Biophysical Journal
Article

Biophysical Society

Modeling bursty transcription and splicing with the chemical master equation

Gennady Gorin[1] and Lior Pachter[2,*]
[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California and [2]Division of Biology and Biological Engineering & Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California

ABSTRACT Splicing cascades that alter gene products posttranscriptionally also affect expression dynamics. We study a class of processes and associated distributions that emerge from models of bursty promoters coupled to directed acyclic graphs of splicing. These solutions provide full time-dependent joint distributions for an arbitrary number of species with general noise behaviors and transient phenomena, offering qualitative and quantitative insights about how splicing can regulate expression dynamics. Finally, we derive a set of quantitative constraints on the minimum complexity necessary to reproduce gene coexpression patterns using synchronized burst models. We validate these findings by analyzing long-read sequencing data, where we find evidence of expression patterns largely consistent with these constraints.

SIGNIFICANCE Understanding mRNA transcription requires an arsenal of tractable and versatile Markov models. Bursty transcription is ubiquitous in mammalian cells; however, only a few such systems have been mathematically characterized. We expand their scope and present full probabilistic solutions for bursty transcription at multiple, potentially synchronized, gene loci, coupled to arbitrary directed acyclic graphs of splicing. We use these general results to derive quantitative constraints on transcript count correlations, motivate physically plausible classes of burst models, and validate the constraints using single-cell sequencing data.

INTRODUCTION

Recent advances in the analysis of single-cell RNA sequencing (scRNA-seq) (1) enable the quantification of pre-mRNA molecules alongside mature mRNA. These experimental data provide an opportunity to infer the topologies and biophysical parameters governing the processes of mRNA transcription, processing, export, and degradation in living cells. In particular, they provide a novel approach to inferring and studying the dynamics of mammalian splicing cascades (2).

Faced with the enormous volume of genome-wide information accessible through this technology (3), we seek to produce models that can represent their data-generating mechanism. If the model matches the physiology, fits can be physically interpreted in terms of biological parameters, analyzed using standard Bayesian machinery, and enormously compress data sets by representing thousands of observations by a few parameters. The choice of models is informed by a variety of considerations: physiologically plausible mechanisms based on orthogonal experiments, the form and granularity of the data, phenomenological observations that we strive to explain, and computational tractability. In this section, we use these considerations to circumscribe the scope of investigation.

We leverage the past two decades of fluorescence transcriptomics to develop consistent models. Results from live-cell profiling are consistent with Markovian dynamics; in vivo reactions, such as mRNA transcription and degradation, are well described by memoryless models dependent only on the instantaneous molecule counts (4,5). This also appears to be the most general class of models that affords straightforward analytical and numerical recipes; as we discuss in section S6, the complementary class of *delay* master equations, which encode molecular memory, resists systematic analysis at this time. If we do adopt the assumption of memorylessness, previous experiments immediately offer plausible models for transcriptional dynamics: mRNA transcription occurs in *bursts* of activity (6,7), with geometrically distributed bursts prevalent in bacterial and

Submitted July 30, 2021, and accepted for publication February 3, 2022.
*Correspondence: lpachter@caltech.edu
Editor: Ramon Grima.
https://doi.org/10.1016/j.bpj.2022.02.004

mammalian cells. We treat the problem in more generality, considering arbitrary burst distributions, which are mandatory for describing coexpression. Since our methods are designed for scRNA-seq, we omit all mechanistic discussion of regulation: feedback is usually modeled through *protein* interactions (8,9). Joint quantification of mRNA and proteins is in its nascence, and requires targeted antibodies, preventing genome-wide quantification (10–12). Thus, we omit stochastic regulation, although we propose that some of our results can be used to describe *deterministic* regulation without feedback.

Commercially available scRNA-seq data are intrinsically discrete: they report integer counts of molecular barcodes in each cell (13). To fit these counts, we need to solve for state probabilities induced by the chemical master equation (CME), which formalizes probability flow between microstates representing copy numbers. In the most general form, the CME is an infinite series of ordinary differential equations represented by the matrix equation $\frac{dp(t)}{dt} = \mathcal{A}p(t)$, where p is a vector of probabilities and $\mathcal{A}$ encodes the state-specific fluxes; the form of the equation, dependent only on $p(t)$ rather than any history, encodes the Markov property of memorylessness. The scRNA-seq protocols also provide base-level information, which can be used to identify specific transcripts. Inspired by the RNA velocity framework, which quantifies *unspliced* RNA along with mature, coding transcripts to infer regulatory dynamics (1), we seek to lay the groundwork for more general problems: the microstates represent joint copy-number quantities of intermediate transcripts with various introns. We do not treat the specifics of technical noise arising from the sequencing chemistry (14) and bioinformatics (15), although our solutions are modular with respect to simple noise models (16).

We would like to use physical models to encode the full scope of available data, particularly gene-gene correlations. At this time, the modeling field either uses single-gene results (treating the marginals and omitting correlations altogether) (17,18) or explicitly treats small systems (explaining correlations using concrete, usually autoregulatory schema) (19,20), whereas the sequencing field uses fully phenomenological descriptions (omitting mechanistic details of coupling) (21–23). We seek a middle ground that retains mechanistic interpretability, but offers straightforward, physically informed, and tractable recipes for representing gene-gene correlations. Finally, scRNA-seq data are particularly appealing because they can provide genome-wide information about numerous cell types and transient processes. We describe a procedure to treat random and deterministic variation in gene parameters, extending and unifying conventional models for extrinsic cell-to-cell noise (24), cell-type heterogeneity (17), and cell-cycle dynamics (25).

To enable inference, the solutions must be computationally facile. In principle, the CME can always be solved by matrix exponentiation: the solution is simply $p(t) = p(0)e^{\mathcal{A}t}$. However, this approach does not scale very well: if the state space size is truncated to $\mathcal{N}$, the matrix exponential has $\mathcal{N}^3$ time complexity. Although reformulating the problem can ameliorate this—for example, by tensor decomposition (26,27) or treating the CME in terms of the reaction counts (28)—these approaches are poorly compatible with bursty systems, which can have an infinite number of possible reaction channels, corresponding to the infinite support of the burst distribution. More problematically, these numerical methods are not amenable to treating simpler subproblems. For example, finding lower moments or computing a single marginal requires solving the entire system, then summing probabilities, so the method complexity is far out of proportion with the difficulty of the problem. On the other hand, true analytical solutions are unavailable for any but the simplest problems, such as constitutive transcription. Inspired by methods in biological and financial stochastic analysis (29,30), we seek to write down generating function solutions that can be evaluated using well-established algorithms, such as numerical quadrature and the fast Fourier transform.

Therefore, we seek to solve systems characterized by the reactions in Eq. 1. We define a set of n transcripts $\{\mathcal{T}_i\}$, whose abundances are represented by the count variables $m_1, \ldots, m_n := \mathbf{m}$. The "parent" transcripts are produced with bursts of size B_i; by intron splicing, they are converted to downstream products and eventually degraded. The splicing reactions can be naturally represented by a directed graph, a discrete structure that encodes causal relationships between states, defined as transcripts in our case. At least one of c_{ij} and c_{ji} must be zero, encoding the intuition that splicing is irreversible and a shorter transcript cannot turn into a longer one. This constraint implies that it is impossible to leave a transcript $\mathcal{T}_i$ and return to it by the process of splicing: the reactions encode an acyclic digraph or *directed acyclic graph* (DAG) (31,32).

$$\begin{aligned} &\varnothing \xrightarrow{k_i} B_i \times \mathcal{T}_i \text{ with the propensity } k_i \\ &\mathcal{T}_i \xrightarrow{c_{i0}} \varnothing \text{ with the propensity } c_{i0}m_i \\ &\mathcal{T}_i \xrightarrow{c_{ij}} \mathcal{T}_j \text{ with the propensity } c_{ij}m_i. \end{aligned} \tag{1}$$

METHODS

In this section, we take a modular approach to the CME, and describe how to construct tractable solutions to systems defined by Eq. 1, with the DAG constraint. As outlined in the discussion of scope, we are interested in broad classes of models, and focus on strategies for developing highly general solutions.

The section is organized as follows. For each class of phenomena, we introduce and solve its most general *modular* formulation, but limit the other system components to their simplest form: for general splicing graphs, we discuss geometric bursts; for more complex burst models, we

discuss only a single RNA species. Finally, we describe specific examples, and use the results derived from the general solution to provide qualitative results and constraints.

Definition of the CME

To solve the systems described by Eq. 1, we need to encode them in a CME, which describes probability flux between microstates. We suppose the system has n species, with counts $m_1, \ldots, m_n := \mathbf{m}$. The random variables describing the distributions over counts are $\{X_i\}$, with joint probability mass function (PMF) $P(\mathbf{m}, t)$. The reactions in Eq. 1 naively correspond to the following fluxes for each species i:

$$\begin{aligned} R_{tx,i} &= k_i \left[\sum_{z=0}^{m_i} p_{i,z} P(m_i - z, t) - P(m_i, t) \right] \\ R_{deg,i} &= c_{i0}\left[(m_i+1)P(m_i+1,t) - m_i P(m_i,t)\right] \\ R_{splic,ij} &= c_{ij}\left[(m_i+1)P\left(m_i+1, m_j-1, t\right) - m_i P\left(m_i, m_j, t\right)\right]. \end{aligned} \tag{2}$$

For convenience of notation, Eq. 2 omits the indices of species not involved in a particular reaction. The derivation is provided in section S1.1. The full master equation takes the following generic form:

$$\frac{dP(\mathbf{m},t)}{dt} = R_{tx} + \sum_{i=1}^{n} \left[R_{deg,i} + \sum_{j=1}^{n} R_{splic,ij} \right], \tag{3}$$

i.e., the net flux into any state is the sum of contributions from all reactions for all species. If the burst processes are mutually independent, they can simply be added to yield $R_{tx} = \sum_i R_{tx,i}$, but this is not true in general. Unfortunately, in all but the simplest cases, the PMF is unwieldy to manipulate. Therefore, we convert the CME to the corresponding probability generating function (PGF) $G(x_1, \ldots, x_n, t) := G(\mathbf{x}, t)$:

$$G(\mathbf{x}, t) : = \sum_{m_1, \ldots, m_n} x_1^{m_1} \ldots x_n^{m_n} P(\mathbf{m}, t). \tag{4}$$

If we suppose the burst processes *are* independent, this partial differential equation (PDE) takes the following form:

$$\begin{aligned} \frac{\partial G(\mathbf{x},t)}{\partial t} = & \sum_{i=1}^{n} k_i (F_i(x_i) - 1) G + \sum_{i=1}^{n} c_{i0}(1 - x_i) \frac{\partial G}{\partial x_i} \\ & + \sum_{i,j=1}^{n} c_{ij}\left(x_j - x_i\right) \frac{\partial G}{\partial x_i}, \end{aligned} \tag{5}$$

where F_i is the PGF of the burst random variable B_i. The derivation of this PDE is provided in section S1.2. It is generally easier to treat the logarithm of G; applying the transformations $u_i := x_i - 1$ and $\varphi := \ln G$ yields the equation:

$$\begin{aligned} \frac{\partial \varphi(\mathbf{u},t)}{\partial t} = & \sum_{i=1}^{n} k_i (M_i(u_i) - 1) - \sum_{i=1}^{n} u_i c_{i0} \frac{\partial \varphi}{\partial u_i} \\ & + \sum_{i,j=1}^{n} \left(u_j - u_i\right) c_{ij} \frac{\partial \varphi}{\partial u_i}, \end{aligned} \tag{6}$$

where $M_i(u_i) = F_i(1 + u_i)$ is the appropriate transformation of the burst PGF. To compute $P(\mathbf{m},t)$, we can evaluate $G(\mathbf{x}, t)$ around the n-dimensional unit sphere and take an inverse fast Fourier transform (30,33).

Certain CMEs are isomorphic to SDEs

Before treating the discrete system defined by Eq. 2, we need to establish a crucial connection to a set of stochastic differential equations (SDEs), allowing us to use SDE tools to solve CMEs. Specifically, it is possible to treat X_i as a random variable with the distribution $Poisson(\Lambda_i)$, where Λ_i is the *intensity* of the discrete process. In the general multivariate case, the following identity can be used to relate a set of stochastic processes $\Lambda_1, \ldots, \Lambda_n$ with joint distribution $F_{\Lambda_1,\ldots,\Lambda_n}$ to the solution of the CME (5,17,34,35):

$$P(\mathbf{m}, t) = \int \prod_{i=1}^{n} \frac{e^{-\Lambda_i} \Lambda_i^{m_i}}{m_i!} dF_{\Lambda_1,\ldots,\Lambda_n}. \tag{7}$$

The *Poisson representation* was introduced by Gardiner (5,35), and successfully applied to modeling regulated gene loci (9,36,37). This representation is always viable, but does not guarantee that the law of Λ_i is a probability: for example, if the variance of marginal random variable X_i is lower than its mean, some probability densities must be negative. Nevertheless, in some cases, we *can* define a relationship between the CME and the "proper" probability laws of SDEs. If the distribution of burst sizes is a Poisson mixture, we can write it down as an equivalent Lévy jump process; for example, the geometric burst size distribution is an exponential-Poisson mixture, and may be interpreted as the Poisson mixture of the underlying compound Poisson process with exponential jumps. This class of models has previously been invoked to explain high trajectory variation observed in living cells (38–40). Thus, if species i undergoes jumps governed by the process $L_{i,t}$, the three prototypical reactions in Eq. 1 take the following form in the continuous worldview:

$$d\Lambda_i = dL_{i,t} - \Lambda_i \sum_{j=0}^{n} c_{ij} dt + \sum_{j=1}^{n} c_{ji} \Lambda_j dt, \tag{8}$$

where the left-hand term denotes the evolution of the process, the first term on the right-hand side represents bursting, the second term represents efflux by isomerization and degradation, and the third represents influx by isomerization. If $dL_{i,t} = k_i dt$, this formulation recovers the deterministic reaction rate equations and the well-known Poisson form for constitutive production (34).

This representation has three features that are occasionally useful to solving CME systems. The first is theoretical: the standard properties of Poisson mixtures mean that the moment-generating function (MGF) of the SDE is simply $G(\mathbf{x} - 1, t)$, allowing easy conversion between the two (41,42). Therefore, CME solutions can be used as SDE solutions and vice versa; although we discuss the CME, the solutions generalize to classes of continuous-valued stochastic models explored in the biological (38) and financial (43) literature.

The second feature is qualitative: if the solution to an SDE is available, we can draw conclusions about the isomorphic CME without actually having to solve this CME. For example, the CME in Eq. 9 is isomorphic to the SDE in Eq. 10, as they share the generating function in Eq. 11 (17,44):

$$\begin{aligned} \frac{dP(m,t)}{dt} = & k_1 \sum_{z=0}^{m} \left[p_{z,1} P(m - z, t) - P(m, t) \right] \\ & - \gamma \left[(m+1) P(m+1, t) - m P(m, t) \right] \end{aligned} \tag{9}$$

$$d\Lambda = dL_t - \gamma \Lambda dt \tag{10}$$

$$G(u,t) = \left(\frac{1 - bue^{-\gamma t}}{1 - bu} \right)^{\frac{k_1}{\gamma}}, \tag{11}$$

where we suppose the burst distribution is geometric with mean b, the jump distribution is exponential with mean b, bursts arrive according to a Poisson process with rate k_1, the degradation rate is γ, and the process begins at $\Lambda = X = 0$. Λ_1 is the gamma Ornstein-Uhlenbeck process (29,43–45). Inspired by a highly general result for constitutive transcription, which states that Poisson distributions always remain Poisson for a birth-death process (34), we may reasonably ask whether equivalent results are available for bursty processes. This intuition turns out to be incorrect, and straightforward to disconfirm using SDE results: $G(u,t)$ is not the gamma MGF, so the distribution of the process is not gamma for *any* finite $t \in (0,\infty)$—although it does approach a gamma law exponentially fast (44). Therefore, the corresponding Poisson mixture is *not* simply a time-varying negative binomial, and adopting such a model to describe gene expression over time (46) is theoretically problematic.

The third feature is quantitative: the Poisson representation facilitates analysis, both in the notation and the computation. The variance $\mathbb{V}[X_i]$ of the CME is simply given by $\mathbb{V}[\Lambda_i] + \mathbb{E}[\Lambda_i]$, allowing for a more compact representation. Finally, since the generating functions do not assume that the entire system is discrete or continuous, we can use the classes of solutions explored throughout this report to solve *hybrid* systems, with the transcription rate of a CME governed by an SDE. For example, if the transcription rate is given by the gamma Ornstein-Uhlenbeck process, we can solve the hybrid SDE-CME system by appending a "virtual" discrete species that undergoes bursty transcription and splicing (47).

General splicing graphs have analytical solutions

In this subsection, we treat the solutions of CMEs induced by general splicing DAGs. For simplicity, we enforce Eq. 6 holds: all burst processes are independent.

We need to solve Eq. 6, a PDE in n complex variables and one time variable. This is impossible in general. However, using the method of characteristics (30,48,49), we can reduce the PDE's complex degrees of freedom to n particularly simple coupled ordinary differential equations, which can be represented by a matrix equation:

$$\frac{d\mathbf{U}}{ds} = C\mathbf{U} \text{ s.t.} \mathbf{U}(s=0) = \mathbf{u}, \tag{12}$$

where C is a matrix containing $C_{ij} = c_{ij}$ for all $i \neq j$ and $C_{ii} = -\sum_{j=1}^{n} c_{ij} - c_{i0}$. We can compute the characteristics by spectral decomposition:

$$U_i(\mathbf{u},s) = \sum_{j=1}^{n} e^{-r_j s}\left(V \cdot Diag\left(V^{-1}\mathbf{u}\right)\right)_{ij} := \sum_{j=1}^{n} e^{-r_j s} A_{ij}(\mathbf{u}), \tag{13}$$

where r_j are the (strictly nonnegative) eigenvalues of $-C$ and each column of matrix V contains an eigenvector of C. The derivation of this solution procedure is provided in section S2. Once the functional form of U_i is available, it is straightforward to compute the solution of Eq. 6 by quadrature:

$$\varphi(\mathbf{u},t) = \int_0^t \sum_{i=1}^{n} k_i M_i(U_i(\mathbf{u},s) - 1) ds. \tag{14}$$

Moments of the splicing graph solutions are tractable by matrix operations

To discuss specific results, such as the correlations between transcripts, we assume that exactly one parent transcript exists, and all other transcripts are produced from it by splicing. Consistently with the conventional model (25,30,50) we assume it is produced in geometrically distributed bursts. This corresponds to defining $\mathcal{T}_1$ as the parent transcript and setting $M_1(u_1) = M(u_1) = \frac{1}{1-bu_1}$ in Eq. 6, with all $k_j, j \neq i$ defined as zero.

To analyze this system, we rewrite Eq. 13 in terms of its marginal components:

$$U_1(\mathbf{u},t) = \sum_{i=1}^{n} u_i \psi_i(t) \text{ s.t.} \psi_i = \sum_{k=1}^{n} a_{i,k} e^{-r_k s}. \tag{15}$$

This form is most convenient for analyzing the summary statistics of a few marginals at a time. If the spectral decomposition is available, converting to this representation amounts to evaluating the coefficient matrix A once: $a_{i,k} = A_{1k}(\boldsymbol{\delta}_{ij}) = (V \cdot Diag(V^{-1}\boldsymbol{\delta}_{ij}))_{1j}$, where $\boldsymbol{\delta}_{ij}$ is the Kronecker delta vector with a 1 in position i and 0 elsewhere.

The moments of the marginal distribution of species i can be computed directly from the derivatives of the marginal MGF of the formal *continuous* system with all complex arguments in $\mathbf{u}$ set to zero. We begin by considering U_1 as the sum in Eq. 15. To marginalize with respect to all $j \neq i$, we simply set all u_j to zero, obtaining $U_1 = u_i \psi_i(t)$. To compute the lower moments of a single species, we take the derivative with respect to u_i and evaluate it at $u_i = 0$ with the help of Eq. 15:

$$\begin{aligned}\mathbb{E}[\Lambda_i] &= e^{\varphi}\frac{\partial}{\partial u_i} k_1 \int_0^{\infty}\left[\frac{1}{1-bu_i\psi_i(s)} - 1\right] ds\Bigg|_{u_i=0} \\ &= k_1 b \int_0^{\infty} \psi_i(s) ds = k_1 b \sum_{k=1}^{n} \frac{a_{i,k}}{r_k}.\end{aligned} \tag{16}$$

Per the standard properties of mixed Poisson distributions (41), the value of μ_i is identical for the underlying continuous process and the derived discrete process.

The covariances can be computed directly from the derivatives of the marginal PGF with $u_q = 0$ for all $q \neq i, l$. By construction, $U_1(u_i, u_l; s) = u_i\psi_i(s) + u_l\psi_l(s)$, where each ψ is the exponential sum corresponding to the marginal of the species in question. Taking the partial derivatives yields the following equation:

$$\text{Cov}(\Lambda_i, \Lambda_l) = 2k_1 b^2 \sum_{j,k=1}^{n} \frac{a_{i,j} a_{l,k}}{r_j + r_k}. \tag{17}$$

From standard identities, the covariance of a mixed bivariate Poisson distribution with no intrinsic covariance forcing is identical to the covariance of the mixing distribution (41). The marginal variances can be found by plugging in $l = i$, and the standard properties of Poisson mixtures (41) allow conversion to the discrete domain, with $\mathbb{V}[X_i] = \mathbb{V}[\Lambda_i] + \mu_i$. Since mixing decreases variance but not covariance, the correlation coefficient of the discrete system will always be lower than that of its continuous or hybrid analog. These lower moments are straightforward to compute, but do not appear to have an easily amenable analytical form for general graphs.

General burst distributions are tractable and encode physics with no free parameters

The relevance of CME to modern transcriptomic experimental data is tempered by the simplicity of tractable models. The model we have presented so far can describe the splicing cascade of a single gene, but does not naturally extend to multigene networks. Yet we know that genes often belong to coexpression modules that are identifiable by similarity metrics (2,21,23). Therefore, we are faced with the challenge of integrating multiple genes in a physically meaningful way.

Instead of building intractable "top-down" models that encode complex networks (51), we may build "bottom-up" models that extend analytical

solutions. For example, we can consider sets of *synchronized* genes that experience bursting events at the same time. This model represents the bursty limit of multiple genes with transcription rates governed by a single telegraph process, up to scaling; a conceptually similar model has previously been used to describe correlations between multiple copies of one gene (52). This model retains the appeal of physical interpretability—for example, gene modules may be regulated by the same molecule—but does not excessively complicate the mathematics, and offers an incremental step toward more detailed descriptions.

