109**, 17454 (2012).

[26] B. T. Donovan, A. Huynh, D. A. Ball, H. P. Patel, M. G. Poirier, D. R. Larson, M. L. Ferguson, and T. L. Lenstra, Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting, The EMBO journal **38**, e100809 (2019).

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

[28] C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. Davis, *et al.*, Dynamic analysis of stochastic transcription cycles, PLoS biology **9**, e1000607 (2011).

[29] B. Zoller, D. Nicolas, N. Molina, and F. Naef, Structure of silent transcription intervals and noise characteristics of mammalian genes, Molecular systems biology **11**, 823 (2015).

[30] C. R. Bartman, N. Hamagami, C. A. Keller, B. Giardine, R. C. Hardison, G. A. Blobel, and A. Raj, Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation, Molecular cell **73**, 519 (2019).

[31] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, *et al.*, A single-molecule view of transcription reveals convoys of rna polymerases and multi-scale bursting, Nature communications **7**, 12248 (2016).

[32] K. Tantale, E. Garcia-Oliver, M.-C. Robert, A. L'hostis, Y. Yang, N. Tsanov, R. Topno, T. Gostan, A. Kozulic-Pirher, M. Basu-Shrivastava, *et al.*, Stochastic pausing at latent hiv-1 promoters generates transcriptional bursting, Nature communications **12**, 4503 (2021).

[33] M. Douaihy, R. Topno, M. Lagha, E. Bertrand, and O. Radulescu, Burstdeconv: a signal deconvolution method to uncover mechanisms of transcriptional bursting in live cells, Nucleic Acids Research **51**, e88 (2023).

[34] I. Brouwer, H. P. Patel, J. V. W. Meeussen, W. Pomp, and T. L. Lenstra, Single-molecule fluorescence imaging in living saccharomyces cerevisiae cells, STAR protocols **1**, 100142 (2020).

[35] T. L. Lenstra and D. R. Larson, Single-molecule mrna detection in live yeast, Current protocols in molecular biology **113**, 14 (2016).

[36] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mrna synthesis in mammalian cells, PLoS biology **4**, e309 (2006).

[37] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation, Proceedings of the National Academy of Sciences **115**, 7533 (2018).

[38] Z. Cao, T. Filatova, D. A. Oyarzún, and R. Grima, A stochastic model of gene expression with polymerase recruitment and pause release, Biophysical Journal **119**, 1002 (2020).

[39] F. Jiao, J. Li, T. Liu, Y. Zhu, W. Che, L. Bleris, and C. Jia, What can we learn when fitting a complex gene expression model to a simple telegraph model?, bioRxiv , 2023.03.09.532005 (2023).

[40] R. Grima and P.-M. Esmenjaud, Quantifying and correcting bias in transcriptional parameter inference from single-cell data, Biophysical Journal 10.1016/j.bpj.2023.10.021 (2023).

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

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

[43] C. Villani *et al.*, *Optimal transport: old and new*, Vol. 338 (Springer, 2009).

[44] K. Öcal, R. Grima, and G. Sanguinetti, Parameter estimation for biochemical reaction networks using wasserstein distances, Journal of Physics A: Mathematical and Theoretical **53**, 034002 (2019).

[45] Z. Cao and R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, Proceedings of the National Academy of Sciences **117**, 4682 (2020).
[46] F. Jiao, Q. Sun, M. Tang, J. Yu, and B. Zheng, Distribution modes and their corresponding parameter regions in stochastic gene transcription, SIAM Journal on Applied Mathematics **75**, 2396 (2015).
[47] P. Thomas, N. Popović, and R. Grima, Phenotypic switching in gene regulatory networks, Proceedings of the National Academy of Sciences **111**, 6994 (2014).
[48] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, Global quantification of mammalian gene expression control, Nature **473**, 337 (2011).
[49] Y. Wang, C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown, Precision and functional specificity in mrna decay, Proceedings of the National Academy of Sciences **99**, 5860 (2002).
[50] D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow, Global rna half-life analysis in escherichia coli reveals positional patterns of transcript degradation, Genome research **13**, 216 (2003).
[51] J. Keizer, On the solutions and the steady states of a master equation, Journal of Statistical Physics **6**, 67 (1972).
[52] L. Ham, M. Jackson, and M. P. Stumpf, Pathway dynamics can delineate the sources of transcriptional noise in gene expression, Elife **10**, e69324 (2021).
[53] S. Braichenko, J. Holehouse, and R. Grima, Distinguishing between models of mammalian gene expression: telegraph-like models versus mechanistic models, Journal of the Royal Society Interface **18**, 20210510 (2021).
[54] G. Li and G. Neuert, Multiplex rna single molecule fish of inducible mrnas in single yeast cells, Scientific data **6**, 94 (2019).
[55] S. Hao and D. Baltimore, Rna splicing regulates the temporal order of tnf-induced gene expression, Proceedings of the National Academy of Sciences **110**, 11934 (2013).
[56] Z. Cao and R. Grima, Linear mapping approximation of gene regulatory networks with stochastic dynamics, Nature communications **9**, 3305 (2018).
[57] P. Thomas and V. Shahrezaei, Coordination of gene expression noise with cell size: analytical results for agent-based models of growing cell populations, Journal of the Royal Society Interface **18**, 20210274 (2021).
[58] L. Ham, R. D. Brackston, and M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression, Physical review letters **124**, 108101 (2020).
[59] G. Gorin, J. J. Vastola, M. Fang, and L. Pachter, Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments, Nature Communications **13**, 7620 (2022).
[60] J. P. Siebrasse, T. Kaminski, and U. Kubitscheck, Nuclear export of single native mrna molecules observed by light sheet fluorescence microscopy, Proceedings of the National Academy of Sciences **109**, 9426 (2012).
[61] M. Oeffinger and D. Zenklusen, To the pore and through the pore: a story of mrna export kinetics, Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1819**, 494 (2012).
[62] G. Gorin, S. Yoshida, and L. Pachter, Assessing markovian and delay models for single-nucleus rna sequencing, Bulletin of Mathematical Biology **85**, 114 (2023).
[63] J. Szavits-Nossan and R. Grima, Charting the landscape of stochastic gene expression models using queueing theory, arXiv , 2307.03253 (2023).
[64] L. Takács, On a coincidence problem concerning telephone traffic, Acta Mathematica Academiae Scientiarum Hungarica **9**, 45 (1958).
[65] C. Commault and S. Mocanu, Phase-type distributions and representations: Some results and open problems for system theory, International Journal of Control **76**, 566 (2003).
[66] C. Commault and J.-P. Chemla, An invariant of representations of phase-type distributions and some applications, Journal of Applied Probability **33**, 368 (1996).
[67] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, The Journal of Chemical Physics **124**, 044104 (2006).
[68] A. Rohatgi, Webplotdigitizer: Version 4.6 (2022).

Appendix A: Mapping between the rate parameters of the telegraph model and the N-state model

In this Appendix we derive the rate parameters of an effective telegraph model by writing these in terms of the rate parameters of the N-state model whose reaction scheme is given by (1). We consider a parameter mapping involving the waiting time distribution between successive mRNA synthesis events.

