Pathway dynamics can delineate the sources of transcriptional noise in gene expression

Lucy Ham,[1*] Marcel Jackson,[2] Michael P.H. Stumpf[1,3*]

[1]School of BioSciences, University of Melbourne, Australia
[2]Department of Mathematics and Statistics, La Trobe University, Australia
[3]School of Mathematics and Statistics, University of Melbourne, Australia

* E-mail: lucy.ham@unimelb.edu.au, mstumpf@unimelb.edu.au

Single-cell expression profiling has opened up new vistas on cellular processes. Among other important results, one stand-out observation has been the confirmation of extensive cell-to-cell variability at the transcriptomic and proteomic level. Because most experimental analyses are destructive we only have access to snapshot data of cellular states. This loss of temporal information presents significant challenges in inferring dynamics, as well as causes of cell-to-cell variability. In particular, we are typically unable to separate dynamic variability from within individual systems ("intrinsic noise") from variability across the population ("extrinsic noise"). Here we mathematically formalise this non-identifiability; but we also use this to identify how new experimental set-ups coupled to statistical noise decomposition can resolve this non-identifiability. For single-cell transcriptomic data we find that systems subject to population variation invariably inflate the apparent degree of burstiness of the underlying process. Such identifiability problems can, in principle, be remedied by dual-reporter assays, which separates total gene expression noise into intrinsic and extrinsic contributions; unfortunately, however, this requires pairs of strictly independent and identical gene reporters to be integrated into the same cell, which is difficult to implement experimentally in most systems. Here we demonstrate mathematically that, in some cases decomposition of transcriptional noise is possible with non-identical and not-necessarily independent reporters. We use our result to show that generic reporters lying in the same biochemical pathways (e.g. mRNA and protein) can replace dual reporters, enabling the noise decomposition to be obtained

from only a single gene. Stochastic simulations are used to support our theory, and show that our "pathway-reporter" method compares favourably to the dual-reporter method.

Introduction

Noise is a fundamental aspect of every cellular process [1]. Frequently it is even of functional importance, for example in driving cell-fate transitions. Sometimes it can afford evolutionary advantages, for example, in the context of bet-hedging strategies. Sometimes, it can be a nuisance, for example, when it makes cellular signal processing more difficult. But noise is nearly ubiquitous at the molecular scale, and its presence has profoundly shaped cellular life. Analysing and understanding the sources of noise, how it is propagated, amplified or attenuated, and how it can be controlled, has therefore become a cornerstone of modern molecular cell biology. Noise arising in gene expression has arguably attracted most of the attention so far (but see e.g. [2] and [3] for the analysis of noise at the signalling level). Generally speaking, gene expression noise is separable into two sources of variability, as pioneered by Swain et al. [4]. Intrinsic noise is generated by the dynamics of the gene expression process itself. The process, however, is often influenced by other external factors, such as the availability of promoters and of RNA polymerase, the influence of long noncoding RNA as a transcriptional regulator [5], as well as differences in the cellular environment. Such sources of variability contribute extrinsic noise, and reflect the variation in gene expression and transcription activity across the cell population. As such, understanding extrinsic noise lies at the heart of understanding cell-population heterogeneity.

So far, elucidating the sources of gene expression noise from transcriptomic measurements alone has proven difficult [6, 7]. The fundamental hindrance lies in the fact that single-cell RNA sequencing, which provides most of the available data, is destructive, so that datasets reflect samples from across a population, rather than samples taken repeatedly from the same cell. As temporal information is lost in such measurements [8], it may be impossible to distinguish temporal variability within individual cells (e.g. burstiness), from ensemble variability across the population (i.e. extrinsic noise). A number of numerical and experimental studies have suggested this confounding effect [9–11], showing that systems with intrinsic noise alone exhibit behaviour that is indistinguishable from systems with both extrinsic and intrinsic noise. This is examined more formally in [12], where we show that the moment scaling behaviour and transcript distribution may be indistinguishable from situations with purely intrinsic noise. The limitations in inferring dynamics from population data are becoming increasingly evident, and a number of studies that seek to address some of these problems have emerged [13, 14].

Here we provide a detailed analysis of the extent to which sources of variability are identifiable from population data. We are able to prove rigorously that it is in general

impossible to identify the sources of variability, and consequently, the underlying transcription dynamics, from the observed transcript abundance distribution alone. Systems with intrinsic noise alone can always present identically to similar systems with extrinsic noise. Moreover, we demonstrate that extrinsic noise invariably distorts the apparent degree of burstiness of the underlying system: data which seems "bursty" is not necessarily generated by a bursty process, if there is cell-to-cell variability in e.g. the transcriptional machinery across cells. This has important ramifications for our analysis of experimental data: we cannot assess causes of transcriptional variability, if we do not simultaneously assess cell-to-cell variability in the transcriptional machinery. Our results highlight (in fact prove mathematically) the requirement for additional information, beyond the observed copy number distribution, in order to constrain the space of possible dynamics that could give rise to the same distribution.

This seemingly intractable problem can at least partially be resolved with a brilliantly simple approach: the dual-reporter method [4]. In this approach, noise can be separated into extrinsic and intrinsic components, by observing correlations between the expression of two independent, but identically distributed fluorescent reporter genes. Dual-reporter assays have been employed experimentally to study the noise contributed by both global and gene-specific effects [15–17]. A particular challenge, however, is that dual reporters are rarely identically regulated [17, 18], and are not straightforward to set up in every system. More recently, it has been shown that the independence of dual reporters cannot be guaranteed in some systems [19]. As a result, there have been efforts in developing alternative methods for decomposing noise [18, 20]. Here we develop a widely applicable generalisation (and simplification) of the original dual-reporter approach [4]. We demonstrate that non-identical and not-necessarily independent reporters can provide an analogous noise decomposition. Our result shows that measurements taken from the same biochemical pathway (e.g. mRNA and protein) can replace dual reporters, enabling the noise decomposition to be obtained from a single gene. This completely circumvents the requirement of strictly independent and identically regulated reporter genes. The results obtained from our "pathway-reporter" method are also borne out by stochastic simulations, and compare favourably with the dual-reporter method. In the case of constitutive expression, the results obtained from our decomposition are identical to those obtained from dual reporters. For bursty systems, we show that our approach provides a satisfactorily close approximation, except in extreme cases.

Modelling Intrinsic and Extrinsic Noise

A simple model for stochastic mRNA dynamics is the Telegraph model: a two-state model describing promoter switching, transcription, and mRNA degradation. In this model, all parameters are fixed, and gene expression variability arises due to the inherent stochasticity of the transcription process. As discussed above, this process will often

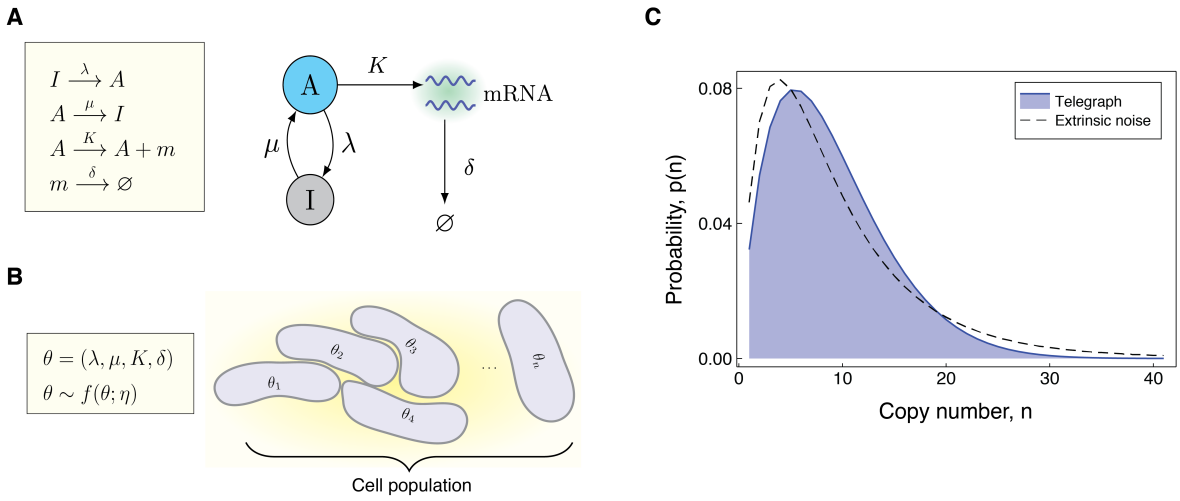


Figure 1: Modelling the effects of both intrinsic and extrinsic noise. (A) A schematic of the Telegraph process, with nodes A (active) and I (inactive) representing the state of the gene. Transitions between the states A and I occur stochastically at rates μ and λ, respectively. The parameter K is the mRNA transcription rate, and δ is the degradation rate. (B) The compound model incorporates extrinsic noise by assuming that parameters θ of the Telegraph model vary across an ensemble of cells, according to some probability distribution $f(\theta; \eta)$. (C) Variation in the parameters across the cell population leads to greater variability in the mRNA copy number distribution.

be influenced by extrinsic sources of variability, and so modifications to account for this additional source of variability are required.

The Telegraph Model

The Telegraph model was first introduced in [21], and since then has been widely employed in the literature to model bursty gene expression in eukaryotic cells [22–25]. In this model, the gene switches probabilistically between an active state and an inactive state, at rates λ (on-rate) and μ (off-rate), respectively. In the active state, mRNAs are synthesized according to a Poisson process with rate K, while in the inactive state, transcription does either not occur, or possibly occurs at some lower Poisson rate, $K_0 \ll K$. Degradation of RNA molecules occurs independently of the gene promoter state at rate δ. Fig. 1A shows a schematic of the Telegraph model. Throughout the discussion here, we will rescale all parameters of the Telegraph model by the mRNA degradation rate, so that $\delta = 1$. The

steady-state distribution for the mRNA copy number can be explicitly calculated as [26],

$$\tilde{p}_T(n;\theta) = \frac{K^n \lambda^{(n)}}{n!(\mu+\lambda)^{(n)}} {}_1F_1(\lambda+n, \lambda+\mu+n, -K). \tag{1}$$

Here θ denotes the parameter vector $(\mu, \lambda, K, \delta)$, the function ${}_1F_1$ is the confluent hypergeometric function [27], and, for real number x and positive integer n, the notation $x^{(n)}$ abbreviates the rising factorial of x (also known as the Pochhammer function). Throughout, we will refer to the probability mass function $\tilde{p}_T(n;\theta)$ as the Telegraph distribution with parameters θ. Constitutive gene expression is just a limiting case of the Telegraph model, which arises when the off-rate μ is 0, so that the gene remains permanently in the active state. In this case, the Telegraph distribution simplifies to a Poisson distribution with rate K; Pois(K).

At the opposite extreme is instantaneously bursty gene expression in which mRNA are produced in very short bursts with potentially prolonged periods of inactivity in between. This mode of gene expression has been frequently reported experimentally, particularly in mammalian genes [17, 22, 24, 25]. Transcriptional bursting may be treated as a limit of the Telegraph model, where the off-rate, μ, has tended toward infinity, while the on-rate λ remains fixed. In this limit, it can be shown [9, 12] that the Telegraph distribution converges to the negative binomial distribution NegBin(λ, $\frac{K}{\mu+K}$).

The Compound Distribution

We can elegantly account for random cell-to-cell variation across a population by way of a compound distribution [28]

$$\tilde{q}(n;\eta) = \int \tilde{p}(n;\theta) f(\theta;\eta)\, d\theta, \tag{2}$$

where $\tilde{p}(n;\theta)$ is the stationary probability distribution of a system with fixed parameters θ and $f(\theta;\eta)$ denotes the multivariate distribution for θ with hyperparameters η. Often we will take $\tilde{p}(n;\theta)$ to be the stationary probability distribution of the Telegraph model ((1)), and refer to (2) as the compound Telegraph distribution. Sometimes $\tilde{p}(n;\theta)$ will be the Poisson distribution or the negative binomial distribution, depending on the underlying mode of gene activity.

The compound distribution is valid in the case of ensemble heterogeneity, that is, when parameter values differ between individual cells according to the distribution $f(\theta;\eta)$, but remain constant over time [2]. This model is also a valid approximation for individual cells with dynamic parameters, provided these change sufficiently slower than the transcriptional dynamics [29]. Fig. 1B gives a pictorial representation of the compound distribution.

In general, the compound Telegraph distribution $\tilde{q}(n;\eta)$ will be more dispersed than a Telegraph distribution to account for the uncertainty in the parameters; see Fig. 1C.

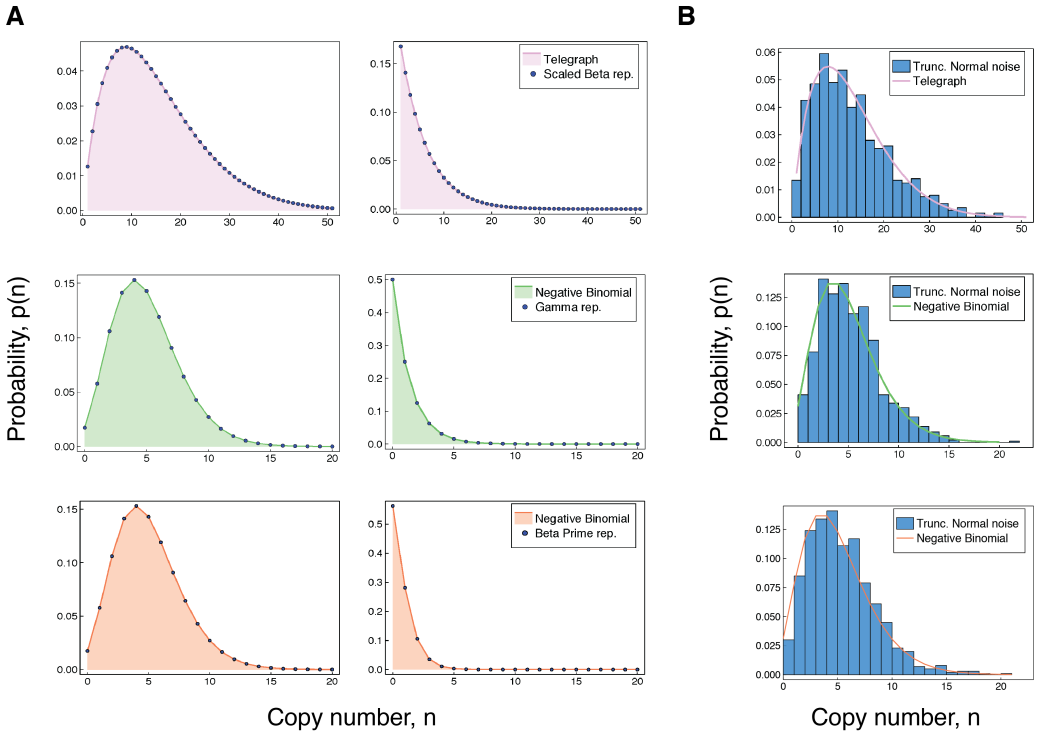


Figure 2: (A) Accuracy of our representations for the Telegraph and negative binomial distribution. For each of the results in (3)-(5), we compare the (fixed-parameter) Telegraph and negative binomial distributions with their respective compound representations for two different sets of parameter values. The top panel (shown in pink) is shows comparisons for (3). Referring to Table 1, the parameter values for the top panel are (left) $\lambda = 2$, $\mu' = 12$, $K' = 100$, $\mu = 3$, and $K \sim \text{Beta}_{K'}(5, 9)$, and (right) $\lambda = 1$, $\mu' = 20$, $K' = 100$, $\mu = 2$ and $K \sim \text{Beta}_{K'}(3, 18)$. The middle panel (green) gives comparisons for (4), with parameter values (left) $\lambda = 10$, $\beta = 2$, $\mu = 2$ and $K \sim \text{Gamma}(12, 2)$ and (right) $\lambda = 1$, $\beta = 1$, $\mu = 2$ and $K \sim \text{Gamma}(3, 1)$. The bottom panel (coral) gives comparisons for (5). The parameter values (left) are $\lambda' = 10$, $\lambda = 15$ and $c = 2$ and (right) are $\lambda' = 2$, $\lambda = 5$ and $c = 3$. (B) The top figure compares a $\text{Telegraph}(2, 4, 60)$ distribution with samples from a compound Telegraph distribution with normal noise $\text{Norm}(37, 10)$ on the transcription rate parameter. The middle figure compares a $\text{NegBin}(5, 0.5)$ with samples from a compound Telegraph distribution with normal noise $\text{Norm}(5.5, 2.3)$ on the transcription rate parameter. The bottom figure compares a $\text{NegBin}(5, 1)$ distribution with samples from a compound negative binomial distribution with normal noise $\text{Norm}(2.3, 0.6)$ on the burst intensity parameter.