For convenience of notation, we omit any discussion of the details of the downstream splicing networks, as the results derived above hold with no change. We have thus far assumed the $R_{tx,i}$ in Eq. 2 are *independent*, i.e., their bursts are unsynchronized. However, more sophisticated models can be built, and encode gene-gene correlations. Most generally, the differential of the log-PGF in Eq. 6 takes the following functional form:

$$\frac{\partial \varphi(\mathbf{u},t)}{\partial t} = \sum_{\ell=1}^{n}\sum_{q=1}^{\omega} k_{\ell,q}\big(M_{\ell,q}(\mathbf{U}(\mathbf{u},t)\,) - 1\big). \tag{18}$$

The full derivation is provided in section S1.4, and relies on iterative application of Cauchy products to conditional PGFs. This rather formal expression states that up to n species in the system may be cotranscribed in a single module. For a particular module of ℓ genes with synchronized transcription times, there are $\omega = \binom{n}{\ell}$ possible combinations of species that can be cotranscribed, indexed by q. The independent case emerges from setting all $k_{\ell,q}$ to zero for ℓ>1. The vector defining $U_1,\ldots,U_n := \mathbf{U}$ is modular, as it is independent of bursting dynamics.

Example: Two-promoter bursty model, no synchronization

To begin, consider a system with two independent promoters that fire reactions $\varnothing \xrightarrow{k_{1,1}} B_1 \times \mathcal{T}_1$ and $\varnothing \xrightarrow{k_{1,2}} B_2 \times \mathcal{T}_2$. If $\mathcal{T}_1$ can be converted to $\mathcal{T}_2$, this model can describe an internal promoter (53) that generates molecules that cover only a part of the gene. In this formulation, the following special case of Eq. 6 holds:

$$\frac{\partial \varphi(\mathbf{u},t)}{\partial t} = k_{1,1}(M_1(U_1(\mathbf{u},t)) - 1) + k_{1,2}(M_2(U_2(\mathbf{u},t)) - 1). \tag{19}$$

In the parlance of Eq. 18, we set $n = 2$ and $k_{2,1} = 0$. Therefore, the stationary log-PGF takes the simple form:

$$\begin{aligned}\varphi(\mathbf{u},t) = \int_0^t &[k_{1,1}(M_1(U_1(\mathbf{u},s)) - 1)\\ &+ k_{1,2}(M_2(U_2(\mathbf{u},s)) - 1)]ds,\end{aligned} \tag{20}$$

i.e., the joint distribution of RNA can be obtained with a single application of quadrature.

If U_1 and U_2 are disjoint (e.g., $\mathcal{T}_1$ and $\mathcal{T}_2$ are products from two different genes), this expression reduces to the trivial case $\varphi(\mathbf{u}) = \varphi_1(u_1) + \varphi_2(u_2)$: intuitively, independent transcription processes produce statistically independent distributions. If only a single parent transcript exists ($U_1 = U_2 = U$), the model can represent a particular class of multistate promoters with two distinct short-lived active states, recapitulating the unsynchronized model in section S1.3. Finally, if the burst distributions are identical ($M_1 = M_2 = M$), the system becomes equivalent to a one-locus system with burst frequency $k_{1,1} + k_{1,2}$. This accords with the superposition property of the Poisson process driving transcription.

Example: Two-gene bursty model, with burst time synchronization

We continue by considering the instructive model of two genes influenced by the same regulator: the active periods of these genes are synchronized in time. However, the burst sizes are not coupled, and may indeed come from different distributions. Such a process has the transcription reaction $\varnothing \xrightarrow{k_{2,1}} B_1 \times \mathcal{T}_1 + B_2 \times \mathcal{T}_2$, where each $\mathcal{T}_i$ may be degraded or processed into further downstream species with total efflux rate r_i. In the parlance of Eq. 18, we set $k_{\ell,q}$ to 0 for $\ell = 1$.

Assuming that $B_1 \sim Geom(b_1)$ and $B_2 \sim Geom(b_2)$—i.e., the marginals are consistent with the standard bursty model—we can exploit the independence of B_1 and B_2 to write down the joint distribution:

$$M(\mathbf{U}(\mathbf{u},t)) = M_1(U_1(\mathbf{u},t))M_2(U_2(\mathbf{u},t)), \tag{21}$$

which follows immediately from standard MGF identities. We propose a physical model leading to this burst size distribution in section S1.3. The expression in Eq. 18, with Eq. 21 used as the transcriptional burst size distribution, has the following solution:

$$\begin{aligned}\varphi &= k_{2,1}\int_0^t [M_1(U_1(\mathbf{u},s))M_2(U_2(\mathbf{u},s)) - 1]ds\\ &= k_{2,1}\int_0^t \left[\frac{1}{(1-b_1U_1(\mathbf{u},s))(1-b_2U_2(\mathbf{u},s))} - 1\right]ds.\end{aligned} \tag{22}$$

This expression is numerically integrable, but does not afford an analytical solution. We can obtain its lower moments by differentiating the log-PGF. Defining $U_i = u_i\psi_i$ and $U_l = u_l\psi_l$ for two transcript species generated from distinct genes, we yield:

$$\begin{aligned}\mathrm{Cov}(\Lambda_i,\Lambda_l) &= k_{2,1}b_1b_2\int_0^\infty \psi_i\psi_l ds\\ &= k_{2,1}b_1b_2\sum_{j,k=1}^{n}\frac{a_{i,j}a_{l,k}}{r_j + r_k}.\end{aligned} \tag{23}$$

Usefully, this expression is agnostic of the actual identity of i,l, so the expression holds for *any* of the species downstream of $\mathcal{T}_1$ and $\mathcal{T}_2$. Finally, if we would like to compute the covariance between the two parent species, we simply plug in $\psi_i = e^{-r_1 s}$ and $\psi_l = e^{-r_2 s}$:

$$\mathrm{Cov}(\Lambda_1,\Lambda_2) = \frac{k_{2,1}b_1b_2}{r_1 + r_2}. \tag{24}$$

This covariance corresponds to the following Pearson correlation coefficient:

$$\begin{aligned}\rho = \frac{\mathrm{Cov}(\Lambda_1,\Lambda_2)}{\sigma_1\sigma_2} &= \frac{\sqrt{r_1/r_2}}{1 + r_1/r_2}\\ &\times \sqrt{\frac{1}{(1+1/b_1)(1+1/b_2)}}.\end{aligned} \tag{25}$$

The first term achieves a global maximum of $1/2$ at $r_1 = r_2$. The second is strictly smaller than 1, but asymptotically approaches 1 as b_1, b_2 jointly approach infinity. All downstream processes are stochastic and desynchronize molecular observables. Therefore, $1/2$ is the supremum of gene-gene correlations in this class of models.

Example: Two-gene bursty model, with perfect burst time and size synchronization

Conversely, we can consider the two-gene problem assuming that the burst distributions are identical and perfectly correlated. Physically, this model may correspond to coupling of *initiation* processes, e.g., this may occur when two genes are controlled by a single promoter. This burst distribution has the following joint PGF:

$$\begin{aligned} F(x_1,x_2) &= \sum_{i,j} x_1^i x_2^j p_{i,j} = \sum_{i,j} x_1^i x_2^j p_i \delta(i-j) \\ &= \frac{1}{1+b-bx_1x_2}. \end{aligned} \tag{26}$$

Upon inserting the characteristics and taking the requisite partial derivatives, we yield the following correlation structure:

$$\rho = \frac{\sqrt{r_1/r_2}}{1+r_1/r_2} \times \left(1+\frac{b}{b+1}\right). \tag{27}$$

As in the case of uncoupled gene sizes reported in Eq. 25, the first term is at most $1/2$. The second term asymptotically approaches 2 as $b \to \infty$. Therefore, there are no intrinsic model constraints on Pearson correlation coefficients of two-gene products; constraints arise as the *effect* of the burst size correlation structure.

Example: Two-gene bursty model, with partial burst time and size synchronization

The models described above are useful for understanding limiting cases, but somewhat restrictive: the model solved in Eq. 19 enforces $\rho = 0$, the synchronized burst time model in Eq. 22 enforces $\rho \in (0, 1/2)$, and the perfectly synchronized burst size model in Eq. 26 requires burst sizes to be identical. Thus, we need to construct models consistent with observations, recapitulating both the multigene correlation structure and the diversity of burst sizes evident from marginals.

This is most easily tractable using the continuous formulation: we can impose perfect correlation between Lévy jump sizes. Conceptually, we suppose the overall driving process L_t for N genes is given by a compound Poisson process with average jump size b and jump distribution $J \sim Exp(b^{-1})$. The driving process for a single gene i is given by $L_{i,t} = w_i L_t$, with average jump size bw_i and $\sum_{i=1}^N w_i = 1$. We propose a physical model leading to this burst size distribution in section S1.3. Since all jump sizes are scalar multiples of each other, they are perfectly correlated. The joint jump distribution has the following MGF:

$$\begin{aligned} M(\mathbf{u}) &= \mathbb{E}\left[\exp\left(\sum_{i=1}^N u_i J_i\right)\right] = \mathbb{E}\left[\exp\left(J\sum_{i=1}^N u_i w_i\right)\right] \\ &= \frac{1}{1-b\sum_{i=1}^N u_i w_i}, \end{aligned} \tag{28}$$

where the third equality stems from recognizing that the joint MGF is simply the univariate exponential MGF evaluated at $\sum_{i=1}^v u_i w_i$. If we only consider the parent transcripts with efflux rates r_i and characteristics $U_i = u_i e^{-r_i t}$, we yield the following stationary log-PGF:

$$\varphi(\mathbf{u}) = k_{N,1}\int_0^\infty \left[\frac{1}{1-b\sum_{i=1}^N u_i w_i e^{-r_i s}} - 1\right] ds. \tag{29}$$

Finally, setting $N = 2$, we can take derivatives of the PGF and compute the correlation structure:

$$\mathrm{Cov}(\Lambda_1,\Lambda_2) = \frac{2k_{2,1}b^2 w_1 w_2}{r_1+r_2} \tag{30}$$

$$\rho = \frac{2\sqrt{r_1/r_2}}{1+r_1/r_2} \times \sqrt{\frac{1}{\left(1+\frac{1}{bw_1}\right)\left(1+\frac{1}{bw_2}\right)}} \in (0,1). \tag{31}$$

As above, this model is appealing since it has no free parameters: the gene-gene correlations are not imposed by an ad hoc correlation parameter, but emerge from the model structure itself. Therefore, we can evaluate the model's performance by comparing true intergene correlations to ones computed based only on the marginals.

Interestingly, single-gene systems with rapid splicing recapitulate the multigene functional form. Consider a source species $\mathcal{T}_0$, which is produced in geometrically distributed bursts and converted to species $\{\mathcal{T}_i\}$, with $i = 1, \ldots, N$, at rates β_i. These transcripts are degraded at rates r_i. Furthermore, suppose all of the $\beta_i \sim \mathcal{O}(\varepsilon^{-1})$ for small ε, i.e., the source transcript is extremely unstable. In this limit, $\mathcal{T}_i$ are produced by a Lévy process with jumps of average size $b\beta_i/\beta$, where b is the underlying $\mathcal{T}_0$ burst size, $\beta := \sum_i \beta_i$, and the ratio β_i/β is $O(1)$. We define corresponding weights $w_i := \beta_i/\beta$; by definition, $\sum_i w_i = 1$. This yields:

$$U_0(\mathbf{u},t) = Ce^{-\beta t} + \sum_{i=1}^N \frac{\beta_i u_i}{\beta - r_i} e^{-r_i t} \to \sum_{i=1}^N u_i w_i e^{-r_i t}, \tag{32}$$

where C is a constant in the solution of the characteristic $U_0(\mathbf{u},t)$ whose value becomes immaterial as β increases. Plugging in the characteristic:

$$M(U_0(\mathbf{u},t)) = \frac{1}{1-bU_0(\mathbf{u},t)} = \frac{1}{1-b\sum_{i=1}^N u_i w_i e^{-r_i t}}, \tag{33}$$

which recapitulates the form of the integrand in Eq. 29. Therefore, the multigene model can be used to describe transcripts arising from a single rapidly processed, unobserved parent molecule, yielding positive, but otherwise unconstrained, correlation between the downstream species.

Example: Anti-correlated two-gene bursty model, with burst time synchronization

Thus far, we have considered the problem of describing synchronized genes, and proposed three models that can produce positive correlations. Putting aside the problem of positing a specific physical mechanism, we can ask whether *any* joint burst distribution can produce negative correlations in molecule counts, *despite* perfect synchronization of burst events. Considering the cross moment of mRNA produced at two synchronized loci, with generic joint burst generating function M, we find:

$$\mathbb{E}[\Lambda_1\Lambda_2] = k_{2,1}\int_0^\infty \frac{d^2M}{du_1 du_2} ds + \mu_1\mu_2, \tag{34}$$

with the partial derivatives evaluated at $u_1 = u_2 = 0$. The second term is strictly positive. The first term is the integral of an exponentially discounted burst cross moment:

$$\frac{d^2M}{du_1 du_2} = \frac{d^2M}{dU_1 dU_2}\frac{dU_1}{du_1}\frac{dU_2}{du_2} = \mathbb{E}[B_1B_2]e^{-(r_1+r_2)s}, \tag{35}$$

where the partial derivatives are yet again evaluated at $u_1 = u_2 = U_1 = U_2 = 0$, and B_1 and B_2 denote the SDE jump sizes at the two-gene loci. Supposing the correlation between the burst sizes is $\rho \in [-1, 0)$, and considering the covariance between the transcripts:

$$\mathrm{Cov}(\Lambda_1, \Lambda_2) = k_{2,1}\frac{\mathbb{E}[B_1 B_2]}{r_1 + r_2} = \frac{k_{2,1}}{r_1 + r_2}\left(\rho\sigma_{B_1}\sigma_{B_2} + \mu_{B_1}\mu_{B_2}\right), \tag{36}$$

which achieves a minimum at $\rho = -1$. Thus, the covariance has a lower limit:

$$\mathrm{Cov}(\Lambda_1, \Lambda_2) \geq \frac{k_{2,1}}{r_1 + r_2}\left(\mu_{B_1}\mu_{B_2} - \sigma_{B_1}\sigma_{B_2}\right). \tag{37}$$

Without constructing the joint distribution explicitly, if we suppose the marginal discrete burst distributions are geometric—i.e., the jump sizes are exponential—then $\mu_{B_i} = \sigma_{B_i} = b_i$, and the lower limit on covariance is zero. This means that negative correlations cannot possibly result from a model with geometrically distributed, synchronized jumps. However, other joint burst laws *can* produce negative correlations, as long as the population correlation coefficient is sufficiently negative and the burst distributions are sufficiently dispersed.

We can demonstrate the existence of processes with negative count correlations induced by synchronized burst events. First, we suppose that the marginal burst distributions are identical and described by a gamma law with shape α and scale θ, enforcing $\mu_{B_1} = \mu_{B_2} = \alpha\theta$ and $\sigma^2_{B_1} = \sigma^2_{B_2} = \alpha\theta^2$. Therefore, the covariance of the Poisson intensities takes the following form:

$$\begin{aligned}\mathrm{Cov}(\Lambda_1, \Lambda_2) &= \frac{k_{2,1}}{r_1 + r_2}\left(\rho\alpha\theta^2 + \alpha^2\theta^2\right) \\ &= \frac{k_{2,1}\theta^2}{r_1 + r_2}\left(\rho\alpha + \alpha^2\right),\end{aligned} \tag{38}$$

which achieves $\mathrm{Cov}(\Lambda_1, \Lambda_2) < 0$ whenever $\rho\alpha + \alpha^2 < 0$. Therefore, for any $\rho \in (-1, 0)$, every $\alpha \in (0, -\rho)$ meets this criterion.

It remains to confirm that a bivariate gamma distribution with a negative correlation can exist. Such a distribution was constructed by Moran, and permits all $\rho \in (-1, 1)$ (54,55). Applying the Cauchy-Schwartz inequality to the marginals guarantees that the joint MGF of the correlated bivariate gamma distribution exists (56). This demonstrates the existence of continuous moving average processes with negative stationary correlation, driven by a common Poisson process arrival process. Finally, the corresponding Poisson mixture has identical covariance, and must also have a negative correlation. Therefore, a CME with marginal negative binomial burst distributions and a carefully chosen joint structure can achieve negative molecular correlations, even if the bursts are synchronized.

Certain models of heterogeneity are analytically tractable

Thus far, we have reported a toolbox of models that can represent Markovian systems with fairly general splicing graphs and burst distributions, and demonstrated that their solutions exist and are computable. By design, these models by themselves describe homogeneous populations of cells, with deterministic parameters. Using standard statistical approaches, they can be reassembled to describe heterogeneity in parameters (57):

$$P(\mathbf{m}, t) = \int_\Theta P(\mathbf{m}, t; \theta) df_\theta = \mathbb{E}_\Theta[P(\mathbf{m}, t; \theta)] \tag{39}$$

$$G(\mathbf{x}, t) = \int_\Theta G(\mathbf{x}, t; \theta) df_\theta = \mathbb{E}_\Theta[G(\mathbf{x}, t; \theta)]. \tag{40}$$

Equation 39 is essentially the definition of a mixture model: the parameter vector θ may itself be distributed according to the statistical law f_θ over a state space Θ. To determine the resulting probability distribution, we need to integrate with respect to this law. By linearity, we can exchange the generating function sum in Eq. 4 and the integral in Eq. 39 to obtain Eq. 40.

This expression is extremely generic; the base case of a homogeneous cell population with parameter vector $\overline{\theta}$ is obtained by setting $df_\theta = \delta(\theta - \overline{\theta})$, where δ is a multivariate Dirac delta functional. Unfortunately, analytical solutions are only available in several highly restrictive cases.

Example: A finite number of cell types

Multimodality in single cells can emerge from the existence of multiple long-lived subpopulations, which are conventionally fit to a finite mixture distribution (58). We can formalize this model for an arbitrary number of populations. Consider a population of cells with J disjoint "cell types," indexed by j. Each cell type has the parameter vector θ_j and relative abundance w_j. The probability distribution of the entire population can be computed by the law of total probability, conditioning on the cell type, or by setting $df_\theta = \sum_{j=1}^J w_j \delta(\theta - \theta_j)$ in Eqs. 39 and 40:

$$P(\mathbf{m}, t) = \sum_{j=1}^J w_j P(\mathbf{m}, t; \theta_j) G(\mathbf{x}, t) = \sum_{j=1}^J w_j G(\mathbf{x}, t; \theta_j). \tag{41}$$

This formulation can produce J-modal distributions. This reduces to the standard model when $J = 1$ or when all θ_j are identical.

Correlations can emerge from the existence of multiple cell types. For simplicity, we set $J = 2$ and $n = 2$, and suppose the molecule distributions are independent Poisson, whether as an effect of constitutive production or as the limit of a bursty distribution. For gene i and cell type j, the average expression is $\lambda_{i,j}$, implying that the stationary distribution takes the following form:

$$\begin{aligned}G(\mathbf{u}) &= \sum_{j=1}^2 w_j G(\mathbf{u}; \lambda_{1,j}, \lambda_{2,j}) = \sum_{j=1}^2 w_j G(u_1, \lambda_{1,j}) G(u_2, \lambda_{2,j}) \\ &= \sum_{j=1}^2 w_j e^{\lambda_{1,j} u_1} e^{\lambda_{2,j} u_2}.\end{aligned} \tag{42}$$

The correlation can be found by differentiation. Setting $w_1 = w_2 = \frac{1}{2}$ and supposing $\lambda_{i,1} = 0$ (i.e., cell type 1 does not express either gene), we yield:

$$\rho = \sqrt{\frac{\lambda_{1,2}\lambda_{2,2}}{(\lambda_{1,2} + 2)(\lambda_{2,2} + 2)}} \in (0, 1). \tag{43}$$

Supposing now $\lambda_{1,2} = \lambda_{2,1} = 0$ (i.e., gene 1 is a perfectly specific marker for cell type 1 and gene 2 is a marker for cell type 2):

$$\rho = -\sqrt{\frac{\lambda_{1,1}\lambda_{2,2}}{(\lambda_{1,1} + 2)(\lambda_{2,2} + 2)}} \in (-1, 0). \tag{44}$$

Therefore, correlations can emerge even in the absence of transcriptional synchronization. To account for these correlations, it is necessary to build mixture models, investigate living systems with low cell type diversity, or use ad hoc filtering to extract putative homogeneous cell populations.

Example: Extrinsic noise in burst frequency

Even within a homogeneous cell type, the parameter values may not be perfectly synchronized. We can thus define *extrinsic* noise, which encodes

the stochastic parameter distribution (24,59). This description presupposes that the noise, if generated by a stochastic process, evolves much slower than the transcriptional dynamics (47). There does not appear to be a general solution for parameter mixing, but a set of solutions is available for systems with random burst frequencies. Suppose there is a single parent transcript with a burst frequency described by the random variable K_i. We can immediately write down a solution for the overall count PGF:

$$\begin{aligned} G(\mathbf{x},t) &= \int_0^\infty G(\mathbf{x},t;k_i)df_{K_i} = \int_0^\infty e^{k_i\varphi^*(\mathbf{x},t)}df_{K_i} \\ &= M_K[\varphi^*(\mathbf{x},t)], \end{aligned} \tag{45}$$

where M_{K_i} is the MGF of the burst frequency distribution and φ^* is a nominal log-PGF with unity burst frequency. The degenerate case is recovered by constraining K_i to a Dirac delta distribution at k_i, with MGF $M_{K_i}(z) = e^{k_i z}$. Equation 45 immediately extends to integrals of Eq. 18 by defining a multivariate burst frequency distribution and computing its MGF.

Example: Extrinsic noise in burst size

Burst size modulation has been heavily implicated in the regulation of cell size (60–62). For systems with a distribution of *static* cell sizes, not governed by the cell cycle, we can at least formally postulate a model:

$$G(\mathbf{x},t) = \int_0^\infty G(\mathbf{x},t;\mathbf{b}(z))df_Z(z), \tag{46}$$

where $\mathbf{b}$ is the vector of average burst sizes, governed by a common *univariate* random variable Z with realizations z, e.g., $b_i = zc_i$. Z thus models the "cell size" in a mechanistic fashion. Unfortunately, despite this relevance to physiology, this general form is intractable: the burst size parameter cannot be "factorized" out. However, certain special cases can be connected to previous work.

Consider the case of a single mRNA species produced in geometric bursts with a random cell-specific burst size. If the inverse burst size distribution is given by the law $\mathcal{Z}$, the distribution of mRNA is given by the $\mathcal{Z}$ -gamma-Poisson PMF. Only a few $\mathcal{Z}$ -gamma distributions have been studied in depth. If $\mathcal{Z}$ is gamma (i.e., the burst size is inverse-gamma distributed), gamma-gamma compounding yields a beta distribution of the second kind (63), with a rather complicated MGF available in terms of Lauricella functions (64).

Transient burst dynamics are solvable

Thus far, we have primarily focused on stationary systems with time-independent parameters. Nevertheless, there are classes of physiological phenomena, such as differentiation and cell cycling, where transient behaviors are crucial, particularly since these processes occur on timescales comparable with the mRNA lifetimes (1,65). Usually, the regulatory events underpinning these processes are modeled by variation in DNA-localized transcriptional parameters (66–69).