1. Moments of the waiting time distribution

Let τ denotes the waiting time between successive mRNA synthesis events in the N-state model, and $f_N(\tau)$ its corresponding probability density function. The moments of $f_N(\tau)$ follow directly from the definition of the Laplace transform of $f_N(\tau)$, denoted by $f_N^*(s)$, since

$$f_N^*(s) = \langle e^{-s\tau} \rangle_N = \sum_{n=0}^{\infty} (-1)^n s^n \frac{\langle \tau^n \rangle_N}{n!}. \tag{A1}$$

Suppose that the gene is in the active state N at $t = 0$. Let $w_{i,j}(t)$ denote the mixed joint probability density function that the gene spends time t in state i before moving to state j. The probability distribution function of the total time to move from state N to 1, and then back to state N through states $2, 3, \ldots, N-1$ is the convolution of the individual distributions $w_{N,1}(t)$,$w_{1,2}(t)$,...,$w_{N-1,N}(t)$. Equivalently, the Laplace transform of the probability density function of the total time is given by

$$w^*_{N,1}(s)w^*_{1,2}(s)\cdot\cdots\cdot w^*_{N-1,N}(s), \tag{A2}$$

where $w^*_{i,j}(s)$ denote the Laplace transform of $w_{i,j}(t)$. Since the mRNA is transcribed only with some probability once the state N is reached, it follows that any number $n \geq 0$ of the cycles (N to 1 to 2 .. to N) is permitted before the mRNA is synthesised. This implies that the Laplace transform of the waiting time distribution between two successive mRNA synthesis events is given by

$$f^*_N(s) = \sum_{n=0}^{\infty}[w^*_{N,1}(s)w^*_{1,2}(s)\cdot\cdots\cdot w^*_{N-1,N}(s)]^n w^*_{N,N}(s) = \frac{w^*_{N,N}(s)}{1 - w^*_{N,1}(s)w^*_{1,2}(s)\cdot\cdots\cdot w^*_{N-1,N}(s)}, \tag{A3}$$

where

$$w^*_{i,i+1}(s) = \frac{k_i}{s+k_i}, \quad i = 1, \ldots, N-1, \quad w^*_{N,1}(s) = \frac{k_N}{s+k_N+\rho}, \quad w^*_{N,N}(s) = \frac{\rho}{s+k_N+\rho}. \tag{A4}$$

Note that $w^*_{N,N}(s)$ is the Laplace transform of the distribution of the time spent in state N before an mRNA transcription event occurs. We also note that the series in Eq. (A3) is guaranteed to converge since $w^*_{N,1}(s)w^*_{1,2}(s)\cdot\cdots\cdot w^*_{N-1,N}(s) < 1$ for non-negative s. Substituting Eq. (A4) into (A3) gives

$$f^*_N(s) = \frac{\rho\prod_{i=1}^{N-1}(s+k_i)}{(s+k_N+\rho)\prod_{i=1}^{N-1}(s+k_i) - \prod_{i=1}^{N}k_i}. \tag{A5}$$

This expression can be inverted to get $f_N(t)$ using partial fraction decomposition. The value of $f_N(0)$, however, can be obtained directly from $f^*_N(s)$ using the initial value theorem, yielding

$$f_N(0) = \lim_{s\to\infty}[sf^*_N(s)] = \rho. \tag{A6}$$

From the expression for $f^*_N(s)$ and using Eq. (A1), we get for the first two (raw) moments,

$$\langle\tau\rangle_N = -\left.\frac{df^*_N}{ds}\right|_{s=0} = \frac{k_N}{\rho}S_{N,1}, \tag{A7}$$

$$\langle\tau^2\rangle_N = \left.\frac{d^2f^*_N}{ds^2}\right|_{s=0} = 2\left[\left(1+\frac{k_N}{\rho}\right)S_{N,1} - \frac{1}{k_N}\right]\frac{k_N}{\rho}S_{N,1} - \frac{k_N}{\rho}(S^2_{N,1} - S_{N,2}), \tag{A8}$$

where

$$S_{N,1} = \sum_{j=1}^{N}\frac{1}{k_j}, \quad S_{N,2} = \sum_{j=1}^{N}\frac{1}{k_j^2}. \tag{A9}$$

From here we find that the variance $\sigma^2_\tau = \langle\tau^2\rangle_N - \langle\tau\rangle^2_N$ is given by

$$\sigma^2_{\tau,N} = \langle\tau^2\rangle_N - \langle\tau\rangle^2_N = \frac{k_N}{\rho}S_{N,2} + \frac{k_N}{\rho}\left(1+\frac{k_N}{\rho}\right)S^2_{N,1} - \frac{2}{\rho}S_{N,1}. \tag{A10}$$

We note that the inverse of $\langle\tau\rangle_N$, which is the effective transcription rate, can be interpreted as the actual transcription rate ρ multiplied by the steady-state probability that the gene is in the Nth state,

$$P_{\mathrm{on},N} = \frac{1}{k_N\sum_{i=1}^{N}\frac{1}{k_i}}. \tag{A11}$$

2. The mRNA count distribution

The probability distribution of mRNA counts in steady-state conditions can be computed by laborious calculations using the CME. Alternatively this can be more easily computed using results from queueing theory [63]. Namely, the N-state model maps to a queueing system with interarrival time τ whose probability density function is $f_N(\tau)$, and the service time is exponentially distributed with a rate parameter d. This queueing system belongs to the $G/M/\infty$ class of queueing systems, where G stands for general or arbitrary interarrival time distribution, M stands for Markovian or exponential service, and there are infinitely many servers. For this class of queueing systems the steady state distribution of the queue length (or in this case, the mRNA count) is known explicitly [64] (see also Ref. [63] for a recent review), and is given by

$$P(m) = \begin{cases} 1 - \sum_{j=1}^{\infty} (-1)^{j-1} \frac{C_{j-1}}{\langle \tau \rangle_N dj}, & m = 0 \\ \sum_{j=m}^{\infty} (-1)^{j-m} \binom{j}{m} \frac{C_{j-1}}{\langle \tau \rangle_N dj}, & m \geq 1 \end{cases}, \tag{A12}$$

where C_i for $i = 0, 1, \ldots$ are defined as

$$C_0 = 1, \quad C_j = \prod_{i=1}^{j} \frac{f_N^*(id)}{1 - f_N^*(id)}, \quad j \geq 1. \tag{A13}$$

Note that since by Eq. (A5) $f_N^*(s)$ is invariant to permutations of k_i, $i = 1, .., N-1$, it follows from Eq. (A12) that the steady-state distribution of mRNA counts also shares this property.

We can further simplify Eq. (A12) as follows. The probability-generating function $G(z)$ is given by

$$G(z) = \sum_{m=0}^{\infty} P(m) z^m = 1 + \sum_{m=1}^{\infty} \frac{C_{m-1}}{\langle \tau \rangle_N dm} (z-1)^m. \tag{A14}$$

We now write the ratio $f_N^*(s)/[1 - f_N^*(s)]$ in the factorized form

$$\frac{f_N^*(s)}{1 - f_N^*(s)} = \frac{\rho \prod_{i=1}^{N-1} (s + k_i)}{\prod_{i=1}^{N} (s + k_i) - \prod_{i=1}^{N} k_i} = \frac{\rho (s + a_1) \ldots (s + a_p)}{s (s + b_1), \ldots (s + b_p)}, \tag{A15}$$

which also serves as an implicit definition of p, $a_1, \ldots, a_p$, and $b_1, \ldots, b_p$; note that $p \leq N-1$. In the last expression of Eq. (A15), it is assumed that the polynomials in the numerator and denominator are coprime, i.e. that their greatest common divisor is 1. This means that p is not necessary equal to $N-1$, as one would expect from the first equality in Eq. (A15), but may in principle be less than $N-1$ if $\rho \prod_{i=1}^{N-1} (s + k_i)$ and $\prod_{i=1}^{N} (s + k_i) - \prod_{i=1}^{N} k_i$ have common factors. If $p \leq 4$, then all b_i can be found explicitly, otherwise they must be computed numerically. Substituting Eq. (A15) into (A14), and recognizing that

$$\langle \tau \rangle_N = \frac{1}{\rho} \left(\frac{b_1 \ldots b_p}{a_1 \ldots a_p} \right), \tag{A16}$$

we get the following expression for the probability generating function $G(z)$,

$$G(z) = {}_pF_p \left(a_1', \ldots, a_p'; b_1', \ldots, b_p'; \rho' (z-1) \right), \tag{A17}$$

where $a_i' = a_i/d$ and $b_i' = b_i/d$ for $i = 1, \ldots, p$, $\rho' = \rho/d$, and ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z)$ is the generalized hypergeometric function defined as

$${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z) = \sum_{m=0}^{\infty} \frac{(a_1)_m \ldots (a_p)_m}{(b_1)_m \ldots (b_q)_m} \frac{z^m}{m!}, \tag{A18}$$

and $(x)_n = x(x+1) \ldots (x+n-1)$ is the rising factorial. The probability $P(m)$ is obtained by taking the mth derivative of $G(z)$ evaluated at $z = 0$, which yields

$$P(m) = \frac{(\rho')^m}{m!} \frac{(a_1')_m \ldots (a_p')_m}{(b_1')_m \ldots (b_p')_m} \, {}_pF_p \left(a_1' + m, \ldots, a_p' + m; b_1' + m, \ldots, b_p' + m; -\rho' \right). \tag{A19}$$