Such dispersion is widely observed experimentally, and as demonstrated in [12], reflects the presence of extrinsic noise.

Identifiability Considerations

Decoupling the effects of extrinsic noise from experimental measurements has been notoriously challenging. In the context of (2), the distribution $f(\theta;\eta)$ reflects the population heterogeneity, but experimental data provides samples only of $\tilde{q}(n;\eta)$. How much can we deduce of the underlying dynamics (that is, $\tilde{p}(n;\theta)$), and the population heterogeneity ($f(\theta;\eta)$), from measurements of transcripts from across the cell population ($\tilde{q}(n;\eta)$)?

Of course, even though we may be able to infer the parameter η from experimental data, the expression $\tilde{p}(n;\theta)$ is really a family of distributions, parameterised by θ. This presents two fundamental challenges. The first is the possibility that there are different families of distributions $\tilde{p}(n;\theta)$ that can yield the same compound distribution, $\tilde{q}(n;\eta)$, but generated by different mechanisms, i.e. noise distributions, $f(\theta;\eta)$. The second, perhaps more subtle, challenge is that, even for a fixed family of distributions $\tilde{p}(n;\theta)$ it may be possible that different choices of the noise distribution $f(\theta;\eta)$ could still yield the same compound distribution $\tilde{q}(n;\eta)$. In these situations, we cannot hope to infer a unique solution for the noise distribution. This belongs to the important class of identifiability problems [30]; it has important ramifications for the interpretability of parameter estimates obtained from experimental data [31]. Indeed, if two or more model parameterisations are observationally equivalent (in this case, in the form of the transcript abundance distribution $\tilde{q}(n;\eta)$), then not only does this cast doubt upon the suitability of the model to represent (and subsequently predict) the system, but it also obstructs our ability to infer mechanistic insight from experimental data.

An example of the first identifiability problem arises from a well-known example of a compound distribution, (2): when $f(\theta;\eta)$ is a gamma distribution and $\tilde{p}(n;\theta)$ is a Poisson distribution, corresponding to constitutive gene expression, the resulting compound distribution $\tilde{q}(n;\eta)$ is a negative binomial distribution [32]. But this is the same distribution as that arising from instantaneously bursty gene expression [12]. Such identifiability instances may be circumvented if there is confidence in the basic mode of gene activity, i.e., if there is reason to believe that a gene is not constitutively active, for example. We find, however, that there are numerous instances that can present insurmountable identifiability problems.

Bursty Gene Expression

We first observe that any Telegraph distribution with fixed parameters can be identically obtained from a Telegraph distribution with parameter variation. As shown in the supplementary material, any Telegraph distribution $\tilde{p}_T(n;\lambda,\mu',K')$ can be written as,

$$\tilde{p}_T(n;\lambda,\mu',K') = \int_0^{K'} \tilde{p}_T(n;\lambda,\mu,t) f_{K'}(t;\lambda+\mu,\mu'-\mu)\,dt, \tag{3}$$

Copy no. distribution $\tilde{q}(n;\eta)$	Underlying distribution $\tilde{p}(n;\theta)$	Noise distribution $f(\theta;\eta)$
$\text{Tele}(\lambda,\mu',K')$	$\text{Tele}(\lambda,\mu,K)$	$K\sim\text{Beta}_{K'}(\lambda+\mu,\mu'-\mu)$
$\text{Tele}(\lambda,\mu',K')$	$\text{Pois}(K)$	$K\sim\text{Beta}_{K'}(\lambda,\mu')$
$\text{NegBin}(\lambda,\frac{\beta}{\beta+1})$	$\text{Tele}(\lambda,\mu,K)$	$K\sim\text{Gamma}(\lambda+\mu,\beta)$
$\text{NegBin}(\lambda,\frac{\beta}{\beta+1})$	$\text{Pois}(K)$	$K\sim\text{Gamma}(\lambda,\beta)$
$\text{NegBin}(\lambda',\frac{b}{b+1})$	$\text{NegBin}(\lambda,\frac{\theta}{\theta+1})$	$\theta\sim\text{BetaPrime}_b(\lambda-\lambda',\lambda')$

Table 1: Summary of the non-identifiability results. The results given in lines 1, 3 and 5 are our contributions, while the remaining representations (lines 2 and 4) are known and can be obtained as special cases of our results. Note that here we use $\text{Tele}(\lambda,\mu,K)$ to denote a Telegraph distribution with parameters λ,μ,K. In lines 3 and 4, the parameter $\beta>0$ can be chosen freely and determines the mean burst intensity in the resulting compound system. In line 5 the parameters $b,\theta>0$ are again mean burst intensities, and b can be chosen freely in the determination of the distribution of θ.

where $\mu<\mu'$ and $f_{K'}(t;\lambda+\mu,\mu'-\mu)$ is the probability density function for a scaled beta distribution $\text{Beta}_K(\lambda+\mu,\mu'-\mu)$ with support $[0,K']$. Thus, any Telegraph distribution can be obtained by varying the transcription rate parameter on a narrower Telegraph distribution (i.e., with a smaller off-rate) according to a scaled beta distribution. Fig. 2A (top panel) compares the representation obtained in (3) with the corresponding fixed-parameter Telegraph distribution for two different sets of parameters. When $\mu=0$ the representation given in (3) simplifies to the well-known Poisson representation of the Telegraph distribution in terms of the scaled beta distribution [33].

Instantaneously Bursty Gene Expression

The previous result extends to instantaneously bursty systems. The copy number distribution of an instantaneously bursty system can be obtained from both bursty and instantaneously bursty dynamics, provided that there is appropriate parameter variation. The supplementary material contains the relevant derivations. In the following, we let $\tilde{p}_{\text{NB}}(n;r,\beta)$ denote the probability mass function of a $\text{NegBin}(r,\frac{\beta}{\beta+1})$ distribution, where $\beta>0$. Then for any negative binomial distribution of the form $\text{NegBin}(\lambda,\frac{\beta}{\beta+1})$ we have,

$$\tilde{p}_{\text{NB}}(n;\lambda,\beta)=\int_0^\infty \tilde{p}_T(n;\lambda,\mu,t)f(t;\lambda+\mu,\beta)\,dt, \tag{4}$$

where $f(t;\lambda+\mu,\beta)$ is the probability density function of a $\text{Gamma}(\lambda+\mu,\beta)$ distribution. This result generalises the aforementioned classic representation of the negative binomial distribution [32], which corresponds to the particular case of $\mu=0$. In Fig. 2A (middle panel), we compare the representation obtained in (4) with the corresponding fixed-parameter negative binomial distribution for two different sets of parameters.

We also have the following representation for a negative binomial distribution in terms of a scaled beta prime distribution,

$$\tilde{p}_{\mathrm{NB}}(n;\lambda',b) = \int_{\frac{1}{b}}^{\infty} \tilde{p}_{\mathrm{NB}}(n;\lambda,\theta) f_b(b\theta - 1; \lambda - \lambda', \lambda')\, d\theta, \tag{5}$$

where $f_b(b\theta - 1; \lambda - \lambda', \lambda')$ is the probability mass function of a scaled beta prime $\mathrm{BetaPrime}_b(\lambda - \lambda', \lambda)$ distribution, where $b > 0$ and $\lambda > \lambda'$. This equivalently corresponds to scaled beta noise $\mathrm{Beta}_b(\lambda - \lambda', \lambda')$ on the inverse of the expected burst intensity. Thus, the distribution of any instantaneously bursty system with mean burst intensity b can be obtained from one with greater burst frequency, by varying the mean burst intensity θ according to a shifted beta prime distribution. Fig. 2A (bottom panel), compares the representation obtained in (5) with the associated fixed-parameter negative binomial distribution for two different sets of parameters.

An Exception: Constitutive Expression

It has long been known [34] that a compound Poisson distribution uniquely determines the compounding distribution. In the context of (2), this means the full extrinsic noise distribution $f(\theta, \eta)$ is identifiable from $\tilde{q}(n;\eta)$. As we demonstrate in related work [35], it is therefore possible to extract the extrinsic noise distribution, $f(\theta, \eta)$, from transcript copy number measurements.

Implications for Parameter Inference

Estimates of kinetic parameters from experimental data suggest that gene expression is often either bursty or instantaneously bursty (i.e., $\mu \gg \lambda$). In turn, the assumption that gene-inactivation events occur far more frequently than gene-activation events is often used to derive other models of stochastic mRNA dynamics [36–38]. The representations given in (3)–(5), however, show that both estimating parameters and the underlying dynamics from the form of the copy number distribution alone can be misleading. Noise on the transcription rate will invariably produce copy number data that is suggestive of a more bursty model. To illustrate this, consider an example in which the underlying process is a (mildly) bursty Telegraph system with distribution $\tilde{p}_T(n; 2, 3, K)$. Now assume that noise on the transcription rate parameter K follows the scaled Beta distribution on the interval $[0, 100]$ with $\alpha = \lambda + \mu = 2 + 3 = 5$ and $\beta = \mu' - \mu = 12 - 3 = 9$. The shape of this noise distribution closely resembles a slightly skewed Gaussian distribution, with the majority of transcription rates between around 10 and 60. This noise on the transcription rate K within the Telegraph system $\tilde{p}_T(n; 2, 3, K)$ will present identically to the significantly burstier system $\tilde{p}_T(n; 2, 12, 100)$.

It is of practical importance to recognise that, while the non-identifiability results (summarised in Table 1) are dependent on specific noise distributions, for practical pur-

poses any similar distribution will produce a similar effect. To demonstrate this, we replace the various noise distributions required for the representations in (3)–(5), with suitable normal distributions truncated at 0. In each case, we sample 1,000 data points from the corresponding compound distribution, and compared this with the associated fixed-parameter copy number distribution. The results are shown in Fig. 2B. The truncated normal distribution is not chosen on the basis of biological relevance, but rather to demonstrate that even a symmetric noise distribution (except for truncation at 0) produces qualitatively similar results to the distributions used in the precise non-identifiability results. In every case, the effect of a unimodal noise distribution on the transcription rate or burst intensity parameter is to produce copy number distributions that are generally consistent with systems that appear burstier.

Finally, we note that our results explain previous empirical observations that static measurements of mRNA are not always sufficient to infer the underlying dynamics of gene activity. Skinner et al. [13] address some of these limitations by quantifying both nascent and mature mRNA in individual cells, as well incorporating cell-cycle effects into their analysis of two mammalian genes. A more developed treatment of model identifiability is given in [14], where it shown how a stochastic model incorporating the downstream processing of mRNA can be used to distinguish a particular instance of non-identifiability. More specifically, the authors consider the non-identifiability problem noted in [12], arising from the Gamma-Poisson compound representation of the negative binomial distribution; a particular case of (4) above. Despite identical distributions at the nascent level, the marginal distributions of the processed (mature) mRNA are found to be substantially different. It is likely that a similar analysis will be valuable in the context of the other identifiability problems we have given in Table 1. In the next section, however, we take a more general approach to resolving non-identifibiality and exploit the properties of complex gene expression dynamics to determine not only the presence of extrinsic noise, but also estimate its magnitude.

Resolving Non-identifiability

The results of the previous section show that additional information, beyond the observed copy number distribution, is required to constrain the space of possible dynamics that could give rise to the same distribution. One way to narrow this space of possibilities, is to determine the intrinsic and extrinsic contributions to the total variation in the system.

The Dual-Reporter Method

The total gene expression noise, as measured by the squared coefficient of variation η^2, can be decomposed exactly into a sum of intrinsic and extrinsic noise contributions [4]. The decomposition applies to dynamic noise [39], and generalises to higher moments in [40].

As shown in [39], the noise decomposition is equivalent to the normalised Law of Total Variance [41]. Indeed, if X is the random variable denoting the number of molecules of a certain species (eg. mRNA or protein) in a given cell, then we can decompose the total noise by conditioning X on the state $\mathbf{Z} = (Z_1, \ldots, Z_n)$ of the environmental variables Z_1,...,Z_n,

$$\eta_X^2 = \frac{\mathrm{E}(\mathrm{Var}(X;\mathbf{Z}))}{\mathrm{E}(X)\,\mathrm{E}(Y)} + \frac{\mathrm{Var}(\mathrm{E}(X;\mathbf{Z}))}{\mathrm{E}(X)\,\mathrm{E}(Y)} \equiv \eta_{\mathrm{int}}^2 + \eta_{\mathrm{ext}}^2. \tag{6}$$

It has been shown [4,39] that if X_1 and X_2 are random variables denoting the expression levels of independent and identically distributed gene reporters, then the extrinsic noise contribution η_{ext}^2 in (6) can be identified by the normalised covariance between X_1 and X_2,

$$\eta_{\mathrm{ext}}^2 = \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{E}(X_1)\,\mathrm{E}(X_2)}. \tag{7}$$

Decomposing Noise with Non-Identical Reporters

The dual-reporter method requires distinguishable measurements of transcripts or proteins from two independent and identically distributed reporter genes integrated into the same cell. In practice, however, dual reporters rarely have identical dynamics, which is widely considered to be a significant challenge to interpreting experimental results [18]. We show that, under certain conditions, the decomposition in (6) can alternatively be obtained from non-identically distributed and not-necessarily independent reporters.