By examining Eq. 18, it is easy to see that the current framework can be extended to *any* deterministic variation in k and M, with solutions available at arbitrary times t:

$$\varphi(\mathbf{u},t) = \int_0^t \sum_{\ell=1}^{n} \binom{n}{\ell} \sum_{q=1} k_{\ell,q}(s)\left(M_{\ell,q}(\mathbf{U}(\mathbf{u},s),s)-1\right)ds. \tag{47}$$

Therefore, burst frequencies, burst distributions, and even the coexpression modules can vary, continuously or discontinuously, as long as the variation is deterministic. This can be exploited to model a variety of phenomena, such as variation in gene copy numbers over the cell cycle (62,66) and the concentration of high-abundance regulators not coupled to the mRNA under investigation. If the reaction rates within $\mathbf{U}$ change over time, the generating function PDE becomes intractable; however, some simple models, such as piecewise constant, can be solved by splitting the integral.

We have further assumed $m_i(t=0) = 0$ for all species i. However, if nonzero molecule counts are present at $t = 0$, it is straightforward to compute the resulting log-PGF by separately defining the homogeneous generating function φ^h with $m_i(t=0) = 0$ and the generating function of the initial condition φ^{init}:

$$\varphi(\mathbf{u},t) = \varphi^h(\mathbf{u},t) + \varphi^{init}(\mathbf{U}(\mathbf{u},t)) \tag{48}$$

$$\varphi(\mathbf{u},t) = \varphi^h(\mathbf{u},t) + \sum_{i=1}^{n} m_i^{(0)}\ln(1+U_i(\mathbf{u},t)), \tag{49}$$

where U_i are the characteristics. Eq. 48 relates the general form of the conditional distribution, whereas Eq. 49 produces the particular form with deterministic initial molecule counts $m_i^{(0)}$, as discussed elsewhere (70). Therefore, the current approach can be used to compute the likelihoods of entire trajectories of observations, and thus perform parameter estimation on live-cell data.

Example: Cell cycling

Using the conditioning relation in Eq. 49, we can solve models of certain cell-cycle phenomena. Suppose that a "cell cycle" consists of two stages with durations Δ_1 and Δ_2 and parameters $b^{(j)}, \gamma^{(j)}, k_1^{(j)}$ in stage j. The PGF of the joint transcript distribution at the end of the first stage, $t = \Delta_1$, is given by Eq. 48, yielding the PGF $G^{(1)}(\mathbf{u})$. Now supposing the stage transition involves the binomial partitioning of all molecules, the distribution of mRNA upon entering the second stage is $G^{(1)}(1+\mathbf{u}/2)$. This identity results from the law of total expectation, and amounts to asserting that the probability of retaining each molecule is $1/2$, with a per-molecule Bernoulli retention PGF of $\frac{1}{2}(1+x)$. This result has been derived in previous models of cell cycling (25,62) and proposed in more generality in our recent discussion of technical noise (16).

Considering the specific example of a single transcript with initial PGF $G^{(0)}$, we can adapt Eq. 11 (44) to yield:

$$G^{(1)}(u,\Delta_1) = G^{(0)}\left(ue^{-\gamma^{(1)}\Delta_1}\right)\left(\frac{1-b^{(1)}ue^{-\gamma^{(1)}\Delta_1}}{1-b^{(1)}u}\right)^{\frac{k_1^{(1)}}{\gamma^{(1)}}}. \tag{50}$$

This is the generating function immediately before partitioning. The generating function immediately after simply inserts $\frac{1}{2}(2+u)$ in place of u to account for partitioning. Finally, to compute the generating function at the end of the cycle, we multiply through by the transient PGF and insert the characteristic $ue^{-\gamma^{(2)}t}$ into the initial condition:

$$\begin{aligned} G^{(2)}(u,\Delta_1+\Delta_2) = {} & G^{(0)}\left(\frac{1}{2}\left(2+ue^{-\gamma^{(2)}\Delta_2}\right)e^{-\gamma^{(1)}\Delta_1}\right) \\ & \times\left(\frac{1-b^{(1)}\frac{1}{2}\left(2+ue^{-\gamma^{(2)}\Delta_2}\right)e^{-\gamma^{(1)}\Delta_1}}{1-b^{(1)}\frac{1}{2}\left(2+ue^{-\gamma^{(2)}\Delta_2}\right)}\right)^{\frac{k_1^{(1)}}{\gamma^{(1)}}} \\ & \times\left(\frac{1-b^{(2)}ue^{-\gamma^{(2)}\Delta_2}}{1-b^{(2)}u}\right)^{\frac{k_1^{(2)}}{\gamma^{(2)}}}. \end{aligned} \tag{51}$$

Although solutions of this type are somewhat unwieldy to write down explicitly, they *can* always be computed by composition of functions for the one-species system.

The general solution, with arbitrary burst distributions, splicing networks, and deterministic regulatory dynamics, can be computed by working backward from the last stage of cell cycle and incrementally adding "initial condition" contributions from previous stages. As an illustration, we define a system with zero molecules at $t = 0$, with a bursty promoter that produces a single parent isoform with the characteristic $U_1(\mathbf{u},s)$. For convenience, we assume the splicing and degradation are time independent. We can immediately write down the generating function for the system state before partitioning at $t = \Delta_1 + \Delta_2$:

$$\varphi^{(2)}(\mathbf{u}, \Delta_1 + \Delta_2) = \left(k_1^{(1)} \int_0^{\Delta_1} \left[\frac{1}{1 - b^{(1)} U_1\left(\frac{1}{2}(2 + \mathbf{U}(\mathbf{u}, \Delta_2)), s\right)} - 1 \right] ds \right) + \left(k_1^{(2)} \int_0^{\Delta_2} \left[\frac{1}{1 - b^{(2)} U_1(\mathbf{u}, s)} - 1 \right] ds \right). \quad (52)$$

This generalizes the systems studied previously (62) by relaxing the assumption of timescale separation between transcripts, while maintaining the assumption of bursty gene dynamics; transient coupling between cell size and burst size (such as the exponential form in (62)) can be incorporated by appropriately defining a time-varying burst distribution in Eq. 47. The solution has previously been extended to describe Erlang-distributed stochastic time intervals (25,71). This requires defining a CME coupled to a multistate Markov chain governing the transcriptional parameters and yields a series of coupled PDEs that take a form reminiscent of multistate promoter equations (72), but are not generally tractable by quadrature.

RESULTS

Ultimately, we would like to use these solutions to fit real data, and represent entire data sets using a small set of physically interpretable parameters for each gene, potentially with some higher-level structure encoding cell types (as in Eq. 41) or gene-gene synchronization (as in Eq. 29). Unfortunately, the experimental and computational infrastructure to do this does not exist yet: as we discuss at length in section S5, single-cell, single-molecule, full-length sequencing methods are in their nascence, the structure of splicing processes is not well characterized on a genome-wide scale, and both biology and sequencing involve obscure sources of noise.

However, we *can* take the first steps in this direction by exploiting some of the simpler results to explain correlations in real data. In Eqs. 33 and 29, we report two models that give positive correlations for coexpressed transcripts. We cannot yet use these models to recapitulate entire data sets. The physics are too complex to explicitly describe: the model is a priori misspecified because we disregard other sources of noise. Furthermore, the inference procedure requires some care, as the joint distributions are too large to compute. If we try to fit distributions two at a time, the results will be inconsistent (e.g., fitting genes 1 and 2 will yield an estimated $\widehat{b}_1$ different from the one obtained by fitting genes 1 and 3), and still suffer from model misspecification.

We can sacrifice some statistical power but improve interpretability by treating the genes one at a time. The models' correlation structure, given in Eq. 31, has no free parameters: it is fully determined by the marginal distributions. Given marginal estimates of gene-specific parameters (straightforward to compute by maximum likelihood estimation with the negative binomial law), we can *predict* the correlation structure and compare predictions to experimental ground truth. We interpret these predictions as *upper bounds* in the absence of all other noise: if additional sources of stochasticity are present, the correlations should degrade relative to the model. As the correlations predicted by Eq. 31 are strictly less than 1, they are nontrivial.

To perform this analysis, we obtained data from the recent FLT-seq (full-length transcript sequencing by sampling) protocol (73). As this experimental technique has molecular and cellular barcodes, the data are interpretable as discrete transcript counts sampled from a distribution. To minimize transient effects, such as cell cycling and differentiation, we selected a data set generated from cultured mouse stem cells. To limit biological heterogeneity due to discrete cell subpopulations (as in Eq. 41), we filtered for cell barcodes corresponding to the activated cell subset (136 barcodes) according to the authors' annotations. In all downstream analyses, we treated this filtered data set as biologically homogeneous up to intrinsic stochasticity.

The FLT-seq protocol produces full-length reads, which can be used to discover new isoforms, but does not reveal causal relationships between those isoforms. Nevertheless, we can use the tools of discrete mathematics to partially infer these relationships. Splicing removes introns, but cannot insert them. We can use this relationship to constrain the splicing DAG: if transcript $\mathcal{T}_j$ can be obtained by removing part of the sequence in transcript $\mathcal{T}_i$, there must be a path from $\mathcal{T}_i$ to $\mathcal{T}_j$. On the other hand, if $\mathcal{T}_i$ contains the sequence I_i but omits the sequence I_j, whereas $\mathcal{T}_j$ contains the sequence I_j but omits the sequence I_i, the transcripts are *mutually exclusive* and must be generated from the parent transcript by disjoint processes.

For each gene, we enumerate the transcripts observed in the data and split them into elementary intervals, contiguous stretches that are either present or absent in each transcript (denoted by the colors in Fig. 1 *a*). These elementary

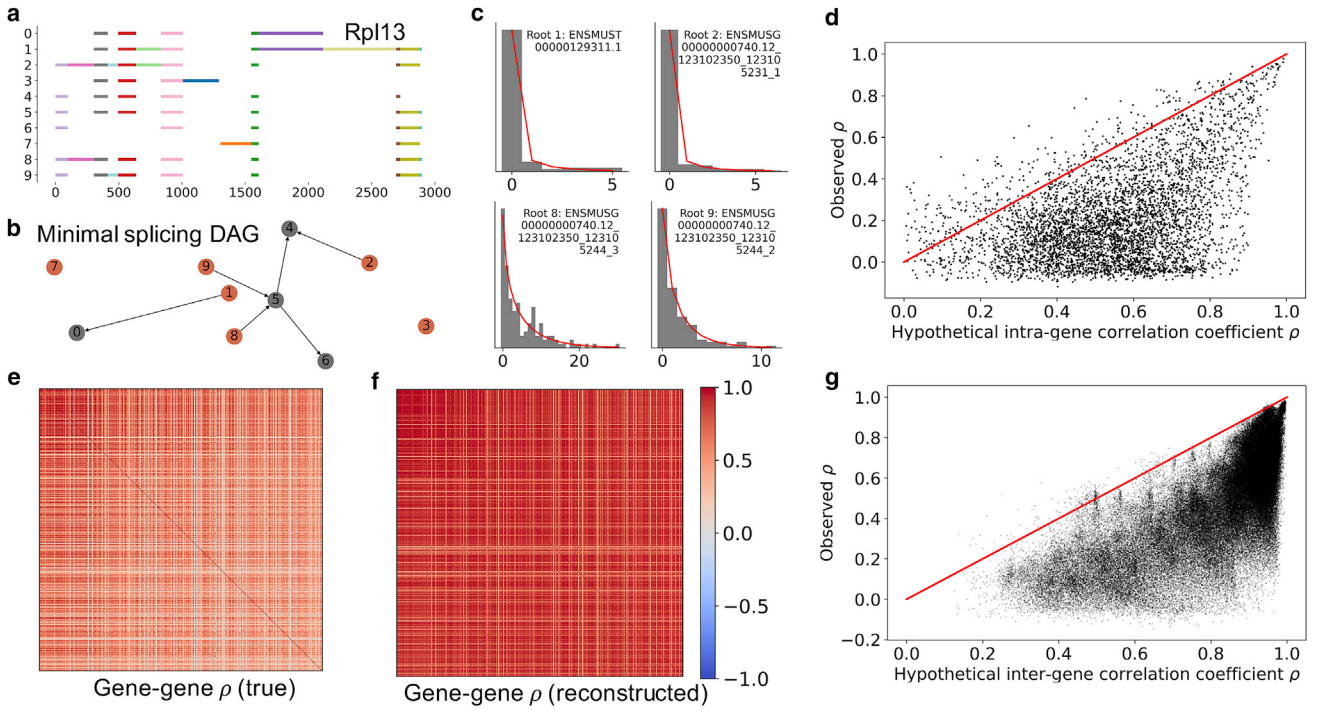

FIGURE 1 Leveraging the synchronized burst model to predict transcript-transcript correlations. (*a*) By inspecting exon coexpression structures in long-read sequencing data, we can split genes into elementary intervals. (*b*) Although sequencing data are not sufficient to identify the relationships between various transcripts, they can provide information about "roots" of the splicing graph (highlighted in orange), which must be produced from the parent transcript by mutually exclusive processes. (*c*) The root transcript copy-number distributions are well described by negative binomial laws (*gray histograms*, raw marginal count data; *red lines*, fits). (*d*) The intrinsic noise model is not sufficient to accurately predict transcript-transcript correlations, but does serve as a nontrivial upper bound: few sample correlations exceed the model-based predictions obtained from Eq. 31 (*points*, transcript-transcript correlation matrix entries for mutually exclusive "root" transcripts of a single gene; *red line*, theory/experiment identity line). (*e*) The highest-expressed transcripts across the top 500 genes show distinctive, and generally positive correlation patterns. (*f*) We can use an analogous model to predict and reconstruct the correlation matrix based solely on marginal data. (*g*) As before, the model (Eq. 31) is not sufficient to accurately predict gene-gene correlations, but provides an effective and nontrivial upper bound (*points*, transcript-transcript correlation matrix entries; *red line*, theory/experiment identity line). To see this figure in color, go online.

intervals constrain the relationships between transcripts, and we can use their presence or absence in each transcript to construct an accessibility graph. The internal structure of this graph is underspecified, but immaterial: the negative binomial model implied by the generating functions in Eqs. 33 and 29 describes the *roots*, mutually exclusive transcripts that must be generated directly from the parent transcript (indicated in orange in Fig. 1 *b*, and solved in Eq. 33). We fit the distributions of these roots, discarding any data that are underdispersed (variance lower than mean), overly sparse (fewer than five molecules in the entire data set), or fail to converge to a fit. The satisfactory fits for the sample gene Rpl13 are shown in Fig. 1 *c*.

The negative binomial fit yields transcript-specific burst sizes bw_i and efflux rates r_i (nondimensionalized by setting burst frequency to unity). We plug these quantities into Eq. 31, compute hypothetical correlations ρ_{theo}, and compare them to sample correlations ρ_{samp} in Fig. 1 *d*. These results represent the 4,885 nontrivial correlation matrix entries between 1,978 transcripts from 500 genes. A total of 302 transcripts were rejected due to underdispersion, 542 due to sparsity, and 100 due to poor fits. The theoretical constraint (sample correlation equal to or lower than predicted correlation) was met in 4,606 cases (94.3%).

The results suggest that the model is not sufficient to recapitulate the full dynamics, but *does* provide an effective, and nontrivial, theoretical constraint. We hypothesize that the "consistent" regime ($\rho_{samp} \in (0, \rho_{theo})$, 3,856 entries) represents the degradation of correlations due to technical noise in the sequencing process and stochastic intermediates. The "inconsistent" regime ($\rho_{samp} \in (\rho_{theo}, 1)$, 279 entries) may stem from model misidentification, and could be explained by coupling between splicing events, resulting in a burst model closer to Eq. 26. Some of these apparently inconsistent correlations may also be due to the small sample sizes, as discussed in section S5.1. Finally, the "negative" regime ($\rho_{samp} < 0$, 750 entries) technically meets the constraint, but cannot actually be reproduced by the model. This does not appear to be an artifact of sample sizes, as evident from the strong negative peak in Figure S8 *a*. Instead, we speculate that enrichment in negative correlations is the signature of a more complicated regulatory schema that preferentially synthesizes some isoforms to the exclusion of others, rather than choosing

the splicing pathway randomly (as encoded in the derivation of Eq. 33).

We can exploit the analogous intergene model, encoded in Eq. 29, to try to predict the gene-gene correlation matrix (Fig. 1 *e*) based solely on the marginals, supposing *all* 500 highest-expressed genes fire simultaneously as a limiting case. For each gene, we consider the highest-abundance root transcript that can be fit by a negative binomial distribution, and identify its marginal burst size and efflux rate. Plugging these parameter estimates into Eq. 26, we obtain theoretical correlations ρ_{theo} and reconstruct the correlation matrix (Fig. 1 *f*). Finally, we compare the intragene sample correlations ρ_{samp} to the theoretical values in Fig. 1 *g*. These results represent the 119,805 nontrivial correlation matrix entries based on the 490 genes with well-fit roots. The theoretical constraint (sample correlation equal to or lower than predicted correlation) was met in 119,503 cases (99.7%).

Yet again, the model provides a nontrivial bound. We hypothesize that the consistent regime ($\rho_{samp} \in (0,\ \rho_{theo})$, 117,542 entries) represents the degradation of correlations due to stochastic effects outside the model, much as before. The correlations in the inconsistent regime ($\rho_{samp} \in (\rho_{theo}, 1)$, 302 entries) lie very close to the identity line, so we hypothesize they are mostly explained by small sample sizes, as discussed in section S5.1. Finally, the "negative" regime ($\rho_{samp}<0$, 1,961 entries) is rare and does not appear to produce a strong signal in the correlation distribution (Figure S8 *b*), so we expect they also emerge from small sample sizes.

CONCLUSIONS

We have described a broad extension of previous work pertaining to monomolecular reaction networks coupled to a bursty transcriptional process. In particular, by exploiting the standard properties of reaction rate equations, we have demonstrated the existence of all moments and cross moments. Furthermore, we have derived the analytical expressions for the generating functions and demonstrated their existence. The following expression gives the general solution for the joint PGF G:

$$G(\mathbf{u},t) \;=\; \mathbb{E}_{\Theta}\left[G^{(0)}(\mathbf{U}(\mathbf{u},t)\,)\exp\left(\int_0^t \sum_{\ell=1}^{n}\sum_{q=1}^{\omega} k_{\ell,q}(s)\big(M_{\ell,q}(\mathbf{U}(\mathbf{u},s),s\,)-1\,\big)ds\right)\right]. \tag{53}$$

This expression is modular with respect to $\mathbf{U}$, the set of characteristics defining the splicing and degradation network, $M_{\ell,q}$, the joint generating functions governing the bursting dynamics for all possible cotranscribing modules, $G^{(0)}$, the initial condition, and f_θ, the law governing the parameter distributions. The burst frequency and distribution can be time dependent, and describe deterministic driving by a latent process or regulator. By iteratively applying the dependence on the initial condition, we can write down analogous expressions for certain cell-cycle processes. The integral in Eq. 53 cannot be solved analytically for any but the simplest $M_{\ell,q}$. However, the form *guarantees* that the dynamics of a general system can be solved using quadrature and do not require matrix-based methods.

Special cases of Eq. 53 *can* be solved to provide useful insights into model complexity necessary to capture the summary statistics of living cells. To attempt to describe high gene-gene correlations, we have investigated several models for synchronized transcription, and found that geometric burst size coordination is *required* to achieve transcript count correlations $\rho>1/2$. Furthermore, we test whether negative correlations are feasible under the assumption of synchronized bursts at multiple gene loci, and find that $\rho<0$ are impossible with geometric bursts, but *can* be achieved with negative binomial bursts. These results substantially constrain and inform the space of models that can recapitulate the combination of bursty dynamics (6) and high absolute gene-gene correlations (2,23) observed in living cells.

We compared the theoretical constraints with experimental data generated by FLT-seq, a recent long-read, single-cell, single-molecule sequencing method. Investigating a set of 500 genes, we found that the constraints were met for 95.3% of the intragene transcript-transcript correlations and 99.7% of the intergene transcript-transcript correlations. Nevertheless, the model was insufficient to recapitulate the precise quantitative details, suggesting that more detailed biophysical models of regulated splicing and technical noise are necessary.

The formal computational complexity of this procedure is $O(\mathcal{N}\log\mathcal{N})$ in state space size. However, this complexity ostensibly arises from the inverse Fourier transform. In the practical regime (with a state space up to approximately 10^3), the scaling is sub-linear; unfortunately, full joint distributions are essentially intractable due to the "curse of dimensionality"—the *space* complexity of holding an array of size $\mathcal{N}$ in memory (potentially requiring petabytes of storage even for small splicing networks, as in section S4.1) for the inverse Fourier transform. Nevertheless, this method can compute marginals without having to solve the entire joint system, which would be intractable with matrix-based methods.

Curiously, this class of analytical solutions to reaction networks can be adapted to a subset of diffusion problems. General diffusion on a multidimensional lattice is not directly solvable, because it violates the assumption of

acyclic graph structure. However, percolation through a DAG coupled to a source and a set of sinks can be described using the current mathematical formalism. Furthermore, such a percolation can represent the incremental movement of RNA polymerase along a DNA strand, integrating discrete copy-number statistics with submolecular details in a single analytical framework (74,75).

Finally, we briefly touch upon the class of delay chemical master equations, and survey several recent advances in the field in section S6. Due to the non-Markovian nature of delayed systems, general probabilistic solutions are rare (76) and represent an open area of study. In our discussion, we motivate delays as a limit of numerous, fast isomerization processes, and clarify the challenges inherent in applying the analysis of delay CMEs to bursty systems.

CODE AVAILABILITY

Google Colab Python notebooks that reproduce the analyses and benchmarking are available at https://github.com/pachterlab/GP_2021_2.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.02.004.

AUTHOR CONTRIBUTIONS

L.P. and G.G. designed the research and wrote the article. G.G. derived and implemented the analytical solutions, validated them against simulations, and performed the sequencing data analysis.

ACKNOWLEDGMENTS

The spectral solution was derived by Meichen Fang. G.G. and L.P. were partially funded by NIH U19MH114830. The DNA illustration used in Figs. S1, S2, and S4, modified from (70), is a derivative of the DNA Twemoji by Twitter, Inc., used under CC-BY 4.0. The directed acyclic graph generation code was adapted from the IPython Parallel reference documentation: https://ipyparallel.readthedocs.io/en/latest/dag_dependencies.html.

SUPPORTING CITATIONS

References (77–85) appear in the supporting material.