If $\rho\prod_{i=1}^{N-1}(s+k_i)$ and $\prod_{i=1}^{N}(s+k_i)-\prod_{i=1}^{N}k_i$ do not have common factors, then $p=N-1$ and $a_i=k_i$ for $i=1,\dots,N-1$, in which case $P(m)$ simplifies to

$$P(m)=\frac{\rho'^m}{m!}\frac{(k_1')_m\dots(k_{N-1}')_m}{(b_1')_m\dots(b_{N-1}')_m}\,{}_{N-1}F_{N-1}\left(k_1'+m,\dots,k_{N-1}'+m;b_1'+m,\dots,b_{N-1}'+m;-\rho'\right), \tag{A20}$$

where we introduced $k_i'=k_i/d$. Note that all rate parameters in Eq. (A20) are normalised by the degradation rate parameter d and hence the steady-state distribution is only a function of non-dimensional parameters.

3. Rate parameters of an effective telegraph model via moment matching

We now consider the case of $N=2$ (the telegraph model). Let $x=k_2/k_1$. Then from Eqs. (A6), (A7) and (A10) we have

$$f_2(0)=\rho,\quad \langle\tau\rangle_2=\frac{1+x}{\rho},\quad \sigma_{\tau,2}^2=\frac{2x^2}{k_2\rho}+\frac{(x+1)^2}{\rho^2}. \tag{A21}$$

To map the N-state model to the effective 2-state telegraph model, we seek ρ^*, k_1^* and k_2^* of the telegraph model such that the following waiting time statistics of the two models match:

$$f_2(0)=f_N(0),\quad \langle\tau\rangle_2=\langle\tau\rangle_N,\quad \sigma_{\tau,2}^2=\sigma_{\tau,N}^2. \tag{A22}$$

From these equations we get the following expressions for ρ^*, k_1^* and k_2^*,

$$\rho^*=\rho,\quad k_1^*=\frac{2\left(\langle\tau\rangle_N-\frac{1}{\rho}\right)}{\sigma_{\tau,N}^2-\langle\tau\rangle_N^2},\quad k_2^*=\frac{2\rho\left(\langle\tau\rangle_N-\frac{1}{\rho}\right)^2}{\sigma_{\tau,N}^2-\langle\tau\rangle_N^2}. \tag{A23}$$

In this mapping, the effective rate parameters depend only on ρ, $k_1,\dots,k_N$ but not on the degradation rate d. It can also be shown that this mapping guarantees that the mean mRNA number is also the same between the models. We also note that both k_1^* and k_2^* are always positive since $\langle\tau\rangle_N>1/\rho$ and $\sigma_{\tau,N}^2>\langle\tau\rangle_N^2$, which can be verified from expressions in Eqs. (A7) and (A10).

The mapping in Eq. (A23) can be interpreted by introducing the random time $\tau_{1\to N}$ it takes the gene to go from state 1 to state N. Since this time is a sum of independent exponentially distributed random variables with rate parameters $k_1,\dots,k_{N-1}$, its mean and variance are given by

$$\langle\tau_{1\to N}\rangle=\sum_{j=1}^{N-1}\frac{1}{k_j},\quad \langle\tau_{1\to N}^2\rangle-\langle\tau_{1\to N}\rangle^2=\sum_{j=1}^{N-1}\frac{1}{k_j^2}. \tag{A24}$$

Substituting these expressions into Eq. (A23), we get

$$k_1^*=\frac{1}{\langle\tau_{1\to N}\rangle}\cdot\frac{2}{\mathrm{CV}_{1\to N}^2+1},\quad k_2^*=k_N\cdot\frac{2}{\mathrm{CV}_{1\to N}^2+1}, \tag{A25}$$

where $\mathrm{CV}_{1\to N}$ is the coefficient of variation of $\tau_{1\to N}$,

$$\mathrm{CV}_{1\to N}^2=\frac{\langle\tau_{1\to N}^2\rangle-\langle\tau_{1\to N}\rangle^2}{\langle\tau_{1\to N}\rangle^2}=\sum_{j=1}^{N-1}\frac{1}{k_j^2}\left(\sum_{j=1}^{N-1}\frac{1}{k_j}\right)^{-2}. \tag{A26}$$

Appendix B: Short-time analysis for the mean mRNA count

Assuming Markovian dynamics, the CME for the N-state reaction system (1) reads:

$$\frac{d}{dt}P_1(n,t)=d(n+1)P_1(n+1,t)-dnP_1(n,t)+k_NP_N(n,t)-k_1P_1(n,t),$$
$$\frac{d}{dt}P_i(n,t)=d(n+1)P_i(n+1,t)-dnP_i(n,t)+k_{i-1}P_{i-1}(n,t)-k_iP_i(n,t),\quad i\in\{2,N-1\}$$

$$\frac{d}{dt}P_N(n,t) = \rho(P_N(n-1,t) - P_N(n,t)) + d(n+1)P_N(n+1,t) - dnP_N(n,t) + k_{N-1}P_{N-1}(n,t) - k_N P_N(n,t), \tag{B1}$$

where $P_i(n,t)$ is the probability of observing n mRNA molecules when the gene is in state i at time t.

We first derive the moment equations from the CME. For compactness and simplicity of the expressions, we non-dimensionalise time and the reaction rates by the mRNA degradation rate: $t \mapsto td$, $k_i \mapsto k_i/d$ and $\rho \mapsto \rho/d$. This convention is used in this Appendix and all the ones that follow (unless otherwise stated).

The rth moment of the mRNA, conditional on the gene being in state i, is given by:

$$(n^r)_i = \sum_n n^r P_i(n). \tag{B2}$$

Note that the case $r = 0$ gives the probability of being in gene state i. In what follows, unless otherwise stated, for notational convenience we suppress the time dependence of the probabilities and of the moments.

By summing the terms on the right and left hand sides of Eq. (B1) over n, we find the time-evolution equations for the probabilities of being in each gene state:

$$\frac{d}{dt}(n^0)_1 = k_N(n^0)_N - k_1(n^0)_1, \tag{B3}$$

$$\frac{d}{dt}(n^0)_i = k_{i-1}(n^0)_{i-1} - k_i(n^0)_i, \quad i \in \{2, N\}. \tag{B4}$$

Similarly by multiplying the right and left hand sides of Eq. (B1) by n and summing over n, we obtain the time-evolution equations for the first moments of mRNA counts (conditional on the gene state):

$$\begin{aligned} \frac{d}{dt}(n)_1 &= -(n)_1 + k_N(n)_N - k_1(n)_1, \\ \frac{d}{dt}(n)_i &= -(n)_i + k_{i-1}(n)_{i-1} - k_i(n)_i, \quad i \in \{2, N-1\}, \\ \frac{d}{dt}(n)_N &= \rho(n^0)_N - (n)_N + k_{N-1}(n)_{N-1} - k_N(n)_N. \end{aligned} \tag{B5}$$

Say the gene is initially in inactive state j where $j \in 1, .., N-1$ and that the number of mRNA is zero. Due to the uni-directional nature of the gene state transitioning processes, the only reaction that can occur for very short time times is $G_j \to G_{j+1}$. Since at $t = 0$, $(n^0)_j = 1$ and $(n^0)_i = 0, \quad \forall i \neq j$, it immediately follows from Eq. (B4) that for short times $(n^0)_{j+1} \approx k_j t$.