Our result relies on the observation that the covariance of any two variables can be decomposed into the expectation of a conditional covariance and the covariance of two conditional expectations (the Law of Total Covariance [41]). If X and Y denote, for example, the numbers of molecules of two chemical species (eg. mRNA and protein) in a given cell, then the covariance of X and Y can be decomposed by conditioning on the state $\mathbf{Z} = (Z_1, \ldots, Z_n)$ of the environmental variables Z_1,...,Z_n,

$$\mathrm{Cov}(X, Y) = \underbrace{\mathrm{E}(\mathrm{Cov}(X, Y; \mathbf{Z}))}_{\text{intrinsic}} + \underbrace{\mathrm{Cov}(\mathrm{E}(X; \mathbf{Z}), \mathrm{E}(Y; \mathbf{Z}))}_{\text{extrinsic}}. \tag{8}$$

We will see that in many cases of interest the random variable $\mathrm{E}(X;\mathbf{Z})$ (as a function of $\mathbf{Z}$) splits across common variables with $\mathrm{E}(Y;\mathbf{Z})$. By this we mean that $\mathrm{E}(X;\mathbf{Z}) = f(\mathbf{Z}_X)h_X(\mathbf{Z}')$ and $\mathrm{E}(Y;\mathbf{Z}) = g(\mathbf{Z}_Y)h_Y(\mathbf{Z}')$, where $\mathbf{Z}_X$ are the variables of $\mathbf{Z}$ that appear in $\mathrm{E}(X;\mathbf{Z})$ but not in $\mathrm{E}(Y;\mathbf{Z})$, and dually, $\mathbf{Z}_Y$ are those in $\mathrm{E}(Y;\mathbf{Z})$ that are not in $\mathrm{E}(X;\mathbf{Z})$. The variables $\mathbf{Z}'$ are those variables from $\mathbf{Z}$ not in $\mathbf{Z}_X \cup \mathbf{Z}_Y$. In these cases, the component of $\mathrm{Cov}(X, Y)$ that is contributed by the variation in $\mathbf{Z}$ (the extrinsic component) may be written as the covariance of the functions $h_X(\mathbf{Z}')$ and $h_Y(\mathbf{Z}')$. Conveniently, in the cases of interest here, the two functions h_X and h_Y coincide, and this is the form we use in the following decomposition principle. The supplementary material contains the proof of this result.

The Noise Decomposition Principle (NDP)

Assume that there are measurable functions f, g and h such that $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$ split across common variables by way of $\mathrm{E}(X;\mathbf{Z}) = f(\mathbf{Z}_X)h(\mathbf{Z}')$ and $\mathrm{E}(Y;\mathbf{Z}) = g(\mathbf{Z}_Y)h(\mathbf{Z}')$. Then, provided that the variables $Z_1, \ldots, Z_m$ are mutually independent, the normalised covariance of $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$ will identify the total noise on $h(\mathbf{Z}')$ (i.e. $\eta^2_{h(\mathbf{Z}')}$).

As we show in the next section, there are many situations where the random variable $\mathrm{E}(X;\mathbf{Z})$ is precisely the common part of $\mathrm{E}(Y;\mathbf{Z})$ and $\mathrm{E}(X;\mathbf{Z})$ (i.e., $h(\mathbf{Z}') = \mathrm{E}(X;\mathbf{Z})$), and the normalised intrinsic contribution to the covariance is either zero or negligible. In these cases, the normalised covariance of X and Y will identify precisely the extrinsic noise contribution η^2_{ext} to the total noise η^2_X. To see this, consider the situation where $\mathrm{E}(Y;\mathbf{Z}) = f(\mathbf{Z}_Y)\,\mathrm{E}(X;\mathbf{Z})$. Then provided $f(\mathbf{Z}_Y)$ and $\mathrm{E}(X;\mathbf{Z})$ are independent random variables, the extrinsic contribution to the covariance of X and Y is given by,

$$\begin{aligned}\mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}), \mathrm{E}(Y;\mathbf{Z})) &= \mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}), f(\mathbf{Z}_Y)\,\mathrm{E}(X;\mathbf{Z})) \\ &= \mathrm{E}(f(\mathbf{Z}_Y))\,\mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}), \mathrm{E}(X;\mathbf{Z}_Y)) \\ &= \mathrm{E}(f(\mathbf{Z}_Y))\,\mathrm{Var}(\mathrm{E}(X;\mathbf{Z})). \end{aligned} \tag{9}$$

If the normalised intrinsic contribution to the covariance is either zero or is negligible, it follows from (8) that

$$\frac{\mathrm{Cov}(X,Y)}{\mathrm{E}(X)\,\mathrm{E}(Y)} \approx \frac{\mathrm{Var}(\mathrm{E}(X;\mathbf{Z}))}{\mathrm{E}(X)^2} = \eta^2_{\mathrm{ext}}. \tag{10}$$

Thus, under certain conditions, measuring the covariance between two non-identically distributed and not-necessarily independent reporters can replace dual reporters.

The Pathway-Reporter Method

We show that for some reporters X and Y belonging to the same biochemical pathway, the covariance of X and Y continues to identify the extrinsic, and subsequently intrinsic, noise contributions to the total noise. The basis of the pathway-reporter method depends on the emergent covariances between the various species (for eg. nascent/mature mRNA and protein) in the gene expression pathway. Qualitatively, this effect can be seen in Figure 3, which compares simulated sample distributions of a simple four-stage model of gene transcription (refer to the model $\mathbf{M}_4$ below) in the case of moderate extrinsic noise to the case with no extrinsic noise. The plots are the bivariate distributions for nascent-mature, nascent-protein and mature-protein levels, respectively. This will be made more precise below, where we find that it is possible to extract quantitative information about the extrinsic noise distribution itself from this data.

Throughout this section, we assume that extrinsic noise sources act independently i.e., the environmental variables $Z_1, \ldots, Z_n$ of $\mathbf{Z}$ are mutually independent. Additionally, our modelling focuses only on a single gene copy, though the same analysis applies to

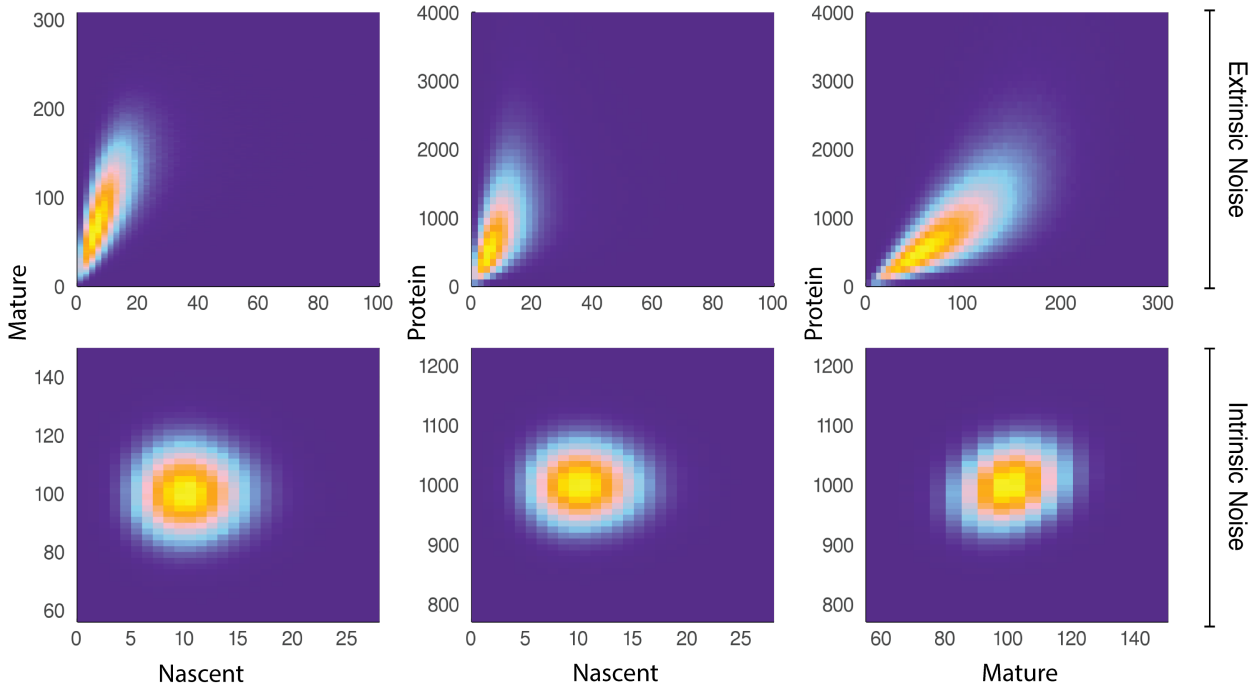
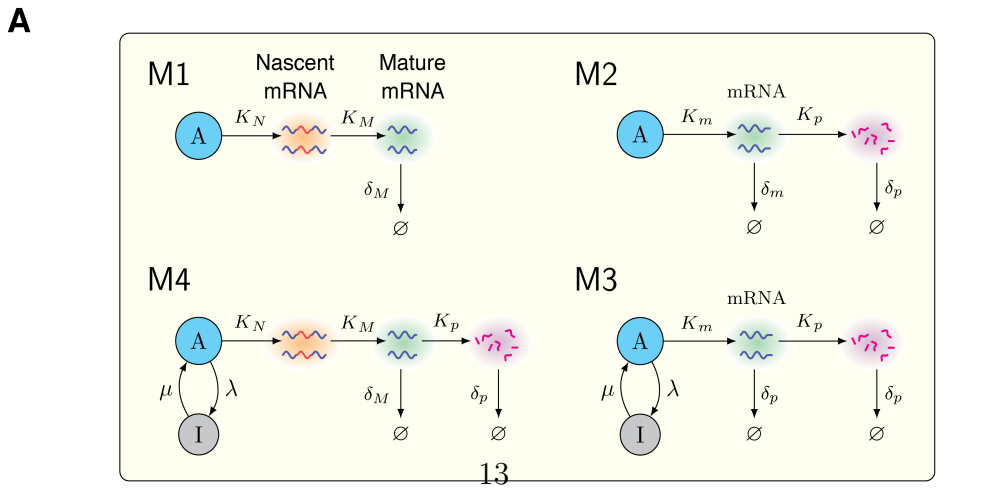


Figure 3: A comparison of joint distributions in the case of moderate extrinsic noise and no extrinsic noise. The plots are generated from a three-stage model of gene transcription, incorporating the production of nascent mRNA, mature mRNA and protein. Details of the model can be found in Fig. 4 (model $\mathbf{M}_4$) and the associated text. The top panel shows nascent-mature, nascent-protein and mature-protein joint distributions in the case of extrinsic noise, while the bottom panel displays the corresponding plots in the case of no extrinsic noise. Extrinsic noise produces a visibly more correlated joint distribution, which forms the basis of the pathway-reporter method.



multiple but indistinguishable gene copies; we refer to the supplementary material for more details.

Measuring Noise from a Constitutive Gene

We consider first the simplest case where the underlying process is constitutive. We begin with a stochastic model of mRNA maturation, which we denote by $\mathbf{M}_1$; Fig. 4A (top left) gives a schematic of the constitutive model. In this model, the gene continuously produces nascent mRNA according to a Poisson process at constant rate K_N, which are subsequently spliced into mature mRNA according at rate K_M. Degradation of mature mRNA occurs as a first-order Poisson process with rate δ_M. The model $\mathbf{M}_1$, together with its extensions, has been considered in a number of recent studies [14, 37, 42–44], and has a known solution for the stationary joint probability distribution [45] given by,

$$\tilde{p}(n, m; \theta) = \frac{e^{-\frac{K_n}{K_M}} (\frac{K_N}{K_M})^n}{n!} \frac{e^{-\frac{K_n}{\delta}} (\frac{K_N}{\delta_M})^m}{m!}. \tag{11}$$

Here n is the number of nascent mRNA, m is the number of mature mRNA, and $\theta = (K_N, K_M, \delta_M)$. We let X_N denote the number of nascent mRNA, X_M the number of mature mRNA produced from the same constitutive gene, and $\mathbf{Z} = (K_N, K_M, \delta_M)$. To simplify notation, we abbreviate the variables in $\mathbf{Z}_{X_N}$ as $\mathbf{Z}_N$, and similarly for $\mathbf{Z}_{X_M}$. It follows immediately from (11) that X_N and X_M are independent conditional on $\mathbf{Z}$, and so the intrinsic contribution to the covariance of X_N and X_M (the first term of (8)) is 0. It is also easy to see from (11) that $\mathrm{E}(X_N; \mathbf{Z}) = f(\mathbf{Z}_N)K_N$ and $\mathrm{E}(X_M; \mathbf{Z}) = g(\mathbf{Z}_M)K_N$, where $f(\mathbf{Z}_N) = \frac{1}{K_M}$ and $g(\mathbf{Z}_M) = \frac{1}{\delta_M}$. Since the extrinsic noise sources are assumed to act independently, it follows that the Noise Decomposition Principle (NDP) of the previous section holds. We then have that $\mathrm{Cov}(X_N, X_M) = \eta^2_{K_N}$, where $\eta^2_{K_N}$ is the total noise on the transcription rate parameter K_N. Thus, measuring $\mathrm{Cov}(X_N, X_M)$ can replace dual reporters to decompose the gene expression noise at the transcriptional level.

To support our mathematical results, we simulate the model $\mathbf{M}_1$ subject to parameter variation using the stochastic simulation algorithm (SSA). Table 2 compares the extrinsic noise contributions found from various simulations with the corresponding theoretical values. In each simulation, the degradation rate δ_m is fixed at 1, with the other parameters scales accordingly. The maturation rate K_M is sampled according to a Gamma$(8, 0.0125)$ distribution, which has coefficient of variation 0.125. We consider different noise distributions on K_N, producing a range of noise strengths. Our theory predicts that pathway-reporters will identify the total noise on K_N. Overall, we observe an excellent agreement between the results obtained by the pathway-reporter method, the dual-reporter method and the theoretical noise. There is consistently slightly more variation in the pathway-reporter results compared with the dual-reporter results.