REFERENCES

1. La Manno, G., R. Soldatov, …, P. V. Kharchenko. 2018. RNA velocity of single cells. *Nature.* 560:494–498. http://www.nature.com/articles/s41586-018-0414-6.
2. Shalek, A. K., R. Satija, …, A. Regev. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 498:236–240. http://www.nature.com/articles/nature12172.
3. Svensson, V., R. Vento-Tormo, and S. A. Teichmann. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13:599–604. http://www.nature.com/articles/nprot.2017.149.
4. Peccoud, J., and B. Ycard. 1995. Markovian modeling of gene product synthesis. *Theor. Popul. Biol.* 48:222–234. https://linkinghub.elsevier.com/retrieve/pii/S0040580985710271.
5. Gardiner, C. 2004. Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, 3rd. Springer, pp. 145–147.
6. Dar, R. D., B. S. Razooky, …, L. S. Weinberger. 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U S A.* 109:17454–17459. http://www.pnas.org/cgi/doi/10.1073/pnas.1213530109.
7. Sanchez, A., and I. Golding. 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science.* 342:1188–1193. http://www.sciencemag.org/cgi/doi/10.1126/science.1242975.
8. Bokes, P. 2021. Exact and WKB-approximate distributions in a gene expression model with feedback in burst frequency, burst size, and protein stability. Discrete Continuous Dynamical Syst. - B 0:0. https://www.aimsciences.org/article/doi/10.3934/dcdsb.2021126.
9. Sugár, I., and I. Simon. 2014. Self-regulating genes. exact steady state solution by using Poisson representation. *Open Phys.* 12.9:615–627. http://www.degruyter.com/view/j/phys.2014.12.issue-9/s11534-014-0497-0/s11534-014-0497-0.xml.
10. M Stoeckius, M., C. Hafemeister, and P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods.* 14:865–868, https://www.nature.com/articles/nmeth.4380.
11. Peterson, V. M., K. X. Zhang, …, J. A. Klappenbach. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–939. http://www.nature.com/articles/nbt.3973.
12. Chung, H., C. N. Parkhurst, …, A. Regev. 2021. Joint single-cell measurements of nuclear proteins and RNA *in vivo*. *Nat. Methods.* 18:1204–1212. https://www.nature.com/articles/s41592-021-01278-1.
13. Zheng, G. X. Y., J. M. Terry, …, J. H. Bielas. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049. http://www.nature.com/articles/ncomms14049.
14. Qiu, P. 2020. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* 11:1169. http://www.nature.com/articles/s41467-020-14976-9.
15. Soneson, C., A. Srivastava, …, M. B. Stadler. 2021. Preprocessing choices affect RNA velocity results for droplet scRNA-seq data. *PLoS Comput. Biol.* 17:e1008585. https://dx.plos.org/10.1371/journal.pcbi.1008585.
16. Gorin, G., and L. Pachter. 2021. Length biases in single-cell RNA sequencing of pre-mRNA. *Preprint at bioRxiv*, https://www.biorxiv.org/content/10.1101/2021.07.30.454514v1.
17. Amrhein, L., K. Harsha, and C. Fuchs. 2019. A mechanistic model for the negative binomial distribution of single-cell mRNA counts, Preprint at *bioRxiv* http://biorxiv.org/lookup/doi/10.1101/657619.
18. Neuert, G., B. Munsky, …, A. van Oudenaarden. 2013. Systematic identification of signal-activated stochastic gene regulation. *Science.* 339:584–587. http://www.sciencemag.org/lookup/doi/10.1126/science.1231456.
19. Jia, C., and R. Grima. 2020. Dynamical phase diagram of an auto-regulating gene in fast switching conditions. *J. Chem. Phys.* 152:174110. http://aip.scitation.org/doi/10.1063/5.0007221.
20. Huang, L., Z. Yuan, …, T. Zhou. 2014. Feedback-induced counterintuitive correlations of gene expression noise with bursting kinetics. *Phys. Rev. E.* 90:052702. https://link.aps.org/doi/10.1103/PhysRevE.90.052702.
21. Pratapa, A., A. P. Jalihal, …, T. M. Murali. 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods.* 17:147–154. http://www.nature.com/articles/s41592-019-0690-6.
22. Ezer, D., V. Moignard, …, B. Adryan. 2016. Determining physical mechanisms of gene expression regulation from single cell gene expression data. *PLoS Comput. Biol.* 12:e1005072. http://dx.plos.org/10.1371/journal.pcbi.1005072.
23. Iacono, G., R. Massoni-Badosa, and H. Heyn. 2019. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.*

20:110. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1713-4.

24. Ham, L., R. D. Brackston, and M. P. H. Stumpf. 2020. Extrinsic noise and heavy-tailed laws in gene expression. *Phys. Rev. Lett.* 124:108101. https://link.aps.org/doi/10.1103/PhysRevLett.124.108101.

25. Beentjes, C. H. L., R. Perez-Carrasco, and R. Grima. 2020. Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. *Phys. Rev. E.* 101:032403. https://link.aps.org/doi/10.1103/PhysRevE.101.032403.

26. Kazeev, V., M. Khammash, …, C. Schwab. 2014. Direct solution of the chemical master equation using quantized tensor trains. *PLoS Comput. Biol.* 10:e1003359. https://dx.plos.org/10.1371/journal.pcbi.1003359.

27. Kazeev, V., and C. Schwab. 2015. Tensor approximation of stationary distributions of chemical reaction networks. *SIAM J. Matrix Anal. Appl.* 36:1221–1247. http://epubs.siam.org/doi/10.1137/130927218.

28. Sunkara, V. 2019. On the properties of the reaction counts chemical master equation. *Entropy.* 21:607. https://www.mdpi.com/1099-4300/21/6/607.

29. Cont, R., and P. Tankov. 2004. *In* Financial Modelling With Jump Processes. Chapman & Hall/CRC.

30. Singh, A., and P. Bokes. 2012. Consequences of mRNA transport on stochastic variability in protein levels. *Biophysical J.* 103:1087–1096. https://linkinghub.elsevier.com/retrieve/pii/S0006349512007904.

31. West, D. B. 2001. Introduction to Graph Theory, 2nd. Prentice Hall, Upper Saddle River.

32. Bondy, J. A., and U. S. R. Murty. 2008. Graph Theory. Springer London https://link.springer.com/book/9781846289699.

33. Bokes, P., J. R. King, M. Loose…, 2012. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J. Math. Biol.* 64:829–854. http://link.springer.com/10.1007/s00285-011-0433-5.

34. Jahnke, T., and W. Huisinga. 2006. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.* 54:1–26. http://link.springer.com/10.1007/s00285-006-0034-x.

35. Gardiner, C. W., and S. Chaturvedi. 1977. The Poisson representation. I. A new technique for chemical master equations. *J. Stat. Phys.* 17:429–468. http://link.springer.com/10.1007/BF01014349.

36. Iyer-Biswas, S., and C. Jayaprakash. 2014. Mixed Poisson distributions in exact solutions of stochastic auto-regulation models. *Phys. Rev. E.* 90:052712. https://pubmed.ncbi.nlm.nih.gov/25493821/.

37. Iyer-Biswas, S., F. Hayot, and C. Jayaprakash. 2009. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E.* 79:031911. https://link.aps.org/doi/10.1103/PhysRevE.79.031911.

38. Friedman, N., L. Cai, and X. S. Xie. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97:168302. https://link.aps.org/doi/10.1103/PhysRevLett.97.168302.

39. Bokes, P. 2021. Heavy-tailed distributions in a stochastic gene autoregulation model. *J. Stat. Mech. Theor. Exp.* 2021:113403. https://iopscience.iop.org/article/10.1088/1742-5468/ac2edb.

40. Jia, C., M. Q. Zhang, and H. Qian. 2017. Emergent Levy behavior in single-cell stochastic gene expression. *Phys. Rev. E.* 96:040402. https://link.aps.org/doi/10.1103/PhysRevE.96.040402.

41. Karlis, D., and E. Xekalaki. 2005. Mixed Poisson distributions. *Int. Stat. Rev./Revue Internationale de Statistique.* 73:35–58. http://www.jstor.org/stable/25472639.

42. Panjer, H. H. 2004–. Mixed Poisson Distributions. *In* Encyclopedia of Actuarial Science. J. L. Teugels, B. Sundt, and G. E. Willmot, eds. John Wiley & Sons, Ltd https://onlinelibrary.wiley.com/doi/10.1002/9780470012505.tam022.

43. Barndorff-Nielsen, O. E., and N. Shephard. 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B.* 63:167–241. https://rss.onlinelibrary.wiley.com/doi/10.1111/1467-9868.00282.

44. Petroni, N. C., and P. Sabino. 2020. Gamma Related Ornstein-Uhlenbeck Processes and their Simulation, Preprint at *arXiv* http://arxiv.org/abs/2003.08810.

45. Barndorff-Nielsen, O. E., and N. Shephard. 2003. Integrated OU processes and non-Gaussian OU-based stochastic volatility models. *Scand. J. Stat.* 30:277–295. https://www.jstor.org/stable/4616764.

46. Papadopoulos, N., P. R. Gonzalo, and J. Söding. 2019. PROSSTT: probabilistic simulation of single-cell RNA-seq data for complex differentiation processes. *Bioinformatics.* 35:3517–3519. https://academic.oup.com/bioinformatics/article/35/18/3517/5305637.

47. Gorin, G., J. J. Vastola, M. Fang, and L. Pachter. 2021. Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments, Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.09.06.459173v1.

48. John, F. 1978. Partial Differential Equations. Springer US https://link.springer.com/book/10.1007/978-1-4684-0059-5.

49. Gans, P. J. 1960. Open first-order stochastic processes. *J. Chem. Phys.* 33:691–694. http://aip.scitation.org/doi/10.1063/1.1731239.

50. Golding, I., J. Paulsson, …, E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036. https://linkinghub.elsevier.com/retrieve/pii/S0092867405010378.

51. Cannoodt, R., W. Saelens, Y. Saeys…, 2021. Spearheading future omics analyses using *dyngen*, a multi-modal simulator of single cells. *Nat. Commun.* 12:3942. http://www.nature.com/articles/s41467-021-24152-2.

52. Xu, H., L. A. Sepúlveda, …, I. Golding. 2015. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nat. Methods.* 12:739–742. http://www.nature.com/articles/nmeth.3446.

53. Huang, P., E. D. Pleasance, S. J. Jones…, 2007. Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. *Genome Res.* 17:1478–1485. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1987351/.

54. Moran, P. 1969. Statistical inference with bivariate gamma distributions. *Biometrika.* 56:627–634. https://academic.oup.com/biomet/article-abstract/56/3/627/233798.

55. Yue, S., T. Ouarda, and B. Bobée. 2001. A review of bivariate gamma distributions for hydrological application. *J. Hydrol.* 246:1–18. https://linkinghub.elsevier.com/retrieve/pii/S0022169401003742.

56. Blitzstein, J. K., and J. Hwang. 2015. Introduction to probability. Texts in statistical science. CRC Press, Taylor & Francis Group.

57. Lindsay, B. G. 1995. Mixture models: theory, geometry and applications. *NSF-CBMS Reg. Conf. Ser. Probab. Stat.* 5:i–163. https://www.jstor.org/stable/4153184.

58. Singer, Z., J. Yong, …, M. Elowitz. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell.* 55:319–331. https://linkinghub.elsevier.com/retrieve/pii/S1097276514005632.

59. Ham, L., D. Schnoerr, …, M. P. H. Stumpf. 2020. Exactly solvable models of stochastic gene expression. *J. Chem. Phys.* 152:144106. http://aip.scitation.org/doi/10.1063/1.5143540.

60. Sun, X.-M., A. Bowman, …, S. Marguerat. 2020. Size-dependent increase in RNA polymerase II initiation rates mediates gene expression scaling with cell size. *Curr. Biol.* 30:1217–1230.e7. https://linkinghub.elsevier.com/retrieve/pii/S096098222030097X.

61. Padovan-Merhar, O., G. Nair, …, A. Raj. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell.* 58:339–352. https://linkinghub.elsevier.com/retrieve/pii/S1097276515001707.

62. Cao, Z., and R. Grima. 2020. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci. U S A.* 117:4682–4692. http://www.pnas.org/lookup/doi/10.1073/pnas.1910888117.

63. Dubey, S. D. 1970. Compound gamma, beta and F distributions. *Metrika.* 16:27–31. http://link.springer.com/10.1007/BF02613934.

64. Pham-Gia, T., and Q. Duong. 1989. The generalized beta- and F-distributions in statistical modelling. *Math. Computer Model.* 12:1613–1625. https://linkinghub.elsevier.com/retrieve/pii/0895717789903373.

65. Milo, R., and R. Phillips. 2015. Cell Biology by the Numbers. *Garland Science* https://www.taylorfrancis.com/books/mono/10.1201/9780429258770/cell-biology-numbers-ron-milo-rob-phillips.

66. Skinner, S. O., H. Xu, …, I. Golding. 2016. Single-cell analysis of transcription kinetics across the cell cycle. *eLife.* 5:e12175. https://elifesciences.org/articles/12175.

67. Dattani, J., and M. Barahona. 2017. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *J. R. Soc. Interf.* 14:20160833. https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0833.

68. Briggs, J. A., C. Weinreb, …, A. M. Klein. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science.* 360:eaar5780. http://www.sciencemag.org/lookup/doi/10.1126/science.aar5780.

69. Zeisel, A., W. J. Kostler, …, E. Domany. 2011. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.* 7:529. http://msb.embopress.org/cgi/doi/10.1038/msb.2011.62.

70. Gorin, G., and L. Pachter. 2020. Special function methods for bursty models of transcription. *Phys. Rev. E.* 102:022409. https://link.aps.org/doi/10.1103/PhysRevE.102.022409.

71. Perez-Carrasco, R., C. Beentjes, and R. Grima. 2020. Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *J. R. Soc. Interf.* 17:20200360. https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0360.

72. Zhou, T., and J. Zhang. 2012. Analytical results for a multistate gene model. *SIAM J. Appl. Mathematics.* 72:789–818. http://epubs.siam.org/doi/10.1137/110852887.

73. Tian, L., J. S. Jabbari, M. E. Ritchie…, 2020. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol.* 22:310. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02525-6.

74. Xu, H., S. O. Skinner, …, I. Golding. 2016. Stochastic kinetics of nascent RNA. *Phys. Rev. Lett.* 117:128101. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.117.128101.

75. Gorin, G., M. Wang, …, H. Xu. 2020. Stochastic simulation and statistical inference platform for visualization and estimation of transcriptional kinetics. *PLoS One.* 15:e0230736. https://dx.plos.org/10.1371/journal.pone.0230736.

76. Leier, A., and T. T. Marquez-Lago. 2015. Delay chemical master equation: direct and closed-form solutions. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* 471:20150049. https://royalsocietypublishing.org/doi/10.1098/rspa.2015.0049.

77. MacDonald, N. 1978. Time Lags in Biological Models. Springer-Verlag, Berlin, Heidelberg http://link.springer.com/10.1007/978-3-642-93107-9.

78. Burrage, K., T. Tian, and P. Burrage. 2004. A multi-scaled approach for simulating chemical reaction systems. *Prog. Biophys. Mol. Biol.* 85:217–234. https://linkinghub.elsevier.com/retrieve/pii/S007961070400029X.

79. Gedeon, T., and P. Bokes. 2012. Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophysical J.* 103:377–385. https://linkinghub.elsevier.com/retrieve/pii/S0006349512006820.

80. Miekisz, J., J. Poleszczuk, …, U. Foryś. 2011. Stochastic models of gene expression with delayed degradation. *Bull. Math. Biol.* 73:2231–2247. http://link.springer.com/10.1007/s11538-010-9622-4.

81. Fatehi, F., Y. N. Kyrychko, and K. B. Blyuss. 2020. A new approach to simulating stochastic delayed systems. *Math. Biosciences.* 322:108327. https://linkinghub.elsevier.com/retrieve/pii/S0025556420300225.

82. Barrio, M., K. Burrage, …, T. Tian. 2006. Oscillatory regulation of hes1: discrete stochastic delay modelling and simulation. *PLoS Comput. Biol.* 2:e117. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0020117.

83. Lafuerza, L. F., and R. Toral. 2011. Exact solution of a stochastic protein dynamics model with delayed degradation. *Phys. Rev. E.* 84:051121. https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.051121.

84. Lafuerza, L. F., and R. Toral. 2011. Role of delay in the stochastic creation process. *Phys. Rev. E.* 84:021128. https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.021128.

85. Jia, T., and R. V. Kulkarni. 2011. Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. *Phys. Rev. Lett.* 106:058102. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.106.058102.

Biophysical Journal, Volume *121*

Supplemental information

Modeling bursty transcription and splicing with the chemical master equation

Gennady Gorin and Lior Pachter

Modeling bursty transcription and splicing with the chemical master equation: Supplementary Note

Gennady Gorin[1] and Lior Pachter[2]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA
[2]Division of Biology and Biological Engineering & Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

January 18, 2022

Contents

S1 CME definition

In the current section, we work through the definition of the CME encoding the reactions in Equation 1 and convert it to the generating function PDE. For the most part, this amounts to rote application of definitions.

S1.1 Converting a reaction network to a CME

The CME is a continuity equation defined with respect to total probability density. For *any* set of states, we can write down the following relation tracking their total probability mass:

$$\text{accumulation} = \text{influx} - \text{efflux} \tag{S1}$$

Probability mass can be neither created nor destroyed, because the total probability mass must add to unity. We can encode the Markovian property by ensuring that accumulation at time t depends only upon the state of the system at t. To compute the likelihood of state $\mathbf{m}$ with m_i molecules of mRNA species $\mathcal{T}_i$, we need to evaluate $P(\mathbf{m}, t)$. This *microstate* $\mathbf{m}$ is the most natural and conventional mathematical object to track using Equation S1. This yields the general form of the CME:

$$\frac{dP(\mathbf{m},t)}{dt} = \sum_{\text{rxn}} \text{influx}(\mathbf{m}, t, \text{rxn}) - \sum_{\text{rxn}} \text{efflux}(\mathbf{m}, t, \text{rxn}) \tag{S2}$$

We can explicitly write down the influx and efflux terms by splitting the reactions into three separate pools of channels: the transcription reactions, the degradation reactions, and the splicing reactions. It is easiest to start with the first-order efflux reactions, as they are only involve a single species i and can be decomposed into an additive form:

$$\begin{aligned} \text{efflux}(\mathbf{m}, t, \text{deg}) &= \sum_{i=1}^{n} \text{efflux}(m_i, t, \text{deg}) = \sum_{i=1}^{n} c_{i0} m_i P(m_i, t) \\ \text{efflux}(\mathbf{m}, t, \text{splic}) &= \sum_{i=1}^{n} \text{efflux}(m_i, t, \text{splic}) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} m_i P(m_i, t), \end{aligned} \tag{S3}$$

a simple form that results immediately from the propensities defined in Equation 1. Since each term in the sums only involves a single i, we use a shorthand with respect to the microstate $\mathbf{m}$, such that $P(m_i)$ is a *multivariate* PMF defined as $P(m_1, ..., m_i, ..., m_n, t)$, and implies that all dimensions

$j \neq i$ are in the state m_j. With this shorthand, we can write down the equations for the influx due to the first-order reactions:

$$\begin{aligned}\text{influx}(\mathbf{m}, t, \text{deg}) &= \sum_{i=1}^{n} \text{influx}(m_i, t, \text{deg}) = \sum_{i=1}^{n} c_{i0}(m_i + 1)P(m_i + 1, t) \\ \text{influx}(\mathbf{m}, t, \text{splic}) &= \sum_{i=1}^{n} \text{influx}(m_i, t, \text{splic}) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(m_i + 1)P(m_i + 1, m_j - 1, t),\end{aligned} \tag{S4}$$

which extends the shorthand to describe two species: $P(m_i + 1, m_j - 1, t) := P(m_1, ..., m_i + 1, ..., m_j - 1, ..., m_n, t)$. Equation S4 encodes the fact that the degradation influx channel removes $\mathcal{T}_i$ and receives probability density from states with $m_i + 1$ counts of $\mathcal{T}_i$, whereas the the splicing influx channel converts $\mathcal{T}_i$ to $\mathcal{T}_j$ and receives probability density from states with $m_i + 1$ counts of $\mathcal{T}_i$ and $m_j - 1$ counts of $\mathcal{T}_j$. We sum over all pairs of indices to encode all possible splicing reaction channels.

Finally, we need to write down the equations that define the transcriptional dynamics. To start, we assume that no co-expression occurs. This implies a simple zeroth-order form for the efflux term:

$$\text{efflux}(\mathbf{m}, t, \text{tx}) = \sum_{i=1}^{n} \text{efflux}(m_i, t, \text{tx}) = \sum_{i=1}^{n} k_{1,i} P(m_i, t), \tag{S5}$$

where $k_{1,i}$ is the burst frequency corresponding to the reaction channel producing the one transcript $\mathcal{T}_i$. The influx is slightly more complicated, as burst sizes are defined as random variables on $\mathbb{N}_0$, so the state $\mathbf{m}$ can be reached from *any* state with $m_z \in [0, 1, ..., m_i - 1, m_i]$.

$$\text{influx}(\mathbf{m}, t, \text{tx}) = \sum_{i=1}^{n} \text{influx}(m_i, t, \text{tx}) = \sum_{i=1}^{n} k_{1,i} \sum_{z=0}^{m_i} p_{i,z} P(m_i - z, t), \tag{S6}$$

where z is the burst size and $p_{i,z}$ is the PMF of the burst size random variable B_i corresponding to the reaction channel producing $\mathcal{T}_i$.

This formulation is sufficient if are content to model independent transcriptional dynamics. However, it is *not* sufficient to recapitulate the gene–gene correlations observed in real datasets. Several avenues are available to model them: for example, it is possible to explicitly describe regulatory interactions [1]. Unfortunately, such detailed schema are not analytically tractable. Instead, we propose that some sets of genes fire simultaneously; physiologically, this can be effected by exposure of neighboring loci on DNA, regulation by a common promoter, or activation by a common inducer [2, 3]. We discuss particular mechanisms that can yield these dynamics in Section S1.3. This formulation lends itself to tractable analytical solutions.