The time-evolution for the next state to be visited, i.e. $j+2$, is given by

$$\frac{d}{dt}(n^0)_{j+2} = k_{j+1}(n^0)_{j+1} - k_{j+2}(n^0)_{j+2}. \tag{B6}$$

Assuming a Taylor series expansion of $(n^0)_{j+2}$ in powers of t, integrating the above differential equation and using the previous result for $(n^0)_{j+1}$ it follows that to leading order, $(n^0)_{j+2} \approx \frac{1}{2}k_j k_{j+1} t^2$.

By iterating this procedure, one finds that generally

$$(n^0)_{j+r} \approx \frac{\prod_{i=j}^{j+r-1} k_i}{r!} t^r, \quad r = 1, ..., N-j. \tag{B7}$$

Now summing Eq. (B5), we obtain a time-evolution equation for the total mRNA count:

$$\frac{d}{dt}\langle m \rangle = \rho(n^0)_N - \langle m \rangle, \tag{B8}$$

where $\langle m \rangle = \sum_{i=1}^{n}(n)_i$. Assuming a Taylor series expansion of $\langle m \rangle$ in powers of t, integrating the above differential equation and using Eq. (B7), we finally obtain:

$$\langle m \rangle \approx \rho \frac{\prod_{i=j}^{N-1} k_i}{(N-j+1)!} t^{N-j+1}. \tag{B9}$$

Next we derive an estimate of how small the time should be such that Eq. (B9) holds. We will do this by deriving the next-to-leading order correction to the mean and then comparing the size of this term with the leading-order term previously derived.

From Eq. (B4) it follows that

$$\frac{d}{dt}(n^0)_{j+r} = k_{j-1+r}(n^0)_{j-1+r} - k_{j+r}(n^0)_{j+r}, \quad j+r \in \{2, N\}. \tag{B10}$$

Substituting $(n^0)_{j+r} = \alpha_{j,r}t^r + \beta_{j,r}t^{r+1} + O(t^{r+2})$ and collecting terms of order t^r, we obtain:

$$\beta_{j,r+1} = \frac{1}{r+2}(k_{j+r}\beta_{j,r} - k_{j+r+1}\alpha_{j,r+1}). \tag{B11}$$

Note that $\alpha_{j,r}$ is known; by Eq. (B7), it immediately follows that it is given by

$$\alpha_{j,r} = \frac{\prod_{i=j}^{j+r-1} k_i}{r!}. \tag{B12}$$

Eq. (B11) is a first-order non-homogeneous recurrence relation with variable coefficients. This can be solved using standard methods yielding:

$$\beta_{j,r} = \left[\prod_{i=0}^{r-1} f_i\right]\left(\beta_{j,0} + \sum_{m=0}^{r-1} \frac{g_m}{\prod_{i=0}^{m} f_i}\right), \tag{B13}$$

where

$$f_r = \frac{k_{j+r}}{r+2}, \; g_r = \frac{-k_{j+r+1}\alpha_{j,r+1}}{r+2}. \tag{B14}$$

Furthermore from Eqs. (B3) and (B4) and the initial conditions $(n^0)_j = 1$, and $(n^0)_i = 0, \quad \forall i \neq j$, it follows that

$$(n^0)_j = 1 - k_j t + O(t^2),$$

which implies that

$$\beta_{j,0} = -k_j. \tag{B15}$$

Hence from Eqs. (B12), (B13), (B14) and (B15) it follows that

$$\begin{aligned}(n^0)_N &= \alpha_{j,N-j}t^{N-j} + \beta_{j,N-j}t^{N-j+1} + O(t^{N-j+2}),\\ &= \frac{\prod_{i=j}^{N-1} k_i}{(N-j)!}t^{N-j} - \frac{\prod_{i=j}^{N-1} k_i}{(N-j+1)!}\left(\sum_{i=j}^{N} k_i\right)t^{N-j+1} + O(t^{N-j+2}).\end{aligned} \tag{B16}$$

Finally substituting this result in the time-evolution for the total mRNA, Eq. (B8), and computing the series solution we obtain:

$$\langle m\rangle = \rho\left[\prod_{i=j}^{N-1} k_i\right]\left[\frac{t^{N-j+1}}{(N-j+1)!} - (1 + \sum_{i=j}^{N} k_i)\frac{t^{N-j+2}}{(N-j+2)!}\right] + O(t^{N-j+3}). \tag{B17}$$

Comparing the magnitudes of the terms proportional to t^{N-j+1} and t^{N-j+2}, we find that the leading-order term, i.e. Eq. (B9), is accurate provided

$$t \ll \frac{N-j+2}{1+\sum_{i=j}^{N} k_i} \leq N+1, \quad j \in \{1, N-1\}. \tag{B18}$$

The result suggests that the leading-order power law is visible over a long period of time provided the number of gene states to be traversed from the the initial inactive state to the active state $(N-j)$ is not too small and provided the rates of hopping from one gene state to another (relative to the mRNA degradation rate) are not too large. Note that Eq. (B18) does not involve ρ which implies that increasing ρ increases the max mRNA count reached at the end of the time interval over which the leading-order power law is valid, without affecting the length of this interval.

1. Post-transcriptional processing of mRNA

We note that in the calculations above we have interpreted the mRNA (M) in the N-state model as being the total mRNA. In some experiments, it is possible to distinguish between spliced and unspliced mRNA [55] and between nuclear and cytoplasmic mRNA [54]. We can model post-processing by adding extra reactions to the scheme (1). For example, if we interpret M as being unspliced nuclear mRNA then we could add a first-order reaction $M \to M_s$ describing the synthesis of spliced nuclear mRNA M_s with some rate k_s. Since for short times, $d/dt\langle m_s\rangle \propto k_s\langle m\rangle$, it immediately follows by integrating Eq. (B9) that the mean spliced mRNA count will follow a power law with an exponent that is one larger than that for unspliced mRNA. By the same argument the mean cytoplasmic mRNA will follow a power law with exponent that is one more than that of nuclear mRNA. If we interpret the mRNA in the N-state model as the cytoplasmic mRNA, then one could add a processing step to describe its translation to protein; in that case the short-time power law exponent for the latter will be one more than that for the cytoplasmic mRNA. On the other hand, if the post-processing of M to a species M_s can be modelled by a deterministic delay reaction (with delay T) then $\langle m_s\rangle(t) = \langle m\rangle(t-T)$ and hence in this case the power law exponent is unaffected by processing.

Appendix C: Short-time analysis for the variance and Fano factor of the mRNA counts

We next compute the perturbation analysis for the variance. The initial conditions are the same as before: gene is in inactive state j and the number of mRNA (and hence its variance) is initially zero.

From the CME, by a similar procedure to how we found the first moments, one can show that the time-evolution equations for the second moments of mRNA counts (conditional on the gene state) are given by

$$\begin{aligned}
\frac{d}{dt}(n^2)_1 &= (n)_1 - 2(n^2)_1 + k_N(n^2)_N - k_1(n^2)_1,\\
\frac{d}{dt}(n^2)_i &= (n)_i - 2(n^2)_i + k_{i-1}(n^2)_{i-1} - k_i(n^2)_i, \quad i \in \{2, N-1\},\\
\frac{d}{dt}(n^2)_N &= \rho[2(n)_N + (n^0)_N] + (n)_N - 2(n^2)_N + k_{N-1}(n^2)_{N-1} - k_N(n^2)_N.
\end{aligned} \tag{C1}$$

The second moment of the total mRNA count is defined to be $\langle m^2\rangle = \sum_{i=1}^{n}(n^2)_i$. Using Eqs. (B5) and (C1), it is straightforward to derive a time-evolution equation for the second moment:

$$\frac{d}{dt}\langle m^2\rangle = \rho[2(n)_N + (n^0)_N] + \langle m\rangle - 2\langle m^2\rangle. \tag{C2}$$