To explore the pathway-reporter method more widely, we consider 60 different parameter combinations to produce a range of mean copy numbers consistent with those

Theory		Simulation	
η^2_{ext}	Noise (K_N)	PR	DR (Mat)
0.00	$K_N = 50$	0.00 ± 0.01	0.00 ± 0.00
0.10	$\text{Beta}_{133.\dot{3}}(6, 10.5)$	0.10 ± 0.01	0.10 ± 0.01
0.20	$\text{Gamma}(5, 10)$	0.20 ± 0.02	0.20 ± 0.01
0.50	$\text{Beta}_{300}(1.5, 7.5)$	0.49 ± 0.04	0.50 ± 0.03

Table 2: A comparison of the pathway-reporter method and the dual-reporter method for constitutive expression under the model $\mathbf{M}_1$. Here PR gives the results of the nascent and mature pathway reporters, while DR (Mat) gives the results of dual reporters calculated from the mature mRNA. We considered noise on both the transcription rate (K_N) and the maturation rate (K_M). The decay rate is fixed at one, with the other parameters scaled accordingly. In each case, the maturation rate K_M is varied according to a $\text{Gamma}(8, 1.25)$ distribution, which has coefficient of variation 0.125. The values given are the average of 100 simulations, each calculated from 500 copy number samples, and the errors are $\pm$ one standard deviation. Our theory predicts that pathway-reporters will identify the noise on the nascent transcription rate K_N (η^2_{ext}). The noise distribution parameters are chosen to produce an average nascent mRNA copy number of approximately 5 and an average mature mRNA copy number of approximately 50.

reported experimentally. We also consider different noise distributions taken from the scaled Beta distribution family in order to produce a range of noise strengths; see the supplementary dataset S1. The pathway-reporter method performa favourably compared to the dual-reporter method calculated from mature mRNA, and consistently outperforms the dual-reporter method on nascent mRNA. Next, we consider the simplest stochastic model of gene expression that includes both mRNA and protein dynamics: the well-known "two-stage model" of gene expression, which, together with its three-stage extension to include promoter switching has been widely studied [16, 17, 46–51]. We denote this model by $\mathbf{M}_2$; see Fig. 4A (top right) for a schematic of this model. In this model, mRNA are synthesized according a Poisson process at rate K_m, which are then later translated into protein at rate K_p. Degradation of mRNA and protein occur as first-order Poisson processes with rates δ_m and δ_p, respectively. If X_m denotes the number of mRNA, X_p the number of proteins produced from the same constitutive gene, and if $\mathbf{Z} = (K_m, K_p, \delta_m, \delta_p)$, then the stationary means and covariance are given by [46, 48]:

$$\text{E}(X_m; \mathbf{Z}) = \frac{K_m}{\delta_m}, \quad \text{E}(X_p; \mathbf{Z}) = \frac{K_p}{\delta_p}\frac{K_m}{\delta_m} \quad \text{and}$$

$$\text{Cov}(X_m, X_p; \mathbf{Z}) = \frac{K_m K_p}{\delta_m(\delta_m + \delta_p)}. \quad (12)$$

It is easily verified that $\mathrm{E}(X_p;\mathbf{Z}) = f(\mathbf{Z}_p)\,\mathrm{E}(X_m;\mathbf{Z})$, where $f(\mathbf{Z}_p) = \frac{K_p}{\delta_p}$. Thus, it follows from the NDP that the normalised contribution of $\mathrm{Cov}(X_m, X_p)$ contributed by $\mathbf{Z}$ will identify the extrinsic noise contribution to the total noise on X_m. Now, if we assume that δ_m is fixed across the cell-population, and all parameters are scaled so that $\delta_m = 1$, we have the following expression for the intrinsic contribution to the covariance of X_m and X_p (refer to the supplementary material for details),

$$\frac{\mathrm{E}(\mathrm{Cov}(X_m, X_p;\mathbf{Z}))}{\mathrm{E}(X_m)\,\mathrm{E}(X_p)} = \frac{\alpha}{\mathrm{E}(K_m)}, \quad \text{where } \alpha = \frac{\mathrm{E}(1/(\delta_p+1))}{\mathrm{E}(1/\delta_p)}. \tag{13}$$

Since mRNA tends to be less stable than protein, we have $\delta_p < 1$, and often $\delta_p \ll 1$ [52, 53]. So, we can expect $\alpha \ll 1$. Further, for many genes we can expect the number of mRNA per cell (K_m) to be in the order of tens, so $1/E(K_m) < 1$. It follows that $\mathrm{E}(\mathrm{Cov}(X_m, X_p;\mathbf{Z})) \ll 1$, so that $\mathrm{Cov}(X_m, X_p)$ will closely approximate the extrinsic noise at the transcriptional level.

We tested our theory using stochastic simulations of the model $\mathbf{M}_2$ subject to parameter variation. Table 3 gives a comparison of the results of the mRNA-protein reporters and dual reporters. In each case, we varied K_p according to a Gamma$(5, 0.4)$ distribution and δ_p according to a Gamma$(8, 0.125)$ distribution; the corresponding noise strengths are 0.20 and 0.125, respectively. We consider different noise distributions on K_m, which produce a range of noise strengths, and the noise distribution parameters are selected to produce a mean mRNA of approximately 50 and a mean number of approximately 1000 proteins in each simulation. As our theory predicts, the mRNA-protein reporters identify the extrinsic noise contribution to the total noise on X_m. Again, we can see from Table 3 that there is excellent agreement between the results of the pathway reporters and the dual reporters, with slightly more variation in the pathway-reporter results. A larger exploration of the parameter space reveals similar results; these are provided in the supplementary dataset S1. Thus, despite mRNA and protein numbers not being strictly independent, they can, for practical purposes, replace dual reporters to decompose the noise at the transcriptional level.

We note that both the pathway-reporter (nascent-mature or mature-protein) and dual-reporter methods show slower convergence when copy numbers are low. Pathway reporters usually show fractionally slower convergence and fractionally more variation than a dual reporter, as suggested by the standard deviations in Tables 2 and 3. A more detailed exploration of convergence is given in the supplementary material.

Measuring Noise from a Facultative (bursty) Gene

The most common mode of gene expression that is reported experimentally is burstiness [9,17,22,24,25,54], in which mRNA are produced in short bursts with periods of inactivity in between. One example is gene regulation via repression, which naturally leads to periods of gene inactivity. Here we consider a four-stage model of bursty gene expression,

Theory		Simulation	
η^2_{ext}	Noise (K_m)	PR	DR (Mat)
0.00	$K_m = 50$	0.00 ± 0.01	0.00 ± 0.00
0.10	$\text{Beta}_{133.\dot{3}}(6, 10.5)$	0.10 ± 0.01	0.10 ± 0.01
0.20	$\text{Gamma}(5, 10)$	0.20 ± 0.02	0.20 ± 0.01
0.50	$\text{Beta}_{300}(1.5, 7.5)$	0.51 ± 0.04	0.50 ± 0.03

Table 3: A comparison of the pathway-reporter method and the dual-reporter method for constitutive expression under the model $\mathbf{M}_2$. Here PR gives the results of the mRNA-protien pathway reporters, while DR (Mat) gives the results of dual reporters calculated from the mature mRNA. We considered noise on the transcription rate (K_m), the protein synthesis rate (K_p), and the protein decay rate (δ_p). The mRNA decay rate is fixed at one. In each case, we varied K_p according to a $\text{Gamma}(5, 0.4)$ distribution and δ_p according to a $\text{Gamma}(8, 0.125)$ distribution; the corresponding noise strengths are 0.20 and 0.125, respectively. We considered different noise distributions on K_m, which produce a range of noise strengths. The noise distribution parameters are selected to produce a mean mRNA of approximately 50 and a mean number of approximately 1000 proteins in each simulation. The values given are the average of 100 simulations, each calculated from 500 copy number samples, and the errors are $\pm$ one standard deviation. As our theory predicts, the mRNA-protein reporters identify the noise on the transcription rate parameter K_m (η^2_{ext}).

which incorporates both promoter switching and mRNA maturation; we denote this model by $\mathbf{M}_4$; see Fig. 4A (bottom left). This model has recently been considered in [37], where the marginal distributions are solved in some limiting cases. In this model, the gene switches probabilistically between an active state (A) and an inactive state (I), at rates λ (on-rate) and μ (off-rate), respectively. In the active state, nascent mRNA is synthesized according to a Poisson process at rate K_N, while in the inactive state transcription does not occur. Nascent mRNA is spliced into mature mRNA at rate K_M; this in turn is later translated into protein, with rate K_P. Degradation of mRNA and protein occur independently of the promoter state at rates δ_M and δ_P, respectively.

For this model we have three natural candidates for pathway reporters: (a) nascent and mature mRNA (b) mature mRNA and protein, and (c) nascent mRNA and protein reporters. Below we show that nascent mRNA–protein reporters yield consistently good estimates of the extrinsic noise contribution η^2_{ext} to the total gene expression noise, while nascent–mature and mature RNA–protein reporters are reliable in some restricted cases. We begin by showing that each of the reporter pairs (a), (b) and (c) satisfy the Noise Decomposition Principle. We then demonstrate computationally, that despite a lack of independence between these reporter pairs, the pathway-reporter method can still be used

to decompose the total gene expression noise at the transcriptional level. Throughout, we let X_N denote the number of nascent mRNA, we let X_M denote the number of mature mRNA, and let X_P denote the number of proteins produced from the same gene. We also let $\mathbf{Z} = \{\lambda, \mu, K_N, K_M, K_P, \delta_M, \delta_P\}$.

Folloqing [17] we assume that the transcription rate K_N is large relative to the other parameters. We further assume that the maturation rate K_M is large (i.e. $K_M > \delta_M$), which is supported by experiments [37]. Then, using the results of [17] and arguments similar to those in [37], it can be shown that the stationary averages for the nascent mRNA, mature mRNA and protein levels are given by,

$$\mathrm{E}(X_N;\mathbf{Z}) = \frac{K_N}{K_M}\frac{\lambda}{(\lambda+\mu)},\ \mathrm{E}(X_M;\mathbf{Z}) = \frac{K_N}{\delta_M}\frac{\lambda}{(\lambda+\mu)} \quad \text{and} \quad \mathrm{E}(X_P;\mathbf{Z}) = \frac{K_P}{\delta_P}\frac{K_N}{\delta_M}\frac{\lambda}{(\lambda+\mu)}, \quad (14)$$

respectively. The averages for the mature mRNA and protein levels are given in [17] for the "three-stage model" of gene expression; see Fig. 4A (bottom right). The supplementary material provides more detail on how these expressions can be obtained.

We consider first the nascent-mature pathway reporters, case (a). From (14), it is easily seen that $\mathrm{E}(X_N;\mathbf{Z}) = f(\mathbf{Z}_N)K_N\frac{\lambda}{(\lambda+\mu)}$ and $\mathrm{E}(X_M;\mathbf{Z}) = g(\mathbf{Z}_M)K_N\frac{\lambda}{(\lambda+\mu)}$, where $f(\mathbf{Z}_N) = \frac{1}{K_M}$ and $g(\mathbf{Z}_M) = \frac{1}{\delta_M}$. So the NDP holds, and the normalised covariance of $\mathrm{E}(X_N;\mathbf{Z})$ and $\mathrm{E}(X_M;\mathbf{Z})$ will identify the extrinsic noise on the transcription component $K_N\frac{\lambda}{(\lambda+\mu)}$. Consider now the mature-protein reporters, case (b). Again, we can see from (14) that $\mathrm{E}(X_M;\mathbf{Z}) = f(\mathbf{Z}_P)\,\mathrm{E}(X;\mathbf{Z})$, where $f(\mathbf{Z}_P) = \frac{K_P}{\delta_P}$. Thus the NDP holds, and so the normalised covariance of $\mathrm{E}(X_M;\mathbf{Z})$ and $\mathrm{E}(X_P;\mathbf{Z})$ will identify the total noise on $\mathrm{E}(X_M;\mathbf{Z})$ (the extrinsic noise on X_M). For the nascent-protein reporters, case (c), it is easy to see that $\mathrm{E}(X_N;\mathbf{Z}) = f(\mathbf{Z}_N)K_N\frac{\lambda}{(\lambda+\mu)}$, where $f(\mathbf{Z}_N) = \frac{1}{K_M}$, and $\mathrm{E}(X_P;\mathbf{Z}) = g(\mathbf{Z}_P)K_N\frac{\lambda}{(\lambda+\mu)}$, where $g(\mathbf{Z}_P) = \frac{K_P}{\delta_M\delta_P}$. Thus, again the NDP holds, and the normalised covariance of $\mathrm{E}(X_N;\mathbf{Z})$ and $\mathrm{E}(X_P;\mathbf{Z})$ will identify the noise on the transcriptional component $K_N\frac{\lambda}{(\lambda+\mu)}$.

In order for the pathway-reporter method to provide a close approximation to the extrinsic noise in cases (a), (b) and (c), we require that the normalised intrinsic contribution to the covariance is either zero or negligible. This condition will hold provided there is sufficiently small correlation between the reporter pairs. In the case of (prokaryotic) mRNA and protein, this lack of correlation has been been verified experimentally in E. Coli [55]. More generally, it is possible to provide an intuition about the conditions under which the lack of correlation might hold. The time series of copy numbers for each of nascent mRNA, mature mRNA and protein broadly follow each other, each with delay from its predecessor (Fig. 4B). Parameter values that reduce this delay will tend to increase correlation, and thereby increase the normalised intrinsic contribution to the covariance. The primary example of this effect is seen when δ_p approaches, or even exceeds δ_m (or for nascent-mature reporters, when δ_M approaches the maturation

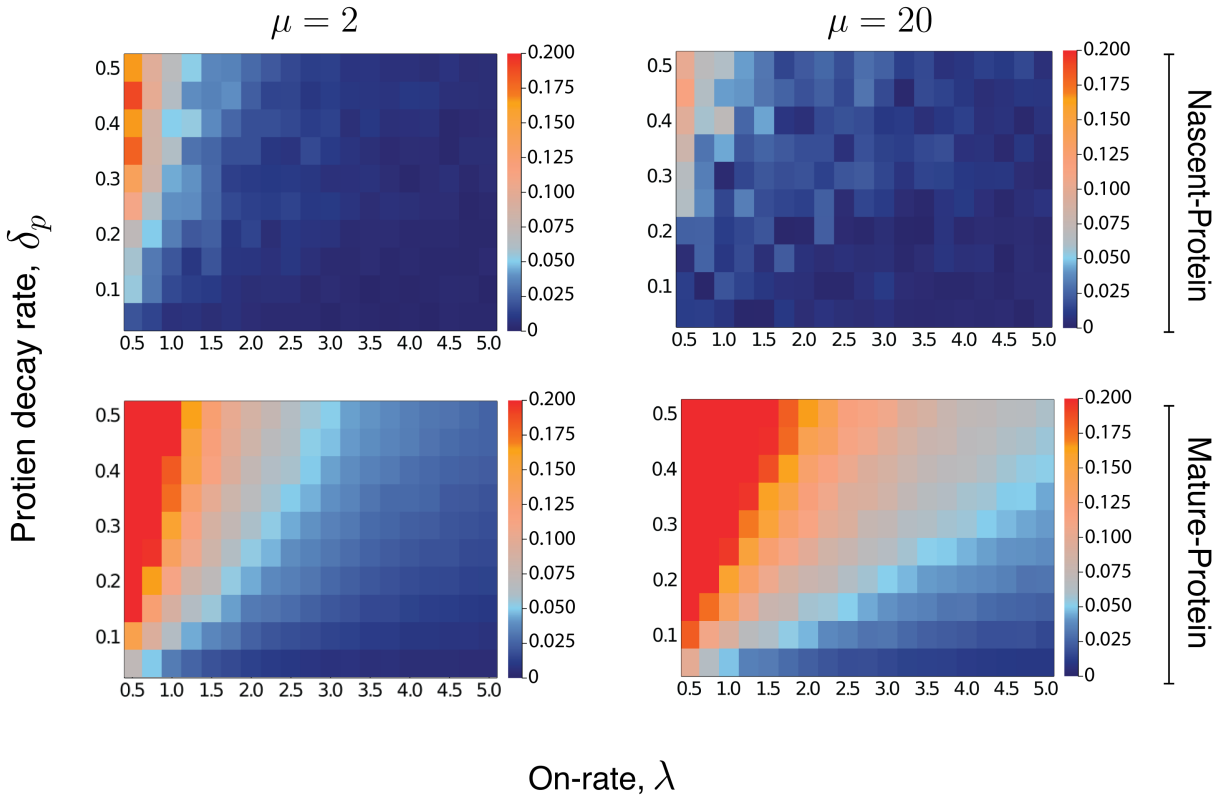


Figure 5: Heatmaps for the intrinsic contribution to the covariance, which estimates the level of overshoot in the pathway-reporter approach for the nascent-protein and mature-protein reporters; blue regions show an overshoot of less than $\simeq 0.05$. Here the intrinsic contribution is calculated using stochastic simulations of the model $\mathbf{M}_4$. In the left panels, the parameter $\mu = 2$ is fixed, while δ_p and the on-rate λ are varied between 0.01 and 1 and 0.5 and 4, respectively. In the right panels, the parameter $\mu = 20$ is fixed, while δ_p and the on-rate λ vary between 0.01 and 1, and 1 and 10, respectively. In all cases, the parameters are scaled so that $\delta_M = 1$. The maturation rate is fixed at 20, with the parameters K_N and K_P are chosen to produce a mean protein level of 1000, a mean nascent mRNA level of 5 and a mean mature mRNA level of 50.

rate). A further contributor to high correlation between mRNA and protein, is when the system undergoes long timescale changes. In this situation the copy numbers tend to drop to very low values for extended periods. The primary parameter influencing this type of behaviour is the active-rate λ, specifically, when λ tends to 0 (Fig. 4C). An illustrative example of this can be seen by considering a Telegraph system in the limit of slow switching, which produces a copy number distribution that converges to a scaled Poisson-Bernoulli compound distribution: even without any extrinsic noise, the pathway reporter

method will identify the η^2 value of the corresponding scaled Bernoulli distribution. An extensive computational exploration of the parameter space (supplementary dataset S2) supports our intuition, though the strength of the effect varies across the three different reporter pairs. This is further corroborated by the heatmaps shown in Fig. 5, which for two fixed values of μ and a broad spectrum of δ_p and λ values, give the intrinsic term η^2_{int} in (8), for fixed $\mathbf{Z}$. Thus, the heatmaps provide an estimate for the overshoot error in the pathway-reporter approach. Note that blue pixels represent an overshoot estimate of less than 0.05. We obtain similar results for the intermediate case of $\mu = 10$ (supplementary material, Fig. S4).