Anywhere between 1 and n species can be simultaneously transcribed in synchronized bursts. Further, for a co-expression module size of ℓ transcripts, there are $\binom{n}{\ell}$ possible combinations of *specific* co-expression modules, indexed by q. For a particular module, defined by the tuple ℓ, q, we can define a function $Q_{\ell,q}(j)$, with $j \in \{1, ..., \ell\}$, which returns the ℓ indices of co-transcribed species. For example, if a system has n transcripts, with species indexed 1 and n transcribed simultaneously, we yield $\ell = 2$, $Q_{2,1}(1) = 1$, and $Q_{2,1}(2) = n$. This function permits us to do the "bookkeeping" for the general multivariate form of burst distributions:

$$\text{efflux}(\mathbf{m}, t, \text{tx}) = \sum_{\ell=1}^{n}\sum_{q=1}^{\binom{n}{\ell}} \text{efflux}(\{m_{Q_{\ell,q}}\}, t, \text{tx}) = \sum_{\ell=1}^{n}\sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} P(\{m_{Q_{\ell,q}}\}, t), \tag{S7}$$

where $\{m_{Q_{\ell,q}}\}$ reports the species indices $m_{Q_{\ell,q}(1)}, ..., .m_{Q_{\ell,q}(\ell)}$ involved in a particular reaction. Finally, we need to write down the conservation equations for the influx reactions. These depend on the multivariate burst probability mass $p_{\ell,q,\mathbf{z}}$, where $\mathbf{z}$ is a vector defining burst size microstates:

$$\begin{aligned}\text{influx}(\mathbf{m}, t, \text{tx}) &= \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} \text{influx}(\{m_{Q_{\ell,q}}\}, t, \text{tx}) \\ &= \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} \sum_{\{\mathbf{z}\}} p_{\ell,q,\mathbf{z}} P(\{m_{Q_{\ell,q}} - z_{Q_{\ell,q}}\}, t) \\ &= \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} \sum_{\{\mathbf{z}\}} p_{\ell,q,\mathbf{z}} P(\mathbf{m} - \mathbf{z}, t)\end{aligned} \tag{S8}$$

where the summation over $\{\mathbf{z}\}$ includes all joint burst sizes with marginal $z_{Q_{\ell,q}}$ up to $m_{Q_{\ell,q}}$. All z_j where $j \notin \{Q_{\ell,q}\}$ are set to zero for consistency. This equation is exact, but somewhat formal. We can illustrate its specific form by returning to the specific case of a system with n transcripts, with co-expressed species $\mathcal{T}_1$ and $\mathcal{T}_n$:

$$\begin{aligned}\text{influx}(\mathbf{m}, t, \text{tx}) &= \text{influx}(m_1, m_n, t, \text{tx}) \\ &= k_{2,1} \sum_{\{\mathbf{z}\}} p_{2,1,\mathbf{z}} P(m_1 - z_1, m_n - z_n, t) \\ &= k_{2,1} \sum_{z_1=0}^{m_1} \sum_{z_n=0}^{m_n} p_{2,1,z_1,z_n} P(m_1 - z_1, m_n - z_n, t)\end{aligned} \tag{S9}$$

Therefore, the most general CME we consider takes the following form:

$$\begin{aligned}\frac{dP(\mathbf{m}, t)}{dt} = &\sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} \left[\sum_{\{\mathbf{z}\}} p_{\ell,q,\mathbf{z}} P(\mathbf{m} - \mathbf{z}, t) - P(\mathbf{m}, t) \right] \\ &+ \sum_{i=1}^{n} c_{i0} \left[(m_i + 1) P(m_i + 1, t) - m_i P(m_i, t) \right] \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \left[(m_i + 1) P(m_i + 1, m_j - 1, t) - m_i P(m_i, t) \right]\end{aligned} \tag{S10}$$

S1.2 Converting a CME to a PDE

With the CME in hand, we can write down the probability-generating function. The multivariate PGF is defined as follows:

$$G(\mathbf{x}, t) = \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} P(m_1, ..., m_n, t) \prod_{i=1}^{n} x_i^{m_i} := \sum_{m_1, ..., m_n} P(\mathbf{m}, t) \prod_{i=1}^{n} x_i^{m_i} \tag{S11}$$

We can sum over the left-hand side of Equation S10:

$$\sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \frac{dP(m_1,...,m_n,t)}{dt} \prod_{i=1}^{n} x_i^{m_i} = \frac{d}{dt} \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} P(m_1,...,m_n,t) \prod_{i=1}^{n} x_i^{m_i} = \frac{\partial G(\mathbf{x},t)}{\partial t} \quad \text{(S12)}$$

As Equation S11 is linear with respect to P, we can treat the interior terms of the summations separately. For a particular combination of ℓ, q, we yield:

$$\sum_{m_1,...,m_n} k_{\ell,q} P(\mathbf{m},t) \prod_{i=1}^{n} x_i^{m_i} = k_{\ell,q} G(\mathbf{x},t), \quad \text{(S13)}$$

i.e., the efflux term of a particular bursting reaction.
Now, taking the PGF and differentiating it with respect to x_i:

$$\begin{aligned} \frac{\partial G(\mathbf{x},t)}{\partial x_i} &= \frac{\partial}{\partial x_i} \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} P(m_1,...,m_n,t) \prod_{k=1}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} P(m_1,...,m_n,t) m_i x_i^{m_i-1} \prod_{k=1,k\neq i}^{n} x_k^{m_k} \end{aligned} \quad \text{(S14)}$$

This implies three useful identities. First, the probability of any state with a negative number of molecules is zero. Therefore, we can reindex the summation over m_i:

$$\begin{aligned} \frac{\partial G(\mathbf{x},t)}{\partial x_i} &= \sum_{m_1=0}^{\infty} ... \sum_{m_i=1}^{\infty} ... \sum_{m_n=0}^{\infty} m_i P(m_1,...,m_n,t) x_i^{m_i-1} \prod_{k=1,k\neq i}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_i=0}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i+1) P(m_1,...,m_i+1,...,m_n,t) x_i^{m_i} \prod_{k=1,k\neq i}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_i=0}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i+1) P(m_1,...,m_i+1,...,m_n,t) \prod_{k=1}^{n} x_k^{m_k}, \end{aligned} \quad \text{(S15)}$$

i.e., the PGF of $(m_i+1)P(m_i+1,t)$ is $\frac{\partial G(\mathbf{x},t)}{\partial x_i}$. This gives us the influx terms of all degradation reactions:

$$\sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} c_{i0}(m_i+1) P(m_1,...,m_i+1,...,m_n,t) \prod_{i=1}^{n} x_i^{m_i} = c_{i0} \frac{\partial G(\mathbf{x},t)}{\partial x_i}. \quad \text{(S16)}$$

Second, we can multiply x_i through the second line of Equation S14:

$$\begin{aligned} x_i \frac{\partial G(\mathbf{x},t)}{\partial x_i} &= x_i \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} m_i P(m_1,...,m_n,t) x_i^{m_i-1} \prod_{k=1,k\neq i}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} m_i P(m_1,...,m_n,t) x_i^{m_i} \prod_{k=1,k\neq i}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} m_i P(m_1,...,m_n,t) \prod_{k=1}^{n} x_k^{m_k}, \end{aligned} \quad \text{(S17)}$$

i.e., the PGF of $m_i P(\mathbf{m})$ is $x_i \frac{\partial G(\mathbf{x},t)}{\partial x_i}$. This gives us the efflux terms of all degradation and splicing reactions:

$$\sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} c_{ij} m_i P(m_1, ..., m_n, t) \prod_{i=1}^{n} x_i^{m_i} = c_{ij} x_i \frac{\partial G(\mathbf{x}, t)}{\partial x_i}. \tag{S18}$$

Third, we can multiply x_j through the relation from the third line of Equation S15:

$$\begin{aligned} x_j \frac{\partial G(\mathbf{x}, t)}{\partial x_i} &= x_j \sum_{m_1=0}^{\infty} ... \sum_{m_j=0}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i + 1) P(m_1, ..., m_i + 1, ..., m_n, t) \prod_{k=1}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_j=0}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i + 1) P(m_1, ..., m_i + 1, ..., m_j, ..., m_n, t) x_j^{m_j+1} \prod_{k=1, k\neq j}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_j=1}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i + 1) P(m_1, ..., m_i + 1, ..., m_j - 1, ..., m_n, t) x_j^{m_j} \prod_{k=1, k\neq j}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} (m_i + 1) P(m_1, ..., m_i + 1, ..., m_j - 1, ..., m_n, t) \prod_{k=1}^{n} x_k^{m_k}, \end{aligned} \tag{S19}$$

where the last step exploits the physical constraint that probabilities of microstates with negative counts are strictly zero, just as in Equation S15. Therefore, the PGF of $(m_i+1)P(m_i+1, m_j-1, t)$ is $x_j \frac{\partial G(\mathbf{x},t)}{\partial x_i}$. This gives us the influx terms of all splicing reactions:

$$\sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} c_{ij} (m_i + 1) P(m_1, ..., m_i + 1, ..., m_j - 1, ..., m_n, t) \prod_{i=1}^{n} x_i^{m_i} = c_{ij} x_j \frac{\partial G(\mathbf{x}, t)}{\partial x_i}. \tag{S20}$$

S1.3 Joint burst distributions can emerge from a model of synchronized gene regulation

The specific form of the burst distribution is governed by the underlying physics. It is conventional [4] to assume that a promoter indexed by i can exist in one of two states, $\mathcal{G}_{i,on}$ or $\mathcal{G}_{i,off}$, where only $\mathcal{G}_{i,on}$ can transcribe. This premise yields the following reactions for each promoter:

$$\begin{aligned} \mathcal{G}_{i,off} &\underset{k_{i,off}}{\overset{k_{i,on}}{\rightleftarrows}} \mathcal{G}_{i,on} \\ \mathcal{G}_{i,on} &\xrightarrow{k_{i,init}} \mathcal{G}_{i,on} + \mathcal{T}_i \end{aligned} \tag{S21}$$

The transcriptional strength of the promoter is described by the *telegraph process* [5], which we denote as K_t. This process is the simplest nontrivial continuous-time, discrete-space Markov chain, which simply switches between two states. If $k_{i,off}, k_{i,init} \to \infty$, the dynamics become bursty, producing multiple transcripts at burst arrival times. The distribution of the number of transcripts is geometric. This result is well-known [5], and has been derived using a variety of mathematical tools [6, 7]. For our purposes, it is easiest to use the Poisson formulation. The following SDE governs the Poisson intensity of $\mathcal{T}_i$:

$$d\Lambda_i = -r_i \Lambda_i dt + K_t dt. \tag{S22}$$

In the bursty limit, K_t is zero almost everywhere, because the off state does not transcribe. It has point masses distributed throughout the time coordinate, with an $Exp(k_{i,on})$ waiting time distribution between their arrivals. The point masses have weights $k_{i,init} \times T$, where T is the duration of the on period, a random variable with an $Exp(k_{i,off})$ distribution. Therefore, the weights are distributed per $Exp(k_{i,off}/k_{i,init})$, i.e., an average of $k_{i,init}/k_{i,off}$. This is precisely the jump size of the *burst subordinator* associated with promoter i. K_t thus takes the following form:

$$K_t = \sum_{q=0}^{N(t)} \delta(t - \tau_q) J_q, \tag{S23}$$

where $N(t)$ is a Poisson counting process with rate $k_{i,on}$, τ_q are the arrival times, and $J_q \sim Exp(k_{i,init}/k_{i,off})$ is the random variable giving the jump sizes or point mass weights. Finally, we note that $L_t := \int_0^t K_s ds$ is a compound Poisson process and thus a legitimate subordinator [8]. Therefore, we can redefine this (somewhat uninformative) form of $K_t dt$ as dL_t, with $k_{1,i} \leftarrow k_{i,on}$ and $b_i \leftarrow k_{i,init}/k_{i,off}$.

More sophisticated models of transcription can be encoded analogously. First, we can suppose that more than two promoter states exist. One state yields no transcription, while the two other states lie in the bursty limit. This is analogous to Figure SN9a of [9] with $k_{INI,1} = 0$, $k_{23} = k_{32} = 0$, and $k_{21}, k_{31}, k_{INI,2}, k_{INI,3} \to \infty$, with no submolecular details of elongation. In this case, the overall subordinator is a *mixture* of the individual states' subordinators, weighted by their relative on rates.

To explain synchronization between burst sizes, we can invoke a slightly more complex model, analogous to Figure SN9b of [9]. Specifically, we can write down equations for a single global regulator and each of the promoters:

$$\begin{array}{c} \mathcal{G}_{G,off} \underset{k_{G,off}}{\overset{k_{G,on}}{\rightleftarrows}} \mathcal{G}_{G,on} \\ \\ \mathcal{G}_{i,on} \underset{?}{\overset{K_G(t)k_{i,on}}{\rightleftarrows}} \mathcal{G}_{i,off}, \end{array} \tag{S24}$$

where $K_G(t)$ is the telegraph process representing the global regulator. If $k_{G,off}, k_{i,on} \to \infty$, the global "on" periods become infinitesimally short, and the activation of the global regulator leads immediately to the activation of the individual promoters.

It remains to specify how the individual promoters turn off. If we suppose that the promoter dynamics are independent, and each promoter shuts off after a delay $\sim Exp(k_{i,off})$, the jump sizes associated with the promoter-specific subordinators are independent and distributed per $Exp(k_{i,off}/k_{i,init})$. This leads to the model described in the section "Example: Two-gene bursty model, with burst time synchronization" of the main text.

On the other hand, if we suppose that the promoters are forced to shut off precisely when the global regulator shuts off, the jump sizes associated with the promoter-specific subordinators are perfectly correlated and distributed per $Exp(k_{G,off}/k_{i,init})$. These are simply copies of the single random variable governing the global regulator "on" time, rescaled by promoter-specific initiation rates to yields the jump sizes. This leads to the model described in the section "Example: Two-gene bursty model, with partial burst time and size synchronization" of the main text.

S1.4 Joint burst distributions are tractable using Cauchy products

It remains to compute the generating function of the bursting influx term in Equation S10. To start, we can treat the simplest case, where $k_{\ell,q} = 0$ for all $\ell > 1$, i.e., no synchronized bursting occurs. First, we define the burst PGF (based on the PMF with weights $p_{1,i,z}$ for species i) and recall the Cauchy product formula [10]:

$$\begin{aligned} F_i(x) &= \sum_{z_i=0}^{\infty} p_{1,i,z} x^{z_i} \\ &\sum_{n=0}^{\infty} a_n x^n \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} c_n x^n \text{ s.t. } c_n = \sum_{k=0}^{n} a_k b_{n-k} \\ &\implies \sum_{n=0}^{\infty} a_n \sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k}, \end{aligned} \tag{S25}$$

Then, we write down the summation over the influx term and rearrange terms:

$$\begin{aligned} &\sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \sum_{z_i=0}^{m_i} p_{1,i,z_i} P(m_1, ..., m_i - z_i, ..., m_n, t) \prod_{k=1}^{n} x_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \prod_{k=1, k\neq i}^{n} x_k^{m_k} \sum_{m_i=0}^{\infty} \sum_{z_i=0}^{m_i} \left[p_{\ell,q,z_i} x_i^{z_i}\right] P(m_1, ..., m_i - z_i, ..., m_n, t) x_i^{m_i - z_i} \\ &= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \prod_{k=1, k\neq i}^{n} x_k^{m_k} \sum_{m_i=0}^{\infty} P(m_1, ..., m_i, ..., m_n, t) x_i^{m_i} \sum_{z_i=0}^{\infty} \left[p_{\ell,q,z_i} x_i^{z_i}\right] \\ &= \sum_{z_i=0}^{\infty} \left[p_{\ell,q,z_i} x_i^{z_i}\right] \times \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} P(m_1, ..., m_n, t) \prod_{k=1}^{n} x_k^{m_k} \\ &= F_i(x_i) G(\mathbf{x}) \end{aligned} \tag{S26}$$

This expression gives us the influx terms for all burst processes with $\ell = 1$. Extending this solution to $\ell > 1$ requires iteratively applying the Cauchy product formula. For simplicity, we demonstrate

this for $\ell = n$.

$$\begin{aligned}
&\sum_{m_1,...,m_n}^{\infty} \sum_{\mathbf{z}} p_{\ell,1,\mathbf{z}} P(\mathbf{m}-\mathbf{z},t) \prod_{k=1}^{n} x_k^{m_k} = \sum_{m_1,...,m_n}^{\infty} \sum_{z_1=0}^{m_1} ... \sum_{z_n=0}^{m_n} p_{\ell,1,\mathbf{z}} P(\mathbf{m}-\mathbf{z},t) \prod_{k=1}^{n} x_k^{m_k} \\
&= \sum_{m_1=0}^{\infty} \sum_{z_1=0}^{m_1} ... \sum_{m_n=0}^{\infty} \sum_{z_n=0}^{m_n} p_{\ell,1,\mathbf{z}} P(\mathbf{m}-\mathbf{z},t) \prod_{k=1}^{n} x_k^{z_k} \prod_{k=1}^{n} x_k^{m_k-z_k} \\
&= \sum_{m_1=0}^{\infty} \sum_{z_1=0}^{m_1} x_1^{z_1} x_1^{m_1-z_1} ... \sum_{m_n=0}^{\infty} \sum_{z_n=0}^{m_n} x_n^{z_n} p_{\ell,1,\mathbf{z}} P(\mathbf{m}-\mathbf{z}) x_n^{m_n-z_n} \\
&= \sum_{m_1=0}^{\infty} \sum_{z_1=0}^{m_1} x_1^{z_1} x_1^{m_1-z_1} ... \sum_{m_n=0}^{\infty} P(m_1-z_1,...,m_{n-1}-z_{n-1},m_n,t) x_n^{m_n} \sum_{z_n=0}^{\infty} [p_{n,1,\mathbf{z}} x_n^{z_n}] \\
&= \sum_{z_n=0}^{\infty} x_n^{z_n} \sum_{m_n=0}^{\infty} x_n^{m_n} \sum_{m_1=0}^{\infty} \sum_{z_1=0}^{m_1} x_1^{z_1} x_1^{m_1-z_1} ... \\
&\times \sum_{m_{n-1}=0}^{\infty} \sum_{z_{n-1}=0}^{m_{n-1}} p_{n,1,\mathbf{z}} x_{n-1}^{z_{n-1}} P(m_1-z_1,...,m_{n-1}-z_{n-1},m_n,t) x_{n-1}^{m_{n-1}-z_{n-1}}
\end{aligned} \tag{S27}$$

We can repeat this process n times, using induction with Equation S26 as the base case and Equation S27 as the inductive step. This finally yields:

$$\begin{aligned}
&\sum_{m_1,...,m_n}^{\infty} \sum_{\mathbf{z}} p_{\ell,1,\mathbf{z}} P(\mathbf{m}-\mathbf{z}) \prod_{k=1}^{n} x_k^{m_k} = \prod_{i=1}^{n} \left(\sum_{z_i=0}^{\infty} x_i^{z_i} \sum_{m_i=0}^{\infty} x_i^{m_i} \right) p_{n,1,\mathbf{z}} P(\mathbf{m},t) \\
&= \left(\sum_{z_1,...,z_n} p_{n,1,\mathbf{z}} x_i^{z_i} \right) \left(\sum_{m_1,...,m_n} P(\mathbf{m},t) \prod_{i=1}^{n} x_i^{m_i} \right) \\
&= F_{n,1}(\mathbf{x}) G(\mathbf{x},t).
\end{aligned} \tag{S28}$$

This relatively involved derivation confirms that the result in S26 generalizes to multivariate burst distributions. For completeness, all cases with $\ell \in [2,...,n-1]$ can be derived as special cases of Equation S27 by defining burst sizes of species not involved in the reaction as zero. Therefore, the full PGF of the CME in Equation S10 takes the following form:

$$\frac{dG(\mathbf{x},t)}{dt} = \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} (F_{\ell,q}(\mathbf{x}) - 1) G + \sum_{i=1}^{n} c_{i0}(1-x_i) \frac{\partial G}{\partial x_i} + \sum_{i,j=1}^{n} c_{ij}(x_j - x_i) \frac{\partial G}{\partial x_i} \tag{S29}$$

Finally, defining $u_i := x_i - 1$, $\phi := \ln G$, and $M_{\ell,q}(\mathbf{u}) = F_{\ell,q}(\mathbf{u}+1)$, we can simplify the PDE:

$$\frac{d\phi(\mathbf{u},t)}{dt} = \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q} (M_{\ell,q}(\mathbf{u}) - 1) - \sum_{i=1}^{n} u_i c_{i0} \frac{\partial \phi}{\partial u_i} + \sum_{i,j=1}^{n} (u_j - u_i) c_{ij} \frac{\partial \phi}{\partial u_i}. \tag{S30}$$

S2 CME solution

It remains to construct a function $\phi(\mathbf{u},t)$ that solves Equation S30. In the current section, we discuss the solution procedure at length, motivating the spectral expression reported in the body of the text.

The integral can be written down in terms of *characteristics*. This amounts to introducing a vector of auxiliary functions $U_1(\mathbf{u}, s), ..., U_n(\mathbf{u}, s) := \mathbf{U}(\mathbf{u}, s)$ that parametrize the solution and taking a total derivative [11, 12]. The characteristics are governed by the following series of n coupled differential equations:

$$\frac{dU_i(\mathbf{u}, s)}{ds} = \sum_{j=1}^{n}(U_j - U_i)c_{ij} - U_i c_{i0} \text{ s.t. } U_i(\mathbf{u}, s = 0) = u_i, \tag{S31}$$

with the ϕ term taking the following form [13]:

$$\phi(\mathbf{u}, t) = \int_0^t \sum_{\ell=1}^{n} \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q}(M_{\ell,q}(\mathbf{U}(\mathbf{u}, s)) - 1)ds \tag{S32}$$

General, closed-form solutions do not exist, as the integral in Equation S32 is typically intractable. Several special cases have been solved. For example, if we set $k_{\ell,q} = 0$ for all $\ell > 1$, and $F_{1,i}(\mathbf{x}) = x_i$, we recover the case of constitutive expression discussed by Jahnke and Huisinga [14]. The ensemble of Ornstein-Uhlenbeck models induced by background driving Lévy processes (BDLP) emerges from setting $n = 1$ and modulating $M_{1,1}$, the jump size distribution; the usual negative binomial bursty model corresponds to exponential Lévy jump sizes, but generalizations have been studied in the financial context by Barndorff-Nielsen and others [15–17]. Unfortunately, systems with more than one species and bursty transcription are intractable; even the case of $n = 2$ and geometric bursts requires quadrature [13], although some approximation methods are available [18]. Instead of unduly restricting analysis to systems that are fully tractable, we seek to reduce the PDE to the single integral in Equation S32, then use numerical quadrature.

We start by computing the characteristics $\mathbf{U}$. As we have assumed the splicing graph is a DAG, there must be at least one vertex with out-degree zero [19]. Therefore, at least one transcript (arbitrarily indexed by n) undergoes degradation only, with no splicing products. This transcript has the following characteristic:

$$\frac{dU_n}{ds} = -U_n c_{n0} \implies U_n = u_n e^{-c_{n0}s}. \tag{S33}$$

We can iterate backwards and write down an equation for each $U_i(\mathbf{u}, s)$. This amounts to iterating over S31 for different i, checking whether all j corresponding to nonzero c_{ij} have already been visited, and solving the ODE if this is the case. Since DAGs have a partial order, it is *always* possible to apply this algorithm and proceed from "downstream" to "upstream" species in a well-defined way.