To obtain the short-time solution for $(n)_N$ we consider its time-evolution equation as given by the last equation of Eq. (B5). Since $(n)_i = 0\ \forall i$ and the short-time solution for $(n^0)_N$ is given by Eq. (B16), it follows that

$$(n)_N = \frac{\rho\alpha_{j,N-j}}{N-j+1}t^{N-j+1} + \frac{\rho}{N-j+2}\left(\beta_{j,N-j} - \frac{(1+k_N)\alpha_{j,N-j}}{N-j+1}\right)t^{N-j+2} + O(t^{N-j+3}). \tag{C3}$$

Substituting the short-time solutions for $(n^0)_N$, $\langle m\rangle$ (given by Eqs. (B16) and (B17), respectively) and the short-time solution for $(n)_N$ (Eq. (C3)) in the time-evolution equation for the second moment of mRNA counts (Eq. (C2)), and integrating we find

$$\langle m^2\rangle = \rho\left[\prod_{i=j}^{N-1} k_i\right]\left[\frac{t^{N-j+1}}{(N-j+1)!} + (2\rho - 1 - \sum_{i=j}^{N} k_i)\frac{t^{N-j+2}}{(N-j+2)!}\right] + O(t^{N-j+3}). \tag{C4}$$

Now the variance of total counts is defined as $\sigma_m^2 = \langle m^2\rangle - \langle m\rangle^2$. Considering Eqs. (B17) and (C4), we immediately see that to next-to-leading order terms, the variance has the same short-time solution as $\langle m^2\rangle$, i.e. Eq. (C4).

We note that the leading-order term in Eq. (C4) is the same as in Eq. (B17) – this implies that to leading-order, the expressions for the mean and the variance are precisely the same. However the next-to-leading order terms differ which implies:

$$\sigma_m^2 - \langle m\rangle = \frac{2\rho^2\prod_{i=j}^{N-1} k_i}{(N-j+2)!}\, t^{N-j+2} + O(t^{N-j+3}). \tag{C5}$$

Since the difference is positive, it follows that the variance deviates faster from the leading-order term than the mean. From Eq. (C5), it also follows that the Fano factor (defined as the variance divided by the mean of the total mRNA counts) is given by:

$$FF_m = \frac{\sigma_m^2}{\langle m \rangle} = 1 + \frac{2\rho}{N-j+2} t + O(t^2). \tag{C6}$$

Appendix D: Generalization of results to a broader class of systems with reversible reactions and gene state switching upon transcription

We consider a general N-state model described by the following reaction scheme:

$$G_i \underset{k_{i,j}}{\overset{k_{j,i}}{\rightleftharpoons}} G_j,\ i,j = 1,\dots,N, \quad G_N \xrightarrow{\rho} G_K + M, \quad M \xrightarrow{d} \emptyset. \tag{D1}$$

We denote the mRNA number at time t by $m(t)$, and assume that there is initially no mRNA present, i.e. $m(0) = 0$, and that the gene starts in state I. This model includes the N-state model presented in the main text, for which $k_{i,i+1} = k_i$ for $i = 1,\dots,N-1$, $k_{N,1} = k_N$, $K = N$, and all other rates $k_{i,j}$ are equal to 0. We note that in this Appendix we keep the absolute values of $k_{i,j}$ and ρ, i.e. we do not rescale them or time by d.

For the general model in Eq. (D1), we derive a power law for the mean number of mRNA for short times, in the form of

$$\langle m(t) \rangle = \frac{A}{n!} t^n + O(t^{n+1}). \tag{D2}$$

We show that A and n are determined from the series expansion of the Laplace transform of the probability density function of the waiting time until the first mRNA synthesis event,

$$f_1^*(s) = \frac{A}{s^n} + O\left(\frac{1}{s^{n+1}}\right), \quad s \to \infty. \tag{D3}$$

Furthermore, the exponent n is equal to the minimal number of states that are visited from the initial state I at time $t = 0$ until the synthesis of the first mRNA molecule. To derive these results, we use a mapping of the reaction scheme in Eq. (D1) to a $G/M/\infty$ queueing system, where G denotes that the arrival (of mRNA) constitute a renewal process with a general (arbitrary) interarrival time distribution, M denotes that the service (mRNA degradation) time is exponentially distributed (Markovian or memoryless), and there are infinitely many servers (meaning that mRNA can be in principle degraded as soon as it synthesised). For more details on how gene expression models map to infinite-server queues, we refer the reader to a recent review in Ref. [63].

We denote by $f(t)$ the probability density function of the interarrival time between two successive mRNA synthesis events. The Laplace transform of $f(t)$ was derived in Appendix A for the N-state model described in the main text, see Eq. (A5). For the general model described by Eq. (D1), $f(t)$ can be computed by solving the master equation for the probability $P_i(t)$ to find the gene in state i at time t,

$$\frac{dP_i}{dt} = \sum_{j \neq i} k_{j,i} P_j - \sum_{j \neq i} k_{i,j} P_i, \quad i = 1,\dots,N-1, \tag{D4}$$

$$\frac{dP_N}{dt} = \sum_{j \neq N} k_{j,N} P_j - \left(\sum_{j \neq i} k_{N,j} + \rho \right) P_N, \tag{D5}$$

where the reaction $G_N \to G_K + M$ is treated as an absorbing state, and the initial condition is

$$P_i(0) = \delta_{i,K}, \quad i = 1,\dots,N. \tag{D6}$$

If we denote $\mathbf{P}(t) = (P_1(t),\dots,P_N(t))$, then the master equation (D4) can be written as

$$\frac{d\mathbf{P}}{dt} = \mathbf{P} D, \quad \mathbf{P}(0) = \mathbf{e}_K \equiv (\underbrace{0,\dots,0}_{K-1}, 1, \underbrace{0,\dots,0}_{N-K}), \tag{D7}$$

where D is $N \times N$ matrix whose matrix elements are defined as

$$D_{ij} = \begin{cases} k_{ij}, & i \neq j, \\ -\sum_{m \neq i} k_{i,m}, & i = j \neq N, \\ -\sum_{m \neq i} k_{i,m} - \rho, & i = j = N. \end{cases} \tag{D8}$$

The solution of Eq. (D7) is

$$\mathbf{P}(t) = \mathbf{e}_K e^{Dt}. \tag{D9}$$

The probability density function of the waiting time between two successive mRNA synthesis events is given by

$$f(t) = \rho P_N(t) = \mathbf{e}_K e^{Dt}(-D\mathbf{1}^T), \tag{D10}$$

where $\mathbf{1}^T$ is $N \times 1$ column vector with every element being equal to 1. This distribution is known as a phase-type distribution [65]. If the initial state $I \neq K$ (the state to which the gene switches after producing an mRNA molecule), the probability density function of the waiting time until the first mRNA synthesis event is not given by $f(t)$, but is instead given by $f_1(t)$, which reads

$$f_1(t) = \mathbf{e}_I e^{Dt}(-D\mathbf{1}^T), \quad \mathbf{e}_I = (\underbrace{0, \ldots, 0}_{I-1}, 1, \underbrace{0, \ldots, 0}_{N-I}). \tag{D11}$$

The waiting time densities $f(t)$ and $f_1(t)$, the initial condition $m(0) = 0$ and the service rate d fully specify the queueing system $G/M/\infty$, for which we are interested in finding the early-time behaviour of the mean mRNA number $\langle m(t) \rangle$. The case in which $f_1(t) = f(t)$, which in our model corresponds to the initial state $I = K$, has been solved in Ref. [64], yielding the following result for the Laplace transform of the mean mRNA number,

$$\mathcal{L}[\langle m(t) \rangle](s) = \frac{f^*(s)}{(s+d)[1 - f^*(s)]}. \tag{D12}$$

Below we generalize this result to $f_1(t) \neq f(t)$, which yields

$$\mathcal{L}[\langle m(t) \rangle](s) = \frac{f_1^*(s)}{(s+d)[1 - f^*(s)]}. \tag{D13}$$