For nascent-protein reporters, the normalised intrinsic contribution to the covariance is satisfactorily small (less than 0.05) for (a) high values of δ_p in unison with (b) low values of λ (less than 1, though lower values are acceptable if δ_p is small). The assumption (a) that $\delta_P < \delta_M$ is known to be true for a large number of genes, and is justified by the difference in the mRNA and protein lifetimes. While there is of course variation across genes and organisms, values of $\delta_P \leq 0.5\delta_M$ and even $\delta_P \leq 0.2\delta_M$ seem reasonable for the majority of genes. In E. Coli [55] and yeast [56], for example, mRNA are typically degraded within a few minutes, while most proteins have lifetimes at the level of 10s of minutes to hours. For mammalian genes [53], it has been reported that the median mRNA decay rate δ_M is (approximately) five times larger than the median protein decay rate δ_P, determined from $4,200$ genes. Assumption (b) requires that the gene is sufficiently active. In a recent paper by Larsson et al. [25], the promoter-switching rates λ, μ, and the transcription rate K of the Telegraph model are estimated from single-cell data for over $7,000$ genes in mouse and human fibroblasts. Of those genes with mean mRNA levels greater than 5, we found that over 90% have a value for λ of at least 0.5, and over 65% have a value for λ greater than 1. In Cao and Grima [37], a comprehensive list of genes (ranging from yeast cells through to human cells) with experimentally-inferred parameter values are sourced from across the literature (see Table 1 in [37]). After scaling the parameter values of the 26 genes reported there, we find that around 88% have a value for λ of at least 0.5, and approximately 58% have a value for λ greater than 1. Thus, the heatmaps given in Fig. 5 (top panel) suggest that nascent-protein reporters will provide a satisfactory estimate of the extrinsic noise level for a substantial fraction of genes.

The mature nRNA–protein reporters are less reliable, with the requirement of slow protein decay and higher on-rate being more pronounced than for the nascent mRNA–protein reporters; this is evident from Fig. 5 (second panel). The performance of the nascent–mature reporter is of course independent of δ_p, but is only viable in the case of a large on-rate (see Fig. S5 of the supplementary material).

We tested our approach for each of the pathway reporter pairs (a), (b) and (c) against dual reporters using stochastic simulations. Table 4 shows the results from six simulations across a spectrum of behaviours from moderately slow switching, to fast switching as well as a range of levels of burstiness. For each of the parameters $\mu, \lambda, K_P, \delta_P$ we selected a scaled Beta$(5, 6)$ distribution, with squared coefficient of variation $\eta^2 = 0.1$; the scaling is

Mean					Simulation		
λ	μ	K_N	K_P	δ_P	PR (MP)	PR (NP)	DR (Mat)
0.5	1	150	2	0.1	0.46 ± 0.06	0.38 ± 0.07	0.32 ± 0.07
1	2	150	2	0.1	0.39 ± 0.05	0.34 ± 0.07	0.32 ± 0.05
1	20	1050	2	0.1	0.66 ± 0.15	0.52 ± 0.22	0.50 ± 0.15
2	2	100	6	0.3	0.35 ± 0.04	0.29 ± 0.05	0.27 ± 0.03
2	20	550	6	0.3	0.61 ± 0.09	0.47 ± 0.15	0.47 ± 0.09
10	10	100	6	0.3	0.29 ± 0.03	0.27 ± 0.04	0.27 ± 0.02

Table 4: A comparison of the pathway-reporter method and dual-reporter method for bursty expression. Here PR (NP) gives the results of the nascent and protein pathway reporters, PR (MP) gives the results of the mRNA and protein reporters, while DR (Mat) gives the results of the dual reporters calculated from the mature mRNA. We consider noise on all of the parameters except for δ_M and K_M; see discussion in main text. The values given are the average of 100 simulations, each calculated from 500 copy number samples, and the errors are $\pm$ one standard deviation. Our theory predicts that pathway-reporters will identify the noise at both the promoter level (λ, μ) and transcriptional level $(K_N,\ \delta_m)$; the total extrinsic noise in each case is given by η^2_{ext}. As before, the noise distribution parameters are chosen to produce an average nascent mRNA copy number of 5 and an average mature mRNA copy number of 50, and an average number of 1000 proteins.

chosen in each case to achieve a mean value equal to the parameter value given in Table 4. It is routinely verified that scaling these distributions does not change the value of η^2. The parameter K_N is given the noise distribution Beta$(3, 6)$, which has a slightly higher coefficient of variation $\eta^2 = 0.2$. In order to achieve direct benchmarking against the dual reporters, the parameter K_M is fixed at 10. This is because the nascent-protein pathway reporter estimates noise on the value of $K_N \frac{\lambda}{\lambda+\mu}$, while the mature mRNA dual-reporter measures noise on $\frac{K_N}{K_M} \frac{\lambda}{\lambda+\mu}$, and these coincide only when K_M is fixed. The mean values of K_N and K_P are chosen to achieve approximate average nascent mRNA levels, mature mRNA levels and protein levels at 5, 50 and 1000 respectively, given the chosen values of λ, μ, δ_P.

The results for the nascent mRNA–protein reporters, case (c), given in Table 4 show comparable performance to dual reporters, with only modest overshoot; even in the worst performing case of $\lambda = 0.5$, $\mu = 1$ the result of the pathway reporters is within one standard deviation, in a very tight distribution. The error heatmaps of Fig. 5 provide an accurate estimate of the overshoot in the nascent-protein results in Table 4. As an example, the first row is most closely matched by the heatmap at top left of Fig. 5, which at $\lambda = 0.5$ and $\delta_P = 0.1$ is suggestive of an error around the boundary between blue and

red (around 0.06). The same accuracy is obtained for the other rows. As predicted, the mature-protein reporters show significantly more overshoot, especially with the less active genes. Improved accuracy can again be obtained by subtracting the estimated overshoot given in the error heatmaps from the obtained value. Thus for example, the error heatmap for $\mu = 2$ (Fig. 5 lower left) gives an error approximately 0.07 for $\lambda = 1, \delta_P = 0.1$, which agrees very closely to the actual overshoot of 0.07 shown in the corresponding row of Table 4. An overshoot of approximately 0.06 is suggested by the heatmap for $\mu = 2$, when $\lambda = 2, \delta_P = 0.3$, which leads to a correction from 0.35 in Table 4 to a value of 0.29. This is quite consistent with the dual reporter benchmark of 0.27. As expected (based on Fig. S5 in the supplementary material), nascent-mature reporters do not perform well on bursty systems except for high λ and so the values are not included in Table 4; only in the case of $\lambda = \mu = 10$ does the result begin to approach the dual reporter value, returning 0.32 ± 0.03 (Fig. S5 in the supplementary material).

Discussion

Despite the proliferation of experimental methods for single-cell profiling, the ability to extract transcriptional dynamics from measured distributions of mRNA copy numbers is limited. In particular, the multiple factors that contribute to mRNA heterogeneity can confound the measured distribution, which hinders analysis. Theoretical innovations that allow us to quantify and help identifying the causes of these observable effects are therefore of great importance. In this work, we have demonstrated, through a series of mathematical results, that it is impossible to delineate the relative sources of heterogeneity from the measured transcript abundance distribution alone: multiple possible dynamics can give rise to the same distribution. Our approach involves establishing integral representations for distributions that are commonly encountered in single-cell data analysis, such as the negative binomial distribution and the stationary probability distribution of the Telegraph model. We show that a number of well-known representations can be obtained from our results. A particular feature of our non-identifiability results is that population heterogeneity inflates the apparent burstiness of the system. It is therefore necessary to collect further information, beyond measurements of the transcripts alone, in order to constrain the number of possible theoretical models of gene activity that could represent the system. In particular, additional work may be required to determine the true level of burstiness of the underlying system.

We have developed a theoretical framework for estimating levels of extrinsic noise, which can assist in resolving the non-identifiability problems. The dual reporter method of Swain et al. [4] already provides one such approach; but it is experimentally challenging to set up in many systems, and requires strictly identical and independent pairs of gene reporters. Our Noise Decomposition Principle directly generalises the theoretical underpinnings of the dual reporter method; here we have used it to identify a practical

approach —the pathway-reporter method— for obtaining the same noise decomposition. Our approach allows us to use measurements of two different species from the transcriptional pathway of a single gene copy instead of having to set up a more cumbersome dual reporter assay. The accuracy of the pathway-reporter method is provably identical for constitutive gene expression, and in the case of nascent-mature mRNA reporters, the measurements are readily obtainable from current single-cell data [13, 42, 57]. For bursty systems, the method in general provides only an approximation of the extrinsic noise. We are, however, able to demonstrate computationally, that one of the proposed pathway reporters provides a satisfactory estimate of the extrinsic noise for most genes. The other pathway reporters also provide viable estimates of the extrinsic noise in some cases.

Despite the generality of our theoretical contribution, our pathway-reporter approach has some caveats. In particular, the approach relies on the assumption that extrinsic noise sources act independently. Experimentally, however, these may be correlated. For example, it has been suggested [58, 59] that the transcription and translation rates in E. coli anticorrelate. Additional work is required to determine degree to which the independence of noise sources is a reasonable assumption.

Recent developments in single-cell profiling now allow simultaneous measurements of transcripts and proteins in thousands of single cells [60–62]. As discussed in [14], experimental improvements that would additionally allow measurements of nascent transcripts, coupled with theoretical developments such as those presented here, will allow for identification of noise sources on a genome-wide scale. Our work reveals that extrinsic noise distorts the multivariate copy number distribution of the different species in the gene expression pathway. We have exploited this to yield reliable estimates of noise strength, which we are confident will assist in setting better practices for model fitting and inference in the analysis of single-cell data. A more nuanced analysis of this multivariate distribution may offer even further insight into model and noise identifiability.

References

[1] V. Shahrezaei, P. S. Swain, The stochastic nature of biochemical networks. Curr. Opin. Biotechnol. 19, 369-374 (2008).

[2] S. Filippi, C. P. Barnes, P. D. W. Kirk, T. Kudo, K. Kunida, S. S. Mc Mahon, T. Tsuchiya, T. Wada, S. Kuroda, M. P. H. Stumpf, Robustness of MEK-ERK Dynamics and Origins of Cell-to-Cell Variability in MAPK Signaling. Cell reports 15, 2524–2535 (2016).

[3] T. Jetka, K. Nienałtowski, S. Filippi, M. P. H. Stumpf, M. Komorowski, An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling. Nature communications 9, 4591 (2018).

[4] P. S. Swain, M. B. Elowitz, E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. 99, 12795–12800 (2002).

[5] J. A. Goodrich, J. F. Kugel, Non-coding-RNA regulators of RNA polymerase ii transcription. Nat. Rev. Mol. Cell. Biol. 7, 612-616 (2006).

[6] J. Paulsson, Summing up the noise in gene networks. Nature 427, 415–418 (2004).

[7] J. M. Pedraza, J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression. Science 319, 339–343 (2008).

[8] M. Komorowski, M. J. Costa, D. A. Rand, M. P. H. Stumpf, Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. Proceedings of the National Academy of Sciences of the United States of America 108, 8645–8650 (2011).

[9] D. Jones, J. Elf, Bursting onto the scene? exploring stochastic mrna production in bacteria. Curr. Opin. in Microbiol. 45, 124–130 (2018).

[10] D. L. Jones, R. C. Brewster, R. Phillips, Promoter architecture dictates cell-to-cell variability in gene expression. Science 346, 1533–1536 (2014).

[11] C. Zopf, K. Quinn, J. Zeidman, N. Maheshri, Cell-cycle dependence of transcription dominates noise in gene expressions. PLoS Comput. Biol. 9, e1003161 (2013).

[12] L. Ham, R. D. Brackston, M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression. Phys. Rev. Lett. 124, 108101 (2020).

[13] S. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. Freire, T. Zwaka, I. Golding, Single-cell analysis of transcription kinetics across the cell cycle. Elife 10.7554/eLife.12175 (2016).

[14] G. Gorin, L. Pachter, Intrinsic and extrinsic noise are distinguishable in a synthesis – export – degradation model of mrna production. bioRxiv 10.1101/2020.09.25.312868 (2020).

[15] M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. Science 297, 1183–1186 (2002).

[16] J. M. Raser, E. K. O'Shea, Noise in gene expression: origins, consequences, and control. Science 304, 1811–1814 (2004).

[17] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4, e309 (2006).

[18] T. Quarton, T. Kang, V. Papakis, K. Nguyen, C. Nowak, Y. Li, L. Bleris, Uncoupling gene expression noise along the central dogma using genome engineered human cell lines. Nucleic Acids Research 10.1093/nar/gkaa668 (2020). Gkaa668.

[19] C. H. Naik, K. Hari, D. Chandel, S. Mandal, M. K. Jolly, S. Gayen, Semi-coordinated allelic-bursting shape dynamic random monoallelic expression in pre-gastrulation embryos. bioRxiv (2021).

[20] A. Singh, Transient changes in intercellular protein variability identify sources of noise in gene expression. Biophys J. 107, 2214-2220 (2014).

[21] M. S. H. Ko, A stochastic model for gene induction. J. Theor. Biol. 153, 181–194 (1991).