The following equality emerges from solving a simple linear ODE with arbitrary parameters a_i, b_i,

and r:

$$\begin{aligned}
\frac{dy}{dx} &= \sum_i a_i e^{-b_i x} - ry \implies y = \sum_i \frac{a_i}{r - b_i} e^{-b_i x} + Ke^{-rx} \\
\frac{dy}{dx} &= \frac{d}{dx} \sum_i \frac{a_i}{r - b_i} e^{-b_i x} + \frac{d}{dx} Ke^{-rx} \\
&= -\sum_i \frac{a_i b_i}{r - b_i} e^{-b_i x} - Kre^{-rx} \qquad \text{(S34)} \\
\sum_i a_i e^{-b_i x} - ry &= \sum_i a_i e^{-b_i x} - \sum_i \frac{a_i r}{r - b_i} e^{-b_i x} - Kre^{-rx} \\
a_i - \frac{a_i r}{r - b_i} &= \frac{a_i r - a_i b_i - a_i r}{r - b_i} = -\frac{a_i b_i}{r - b_i},
\end{aligned}$$

which confirms that the expression for y solves the ODE. Therefore, we can write down Equation S31 in an analogous form:

$$\frac{dU_i}{ds} = \sum_{j=1}^{n} U_j c_{ij} - U_i \left(c_{i0} + \sum_{j=1}^{n} c_{ij} \right). \qquad \text{(S35)}$$

Since the terminal characteristics, as derived in Equation S33, are exponential, and iterating backwards maintains the functional form, each characteristic must be expressible as a weighted sum of exponentials $U_i = \sum_j A_{ij} e^{-r_j s}$, with exponents defined as $r_i = \sum_{j=1}^{n} c_{ij} + c_{i0}$, the total efflux rate. Therefore,

$$U_i = \sum_{j=1}^{n} \frac{a_j}{r_i - r_j} e^{-r_j t} + Ke^{-r_i t} \qquad \text{(S36)}$$

The initial condition is $U_i(\mathbf{u}, 0) = u_i$ implies $K = u_i - \sum_j \frac{a_j}{r_i - r_j}$. Care must be taken when the downstream paths converge. Duplicate terms in product characteristics U_j need to be aggregated:

$$\sum_j U_j c_{ij} = \sum_j c_{ij} \sum_k A_{jk} e^{-r_k s} = \sum_k \left(\sum_j c_{ij} A_{jk} \right) e^{-r_k s} \qquad \text{(S37)}$$

Although the derivation is somewhat technical, it lends itself to automation. In a n-species system with as many distinct efflux rates, each characteristic can be defined with respect to n basis functions $e^{-r_i s}$. Therefore, we iterate over the splicing graph from "downstream" to "upstream", and reweight the coefficients of the bases for each characteristic.

S2.1 Example: path graph splicing

In this and the two following subsections, we solve the ODEs in Equation S31 by hand for several small systems.

Consider the system consisting of a bursting gene coupled to a n-step birth-death process, characterized by the path graph in Figure S1, where $B \sim Geom(b)$, and all reactions occur after exponentially-distributed waiting times. The bursts occur with rate k_1, the conversion of adjacent

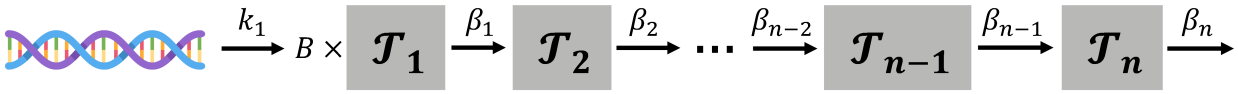

Figure S1: Graph representation of the generic path graph model. The source transcript $\mathcal{T}_1$ is synthesized at the gene locus in random geometrically distributed bursts (according to a distribution B with burst frequency k_1). Each molecule proceeds to isomerize in a chain of splicing reactions governed by successive rates $\beta_1, \beta_2, \ldots, \beta_{n-1}$, until reaching the form $\mathcal{T}_n$, which is ultimately degraded at rate β_n.

transcripts $\mathcal{T}_i$ to $\mathcal{T}_{i+1}$ occurs with rate β_i, and the degradation of $\mathcal{T}_n$ occurs with rate β_n. We assume the rates of conversion and degradation are all distinct. The amount of species $\mathcal{T}_i$ can be described by the non-negative discrete random variable m_i. We assume no molecules are present at $t = 0$.
Analysis yields the following PDE governing the generating function:

$$\frac{\partial \phi}{\partial t} = k_1(M(u_1) - 1) + \sum_{i=1}^{n-1} \beta_i (u_{i+1} - u_i) \frac{\partial \phi}{\partial u_i} - \beta_n u_n \frac{\partial \phi}{\partial u_n}, \tag{S38}$$

where $M(u) := F(1 + u)$. This equation can be solved using the method of characteristics, with formal solution $\phi = k_1 \int_0^t [M(U_1(s)) - 1] ds$. The characteristics U_i, $i < n$ satisfy $\frac{dU_i}{ds} = \beta_i(U_{i+1} - U_i)$, with $U_n(\mathbf{u}, s) = u_n e^{-\beta_n s}$.
The functional form of $\frac{dU_i}{ds}$, combined with Equation S36, implies that $U_1(s)$ is the weighted sum of exponentials $\sum_{i=1}^{n} A_{1i} e^{-\beta_i s}$. The weights A_{1i} can be computed through a simple iterative procedure, which proceeds from the terminal species and successively incorporates dependence on upstream rates:

$A_{1i} \leftarrow 0 \; \forall \; i < n$
$A_{1n} \leftarrow u_n$
$i \leftarrow n - 1$
while $i > 0$ **do**
  for $j > i$ **do**
    $A_{1j} \leftarrow A_{1j} \times \frac{\beta_i}{\beta_i - \beta_j}$
  end for
  $A_{1i} \leftarrow u_i - \sum_{j>i} A_{1j}$
  $i \leftarrow i - 1$
end while

This algorithm iteratively applies the solution from Equation S34 and its particular form in Equation S36 to explicitly compute the basis coefficients.

S2.2 Example: alternative splicing

Suppose the downstream dynamics are given by a directed rooted tree. The solution procedure is analogous to that used for the path graph. First, starting at the leaves, the path subgraph solutions are produced by the procedure above, yielding a sum of exponentials. Then, at a node of out-degree

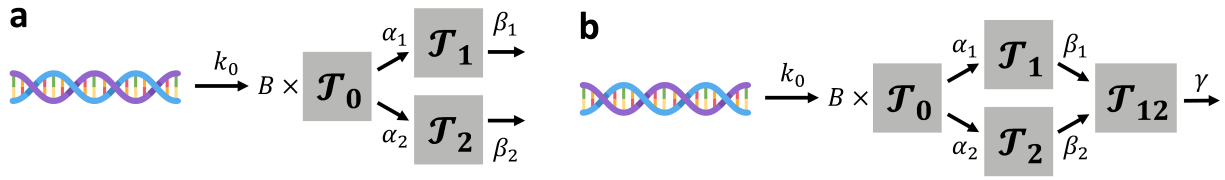

Figure S2: Graph representations of simple directed acyclic graph models. (a) Tree splicing graph with two terminal isoforms. (b) Convergent splicing graph with a single terminal isoform and two intermediate transcripts.

> 1 (i.e., molecular species with several potential products), the associated ODE has a functional form identical to that of a path graph. Therefore, the solutions are analogous.

As an illustration, consider the simplest tree graph, shown in Figure S2a, where the splicing reactions occur at rates α_1 and α_2 and degradation reactions occur at rates β_1 and β_2. Physically, this graph can be interpreted as a single source mRNA being directly and stochastically converted to one of two terminal isoforms by removal of intron 1 or intron 2. Clearly, $U_i = u_i e^{-\beta_i s}$ for $i \in \{1, 2\}$. The ODE governing the source species is:

$$\begin{aligned} \frac{dU_0}{ds} &= \alpha_1(U_1 - U_0) + \alpha_2(U_2 - U_0) = \alpha_1 U_1 + \alpha_2 U_2 - (\alpha_1 + \alpha_2)U_0 \\ \implies U_0 &= \frac{\alpha_1}{\alpha_1 + \alpha_2 - \beta_1} u_1 e^{-\beta_1 s} + \frac{\alpha_2}{\alpha_1 + \alpha_2 - \beta_2} u_2 e^{-\beta_2 s} + K e^{-(\beta_1 + \beta_2)s} \\ \text{s.t. } K &= u_0 - \frac{\alpha_1}{\alpha_1 + \alpha_2 - \beta_1} u_1 - \frac{\alpha_2}{\alpha_1 + \alpha_2 - \beta_2} \end{aligned} \tag{S39}$$

Finally, the expression for U_0 can be directly plugged into the burst generating function and integrated.

S2.3 Example: two-intron splicing with non-deterministic order

Consider the same tree graph as in the example above, and suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are converted to product $\mathcal{T}_{12}$ at rates β_1 and β_2, as shown in Figure S2b. Afterward, $\mathcal{T}_{12}$ is degraded at rate γ. Physically, this graph can be interpreted as a single source mRNA being converted to one of two intermediate isoforms by the removal of one of two introns, then to a single terminal isoform by the removal of the other intron. Clearly, $U_{12} = u_{12} e^{-\gamma s}$. Setting $f_i := \frac{\beta_i}{\beta_i - \gamma}$, we find $U_i = (u_i - f_i u_{12}) e^{-\beta_i s} + f_i u_{12} e^{-\gamma s}$. Finally, the dynamics of the source molecule $\mathcal{T}_0$ are governed by the following ODE:

$$\frac{dU_0}{ds} = \alpha_1(u_1 - f_1 u_{12}) e^{-\beta_1 s} + \alpha_2(u_2 - f_2 u_{12}) e^{-\beta_2 s} + (\alpha_1 f_1 + \alpha_2 f_2) u_{12} e^{-\gamma s} - (\alpha_1 + \alpha_2) U_0$$

Yet again, the functional form affords a straightforward analytical solution:

$$U_0 = K e^{-cs} + \frac{C_1}{c - \beta_1} e^{-\beta_1 s} + \frac{C_2}{c - \beta_2} e^{-\beta_2 s} + \frac{C_3}{c - \gamma} e^{-\gamma s},$$

where $c := \alpha_1 + \alpha_2$, $C_1 := \alpha_1(u_1 - f_1 u_{12})$, $C_2 := \alpha_2(u_2 - f_2 u_{12})$, and $C_3 := \alpha_1 f_1 + \alpha_2 f_2$. From the initial condition $U_0(s = 0) = u_0$, we yield $K = u_0 - \frac{C_1}{c - \beta_1} - \frac{C_2}{c - \beta_2} - \frac{C_3}{c - \gamma}$. The computation procedure is demonstrated in Figure S3.

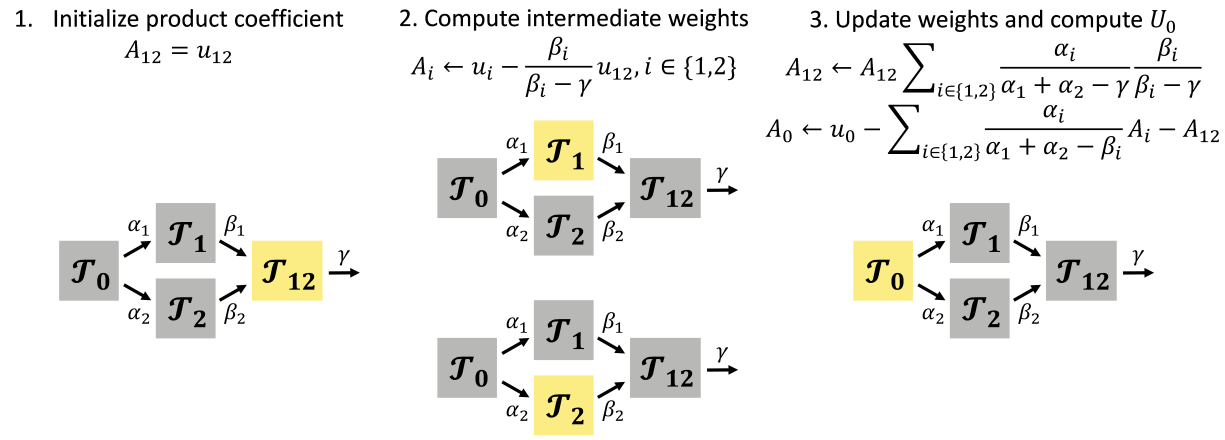

Figure S3: Illustration of the solution algorithm. The differential equation structure requires the backward propagation of downstream species' solutions, weighted by ratios of rates.

S2.4 General spectral solution

The solutions described above, although exact, are cumbersome to compute by hand. Ultimately, we seek to evaluate the characteristics of an arbitrary graph at an arbitrary set of $\mathbf{u}$ for the Fourier transformation and an arbitrary set of t for the quadrature.

In principle, we can exploit previous results. For example, the case of constitutive expression is known to have a multivariate Poisson distribution with intensities given by the reaction rate equations [14]. Therefore, to obtain the functions $\mathbf{U}$, we need to write down the generating function G^* of the solution to the constitutive case, differentiate the log-PGF $\phi^* = \int_0^t \sum_{i=1}^n k_{1,i} U_i(\mathbf{u}, s) ds$, and obtain the characteristics. The solution to this system has been well-characterized for sixty years, with cornerstone work by Gans [20] and McQuarrie [21], as well as recent extensions by Gadgil [22] and Reis [23], and can be obtained using the eigenvalues of the transition rate matrix with entries c_{ij}. Equivalently, we know that the constitutive solution consists of *independent* Poisson distributions [14]; therefore, it it must be possible to write down each $\frac{\partial \phi^*}{\partial t}$ as a sum of independent species-specific functions $\sum_{i=1}^n u_i \psi_i(s)$, where ψ_i is the derivative of the species average obtained by solving the reaction rate equations.

Ultimately, it is easiest to apply spectral methods directly to the characteristic relations. We can write down Equation S40 as a matrix equation:

$$\frac{d\mathbf{U}}{ds} = C\mathbf{U} \text{ s.t. } \mathbf{U}(s=0) = \mathbf{u}, \tag{S40}$$

where C is a matrix containing $C_{ij} = c_{ij}$ for all $i \neq j$ and $C_{ii} = -\sum_{j=1}^n c_{ij} - c_{i0}$. We can decompose it as $C = V\Lambda V^{-1}$, where Λ is a matrix containing the eigenvalues $\lambda_1, ..., \lambda_n$ on the diagonal. The general solution to this system takes the following form:

$$\mathbf{U}(s) = \sum_{i=1}^{n} a_i e^{\lambda_i s} \mathbf{v}_i, \tag{S41}$$

where $\mathbf{v}_i$ are the eigenvectors contained in V. To compute the coefficients a_i, we need to plug in the initial condition:

$$\begin{aligned} \mathbf{u} = \mathbf{U}(0) = [\mathbf{v}_1, ..., \mathbf{v}_1][a_1, ..., a_n]^T = V[a_1, ..., a_n]^T \\ \implies [a_1, ..., a_n]^T = V^{-1}\mathbf{u}. \end{aligned} \tag{S42}$$

Plugging these coefficients into the solution:

$$\mathbf{U} = [e^{\lambda_1 s}\mathbf{v}_1, ..., e^{\lambda_n s}\mathbf{v}_n]V^{-1}\mathbf{u} \tag{S43}$$

Inspecting each individual characteristic U_i:

$$U_i = \sum_{j=1}^{n}\sum_{k=1}^{n} V_{ij}e^{\lambda_j s}V^{-1}_{jk}u_k = \sum_{j=1}^{n}\sum_{k=1}^{n} V_{ij}V^{-1}_{jk}u_k e^{\lambda_j s} \tag{S44}$$

To actually compute the integral in Equation S32, we need to know $\mathbf{U}$ as a function of time. This is easiest to represent by writing the characteristics as weighted sums of the time-dependent spectral terms:

$$\begin{aligned} \mathbf{U} &= A[e^{\lambda_1 s}, ..., e^{\lambda_n s}] \\ A_{ij} &= \sum_{k=1}^{n} V_{ij}V^{-1}_{jk}u_k \\ \mathbf{w} &:= V^{-1}\mathbf{u} \implies A = V \cdot Diag(\mathbf{w}). \end{aligned} \tag{S45}$$

Therefore, for each reaction system we investigate, we need only compute V, V^{-1}, and Λ once, by a single spectral decomposition and a single inversion. Thereafter, we can compute $\mathbf{w}$ for each value of $\mathbf{u}$ of interest, by two matrix multiplications per $\mathbf{u}$.
Finally, if we are interested in a single characteristic, we simply extract the corresponding coefficients. For example, if only a single transcript with index i is being produced in bursts, the following equation gives its characteristic:

$$U_i(\mathbf{u}, s) = \sum_{j=1}^{n} e^{\lambda_j s}(V \cdot Diag(V^{-1}\mathbf{u}))_{ij}. \tag{S46}$$

By construction, the diagonal elements of C give the eigenvalues λ_i. Therefore, $r_i = -\lambda_i$. If multiple eigenvalues coincide, it is necessary to use a version of the spectral solution with mixed exponential-polynomial terms [13, 24].

S3 Bursty systems have well-behaved moments

In the current section, we use the standard geometric burst distribution and demonstrate qualitative properties of the solutions: all solutions exist, have all moments, and are unimodal. For convenience, we adopt the assumptions and notation of the section "Moments of the splicing graph solutions are tractable by matrix operations" in the main text.

S3.1 The exponential sum is always positive

First, we demonstrate that the downstream processes yield a strictly positive functional form of time dependence. Noting that the marginal of species i (with $u_j = 0$ for all $j \neq i$) yields the functional form $U_1(u_i, s) = u_i \sum_{k=1}^{n} a_{i,k} e^{-r_k s} := u_i \psi_i(s)$, this condition translates to $\psi_i(s) > 0$ for all $s > 0$.

Consider $F(x) = x$, corresponding to constitutive production of the source species (i.e. a Poisson birth process), with no molecules present at $t = 0$. Focusing on the marginal of species i, this assumption yields $\phi(u_i, t) = k_1 \int_0^t U_1(u_i, s) ds = k_1 \int_0^t u_n \psi_n(s) ds$. Evaluating e^{ϕ} at $x_i = 0$, i.e. $u_i = -1$, marginalizes over all $j \neq i$ and yields the probability of observing zero counts of species i: $G(u_i, t) = P(m_i = 0, t) = P_0(t) = \exp(-k_1 \int_0^t \psi_i(s) ds)$. The corresponding time derivative is $\frac{dP_0}{dt} = -k_1 \psi_n(t) \exp(-k_1 \int_0^t \psi_i(s) ds)$. Simultaneously, we know that $P_0(t) = e^{-\lambda_i(t)}$, where $\lambda_i(t)$ is the solution of the reaction rate equation for species i [14]. Clearly, $\frac{dP_0}{dt} = -P_0(t) \frac{d\lambda_i(t)}{dt}$. The reaction rate $\frac{d\lambda_i(t)}{dt} > 0$ at $t = 0$ under the given initial conditions. Furthermore, $\frac{d\lambda_i(t)}{dt}$ is strictly positive. This follows from the reaction rate equations. By the continuous formulation, λ_i is a weighted moving average of some set of processes $\{\lambda_k\}$. λ_1 is a strictly increasing function governed by $\frac{k_1}{r_1}(1 - e^{-r_1 s})$. The property of being strictly increasing is retained under moving average and rescaling. Therefore, each successive moving average must be strictly increasing.

Finally, $P_0 \in (0, 1) > 0$, because the Poisson distribution has support on all of $\mathbb{N}_0$. Therefore, $\frac{dP_0}{dt}$ is strictly negative. As the exponential term and k_1 are positive, this implies $\psi_i(s)$ is strictly positive for all $s > 0$.

S3.2 All generating functions and marginals exist

Next, we show that $G(u_i, t)$, the generating function of the ith marginal, is finite for the geometric burst system. This follows from the construction of the original PGF: the marginal PGF is guaranteed to be finite if $1 - b u_i \psi_i$ is never zero. But for the relevant domain $\Re(u_n) \leq 0$, on the shifted complex unit circle, $\Re(1 - b u_i \psi_i) \geq 1$, except at the degenerate initial case. The existence of the marginal moments of X_i is implied by the existence of the generating function. The existence of all cross moments follows from the Cauchy-Schwartz inequality. Per standard properties, this existence property holds for both X_i and Λ_i.

The tails of the stationary discrete marginals decay no slower than the geometric distribution. This follows immediately from the lower bound on $\Re(1 - b u_i \psi_i)$, which in turn gives an upper bound on x_i [13]. Equivalently, this follows from the existence of all moments [25]. An analytical radius of convergence has been given previously for $n = 2$ [13], but numerical optimization is necessary to establish rates of tail decay for $n > 2$.

S3.3 All marginals are infinitely divisible

An infinitely divisible distribution of a random variable X can be represented as the sum of q independent random variables $X_{q1}, ..., X_{qq}$ for every integer q [26]. This follows from the functional form of the PGF in Equation 14 (setting $k_j = 0$ when $j \neq 1$ for consistency): the system can be decomposed into q independent systems with burst frequencies k_1/q.

S3.4 Only the first marginal is self-decomposable

The law of a self-decomposable (sd) random variable X is equal in distribution to the random variable $cX + X_c$, where X_c is independent of X [26]. By definition, a random variable is sd if such a representation can be found for *any* $c \in (0, 1)$. Distributions in this class are frequently invoked in the stochastic process literature, as they are amenable to analysis through the Lévy-Khinchine representation and guarantee useful properties, chief among them unimodality [17, 27]. Unfortunately, we cannot rely on self-decomposability to prove unimodality in our case, because all of the downstream processes are non-sd.

A random variable has a self-decomposable law if and only if it *also* offers a representation of the form $Y = \int_0^\infty e^{-t} dX_t$, with Lévy X_t [26]. By considering Equation 8, we find that downstream processes *can* be represented as moving averages of upstream processes. However, only $L_{1,t}$, the jump driver of the transcriptional process, is Lévy. All downstream intensity processes have nontrivial, almost-everywhere C^∞ trajectories, which implies they cannot be represented by a Lévy triplet: the only permitted continuous Lévy processes are linear combinations of the (non-differentiable) Brownian motion W_t and the trivial process t. Therefore, Λ_i is sd for $i = 1$ and non-sd for all $i > 1$.

S3.5 All stationary marginals are unimodal

Even though we cannot rely on the sd property, we can still invoke the properties of the transcriptional process to prove that all marginals are unimodal.

First, we consider a single trajectory of the intensity process Λ_i governed by Equation 8. The trajectory of Λ_1 is a realization of the gamma Ornstein-Uhlenbeck process, a deterministically transformed version of a realization of $L_{1,t}$ [28]. The trajectories of all other species are deterministically transformed versions of Λ_1. For example, if the transcript processing reactions $\mathcal{T}_1 \xrightarrow{r_1} \mathcal{T}_2 \xrightarrow{r_2} ...$ take place, applying variation of parameters to the Poisson representation yields:

$$\begin{aligned} \Lambda_1(t) &= \sum_{q=0}^{N(t)} e^{-r_1(t-\tau_q)} J_q \\ \Lambda_2(t) &= \int_0^t r_2 \Lambda_1(s) e^{-r_2(t-s)} ds. \end{aligned} \tag{S47}$$

Therefore, the trajectory of every $\mathcal{T}_i$ is an iterated and rescaled moving average of the underlying process $L_{1,t}$, with exponential jumps J_q at times τ_q generated by the Poisson process $N(t)$.

Multimodality in the distribution of a trajectory can result from the definition of the moving average or from multimodality of the underlying process's trajectory (e.g., Λ_1 in Equation S47). The one-parameter exponential moving average cannot induce multimodality. The stationary law of many realizations of Λ_1 is gamma; by ergodicity, a single realization of its trajectory over a long enough period of time converges to this law. Therefore, the stationary distributions of all downstream species are unimodal in the continuous worldview. By standard properties of Poisson mixtures [29], the CME marginals are likewise unimodal.