We note that $f_1^*(s)$ is a rational function of s, i.e. it can be written as a ratio of two coprime polynomials $p(s)$ and $q(s)$,

$$f_1^*(s) = \frac{p(s)}{q(s)} = \frac{As^{\deg(p)} + \ldots}{s^{\deg.(q)} + \ldots}, \tag{D14}$$

where we have written only the highest-order terms in each of the polynomials. To get an early-time behaviour of $\langle m(t) \rangle$, we expand $f_1^*(s)$ in s as $s \to \infty$,

$$f_1^*(s) = \frac{A}{s^n} + O\left(\frac{1}{s^{n+1}}\right), \quad s \to \infty, \quad n = \deg(p) - \deg(q). \tag{D15}$$

and substitute back into Eq. (D13), obtaining

$$\mathcal{L}[\langle m \rangle](s) = \frac{A}{s^{n+1}} + O\left(\frac{1}{s^{n+2}}\right), \tag{D16}$$

where we have used the fact that $f^*(s) \to 0$ for $s \to \infty$ (hence, there is no contribution from $f^*(s)$ to the lowest order). The power law in Eq. (D2) follows from the inverse Laplace transform of $1/s^{n+1}$, which gives $t^n/n!$. According to Theorem 3.1 in Ref. [66], given a phase-type probability density function whose Laplace transform is $p(s)/q(s)$, where $p(s)$ and $q(s)$ are coprime polynomials, the difference $n = \deg(q) - \deg(p)$ is equal to the minimal number of states that are visited before the absorption (in our case the synthesis of mRNA). This completes the derivation of the power law in Eq. (D2), and the interpretation of A and n.

In the rest of this Appendix, we derive Eq. (D13) by reiterating the steps in Ref. [64], adjusted to the first waiting time described by $f_1(t)$ rather than $f(t)$. Let T_n denotes the arrival time of the nth mRNA, and χ_n its degradation time. The number of mRNA molecules at time t can be written as

$$m(t) = \sum_{T_n \leq t} \theta(t - T_n, \chi_n), \tag{D17}$$

where the function $\theta(u, x)$ is defined as

$$\theta(u, x) = \begin{cases} 1, & 0 \leq u \leq x, \\ 0, & \text{otherwise.} \end{cases} \tag{D18}$$

Conditioned on $T_1 = x$, $m(t)$ can be written as

$$m(t) = \begin{cases} \theta(t - T_1, \chi_1) + m_0(t - x), & x \leq t, \\ 0, & x > t, \end{cases} \tag{D19}$$

where $m_0(t)$ is the number of mRNA molecules at time t in a process in which the time until the first mRNA synthesis event is distributed according to f instead of f_1 (another way to say this is that time $t = 0$ in this process is an arrival epoch). For $x \leq t$, the conditional probability to find $m(t) = k$ mRNA molecules at time, given that the first arrival occurred at time $t = x$, is given by

$$P(m(t) = 0|T_1 = x) = P(m_0(t) = 0)P(\chi_1 \leq t - x), \tag{D20a}$$

$$P(m(t) = k|T_1 = x) = P(m_0(t - x) = k)P(\chi_1 \leq t - x) + P(m_0(t - x) = k - 1)P(\chi_1 > t - x), \quad k \geq 1, \tag{D20b}$$

where

$$P(\chi_1 \leq t - x) = 1 - e^{-d(t-x)}, \quad P(\chi_1 > t - x) = e^{-d(t-x)} \tag{D21}$$

On the other hand, for $x > t$,

$$P(m(t) = k|T_1 = x) = \delta_{k,0}. \tag{D22}$$

We now define the probability generating function

$$G(z, t) = \sum_{k=0}^{\infty} P(m(t) = k)z^k = \sum_{k=0}^{\infty} \int_0^{\infty} dx\ f_1(x)P(m(t) = k|T_1 = x)z^k. \tag{D23}$$

Inserting Eqs. (D20), (D21) and (D22) into Eq. (D23) yields

$$G(z, t) = 1 - F_1(t) + \int_0^t dx\ f_1(x)G_0(z, t - x)\left[1 + (z - 1)e^{-d(t-x)}\right], \tag{D24}$$

where

$$G_0(z, t) = \sum_{k=0}^{\infty} P(m_0(t) = k)z^k. \tag{D25}$$

By taking the Laplace transform on both sides of Eq. (D24), we get

$$G^*(z, s) = \frac{1 - f_1^*(s)}{s} + f_1^*(s)G_0^*(z, s) + (z - 1)f_1^*(s)G_0^*(z, s + d), \tag{D26}$$

where

$$G^*(z, s) = \int_0^{\infty} dt\ G(z, t)e^{-st}, \quad G_0^*(z, s) = \int_0^{\infty} dt\ G_0(z, t)e^{-st}. \tag{D27}$$

Similarly, for the process in which the initial time $t = 0$ is an arrival epoch we get

$$G_0^*(z, s) = \frac{1 - f^*(s)}{s} + f^*(s)G_0^*(z, s) + (z - 1)f^*(s)G_0^*(z, s + d). \tag{D28}$$

By combining these two equations we get

$$G^*(s) = \frac{1}{s} + \frac{f_1^*(s)}{f^*(s)}\left[G_0^*(z,s) - \frac{1}{s}\right]. \tag{D29}$$

The Laplace transform of the mean mRNA number $\langle m(t)\rangle$ is given by

$$\mathcal{L}[\langle m(t)\rangle](s) = \left.\frac{dG}{dz}\right|_{z=1} = \frac{f_1^*(s)}{f^*(s)}\mathcal{L}[\langle m_0(t)\rangle](s) = \frac{f_1^*(s)}{(s+d)[1-f^*(s)]}, \tag{D30}$$

where in the last we equality used the expression for the Laplace transform of $\langle m_0(t)\rangle$ in Eq. (D12).

Let us now apply this result to the N-state model in the main text, with state j as the initial state as time $t = 0$. For $j \neq N$, The Laplace transform $f_1^*(s)$ reads

$$f_1^*(s) = \left(\prod_{i=j}^{N-1} \frac{k_i}{s+k_i}\right) f_N^*(s), \tag{D31}$$

where $f_N^*(s)$ is given by Eq. (A5). Expanding $f_1^*(s)$ in s at $s \to \infty$, we get

$$f_1^*(s) = \frac{\rho \prod_{i=j}^{N-1} k_i}{s^{N-j+2}} + O\left(\frac{1}{s^{N-j+3}}\right), \tag{D32}$$

from which we get that

$$A = \rho \prod_{i=j}^{N-1} k_i, \quad n = N - j + 1, \tag{D33}$$

exactly as in Eq. (3).

Appendix E: Details of calculations performed to generate the main text figures

1. Fig. 2

The rate parameter values of the N-state model were sampled to sufficiently cover the parameter space suggested by experimental data. Next we define the effective gene activation rate (σ_u), gene inactivation rate (σ_b), and transcription rate (ρ_u) of the N-state model and choose them to lie approximately within the bounds estimated from experimental data (see Table 1 in [45]). This was done to reflect the diversity of observed gene expression signatures in a variety of eukaryotic cells.

For the N-state model, the effective gene activation rate is roughly given by the inverse of the time it takes to transition from gene state 1 to N, i.e. $\sigma_u = (\sum_{i=1}^{N-1} k_i^{-1})^{-1}$, the effective gene inactivation rate is simply the inverse of the time it takes to leave the active state N, i.e. $\sigma_b = k_N$ and the effective transcription rate $\rho_u = \rho$. Note that these formulae are only used to choose meaningful rate parameter ranges for the N-state model.