[22] K. B. Halpern, S. Tanami, S. Landen, M. Chapal, L. Szla, A. Hutzler, A. Nizhberg, Bursty gene expression in the intact mammalian liver. Mol. Cell 58, 147–156 (2015).

[23] L.-H. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, I. Golding, General properties of transcriptional time series in escherichia coli. Nature Genetics 43, 554–560 (2011).

[24] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics. Science 332, 472–474 (2011).

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

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

[27] M. Abramowitz, I. A. Stegun, Handbook Of Mathematical Functions With Formulas, Graphs, And Mathematical Tables (U.S. Govt. Print. Off., 1965).

[28] L. Ham, R. D. Schnoerr, D.mand Brackston, M. P. H. Stumpf, Exactly solvable models of stochastic gene expression. J. Chem. Phys. 152, 144106 (2020).

[29] O. Lenive, P. D. W. Kirk, M. P. H. Stumpf, Inferring extrinsic noise from single-cell gene expression data using approximate bayesian computation. BMC Syst. Biol. 10, 81 (2016).

[30] Villaverde, Alejandro Fernández, Banga, Julio R, Dynamical compensation and structural identifiability of biological models: Analysis, implications, and reconciliation. PLoS computational biology 13, e1005878 (2017).

[31] P. J. Ingram, M. P. H. Stumpf, J. Stark, Nonidentifiability of the source of intrinsic noise in gene expression from single-burst data. PLoS Comput. Biol. 4, e1000192 (2008).

[32] M. Greenwood, G. U. Yule, An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society 83, 255-279 (1920).

[33] I.-B. Srividya, F. Hayot, C. Jayaprakash, Transcriptional pulsing and consequent stochasticity in gene expression. Phys. Rev. E 79, 031911 (2009).

[34] W. Feller, On a general class of contagious distributions,. Annals of Mathematical Statistics 14, 389-400 (1943).

[35] L. Ham, M. Jackson, R. D. Brackston, Decoupling extrinsic noise distributions from single-cell gene expression data (2020). Unpublished manuscript.

[36] C. Jia, R. Grima, Small protein number effects in stochastic models of autoregulated bursty gene expression. The Journal of Chemical Physics 152, 084115 (2020).

[37] Z. Cao, R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proceedings of the National Academy of Sciences 117, 4682–4692 (2020).

[38] C. H. L. Beentjes, R. Perez-Carrasco, R. Grima, Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. Phys. Rev. E 101, 032403 (2020).

[39] A. Hilfinger, J. Paulsson, Separating intrinsic fluctuations in dynamic biological systems. PNAS 108, 12167-12172 (2011).

[40] A. Hilfinger, M. Chen, J. Paulsson, Using temporal correlations and full distributions to separate intrinsic and extrinsic fluctuations in biological systems. Phys Rev Lett. 109, 248104 (2012).

[41] S. Ross, A first course in probability 9th edition (Pearson, 2014).

[42] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. Nature 560, 494–498 (2018).

[43] G. Gorin, L. Pachter, Special function methods for bursty models of transcription. Phys. Rev. E 102, 022409 (2020).

[44] V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing rna velocity to transient cell states through dynamical modeling. Nature Biotechnology 10.1038/s41587-020-0591-3 (2020).

[45] T. Jahnke, W. Huisinga, Solving the chemical master equation for monomolecular reaction systems analytically. J. Math. Biol. 54, 1–26 (2007).

[46] M. Thattai, A. van Oudenaarden, Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences 98, 8614–8619 (2001).

[47] N. Friedman, L. Cai, X. S. Xie, Linking stochastic dynamics to population distribution: An analytical framework of gene expression. Phys. Rev. Lett. 97, 168302 (2006).

[48] A. Singh, J. P. Hespanha, Stochastic analysis of gene regulatory networks using moment closure. 2007 American Control Conference 10.1109/ACC.2007.4282604, 1299-1304 (2007).

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

[50] P. Bokes, J. R. King, A. T. A. Wood, M. Loose, Exact and approximate distributions of protein and mrna levels in the low-copy regime of gene expression. J. Math. Biol. 64, 829–854 (2012).

[51] N. Molina, D. M. Suter, R. Cannavo, B. Zoller, I. Gotic, F. Naef, Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proceedings of the National Academy of Sciences 110, 20563–20568 (2013).

[52] J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, S. N. Cohen, Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. Proceedings of the National Academy of Sciences 99, 9697–9702 (2002).

[53] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control. Nature 473, 337–342 (2011).

[54] I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, Real-time kinetics of gene activity in individual bacteria. Cell 123, 1025–1036 (2005).

[55] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, Quantifying E. Coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538 (2010).

[56] A. Belle, A. Tanay, L. Bitincka, R. Shamir, E. K. O'Shea, Quantification of protein half-lives in the budding yeast proteome. PNAS 103, 13004–13009 (2006).

[57] S. Shah, T. Y., W. Zhou, E. Lubeck, J. Yun, C. Eng, N. Koulena, C. Cronin, C. Karp, E. Liaw, M. Amin, L. Cai, Dynamics and spatial genomics of the nascent transcriptome by intron seqfish. Cell. 174, 363-376 (2018).

[58] A. Hilfinger, T. Norman, J. Paulsson, Fluctuations to identify kinetic mechanisms in sparsely characterized systems. Cell Syst. 2, 251–259 (2016).

[59] J. Cole, Z. Luthey-Schulten, A careful accounting of extrinsic noise in protien expression reveals correlations among its sources. Phys. Rev. E 95, 062418 (2017).

[60] M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells. Nature Methods 14, 865–868 (2017).

[61] V. Peterson, K. Zhang, N. Kumar, J. Wong, L. Li, D. Wilson, R. Moore, T. McClanahan, S. Sadekova, J. Klappenbach, Multiplexed quantification of proteins and transcripts in single cells. Nat. Biotechnol. 35, 936–939 (2017).

[62] J. Reimegård, M. Danielsson, M. Tarbier, J. Schuster, S. Baskaran, S. Panagiotou, N. Dahl, M. R. Friedländer, C. J. Gallant, Combined mrna and protein single cell analysis in a dynamic cellular system using sparc. bioRxiv 10.1101/749473 (2019).

[63] F. W. Olver, D. W. Lozier, R. F. Boisvert, C. W. Clark, NIST Handbook of Mathematical Functions (Cambridge University Press, 2010).

[64] N. Saad, R. L. Hall, Integrals containing confluent hypergeometric functions with applications to perturbed singular potentials. Journal of Physics A: Mathematical and General 36, 7771–7788 (2003).

[65] A. Singh, P. Bokes, Consequences of mRNA transport on stochastic variability in protein levels. Biophys J. 103, 1087-1096 (2012).

Acknowledgments

We gratefully acknowledge Rowan D. Brackston helpful discussions in the early stages of this research. We also wish to thank Arjun Raj for providing valuable feedback on this work. L.H. and M.P.H.S. were supported by the University of Melbourne DVCR fund.

Data Availability

Simulations of the models used in the paper are performed using Gillespie's Stochastic Simulation Algorithm (SSA) implemented in Julia. The simulation code is available in the GitHub repository https://github.com/leham/PathwayReporters. The data used in the paper are provided in the supplementary datasets.

Author Contributions

L.H. and M.J. conceptualised the research, with support from M.P.H.S. The theoretical framework was developed by L.H. Formal analysis was conducted by L.H. and M.J. Original draft was written by L.H., with editing and review contributed by M.J. and M.P.H.S. All authors provided critical feedback and helped shape the research.

Competing Interests

The authors declare no competing interests.

Supplementary Text

Derivations of the Non-Identifiability Results

In the main text we provide a number of examples of when the compound distribution does not have a unique representation. We here provide the full details of the derivations of these results.

Bursty Expression: Telegraph Representation

Consider first a Telegraph distribution $\tilde{p}_T(n;\lambda,\mu',K')$ and the probability density function of the scaled beta distribution $\text{Beta}_{K'}(\lambda+\mu,\mu'-\mu)$, given by

$$f_{K'}(t;\lambda+\mu,\mu'-\mu)=\frac{\Gamma(\lambda+\mu')}{\Gamma(\lambda+\mu)\Gamma(\mu'-\mu)}K'^{1-\lambda-\mu'}t^{\lambda+\mu-1}(K'-t)^{\mu'-\mu-1}.$$

Note that this distribution has support $[0,K']$. In the main text we claim that the Telegraph distribution $\tilde{p}_T(n;\lambda,\mu',K')$ can be obtained from compounding a Telegraph distribution by a scaled beta distribution with pdf $f(t;\lambda+\mu,\mu'-\mu)$. In other words, that $\tilde{p}_T(n;\lambda,\mu',K')$ can be written as:

$$\tilde{p}_T(n;\lambda,\mu',K')=\int_0^{K'}\tilde{p}_T(n;\lambda,\mu,t)f_{K'}(t;\lambda+\mu,\mu'-\mu)\,dt. \tag{15}$$

This is Eq. (3) in the main text. Starting from the right hand side of (15), we have

$$\begin{aligned}\int_0^{K'}&\tilde{p}_T(n;\lambda,\mu,t)f_{K'}(t;\lambda+\mu,\mu'-\mu)\,dt\\ &=\frac{1}{n!}\frac{\Gamma(\lambda+\mu')}{\Gamma(\lambda+\mu)\Gamma(\mu'-\mu)}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{\Gamma(\lambda+\mu)}{\Gamma(\lambda+\mu+n)}\\ &\int_0^{K'}{}_1F_1(\lambda+n,\lambda+\mu+n,-t)(K')^{1-\lambda-\mu'}t^{\lambda+\mu+n-1}(K'-t)^{\mu'-\mu-1}\,dt.\end{aligned} \tag{16}$$

Substituting $t=K'T$ and simplifying, the right hand side of (16) becomes

$$\begin{aligned}&\frac{1}{n!}\frac{\Gamma(\lambda+\mu')}{\Gamma(\mu'-\mu)}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{(K')^n}{\Gamma(\lambda+\mu+n)}\\ &\qquad\int_0^1{}_1F_1(\lambda+n,\lambda+\mu+n,-KT)T^{\lambda+\mu+n-1}(1-T)^{\mu'-\mu-1}\,dT.\end{aligned}$$

Now, using the integral representation [63][13.4.2] of the confluent hypergeometric function (with $a = \lambda + n$, $b = \lambda + \mu' + 1$ and $c = \lambda + \mu + 1$), this becomes

$$\frac{1}{n!}\frac{\Gamma(\lambda+\mu')}{\Gamma(\mu'-\mu)}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{(K')^n}{\Gamma(\lambda+\mu+n)}\frac{\Gamma(\lambda+\mu+n)\Gamma(\mu'-\mu)}{\Gamma(\mu'+\lambda+n)}{}_1F_1(\lambda+n,\lambda+\mu'+n,-K')$$
$$= \frac{\Gamma(\lambda+\mu')}{\Gamma(\lambda+\mu'+n)}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{(K')^n}{n!}{}_1F_1(\lambda+n,\lambda+\mu'+n,-K') = \tilde{p}_T(n;\lambda,\mu',K'),$$

which is the left hand side of (15). Hence, we have that (15) holds.

Instantaneously Bursty Expression: Negative Binomial Representations

We consider first the representation given in Eq. (4) in the main text. Recall that we let $\tilde{p}_{\mathrm{NB}}(n;r,\beta)$ denote the probability mass function of a NegBin$(r, \frac{\beta}{\beta+1})$ distribution, where $\beta > 0$. In the main text we claim that for any negative binomial distribution of the form NegBin$(\lambda, \frac{\beta}{\beta+1})$ (where $\beta > 0$) we have,

$$\tilde{p}_{\mathrm{NB}}(n;\lambda,\beta) = \int_0^\infty \tilde{p}_T(n;\lambda,\mu,t)f(t;\lambda+\mu,\beta)\,dt, \tag{17}$$

where $f(t;\lambda+\mu,\beta)$ is the probability mass function of a Gamma$(\lambda+\mu,\beta)$ distribution. Beginning with the right hand side of (17) we have,

$$\int_0^\infty \tilde{p}_T(n;\lambda,\mu,t)f(t;\lambda+\mu,\beta)\,dt$$
$$= \frac{1}{n!}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{\Gamma(\lambda+\mu)}{\Gamma(\lambda+\mu+n)}\frac{\beta^{(\lambda+\mu)}}{\Gamma(\lambda+\mu)}\int_0^\infty {}_1F_1(\lambda+n,\lambda+\mu+n,-t)t^{\lambda+\mu+n-1}e^{-\beta t}\,dt. \tag{18}$$

We now apply the following identity given in [64][Appendix 1] with $a = \lambda+n$, $b = \lambda+\mu+n$, $k = -1$, $d = \lambda+\mu+n$ and $h = \beta$.

$$\int_0^\infty {}_1F_1(a,b,kt)t^{d-1}e^{-ht} = \frac{\Gamma(d)}{h^d}(1-\frac{k}{h})^{-a}. \tag{19}$$

So the left hand side of (18) becomes

$$\frac{1}{n!}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{\Gamma(\lambda+\mu)}{\Gamma(\lambda+\mu+n)}\frac{\beta^{(\lambda+\mu)}}{\Gamma(\lambda+\mu)}\frac{\Gamma(\lambda+\mu+n)}{\beta^{\lambda+\mu+n}}(1+\frac{1}{\beta})^{-(\lambda+n)}$$
$$= \frac{1}{n!}\frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}\frac{\beta^\lambda}{(\beta+1)^{\lambda+n}} = \tilde{p}_{\mathrm{NB}}(n;\lambda,\beta),$$

which is the right hand side of (17). Hence, we have that (17) holds. We now consider the representation given in Eq. (5) of the main text. Here we claim that any negative binomial distribution of the form $\text{NegBin}(\lambda', \frac{b}{1+b})$ (where $b > 0$) can be written as,

$$\tilde{p}_{\text{NB}}(n; \lambda', b) = \int_{\frac{1}{b}}^{\infty} \tilde{p}_{\text{NB}}(n; \lambda, \theta) f_b(b\theta - 1; \lambda - \lambda', \lambda')\, d\theta, \tag{20}$$

where $f_b(b\theta - 1; \lambda - \lambda', \lambda')$ is the probability mass function of a scaled beta prime $\text{BetaPrime}_b(\lambda - \lambda', \lambda)$ distribution, where $b > 0$ and $\lambda > \lambda'$. Starting from the right hand side of (20), we have

$$\begin{aligned}
&\int_{\frac{1}{b}}^{\infty} \tilde{p}_{\text{NB}}(n; \lambda, \theta) f_b(b\theta - 1; \lambda - \lambda', \lambda')\, d\theta \\
&\quad = \int_{\frac{1}{b}}^{\infty} \frac{\Gamma(\lambda + n)}{\Gamma(n+1)\Gamma(\lambda)} \left(1 - \frac{\theta}{\theta + 1}\right)^n \left(\frac{\theta}{\theta + 1}\right)^{\lambda} \frac{b\Gamma(\lambda)}{\Gamma(\lambda - \theta)\Gamma(\theta)} \frac{(b\theta - 1)^{\lambda - \lambda' - 1}}{(b\theta)^{\lambda}} \\
&\qquad = \frac{b^{1-\lambda}\Gamma(\lambda + n)}{\Gamma(n+1)\Gamma(\lambda')\Gamma(\lambda - \lambda')} \int_{\frac{1}{b}}^{\infty} \frac{(b\theta - 1)^{\lambda - \lambda' - 1}}{(1 + \theta)^{\lambda + n}}.
\end{aligned}$$

Now substituting $H = b\theta - 1$ and simplifying, we obtain

$$\frac{b^n \Gamma(\lambda + n)}{\Gamma(n+1)\Gamma(\lambda')\Gamma(\lambda - \lambda')} \int_0^{\infty} \frac{H^{\lambda - \lambda' - 1}}{(H + b + 1)^{\lambda + n}}\, dH, \tag{21}$$

and letting $y = b + 1$ and simplifying we have that

$$\int_0^{\infty} \frac{H^{\lambda - \lambda' - 1}}{(H + b + 1)^{\lambda + n}}\, dH = \frac{1}{(b+1)^{\lambda' + n}} \int_0^{\infty} \frac{y^{\lambda - \lambda' - 1}}{(y + 1)^{\lambda + n}}\, dy \tag{22}$$

$$= \frac{1}{(b+1)^{\lambda' + n}} \frac{\Gamma(\lambda - \lambda')\Gamma(\lambda' + n)}{\Gamma(\lambda + n)}. \tag{23}$$

Here we used the fact that the integral in the variable y is the probability density function of a $\text{BetaPrime}(\lambda - \lambda', \lambda' + n)$ distribution. Thus, Eq. 21 simplifies to

$$\frac{\Gamma(\lambda' + n)}{\Gamma(n+1)\Gamma(\lambda')} \frac{b^n}{(b+1)^{\lambda' + n}} = \tilde{p}_{\text{NB}}(\lambda', b), \tag{24}$$

which is the right hand side of (20). Thus, we have that (20) holds.