S4 Simulation

To compare the analytical solutions with simulation, we generated a random directed acyclic graph, shown in Figure S4a. The numbers of species (7) and isomerization reactions (11) were chosen

arbitrarily. We enforced the existence of a single unique source node (a) and the weakly connected property to ensure only a single source mRNA would be present and all isoforms would be reachable from it, but did not impose any other conditions. The number of degraded species (3) was chosen arbitrarily; we assigned degradation reactions to the two sink species (c, e) and randomly chose a degraded intermediate (b) from a uniform distribution over the molecular species.
All reaction rates were drawn from a log-uniform distribution on $[10^{-0.5}, 10^{0.5}]$; we chose to sample them from a single order of magnitude to avoid the trivial degenerate cases that occur in cases of very slow or very fast export [13]. This process produced the parameter values $k_a = 0.44$, $\beta = [0.48, 2.12, 1.31, 2.21, 1.16, 2.41, 0.4, 1.19, 0.37, 1.19, 0.53]$, and $\gamma = [0.94, 2.38, 0.72]$, with the indices corresponding to those in Figure S4a. Finally, we chose the geometric burst model with $b = 10$.

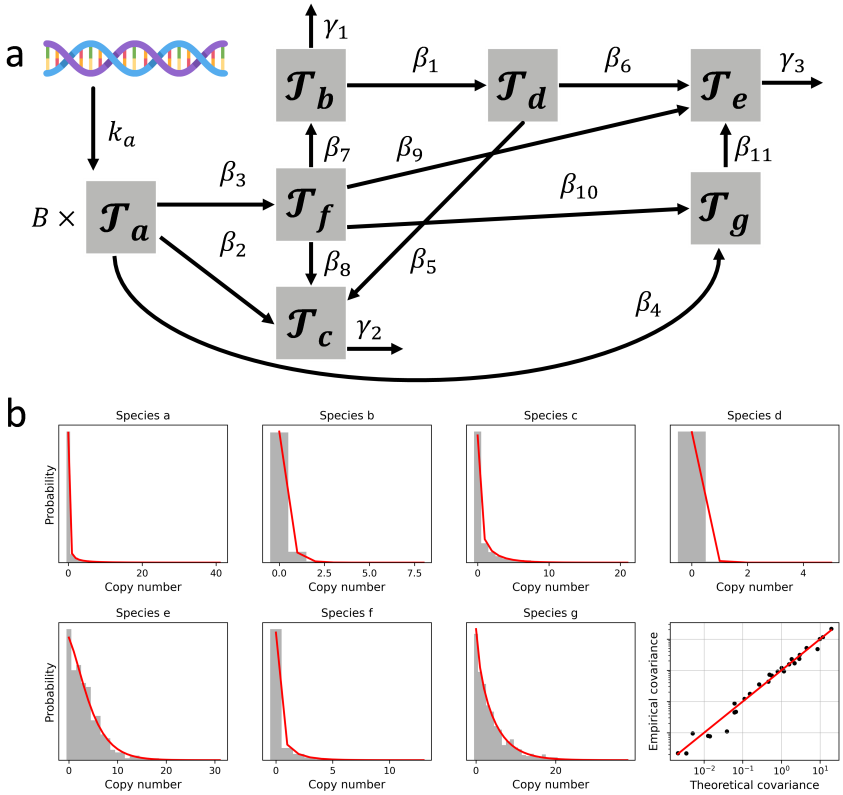

Figure S4: Validation of the solution algorithm using stochastic simulations. **a.** Graph representation of the randomly generated transcription, splicing, and degradation model. A single source isoform $\mathcal{T}_a$ is converted to a variety of downstream isoforms $\mathcal{T}_b, ..., \mathcal{T}_g$, which isomerize according to a randomly generated directed acyclic graph. **b.** The simulated marginal distributions and lower moments match the analytical solutions (gray bars: histograms obtained from simulations; red lines: analytical solutions; black points: covariance matrix entries from simulated data).

We applied the algorithm to compute the exponents and coefficients, and computed the stationary distributions of all species. Further, we simulated 1000 cells up to $T = 5r_{min}^{-1}$, where r_{min} was the minimum rate of k_a, all β, and all γ. The simulated distributions matched the quantitative results

for the marginals, as shown in Figure S4b. Furthermore, the 49 entries of the covariance matrix were effectively predicted by the procedure for moment calculation.

S4.1 Benchmarking

To benchmark the performance of the algorithm and identify bottlenecks, we generated 100 splicing networks with the same number of nodes, edges, and degradation channels as above, and with rates and graphs drawn from the same distribution. Log-burst sizes $\log_{10} b$ were drawn from a normal distribution with mean and variance of 1, clipped at $[-1, 2]$. For each graph, we simulated 1000 cells and timed each step of the procedure. The results are shown in Figure S5. As evident from Figure S5a, all marginal PMF computation runtimes (including computation of coefficients, numerical integration, and inverse Fourier transformation) were below 100 ms, and demonstrated time complexity of $\mathcal{O}(\mathcal{N}^{0.414})$ in the relevant domain (up to approximately 1000 counts).

We can use this benchmark to demonstrate the impracticality of matrix methods for computing the marginals of highly multimodal systems. If we are interested in the marginal i, this semi-analytical solution requires an array of size $\mathcal{N} = \max m_i$, with the maximum taken over all cells. However, matrix methods involve a state space size of $\mathcal{N} = \prod_i \max m_i$: for example, if we wish to find the marginal of species e in Figure S4a, we need to compute the entire joint distribution of species a, b, d, f, g, and e and sum over the dimensions corresponding to a, b, d, f, and g. To understand how feasible this is, we can examine the "latent" dimension of the systems by computing $\mathcal{N} = \prod_i (\max m_i + 6)$, the state space size used throughout the benchmarking to limit ringing artifacts. The distribution of overall system sizes is shown in Figure S5b, and ranges from 10^6 to 10^{16}. The upper range of this size requires 281 petabytes to define each state in 8-byte floating point format, making it impossible to even store an array of size $\mathcal{N}$, much less process it in the conventional matrix time of $\mathcal{O}(\mathcal{N}^3)$.

Inspecting the predictive performance of all 100 systems is challenging. However, we can at least visualize the covariances, in the vein of the last panel of Figure S4b. The results are shown, without axes, in Figure S5c: overall, the theory agrees with the simulations, although some negative sample covariances are observed in simulations when the entries are close to zero.

Finally, in Table S1, we report the computation times necessary to generate the marginals for the 100 systems benchmarked in Figure S5. The simulation is by far the most computationally intensive part of the process, followed by numerical integration. The spectral decomposition requires approximately 0.1 ms per *system* (rather than per marginal). The computation of coefficients (matrix multiplication) takes on the order of 1 ms per marginal, whereas the Fourier inversion takes on the order of 0.1 ms per marginal. Therefore, the computational cost of the forward problem is dominated by the numerical integration, which takes on the order of tens of ms for each PMF. The runtime can be halved by exploiting the Hermitian property of the Fourier transform of a probability mass function [30]. Some further speed improvements can be obtained by approximation; we have reported gains by using Gaussian rather than adaptive quadrature [31] and by using special functions for certain narrow sets of splicing networks [18], but the performance reported here appears to be sufficient for most practical use.

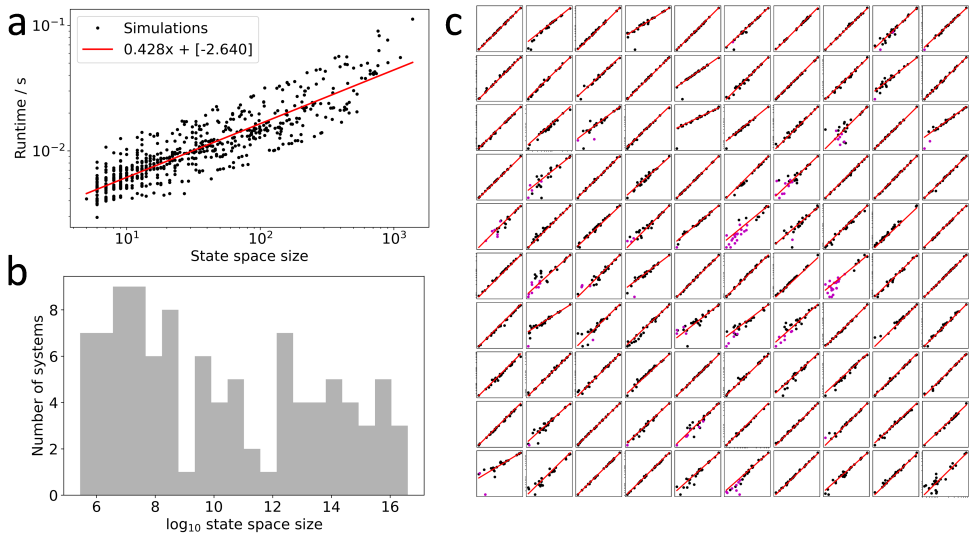

Figure S5: Benchmarking of the solution algorithm using stochastic simulations. **a.** The runtime varies between 3 and 150 ms for all marginals, with sub-linear state space size scaling in the low-copy number regime. **b.** Although the marginals are tractable, the full joint distributions are not: the latent space sizes range from 10^6 to 10^{16}. **c.** The simulated lower moments match the analytical solutions for all 100 simulated graphs (x-axis: analytical solution; y-axis: black points: covariance matrix entries from simulated data; red lines: identity; magenta points: negative of sample covariance with incorrect sign).

S5 Inference

The current framework provides quadrature-based solutions to the forward problem of PMF prediction for a broad set of transcriptional processes. More broadly, we would like to treat the *inverse* problem of identifying parameters from sequencing data. A wide range of statistical approaches are available; however, in practice, even the simplest, ergodic version of the inverse problem depends on the following prerequisites:

1. Single-cell, single-molecule data for a set of cells in local equilibrium. This information permits the application of the ergodic model.

2. Full annotation of intermediate transcripts, *including causal relationships*, such as the splicing graph and the identities of degraded molecules. This information permits mapping between experimental data and theoretical quantities.

3. Transcriptome-wide quantification of all transcripts, ideally unbiased and fully saturated – or, at the very least, imperfect quantification combined with a statistical model of sequencing. This information permits the inclusion of technical noise.

No perfect experimental solution exists. The collection of single-cell, single-molecule data is enabled by barcoding [32]. Characterization of splicing graphs has been treated in experimental [33,34] and computational [35] contexts. However, these necessarily rely on comprehensive single-molecule annotations – which distinct intron/exon combinations occur? – which have only become feasible

Process	Time
Graph construction	607 ms
Simulation	14.1 min
Spectral decomposition	89 ms
Covariance computation	475 ms
Coefficient computation	1038 ms
Numerical integration	7272 ms
Inverse fast Fourier transform	113 ms
Total PMF computation	8482 ms

Table S1: Timing of computation steps for the 700-marginal computation needed to generate Figure S5.

on a genome-wide scale since the introduction of molecular barcoding. Fully saturated sequencing is infeasible due to cost, and potentially due to thermodynamic constraints. Finally, standard sequencing capture protocols are, by design, biased toward polyadenylated regions [32]. This effect has been exploited to capture natural intronic sequences [36] and synthetic antibody-conjugated oligonucleotides [37, 38], but the quantitative effects of these biases are as of yet unclear; we hypothesize that they can lead to systematic false negatives and false positives in the sequencing process [31]. Naturally, these data quality challenges are compounded with standard statistical challenges, such as the often considerable computational expense of determining confidence regions. Formally, we *can* use a sequencing protocol, count the number of molecular barcodes assigned to a particular transcript, and produce a data matrix, which we then treat as a set of realizations of a random variable. However, we do not know the exact form of this random variable. Even if we assume that we *can* describe the physics of the system by a Markov model, the system may contain contributions from the following phenomena:

1. Intrinsic stochasticity due to factors incorporated in the model.
2. Intrinsic stochasticity due to factors outside the model scope.
3. Various regulatory effects that counteract stochasticity.
4. Extrinsic stochasticity due to random variability in biological parameters.
5. Extrinsic effects due to multiple cell types or trajectories.
6. Technical effects due to imperfect sequencing.
7. Ambiguities or bioinformatic uncertainties in identifying transcripts.

We can counteract some of these sources of uncertainty by choosing our data appropriately. The challenges of point 7 can be mitigated by choosing technologies that perform long-read sequencing. This essentially restricts us to two candidates with long-read sequencing, cell barcodes, and molecular barcodes: FLT-seq (full-length transcript sequencing by sampling) [39] and Smart-seq3 [40, 41]. We chose the former due to the ready availability of processed, isoform-resolved data. In brief, FLT-seq synthesizes a cDNA library using 10X gel beads and primers, amplifies it, then applies

nanopore sequencing to generate long reads with associated cell and molecule barcodes. The accompanying bioinformatic pipeline *FLAMES* (full-length analysis of mutations and splicing) produces novel annotations characterizing the exons present in each molecule.
The challenges of point 5 can be mitigated by analyzing experimental conditions where the cells can be plausibly assumed stationary and homogeneous, up to stochasticity. We used a cultured mouse stem cell dataset and filtered for a subset of activated cells. Summary visualizations reported by the authors suggest that this subset is fairly internally homogeneous [39].
We analyzed the transcripts of the top 500 genes (based on total expression) observed in the data to constrain causal relationships between them. We used *gffutils* 0.10.1 to construct a database of identified intermediate and terminal isoforms, based on the accompanying annotations generated by the *FLAMES* pipeline. We used the *intervaltree* 2.1.0 Python package to split the transcript-specific exons into elementary intervals, defined as the set of largest contiguous stretches that are either present or absent in each isoform. The distributions of isoforms and introns (elementary intervals) per gene are shown in Figure S6. With these annotations in hand, we used graph tools from the *NetworkX* 2.5.1 Python package to generate directed acyclic graphs induced by accessibility relations between transcripts: if transcript $\mathcal{T}_j$ is accessible from transcript $\mathcal{T}_i$ by removing some portions of sequence, a path must lead from $\mathcal{T}_i$ to $\mathcal{T}_j$. By filtering for transcripts with in-degree zero, we identified "root" transcripts that cannot be obtained by removing regions of any other transcript.

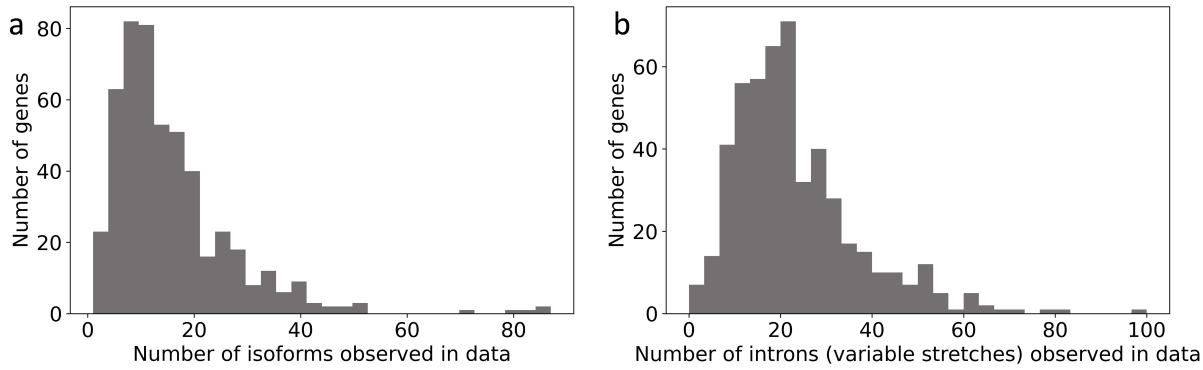

Figure S6: Isoform diversity of the FLT-seq dataset. **a.** The top 500 highest-expressed genes demonstrate a high diversity of transcript forms, with most genes having up to 20, but a significant tail having 20-90. **b.** By definition, the isoform diversity stems from presence and absence of variable regions. An average gene has approximately 20 such regions.

The hypothetical source transcript covering the entire locus was not observed for any of the genes. As their exonic patterns are mutually exclusive, we model each gene's root transcripts as products of a single rapidly processed source species. The theoretical results given in Equation 29 immediately imply that the root transcripts must be distributed according to a negative binomial law. Therefore, by fitting the transcript distributions, we can estimate effective burst sizes bw_i and normalized efflux rates r_i. These marginal parameters can be plugged into the formula for Pearson correlation in Equation 31, and compared to the empirical correlation coefficients.
The procedure assumes that all of the root transcripts are rapidly, and stochastically, produced from a single parent transcript, and attempts to use this model to predict their sample correlation

structure without foreknowledge of anything but the marginal distributions. As outlined in the "Results" section, the theoretical correlations are to be understood as upper limits. The noise sources outlined in points 2, 4, and 6 – and potential residual contributions from 5 and 7 – inevitably reduce transcript–transcript correlations. By treating correlations as upper limits in the absence of all noise outside the intrinsic stochasticity in point 1, we side-step the error sources outside the model: their specific form can be omitted; we are only concerned with the fact that they *decrease* mutual information.

To facilitate the fitting process, we performed several steps of filtering. After identifying the root transcripts, we filtered out "sparse" transcripts with fewer than five molecules in the entire dataset, as their fits are unlikely to be informative. To account for point 3 in an *ad hoc* way, we removed all transcripts with variance lower than the mean: they cannot be effectively fit to a negative binomial model, and underdispersion is a known to be a hallmark of feedback regulation [42]. We fit the remaining root transcript marginal distributions using the *statsmodels* 0.10.2 Python package, and filtered out all genes that were not successfully fit.

542 transcripts were rejected due to sparsity. 302 transcripts were rejected due to underdispersion. 100 transcripts were rejected due to failure to converge to a satisfactory fit; all of these had average expression below 1 mRNA per cell, with the distribution of average transcript counts for this subset shown in Figure S7. The quality of fits is demonstrated for a sample gene in Figure 1c. The interactive Python notebook used to perform the analysis can be used to investigate the exon structure, elementary intervals, putative splicing DAGs, and fits for any gene in the dataset.

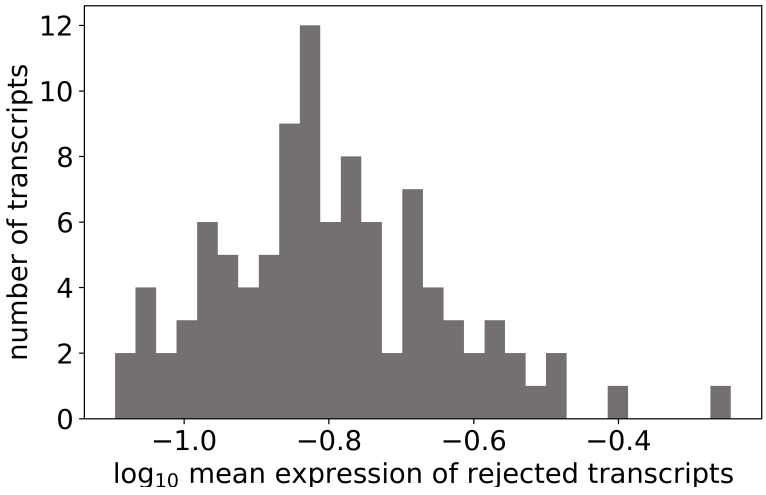

Figure S7: Distribution of average copy numbers of transcripts rejected by the fitting procedure.

This analysis produced 4885 nontrivial correlations for pairs among 1978 transcripts. The sample correlations are visualized against the theoretical correlations in Figure 1d. The points are aggregated in the lower right corner of the plot, as we expect: if the theoretical correlations are computed in the absence of noise, any additional stochasticity will degrade the correlations actually observed in the data. Some points lie above the line of identity, possibly reflecting tight regulation. A cluster of points (evident as the sharp peak on the histogram in Figure S8) has small negative correlations, possibly indicating mutually exclusive splicing of certain isoforms. However, it is not yet clear that

either of these deviations results from meaningful model differences rather than the relatively small sample size of the FLT-seq datasets, filtered for a single cell cluster.

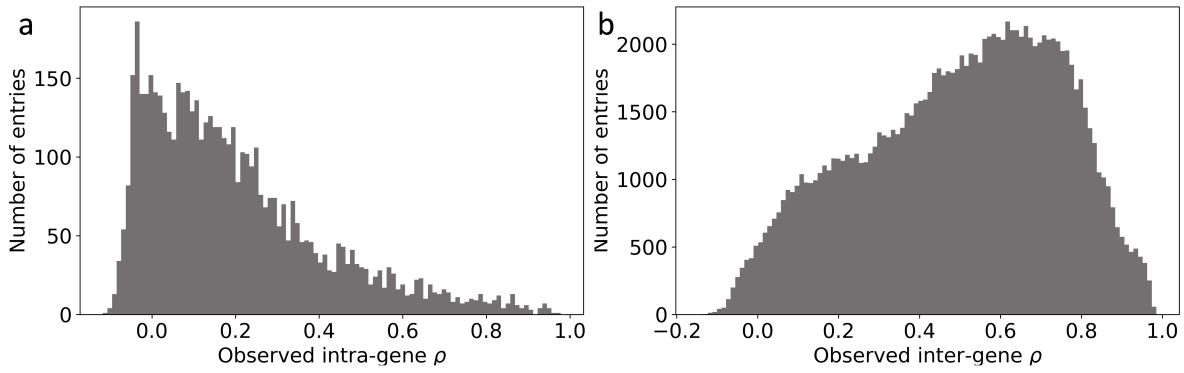

Figure S8: Histograms of sample correlations between root transcript counts in the FLT-seq dataset, computed for isoforms that passed filtering and were fit according to the procedure described in Section S5. **a.** The distribution of intra-gene correlations demonstrate a peak near -0.05. This visualization can be obtained by marginalizing Figure 1d. **b.** The distribution of inter-gene correlations demonstrate a peak near 0.7 and a shoulder near 0.2. This visualization can be obtained by marginalizing Figure 1g.

We can use the same procedure to try and predict inter-gene correlations: in the simplest model of co-expression, *all* genes fire simultaneously, with correlated burst sizes given by the model in Equation 28. Thus, their root transcripts are generated at the same time. To analyze the data, we need to select a single isoform for each gene. We cannot simply aggregate all isoforms, or even all root transcripts, as the distribution of sum of these random variables is not necessarily negative binomial. Therefore, we choose a single "dominant" isoform for every gene, based on highest expression. In the vast majority of cases, this isoform predominates and accounts for most of the expression among the root isoforms (Figure S9).
After processing the same set of 500 genes, 490 root transcripts remained after filtering and fitting. The analysis produced 119805 nontrivial correlations for pairs of transcripts. The sample correlations are visualized against the theoretical correlations in Figure 1g. As before, the points are aggregated in the lower right corner of the plot, suggesting that the upper limit on correlation obtained from the intrinsic-only noise model holds. Some points lie above the line of identity, but there does not appear to be a distinctive population of genes with negative correlations (Figure S8b).