From Table 1 in [45], after normalising each of the estimated rate parameters by the mRNA degradation rate parameter, the minimum and maximum values were: 0.22-11.3 (σ_u), 3.18-243.6 (σ_b), and 28.9-6300 (ρ_u), respectively. Hence the N-state rate parameters were chosen such that the observed rate parameters σ_u, σ_b, and ρ_u were log-uniformly distributed in the following ranges: $\sigma_u \in [0.025, 20]$, $\sigma_b \in [0.1, 250]$, and $\rho_u \in [10, 6500]$. Specifically k_i was chosen to be log-uniformly distributed on $[0.1, 40]$; k_N was chosen to be log-uniformly distributed on $[0.1, 250]$; and ρ was chosen to be uniformly distributed on $[10, 6500]$, for each $N = 3, 4, 5$.

Steady-state distributions of mRNA counts were computed using the finite-state projection algorithm (FSP) [67]. Rate parameter sets which gave steady-state mean mRNA count < 2 and > 200 were filtered out, as these distributions are either: in the former case, trivial to approximate due to very low copy numbers; or, in the latter case, have excessively large copy numbers, beyond those of common experimental distributions. The maximum mRNA count to truncate the FSP state space was selected to be the minimum of either the steady-state mean plus 10 steady-state standard deviations or 800. The absolute minimum size of the state space was chosen to be 20 mRNA.

In order for summary statistics to be comparable between the $N = 3, 4, 5$ state models, rate parameter sets were sampled such that the effective sampling density was constant for all values of N. This effectively enforces an

approximate equal spread of samples across the different N-state parameter spaces. If L represents the total number of samples, y the average number of samples per dimension (sampling density), and q the dimensionality of the parameter space (this equals $N+1$ since we have N gene state transition parameters $k_1, .., k_N$ and ρ; the degradation rate parameter d is fixed to 1), then they are connected by the following equality: $y = L^{1/q}$. Therefore the total number of samples for each model (L) was approximately: 1,200 (3-state), 8,200 (4-state), and 38,000 (5-state); preserving a sampling density of approximately 6.0. These procedures were used to construct Fig. 2A-G and I-K.

The rate parameter values used in Fig. 2D-G for the 5-state model steady-state mRNA count distribution shapes are given in Table II.

Fig. 2H used the same procedure as detailed previously, with one exception. Since shapes I and IV were sampled highly infrequently for each value of N, additional rate parameter sets were generated to ensure a sufficient sample size to draw summary statistics from. Note that these additional samples were not included in the results seen in Fig. 2I-K. For sampling parameters that resulted in distribution shape IV, k_i was chosen to be uniformly distributed in [1, 4], for each $N = 3, 4, 5$, and k_N was uniformly distributed on [0.1,1]. The sampling range of ρ remained unchanged. Summary statistics for Fig. 2H are shown in Table III.

Parameter	Shape			
	I	II	III	IV
k_1	29	7.2	1.2	2.94
k_2	16.3	17.3	0.9	4.21
k_3	37.5	12.1	1.1	2.6
k_4	36.7	32	1.4	1.7
k_5	0.1	5.4	0.25	0.58
ρ	35	81	40	80
d	1	1	1	1
Wasserstein distance	0.00588	0.235	0.946	1.44

TABLE II. Rate parameter values used for steady-state mRNA count distribution shapes I-IV in Fig. 2D-G.

Distribution Shape	N-state model	Percentile					Number of samples
		10th	25th (LQ)	50th (median)	75th (UQ)	90th	
I	3	0.00161	0.00442	0.0106	0.0246	0.0471	151
II		0.00196	0.00955	0.0520	0.161	0.400	1053
III		0.00179	0.0106	0.0792	0.308	0.716	183
IV		0.270	0.500	1.02	1.86	2.65	243
I	4	0.00670	0.0115	0.0280	0.0634	0.108	313
II		0.0159	0.0552	0.172	0.413	0.834	6681
III		0.0212	0.0948	0.302	0.783	1.25	1313
IV		0.330	0.571	1.09	1.68	2.24	509
I	5	0.0124	0.0230	0.0508	0.0862	0.129	963
II		0.0409	0.110	0.262	0.614	1.13	30758
III		0.0753	0.207	0.471	0.952	1.60	7023
IV		0.427	0.692	1.25	1.96	2.66	1464

TABLE III. Statistics summarising the WD distances computed between the steady-state distributions of the effective telegraph and the $N = 3, 4, 5$ state models using randomly chosen rate parameter values covering the physiologically relevant parameter space. See Fig. 2H for the box plots displaying these results.

2. Fig. 3

The reaction network used in Fig. 3A-C was:

$$\begin{aligned}
&G_1 \xrightarrow{k_1} G_2 \xrightarrow{k_2} G_3 \xrightarrow{k_3} G_4 \xrightarrow{k_4} G_5 \xrightarrow{k_5} G_1, \\
&G_5 \xrightarrow{\rho} G_5 + M, \\
&M \xrightarrow{d} \emptyset.
\end{aligned} \tag{E1}$$

The reaction network used in Fig. 3D was:

$$\begin{aligned}
&G_1 \xrightarrow{k_1} G_2 \xrightarrow{k_2} G_3 \xrightarrow{k_3} G_4 \xrightarrow{k_4} G_5 \xrightarrow{k_5} G_1, \\
&G_5 \xrightarrow{\rho} G_5 + M, \\
&M \xrightarrow{\gamma} M + P, \\
&M \xrightarrow{d} \emptyset, P \xrightarrow{d_p} \emptyset.
\end{aligned} \tag{E2}$$

The reaction networks used in Fig. 3F were the following. For the blue (t^3) line:

$$\begin{aligned}
&G_1 \xrightarrow{k_1} G_2 \xrightarrow{k_2} G_3 \xrightarrow{k_3} G_4 \xrightarrow{k_4} G_5 \xrightarrow{k_5} G_1, \\
&G_2 \xrightarrow{k_{2,5}} G_5 \\
&G_5 \xrightarrow{\rho} G_5 + M, \\
&M \xrightarrow{d} \emptyset,
\end{aligned} \tag{E3}$$

and for the pink (t^4) line:

$$\begin{aligned}
&G_1 \underset{k_{2,1}}{\overset{k_1}{\rightleftharpoons}} G_2 \xrightarrow{k_2} G_3 \xrightarrow{k_3} G_4 \xrightarrow{k_4} G_5 \xrightarrow{k_5} G_1, \\
&G_5 \xrightarrow{\rho} G_2 + M, \\
&M \xrightarrow{d} \emptyset.
\end{aligned} \tag{E4}$$

The rate parameter values used in Fig. 3A-D, and F are given in Table IV. The rate parameter values and gamma distribution parameters used in Fig. 3E for the static extrinsic noise SSA mean mRNA count results are given in Table V.

Parameter	Value			
	Panels A, B, C	Panel D	Panel F (t^3)	Panel F (t^4)
k_1	18.5	18.5	18.5	20.5
k_2	12.5	12.5	22.5	35.5
k_3	14.5	14.5	14.5	44.5
k_4	13.5	13.5	17.5	28.0
k_5	1.7	1.7	4.0	0.5
ρ	850	850	1000	1600
d	1.0	1.0	1.0	1.0
γ	-	100	-	-
d_p	-	0.2	-	-
$k_{2,5}$	-	-	9.0	-
$k_{2,1}$	-	-	-	0.5

TABLE IV. Rate parameter values used in Fig. 3A-D, and F.

Rate parameters	Gamma distribution						
	Mean value	t^2		t^4		t^5	
		Shape	Scale	Shape	Scale	Shape	Scale
k_1	18.5	0.25	74	4	4.625	1	18.5
k_2	12.5	0.25	50	4	3.125	1	12.5
k_3	14.5	0.25	58	4	3.625	1	14.5
k_4	13.5	0.25	54	4	3.375	1	13.5
k_5	1.7	0.25	6.8	4	0.425	1	1.7
ρ	850	0.25	3400	4	212.5	1	850
d	1.0	-	-	-	-	-	-

TABLE V. Rate parameter and gamma distribution parameter values used in Fig. 3E.