Proof of the Noise Decomposition Principle

Here we provide the proof of the Noise Decomposition Principle given in the main text. For convenience we restate the principle below.

The Noise Decomposition Principle (NDP)

Assume that there are measurable functions f, g and h such that $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$ split across common variables by way of $\mathrm{E}(X;\mathbf{Z}) = f(\mathbf{Z}_X)h(\mathbf{Z}')$ and $\mathrm{E}(Y;\mathbf{Z}) = g(\mathbf{Z}_Y)h(\mathbf{Z}')$. Then provided that the variables $Z_1,\ldots,Z_m$ are mutually independent, the normalised covariance of $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$ will identify the total noise on $h(\mathbf{Z}')$ (i.e., $\eta^2_{h(Z')}$).

Consider first the covariance of $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$. We have

$$\begin{aligned}\mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}),\mathrm{E}(Y;\mathbf{Z})) &= \mathrm{Cov}(f(\mathbf{Z}_X)h(\mathbf{Z}'),g(\mathbf{Z}_Y)h(\mathbf{Z}'))\\ &= \mathrm{E}(f(\mathbf{Z_X})g(\mathbf{Z_Y})(h(\mathbf{Z}'))^2) - \mathrm{E}(f(\mathbf{Z_X})h(\mathbf{Z}'))\,\mathrm{E}(g(\mathbf{Z_Y})h(\mathbf{Z}'))\\ &= \mathrm{E}(f(\mathbf{Z_X}))\,\mathrm{E}(g(\mathbf{Z_Y}))\left[\mathrm{E}((h(\mathbf{Z}'))^2) - (\mathrm{E}(h(\mathbf{Z}')))^2\right]. \qquad (25)\end{aligned}$$

Note that here we used the fact that the variables in $\mathbf{Z} = (Z_1,\ldots,Z_n)$ are mutually independent. Now using the fact that $\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X;\mathbf{Z}))$ and $\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(Y;\mathbf{Z}))$ (the Law of Total Expectation), and then normalising we obtain

$$\begin{aligned}\frac{\mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}),\mathrm{E}(Y;\mathbf{Z}))}{\mathrm{E}(X)\,\mathrm{E}(Y)} &= \frac{\mathrm{E}(f(\mathbf{Z_X}))\,\mathrm{E}(g(\mathbf{Z_Y}))\,[\mathrm{E}((h(\mathbf{Z}'))^2) - (\mathrm{E}(h(\mathbf{Z}')))^2]}{\mathrm{E}(f(\mathbf{Z_X}))\,\mathrm{E}(g(\mathbf{Z_Y}))(\mathrm{E}((h(\mathbf{Z}')))^2}\\ &= \frac{\mathrm{Var}(h(\mathbf{Z}'))}{(\mathrm{E}(h(\mathbf{Z}')))^2}\\ &= \eta^2_{h(\mathbf{Z}')}. \qquad (26)\end{aligned}$$

Hence, under certain conditions the normalised covariance of $\mathrm{E}(X;\mathbf{Z})$ and $\mathrm{E}(Y;\mathbf{Z})$ will identify the total noise on $h(\mathbf{Z}')$.

Pathway Reporters

Justification for Simulating Only One Copy of the Gene

Our simulations and theory have been based over reporters from a single gene copy, whereas in practice there may be multiple copies of the gene that contribute to the overall mRNA. If there are mechanisms in place to distinguish the mRNA or protein of one gene copy from another, then the theory and analysis we have developed in main paper holds without change. In the case where it is not possible to distinguish mRNA and protein (respectively) from the multiple gene copies, then we now observe that the general theory continues to hold, provided there is independence between the gene copies; this assumption has been verified experimentally [13]. Here we demonstrate this in the case of two gene copies, though the general case for more than two genes is essentially identical but is notationally cumbersome. We are considering a situation where the variable X in the Noise Decomposition Principle is the sum of two independent variables X_1, X_2 and Y is the sum of two independent variables Y_1, Y_2. We assume common dependence on the

environmental variables $\mathbf{Z}$ so that $E(X_1;\mathbf{Z}) = E(X_2;\mathbf{Z})$, $E(Y_1;\mathbf{Z}) = E(Y_2;\mathbf{Z})$. Using these equalities and the independence of X_1, X_2 and Y_1, Y_2 in $X = X_1 + X_2$, $Y = Y_1 + Y_2$, we find the numerator of

$$\frac{\mathrm{Cov}(\mathrm{E}(X;\mathbf{Z}), \mathrm{E}(Y;\mathbf{Z}))}{\mathrm{E}(X)\,\mathrm{E}(Y)}$$

is simply $4\,\mathrm{Cov}(\mathrm{E}(X_1;\mathbf{Z}), \mathrm{E}(Y_1;\mathbf{Z}))$, while the denominator is $4\,\mathrm{E}(X_1)\,\mathrm{E}(Y_1)$. Thus the noise decomposition coincides with that for the single copy X_1, Y_1. Further work may be required to consider systems where there is independence between, or there is significant deviations in the gene copies.

Upper Bound for the Intrinsic Contribution to the Covariance: Constitutive Expression

In the main text, we claim that the error in the pathway-reporter approach in the case of mRNA-protein reporters is negligible (i.e. the error is $\ll 1$). Our argument relies on the expression given in Eq. 13 in the main text. We here provide full details of the derivation of this expression. First let X_m and X_p be the number of mRNA and protein produced from the same constitutive gene modelled by the "two-stage" model, $\mathbf{M}_2$ (see Figure 3A (top right) of the main text). Also, let $\mathbf{Z} = \{K_m, \delta_m, K_p, \delta_p\}$.

We restate here the expression for the intrinsic contribution to the covariance of X_m and X_p given in Eq. 13 of the main text.

$$\frac{\mathrm{E}(\mathrm{Cov}(X_m, X_p;\mathbf{Z}))}{\mathrm{E}(X_m)\,\mathrm{E}(X_p)} = \frac{\alpha}{\mathrm{E}(K_m)}, \quad \text{where } \alpha = \frac{\mathrm{E}(1/(\delta_p+1))}{\mathrm{E}(1/\delta_p)}. \tag{27}$$

We require the following expressions for the stationary mean mRNA level and protein level of the two-stage model [46, 48].

$$\mathrm{E}(X_m;\mathbf{Z}) = \frac{K_m}{\delta_m}, \quad \mathrm{E}(X_p;\mathbf{Z}) = \frac{K_p}{\delta_p}\frac{K_m}{\delta_m} \quad \text{and} \quad \mathrm{Cov}(X_m, X_p;\mathbf{Z}) = \frac{K_m K_p}{\delta_m(\delta_m+\delta_p)}. \tag{28}$$

Assuming that δ_m is fixed across the cell-population, and all parameters are scaled so that $\delta_m = 1$, it follows from (28), that

$$\mathrm{E}(\mathrm{Cov}(X_m, X_p;\mathbf{Z})) = \mathrm{E}\left(\frac{1}{\delta_p+1}\right)\mathrm{E}(K_m)\,\mathrm{E}(K_p). \tag{29}$$

Using the Total Law of Expectation, we also have,

$$\mathrm{E}(X_m) = \mathrm{E}(K_m) \quad \text{and} \quad \mathrm{E}(X_p) = \mathrm{E}\left(\frac{1}{\delta_p}\right)\mathrm{E}(K_m)\,\mathrm{E}(K_p). \tag{30}$$

Thus, it follows that

$$\frac{\mathrm{E}(\mathrm{Cov}(X_m, X_p;\mathbf{Z}))}{\mathrm{E}(X_m)\,\mathrm{E}(X_p)} = \frac{\alpha}{\mathrm{E}(K_m)}, \quad \text{where } \alpha = \frac{\mathrm{E}(1/(\delta_p+1))}{\mathrm{E}(1/\delta_p)}. \tag{31}$$

Determination of the Marginal Means in the Model M4

In order to establish the Noise Decomposition principle in the case of bursty gene expression, we rely on expressions for the marginal stationary means of the model $\mathbf{M}_4$; see Fig. 3A (bottom left) and the associated caption for more details of this model. To derive the marginal means for nascent and mature mRNA and for protein (Eq. 14 of the main text), we first observe that the nascent mRNA population may treated identically to that of mRNA in general (that is, no distinction between nascent and mature), as in [26], except that mRNA decay is replaced by the sum of decay and maturation. As in the work of Cao and Grima [37], the assumption of fast maturation allows us to ignore decay completely in the nascent phase, so that the marginal distribution is identical to that of [26], except with decay replaced by maturation. This leads to a marginal nascent mRNA mean of

$$\mathrm{E}(X_N;\mathbf{Z}) = \frac{K_N}{K_M}\frac{\lambda}{(\lambda+\mu)}.$$

The marginal means for mature mRNA and protein are derived in [17] under the assumption that the transcription rate parameter K_N is large relative to the other parameters. The expressions are given by

$$\mathrm{E}(X_M;\mathbf{Z}) = \frac{K_N}{\delta_M}\frac{\lambda}{(\lambda+\mu)} \quad \text{and} \quad \mathrm{E}(X_P;\mathbf{Z}) = \frac{K_P}{\delta_P}\frac{K_N}{\delta_M}\frac{\lambda}{(\lambda+\mu)}.$$

Formally, the marginal means in [17] are for the three-stage model $\mathbf{M}_3$, which ignores the downstream processing of mRNA, such as splicing. The assumption of fast maturation however, justifies the treatment of the nascent phase of mRNA as an ephemeral step within the Poissonian modelling of mRNA transcription.

There are a number of possibly compounding assumptions on the parameters here, but simulations show that there is a lot of tolerance, with even only moderate maturation and transcription still returning sample means consistent with the formulas.

Measuring Noise from an Instantaneously Bursty Gene

Here we use the simple stochastic model of [65] that includes both instantaneous transcriptional bursting and mRNA maturation. This model can be obtained as a limiting case of the model $\mathbf{M}_4$ (Fig. 3A (top left) in the main text), where the off-rate μ has tended toward infinity, while the on-rate λ remains fixed. Under this condition, transcription is rare enough to be considered instantaneous, leading to "bursts" of transcriptional activity. The intensity of the bursts M is known to be geometrically distributed with mean burst size $B = \frac{K_N}{\mu}$.

If we let X_N and X_M denote the number of nascent and mature mRNA, respectively, then according to [65], the steady-state marginal means and covariance are given by

$$\mathrm{E}(X_N;\mathbf{Z}) = \frac{\lambda B}{K_M}, \quad \mathrm{E}(X_M;\mathbf{Z}) = \frac{\lambda B}{\delta_M} \quad \text{and} \quad \mathrm{Cov}(X_N, X_M) = \frac{\lambda B^2}{K_M+\delta_M}. \tag{32}$$

It is easily seen from (32) that $\mathrm{E}(X_N;\mathbf{Z}) = f(\mathbf{Z}_N)B\lambda$ and $\mathrm{E}(X_M;\mathbf{Z}) = g(\mathbf{Z}_M)B\lambda$, where $f(\mathbf{Z}_N) = \frac{1}{K_M}$ and $g(\mathbf{Z}_M) = \frac{1}{\delta_M}$. Thus, the Noise Decomposition Principle holds. Using arguments similar to those given in Section above, it is straightforward to derive the following expression for the error in the pathway reporter approach:

$$\frac{\mathrm{E}(\mathrm{Cov}(X_N, X_M;\mathbf{Z}))}{\mathrm{E}(X_N)\,\mathrm{E}(X_M)} = \frac{\alpha}{\mathrm{E}(\lambda)}(\eta_B^2 + 1), \quad \text{where } \alpha = \frac{\mathrm{E}(1/(K_M+1))}{\mathrm{E}(1/K_M)}. \tag{33}$$

We can expect that $\alpha \approx 1$, so that unless the burst frequency is substantial, that is $\lambda >> 1$, the overshoot in the the nascent-mature pathway-reporter approach is significant.

Convergence for the mature-protein reporters is considered in Figure S2. Here Eq. (13) shows that we should expect an overshoot in comparison to the mature-mature dual reporter. We have again compared a low output gene with a high output gene, and with all parameters (except δ_M) experiencing noise. The transcription is again given scaled Beta$(3, 6)$ distribution, having squared coefficient of variation 0.2, with scaling to achieve $E(K_N) = 5$ in low output and $E(K_N) = 50$ in the high output. The noise distribution for δ_P is a Beta$(8, 6)$ distribution, scaled to achieve a mean value of 0.2. Computational sampling from 10^7 samples finds the value of $\frac{\mathrm{E}(1/(\delta_p+1))}{\mathrm{E}(1/\delta_p)}$ as approximately 0.1573. The high output gene then expects an overshoot of 0.003146 from the mature-protein reporters, while the low output gene expects an overshoot of approximately 0.03146. Figure S2.A shows comparison of convergence over the first 2000 samples, with the theoretical values accommodating the calculated overshoot. Figure S2.B shows the same over $20,000$ samples; this time two low-output two high-output genes are considered, so that the variation around the expected long term value can be seen. It is evident that reasonable estimates are given after a relatively modest number of samples, but there is a very long delay to a highly accurate convergence for the pathway reporters in the low output gene.

Convergence of Pathway and Dual Reporters

Convergence of both pathway reporters and dual reporters in low output genes is slower than for high output genes, and this affect is most pronounced for reporters taken from nascent mRNA. Figure S1 compares convergence in some low output genes and high output genes for nascent-mature pathway reporters and the mature-mature dual reporter, in the case of constitutive expression. From Eq. (11) and the NDP, these should measure precisely the extrinsic noise on transcription. In the low output gene, both the dual reporter and pathway reporters are yet to show accurate measurement of the overall extrinsic noise after 1000 samples (Figure S1.A), though they are providing rough estimates after as little as a few hundred samples. Pathway reporters for nascent-mature typically exhibit slightly slower convergence than mature-mature dual reporter; the difference is marginal, but can be seen for both the high output gene (compare the values after 500 samples) and the low output gene (compare the values after 1000 samples). In this figure, all parameters (except the reference parameter δ_M) were given scaled Beta distribution

noise. The noise on K_N is a Beta$(3, 6)$ distribution scaled to achieve $E(K_N) = 5$ in the low output case and $E(K_N) = 50$ in the high output case. The squared coefficient of variation is 0.2, and is shown in green in the figure.

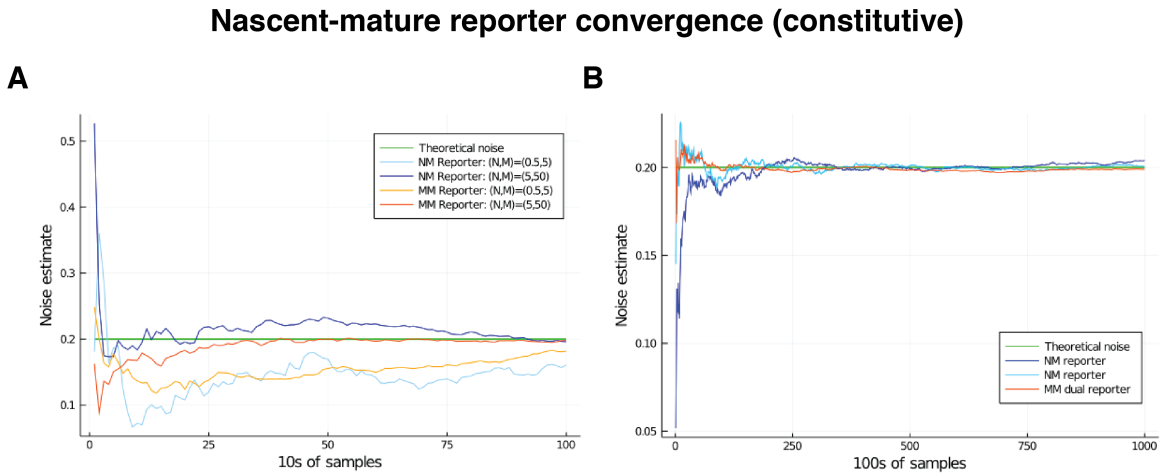


Figure S1: Comparison of convergence of η^2 estimates for low and high mRNA levels by way of nascent-mature reporters and mature-mature reporters. Low level corresponds to mean nascent mRNA level of 0.5, and mean mature mRNA level of 5. High level corresponds to mean nascent mRNA level of 5, and mean mature mRNA level of 50. In both cases the simulated genes are constitutive and noise is on all parameters except for $\delta_M = 1$. The green line gives the squared coefficient of variation for K_N, set to 0.2, which is the value the various reporters are expected to estimate. A. Convergence of the η^2 estimate over the first 2000 samples in the low- and high-output genes. B. Convergence of the η^2 estimate over 100,000 samples in the low-output gene only.

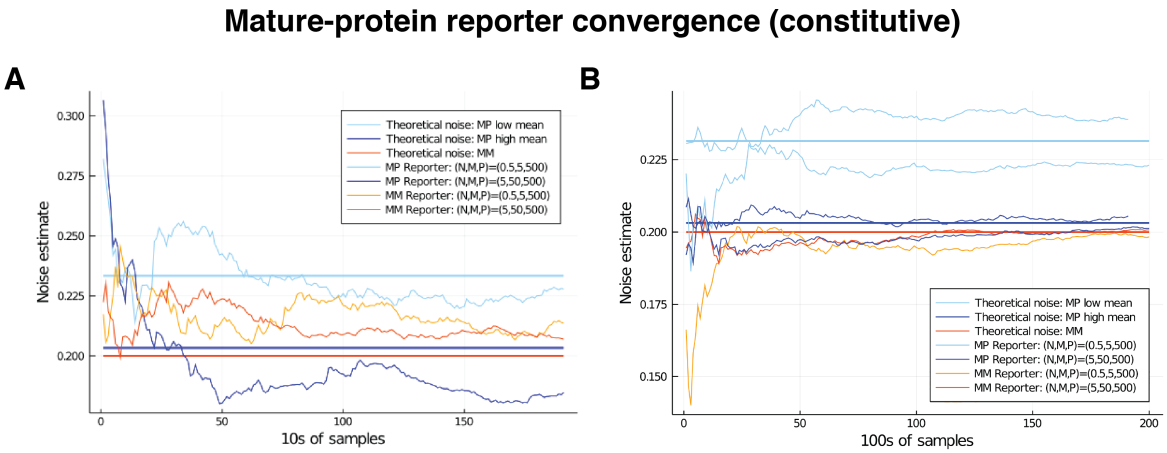


Figure S2: Comparison of convergence for low and high mRNA levels by way of mature-protein and mature-mature reporters. Low level corresponds to mean nascent mRNA level of 0.5, and mean mature mRNA level of 5. High level corresponds to mean nascent mRNA level of 5, and mean mature mRNA level of 50. In both cases the simulated genes are constitutive and noise is on all parameters except for $\delta_M = 1$. The noise on K_N has squared coefficient of variation equal to 0.2, which is shown as the red horizontal line. Our theory shows that mature-protein reporters will return an overshoot that is negligible in the high-output gene (the blue horizontal line), but larger in the low output gene (light blue horizontal line); these values are calculated in the text. A. Comparison of convergence for low and high mRNA levels over the first 2000 samples. B. Convergence of the η^2 estimate over 20,000 samples in the case of the low-output gene only. Two examples of each are given, to show the variation in behaviour.

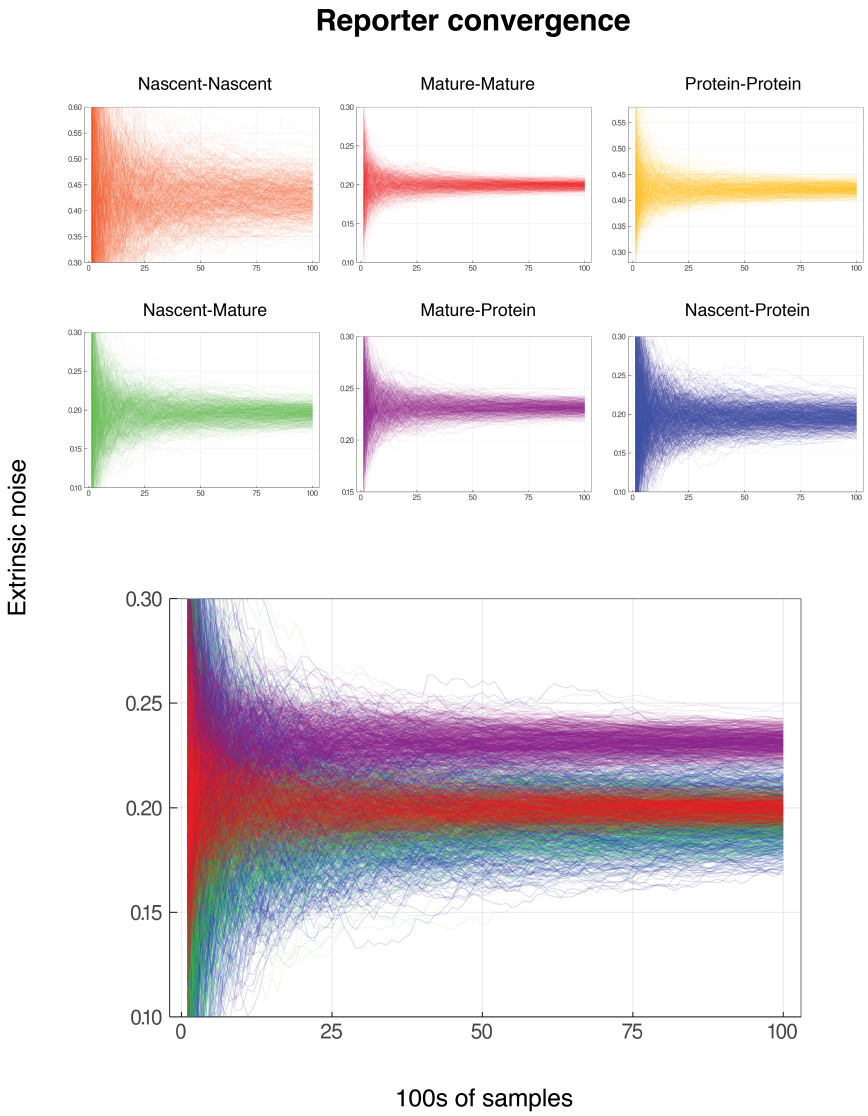


Figure S3: Convergence for reporter pairs for low gene activity. In each case, the mean nascent mRNA level is 0.5, the mean mature mRNA level of 5, and the mean protein level is 500. The simulated genes are constitutive and noise is on all parameters except for $\delta_M = 1$. Each graph shows the convergence of 600 individual reporter simulations, for each combiantion of reporters from nascent mRNA, mature mRNA and protein. Each reporter simulation is from 10,000 samples, with the reporter estimates calculated at intervals of 100. The noise on K_N has squared coefficient of variation equal to 0.2, which should be identified by both the nascent-mature reporter and mature-mature dual reporter. As in Figure S2, the mature-protein reporter should converge to an estimate of approximately 0.2315. Nascent-nascent and protein-protein reporters identify combined noise on more parameters, so do not converge to 0.2. The lower graph shows each of nascent-mature, mature-mature, mature-protein and nascent-protein in the same plot for direct comparison.

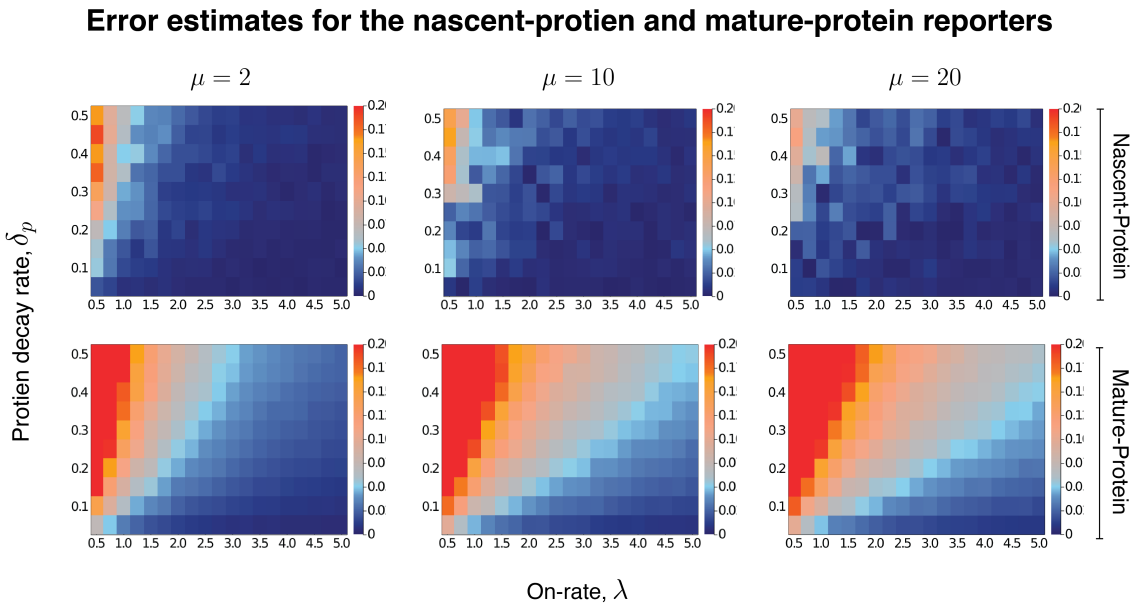


Figure S4: An extension of Fig. 3D of the main text. The heatmaps shown estimate the level of overshoot in the pathway-reporter approach, for both the nascent-protein and mature-protein reporters. For more details refer to the caption given in Fig. 3 in the main text. Here we include the additional case of $\mu = 10$. The remaining parameter values are the same as in the main text. Each individual pixel is generated from a sample of size 3000, though there is still some instability in the convergence for the nascent-protein reporter, particularly as the overshoot estimation starts to increase, and particularly as μ is larger. To produce more accurate values, the case of $\mu = 2$ was averaged over two full experiments while $\mu = 20$ was averaged over three. This was also done for the mature-protein reporter, however for these images there was almost no visible difference between the various runs of the experiment and their averages. Each of the three μ values takes roughly 7–10 hours of computation, depending on lead in time before sampling within a simulation.

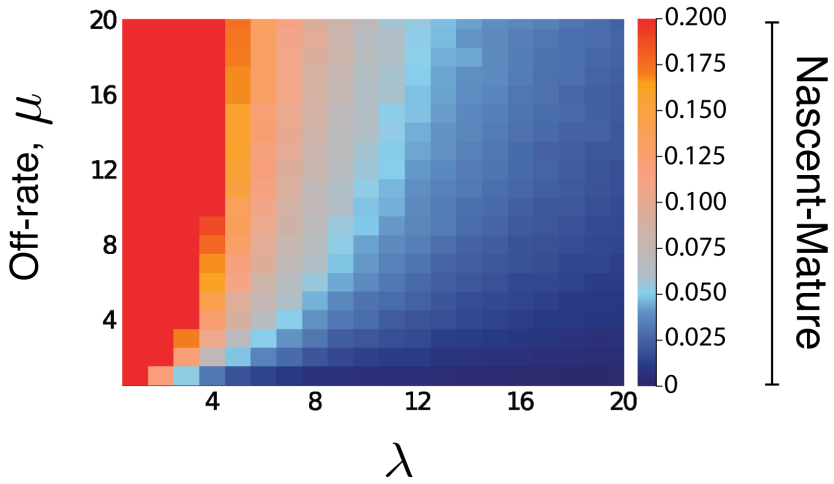


Figure S5: Heatmap for the intrinsic contribution to the covariance for nascent-mature pathway reporters. The nascent-mature reporter concerns only mRNA and so is independent of all protein-related parameters. The heatmap shows the intrinsic contribution for values of λ and μ between 0.1 and 20, with the same parameter selections for K_N, K_M as in Fig. 4 of the main text. Similar simulations for average nascent mRNA levels of 3 and of 8, and mature mRNA levels of 30 and of 160 produced almost identical heatmaps.

Dataset S1

Simulation results of the pathway-reporter method for constitutive genes across 60 different parameter values. We consider noise on all of the parameters except mRNA decay in a constitutive model with mRNA maturation and protein translation. Refer to the excel spreadsheet Constitutive_Results.xlsx for full details of the simulation, including the chosen noise distributions and parameters.

Dataset S2

Simulation results for the overshoot estimate in the pathway-reporter method for bursty genes across 448 different parameter values. Refer to the excel spreadsheet NoiseFree_Results.xlsx for full details of the simulation, including the chosen noise distributions and parameters.