3. Fig. 4

The rate parameters for the N-state gene model were chosen as $k_i = (N-1)a_i$ for $i = 1, .., N-1$, $k_N = b$ and $\rho = c$ where a_i and b are uniform random numbers generated over $(0, 10]$ and c is a uniform random number over the interval $[100, 2000]$. The rate degradation parameter d is fixed to 1. Note that the rescaling of k_i by $N-1$ is such that the effective activation rate of the N-state model (as defined earlier for Fig. 2) is independent of N. The effective activation rates can be smaller or larger than the effective inactivation rate thus the procedure enables sampling of rate parameters leading to both constitutive and bursty gene expression, the two commonest modes of gene expression.

4. Fig. 5

For the non-linear regression analysis, we used the "nlinfit" and "lsqcurvefit" functions in MATLAB. We fitted the model $c_1 t^{c_2}$ to each data set of the mRNA counts versus time using a least-squares approach, thus obtaining estimates of the unknown parameters c_1 and c_2. Multiple different starting values of c_1 and c_2 were chosen to ensure that these functions returned estimates which satisfied the least-squares criteria.

a. Yeast *(S. cerevisiae)*, osmotic stress - Fig. 5A-D

Data was taken from [54] and pre-processed by summing the nuclear (respectively cytoplasmic) mRNA count over all cells and dividing by the total number of cells to obtain the mean mRNA count across all repetitions of each experimental condition. Where NaN values were present then the mean mRNA at that time was computed with the remaining values. Non-linear regressions were performed on the first 5-7 data points for each induction curve, as seen in the figure. All R-squared values were greater than 0.996 and all p-values were less than 10^{-5}.

b. Mouse, inflammation - Fig. 5E-H

Data was extracted using WebPlotDigitizer [68] to obtain the pre-mRNA and mRNA counts and times from Fig. S2 of Ref. [55]. Both Tnfaip3 and Icam1 induction curves were extracted from Fig. S2A, and both Cxcl1 and IL-1b induction curves were extracted from Fig. S2B. Non-linear regressions were performed on the first 5-7 data points for each induction curve, as seen in the figure, except for the IL-1b mRNA count curve (Fig. 5H, blue) which excluded the first data point, (0,0), and used the 2nd through 6th data points instead. All R-squared values were greater than 0.977 and all p-values were less than 10^{-3}.

Appendix F: Additional steady-state mRNA count distribution shapes

Fig. S1 shows various additional steady-state mRNA count distribution shapes for each of the shape classes I-IV described in Fig. 2D-G. Here we compare the distributions from the N-state model with $N = 5$ with those from the

effective telegraph model. The corresponding rate parameters for the 12 plots can be found in Table VI.

Fig. S1C displays a notable distribution shape, representing a very slow-switching regime of the telegraph model where the timescales of gene activation and inactivation are much longer than those of mRNA degradation leading to a distribution that is well approximated by the sum of a Dirac delta distribution at 0 mRNA (due to the inactive state) and a Poisson distribution (due to the active state). Fig. S1L displays the limiting case when the effective telegraph model gives a distribution with a long plateau starting at 0 molecules. This distribution is obtained when the parameters of the effective telegraph model are $k_1 = k_2 = d$ and $\rho \geq d$. Other notable shapes include Fig. S1F showing a left-skewed distribution with a long plateau, and Fig. S1J showing a distribution with high probability at very low copy numbers, representing mRNA that is degraded rapidly and/or a gene that is highly inactive.

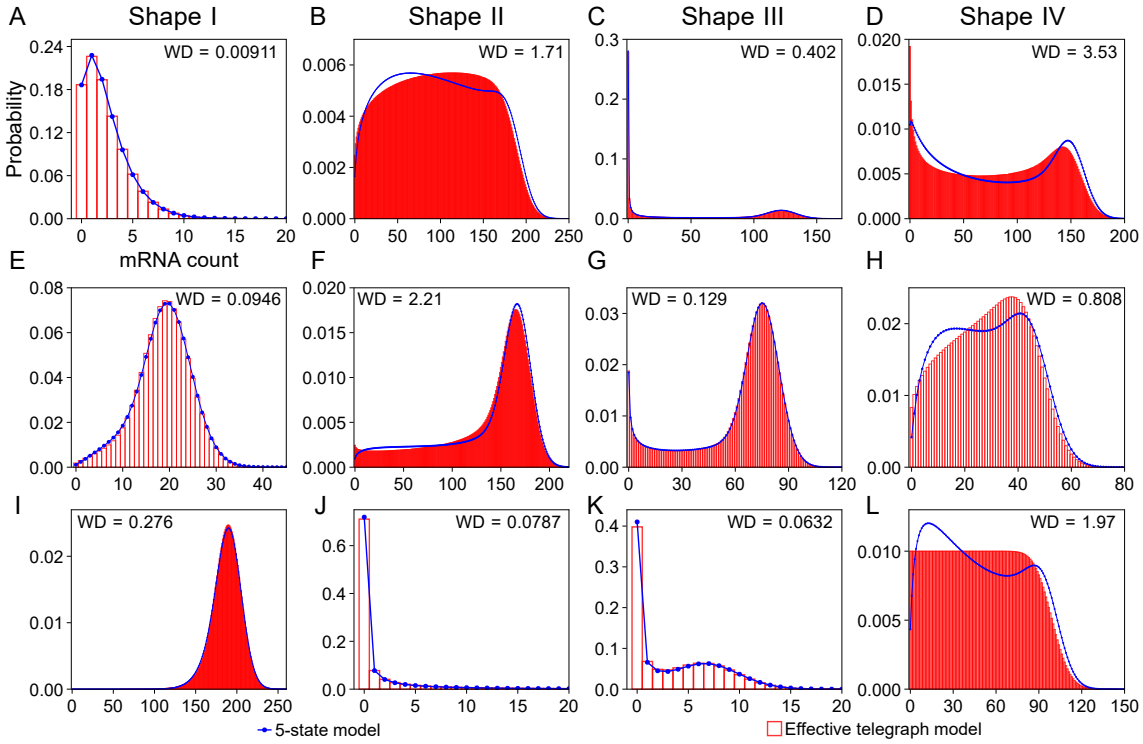

FIG. S1. Steady-state mRNA count distributions from the N-state model with $N = 5$ come in a variety of shapes, many of which can be well approximated by an effective telegraph model. Each panel displays the Wasserstein distance (WD) between the 5-state model (solid blue lines) and the effective telegraph model (histograms) distributions for three cases of each shape class. The parameters of the effective telegraph model are determined using Eq. (2). Histogram bar widths represent a single mRNA value. (A, E, I) Shape I - unimodal with Fano factor < 2; (B, F, J) Shape II - unimodal with Fano factor ≥ 2; (C, G, K) Shape III - bimodal with one mode at zero value and one mode at a non-zero value; (D, H, L) Shape IV - bimodal with both modes at non-zero values. Table VI shows the rate parameter values for each of these distributions.

Parameter	Value											
k_1	2.6	3.8	24.5	1.35	1.72	4.3	19.7	0.52	0.16	1.67	3.28	2.5
k_2	27	2.2	27.2	5.6	4.7	0.12	0.95	38.2	6	2.44	3.25	2.5
k_3	19.6	9.1	25.7	5.1	5.2	1.45	29.4	7.97	2.41	1.97	2.36	2.5
k_4	14	7.1	27.5	25	1.4	1.36	0.12	6.56	0.54	1.24	3.66	2.5
k_5	101	0.13	0.27	0.82	0.15	15.5	0.11	0.1	0.14	0.37	0.54	0.625
ρ	135	21	195	192	170	343	124	77	8	155	49	100
d	1	1	1	1	1	1	1	1	1	1	1	1
WD	0.00911	0.0946	0.276	1.71	2.21	0.0787	0.402	0.129	0.0632	3.53	0.808	1.97
Shape	I			II			III			IV		
Fig. S1 Panel	A	E	I	B	F	J	C	G	K	D	H	L

TABLE VI. Rate parameter values of the N-state model ($N = 5$) used for the steady-state mRNA count distributions in Fig. S1.