S5.1 Quantifying uncertainty in correlation coefficients

To understand whether the model constraints hold, we have compared predicted "theoretical" correlations to sample correlations. However, with the relatively small sample sizes available by FLT-seq, both the predictions and the sample correlations themselves have non-negligible uncertainty. To understand whether violations of the constraints are systematic or artifactual, we need at least a qualitative understanding of the uncertainty involved. The prediction procedure is intended to be *ad hoc*: it sacrifices statistical power by predicting rather than fitting joint distributions. There-

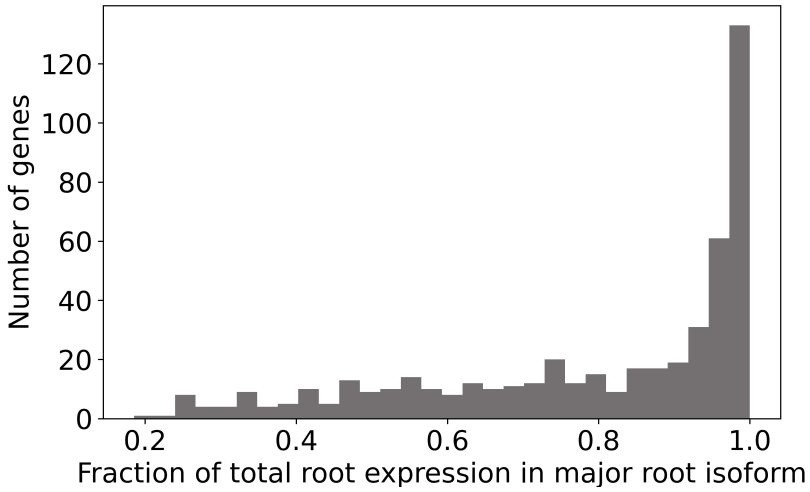

Figure S9: Histogram of the relative prevalence of the dominant root isoform. In the vast majority of cases, a single isoform predominates.

fore, we do not treat uncertainty in ρ_{theo}: its interpretation and computation are rather obscure. However, we *can* analyze the distributions of the sample correlations and attempt to quantify their stability.

The distribution of correlations can be computed analytically, through the bootstrap, or through approximations. Unfortunately, it appears that analytical solutions are only available for a very narrow set of distributions, such as the multivariate normal [43]. For a rough estimate, we use an approximation and adapt the Fisher z-transformation $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$. If the number of observed cells is N, the 95% confidence interval for the z-transformed correlation has the half-width $1.96/\sqrt{N-3}$. If the true correlation $\rho = 0$, the confidence interval (CI) for the sample correlation takes the following form [44]:

$$\left[\tanh(-1.96/\sqrt{N-3}), \tanh(1.96/\sqrt{N-3})\right] \tag{S48}$$

For the sample size of 137 cells we use, the 95% CI is 0 ± 0.168. Although this estimate is very approximate, it appropriately conveys the high uncertainty in the observed correlations.

We can compute 95% CIs for the sample correlation directly, using the bootstrap. We demonstrate these confidence intervals for the intra-gene correlations in Figure S10a. Compared with the 750 correlations in the "negative" regime obtained by point estimation, only 144 entries have a CI that lies entirely below the zero-correlation line (2.9% of all correlations). Further, compared with the 279 "inconsistent" entries, only 29 have a CI that lies entirely above the line predicted by Equation 31 (0.59% of all correlations). Therefore, the qualitative picture of the data as broadly consistent with the proposed model constraints appears to be supported by the analysis of uncertainty. We visualize the dependence of the uncertainty on the sample correlation in Figure S10. The CIs are in line with the approximation obtained from Equation S48: the correlations can be estimated up to an error of 0.1-0.2.

We perform an analogous analysis of inter-gene correlations in Figure S11a. Compared with the 1961 correlations in the "negative" regime, only 262 entries have a CI that lies entirely below the

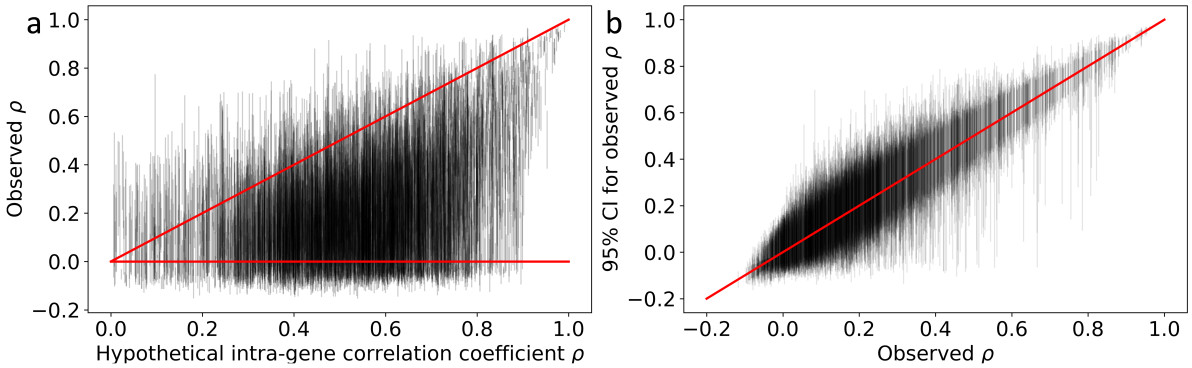

Figure S10: Bootstrap 95% confidence intervals for the intra-gene sample correlations. **a.** A variant of Figure 1d with uncertainty in the sample correlations (points: sample correlations; error bars: confidence intervals for the sample correlations; horizontal line: zero correlation; diagonal line: identity). **b.** Confidence intervals visualized as a function of the sample correlation (error bars: confidence intervals corresponding to a particular correlation matrix entry; red line: identity).

zero-correlation line (0.0022% of all correlations). Further, compared with the 302 "inconsistent" entries, only 5 have a CI that lies entirely above the line predicted by Equation 31 (0.0042% of all correlations). Therefore, data appear to be broadly consistent with proposed model constraints, potentially to a greater extent than in the foregoing investigation of intra-gene correlations. We visualize the dependence of the uncertainty on the sample correlation in Figure S11; as before, the CIs are in line with the Fisher approximation.

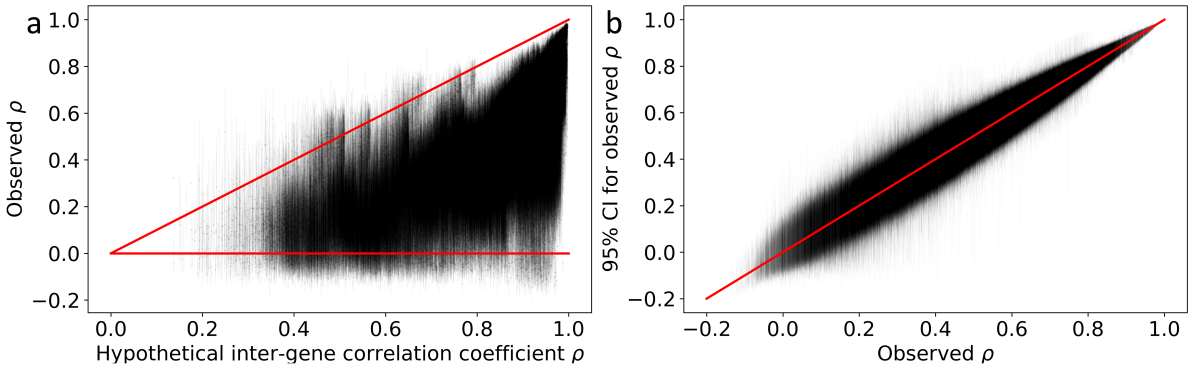

Figure S11: Bootstrap 95% confidence intervals for the inter-gene sample correlations. **a.** A variant of Figure 1g with uncertainty in the sample correlations (points: sample correlations; error bars: confidence intervals for the sample correlations; horizontal line: zero correlation; diagonal line: identity). **b.** Confidence intervals visualized as a function of the sample correlation (error bars: confidence intervals corresponding to a particular correlation matrix entry; red line: identity).

S6 Delay chemical master equations

In the current supplement, we detour from the Markovian framework to consider *delay* systems, which have deterministic, rather than stochastic, state transitions. Certain degenerate cases – for example, the problem of incremental, linear movement with identical transition rates – directly bear upon the class of delay chemical master equations (DCMEs).

S6.1 Example: constitutive production, one species

To begin, we can model the simple linear chain of reactions with constitutive production:

$$\varnothing \xrightarrow{k_1} \mathcal{T}_1 \xrightarrow{\beta/n} ... \xrightarrow{\beta/n} \mathcal{T}_n \xrightarrow{\beta/n} \varnothing, \tag{S49}$$

where the total delay between production of $\mathcal{T}_1$ and degradation of $\mathcal{T}_n$ is Erlang-distributed, with shape n and rate β. As $n \to \infty$, the Erlang distribution reduces to a point mass at $\beta^{-1} := \tau$. This implies that we can treat an *aggregated* species $\mathcal{T}$, produced at rate k and degraded after a deterministic delay τ:

$$\varnothing \xrightarrow{k_1} \mathcal{T} \overset{\tau}{\Rightarrow} \varnothing \tag{S50}$$

This is precisely the "linear chain trick" introduced by MacDonald in 1978 [45]. The study of delayed dynamical systems, such as delay differential equations, dates back to the eighteenth century [46], with cornerstone biological models by Lotka and Volterra [45, 46]. Recent work has focused on developing exact solutions [47–49] and simulation methods [50, 51]. In particular, studies by Lafuerza and Toral [52, 53] report full analytical solutions for constitutive systems with isomerization, while a contemporary study by Jia and Kulkarni [54] reports lower moments for a system with bursty mRNA production and catalysis. The distribution of $\mathcal{T}$ is $Poisson(k_1\tau)$, as this is the number of transcriptional events in a window of length τ.

S6.2 Example: constitutive production, two species

As a secondary illustration, we consider the constitutive two-stage system described by Lafuerza and Toral [52]. If we assume that no stochastic degradation reactions occur, the reaction equations and generating function relations take the following form:

$$\varnothing \xrightarrow{k_1} \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \overset{\tau}{\Rightarrow} \varnothing$$

$$\frac{\partial G}{\partial t} = k_1(F(x_1) - 1)G + \beta(x_2 - x_1)\frac{\partial G}{\partial x_1} + \beta(1 - x_2)\sum_{m_1=0}^{\infty} G^*(x_1, x_2, \tau)m_1 P(m_1, t - \tau),$$

where G^* is a conditional generating function for an auxiliary *non-degrading* system, initialized at $m_1 - 1$ molecules of the parent transcript $\mathcal{T}_1$. This auxiliary system has no degradation reactions, and allows us to incorporate the non-Markovian effects of delays. Assuming constitutive production, and using the shifted variables u_i for convenience, we find:

$$\frac{\partial G}{\partial t} = k_1 u_1 G + \beta(u_2 - u_1)\frac{\partial G}{\partial u_1} - \beta u_2 \sum_{m_1=0}^{\infty} G^*(u_1, u_2, \tau)m_1 P(m_1, t - \tau) \tag{S51}$$

The final term is *not* proportional to G, so no convenient exponential *ansatz* is available. However, the sum affords an alternative representation, which exploits the separability of the initial condition and the dynamics on $[0, \tau]$:

$$\begin{aligned} G^*(u_1, u_2, \tau) &= [1 + U_1(\tau)]^{m_1 - 1} e^{\phi^*(\tau)} \\ \sum_{m_1=0}^{\infty} G^*(u_1, u_2, \tau) m_1 P(m_1, t-\tau) &= e^{\phi^*(\tau)} \sum_{m_1=0}^{\infty} [1 + U_1(\tau)]^{m_1 - 1} m_1 P(m_1, t-\tau), \end{aligned} \tag{S52}$$

where ϕ^* is the factorial-cumulant generating function of the auxiliary system, started at zero molecules. This sum may be treated as the first derivative of the stationary $\mathcal{T}_1$ PGF, evaluated at $1 + U_1$, where U_1 is a function computed by solving the non-degrading system with the method of characteristics.

We start by computing the auxiliary U_1 by using the method of characteristics and enforcing $U_2(0) = u_2$ and $U_1(0) = u_1$.

$$\begin{aligned} \varnothing &\xrightarrow{k} \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \\ \frac{\partial U_2}{\partial s} = 0 &\implies U_2 = u_2 \\ \frac{\partial U_1}{\partial s} = \beta(U_2 - U_1) &\implies U_1 = u_2 + (u_1 - u_2) e^{-\beta s} \end{aligned} \tag{S53}$$

Now, we compute the generating function of the subsystem:

$$\begin{aligned} \phi^*(t) &= k_1 \int_0^t U_1(s) ds = k_1 \int_0^\tau \left[u_2 + (u_1 - u_2) e^{-\beta s} \right] ds \\ &= k_1 u_2 \tau + \frac{k_1}{\beta} (u_1 - u_2) \left[1 - e^{-\beta \tau} \right] \end{aligned} \tag{S54}$$

We compute the derivative of the $\mathcal{T}_1$ Poisson PGF:

$$\begin{aligned} H(x_1) &= e^{k_1 (x_1 - 1)/\beta} \\ H(u_1) &= e^{k_1 u_1 / \beta} \\ H'(U_1) &= \frac{k_1}{\beta} e^{k_1 U_1 / \beta} \end{aligned} \tag{S55}$$

This construction is slightly simpler than in the original: we do not use the full time-dependent Poisson distribution, but presuppose that the system starts with $\mathcal{T}_1$ in equilibrium. Since it approaches this distribution exponentially fast regardless of initial conditions, the error is minimal, and the simplification eliminates the time dependence in the degradation term.

Plugging in and evaluating the non-Markovian term:

$$e^{\phi^*(\tau)} H'(U_1(\tau)) = \frac{k_1}{\beta} \exp(k_1 u_2 \tau + k_1 u_1 / \beta) \tag{S56}$$

Finally, considering the full generating function expression:

$$\frac{\partial G}{\partial t} = k_1 u_1 G + \beta(u_2 - u_1) \frac{\partial G}{\partial u_1} - k_1 u_2 e^{u_1 \frac{k}{\beta} + u_2 k_1 \tau} \tag{S57}$$

Lafuerza and Toral report a solution [52], though computed through an *ansatz* rather than directly – this PDE is not quite as simple as that of the Markovian system. We restrict ourselves to the stationary solution, which can be solved with a rather mechanical application of the integrating factor method, or by noticing that the uncorrelated Poisson PMF solves the equation:

$$\begin{aligned} G &= e^{u_1 \frac{k_1}{\beta} + u_2 k_1 \tau} \\ \frac{\partial G}{\partial u_1} &= \frac{k_1}{\beta} G \\ \frac{\partial G}{\partial t} &= k_1 u_1 G + \beta(u_2 - u_1)\frac{k_1}{\beta} G - k_1 u_2 G = 0 \end{aligned} \tag{S58}$$

Of course, this result can be derived just as well without writing down anything at all – by using the standard results for constitutive production [14], the linear chain trick, and the fact that sums of Poisson random variables are Poisson.

S6.3 Example: bursty production, one species

We can use such an approach to treat the simplest bursty delay reaction system:

$$\varnothing \xrightarrow{k_1} B \times \mathcal{T} \overset{\tau}{\Longrightarrow} \varnothing \tag{S59}$$

Instead of writing down a DCME, we can notice that the system at time t has no memory beyond the last period of length τ. Therefore, we can immediately write down the distribution of $\mathcal{T}$: there are $Poisson(k_1\tau)$ burst events in each interval of length τ, and each burst event gives rise to $Geom(b)$ molecules. Therefore, the stationary distribution of molecules remaining in the system is a Poisson-geometric mixture. This approach has recently been used in a more general treatment of DCMEs [55].

S6.4 Example: bursty production, two species

Unfortunately, more general systems with a combination of bursting and downstream processing are not amenable to either the *ad hoc* or the rigorous approach.
In particular, we attempt to solve the delayed analog of the two-stage bursty system [13]:

$$\begin{aligned} \varnothing \xrightarrow{k_1} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \overset{\tau}{\Longrightarrow} \varnothing \\ \frac{\partial G}{\partial t} = k_1 \left[\frac{1}{1 - b u_1} - 1\right] G + \beta(u_2 - u_1)\frac{\partial G}{\partial u_1} - \beta u_2 e^{\phi^*(\tau)} H'(U_1(\tau)), \end{aligned} \tag{S60}$$

where the auxiliary system is now bursty.
First, we compute the factors of the non-Markovian term. The PGF derivative is found by evaluating the $\mathcal{T}_1$ marginal:

$$k_1 \int_0^T \left[\frac{1}{1 - b u_1 e^{-\beta s}} - 1\right] ds = \frac{k_1}{\beta} \ln\left(\frac{b u_1 e^{-\beta T} - 1}{b u_1 - 1}\right), \tag{S61}$$

which coincides with the relevant result for the gamma Ornstein–Uhlenbeck SDE [28]. However, this form is needlessly challenging to work with, and it is more straightforward to assume $T \gg 0$,

or the system starts in the equilibrium distribution of $\mathcal{T}_1$. Again, due to exponential convergence, the error is minimal. Differentiating with respect to $x_1 = u_1 + 1$:

$$H'(x_1) = \frac{d}{dx_1}\left(\frac{1}{1-b(x_1-1)}\right)^{k_1/\beta} = \frac{k_1 b}{\beta}\left(\frac{1}{1-b(x_1-1)}\right)^{k_1/\beta+1} \tag{S62}$$

$$H'(u_1) = \frac{\mu_1 H(u_1)}{1-bu_1}, \tag{S63}$$

where we define $\mu_1 := k_1 b/\beta$ for simplicity. This yields a straightforward expression for the summation:

$$\sum_{m_1=0}^{\infty} [1+U_1]^{m_1-1} m_1 P(m_1, t-\tau) = \frac{\mu H(U_1)}{1-bU_1} \tag{S64}$$

We reuse U_1 and U_2 from the derivation of the constitutive system, as the downstream components of the auxiliary systems match:

$$\begin{aligned}
&\varnothing \xrightarrow{k_1} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \\
\phi^*(\tau) &= k_1 \int_0^\tau (M(U_1)-1)ds = k_1 \int_0^\tau \left[\frac{1}{1-bU_1} - 1\right] ds \\
&= k_1 \int_0^\tau \left[\frac{1}{1-bu_2 - b(u_1-u_2)e^{-\beta s}} - 1\right] ds \\
\theta &:= \frac{b(u_1-u_2)}{1-bu_2} \\
\phi^* &= k_1 \int_0^\tau \left[\frac{(1-bu_2)^{-1}}{1-\theta e^{-\beta s}} - 1\right] ds \\
&= k_1\tau\left(\frac{bu_2}{1-bu_2}\right) + \frac{k_1}{\beta(1-bu_2)} \ln\left(\frac{\theta e^{-\beta\tau}-1}{\theta-1}\right) \\
&= k_1\tau\left(\frac{bu_2}{1-bu_2}\right) + \frac{k_1}{\beta(1-bu_2)} \ln\left(\frac{bU_1(\tau)-1}{bu_1-1}\right)
\end{aligned} \tag{S65}$$

which follows from the derivation of the PGF of the nascent marginal.
Now, considering the full generating function relation:

$$\begin{aligned}
\frac{\partial G}{\partial t} &= k_1\left[\frac{1}{1-bu_1}\right] G + \beta(u_2-u_1)\frac{\partial G}{\partial u_1} \\
&\quad - \beta u_2 e^{-k_1\tau} \exp\left(\frac{k_1\tau}{1-bu_2}\right)\left(\frac{bU_1(\tau)-1}{bu_1-1}\right)^{k_1\beta^{-1}(1-bu_2)^{-1}} \times \frac{k_1 b}{\beta}\left(\frac{1}{1-bU_1(\tau)}\right)^{k_1/\beta+1}
\end{aligned} \tag{S66}$$

This PDE is not easily tractable by standard analytical or numerical methods. The form of the equation is rather complicated and not amenable to analysis by characteristics. In principle, a numerical PDE or ODE solver can be used: we may fix u_2 and solve for $G(u_1, u_2)$ over a mesh of u_1. By repeating this for many values of u_2, we can compute the Fourier transform of the joint distribution. However, this requires solvers that can integrate over the complex plane, as well as initial conditions $G(0, u_2)$ for each u_2. These are the very values we seek, so even numerical approaches require some ingenuity.

In short, the stochastically delayed systems reduce to deterministically delayed systems in some well-studied regimes. However, in spite of the formal connection between the CME and the DCME, the former is far simpler to analyze: the DCME is non-Markovian, and generally resistant to exact analysis. Although much recent progress has been made, regulated transcriptional systems do not yet have full probabilistic solutions.

References

[1] Robrecht Cannoodt, Wouter Saelens, Louise Deconinck, and Yvan Saeys. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nature Communications*, 12(1):3942, December 2021.

[2] Gennady Gorin, John J Vastola, Meichen Fang, and Lior Pachter. Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments. Preprint, bioRxiv: 10.1101/2021.09.06.459173, September 2021.

[3] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. Using Gene Expression Noise to Understand Gene Regulation. *Science*, 336(6078):183–187, 2012.

[4] Jean Peccoud and Bernard Ycard. Markovian Modeling of Gene Product Synthesis. *Theoretical Population Biology*, 48(2):222–234, 1995.

[5] Niraj Kumar, Abhyudai Singh, and Rahul V. Kulkarni. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. *PLoS Computational Biology*, 11(10), October 2015.

[6] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell*, 123(6):1025–1036, December 2005.

[7] Lisa Amrhein, Kumar Harsha, and Christiane Fuchs. A mechanistic model for the negative binomial distribution of single-cell mRNA counts. Preprint, bioRxiv: 657619, June 2019.

[8] Rama Cont and Peter Tankov. *Financial Modeling with Jump Processes*. Financial Mathematics. Chapman & Hall, 2004.

[9] Heng Xu, Leonardo A Sepúlveda, Lauren Figard, Anna Marie Sokac, and Ido Golding. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nature Methods*, 12(8):739–742, August 2015.

[10] Herbert S Wilf. *generatingfunctionology*. Academic Press, second edition, 1994.

[11] Erich Zauderer. *Partial differential equations of applied mathematics*. Pure and applied mathematics. Wiley-Interscience, Hoboken, N.J, 3rd ed edition, 2006. OCLC: ocm70158521.

[12] Fritz John. *Partial Differential Equations*. Springer US, New York, NY, 1978. OCLC: 859156366.

[13] Abhyudai Singh and Pavol Bokes. Consequences of mRNA Transport on Stochastic Variability in Protein Levels. *Biophysical Journal*, 103(5):1087–1096, September 2012.

[14] Tobias Jahnke and Wilhelm Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, December 2006.

[15] Ole E. Barndorff-Nielsen and Neil Shephard. Integrated OU Processes and Non-Gaussian OU-based Stochastic Volatility Models. *Scandinavian Journal of Statistics*, 30(2):277–295, June 2003.

[16] Ole E Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in Financial economics. *Journal of the Royal Statistical Society: Series B*, 63:167–241, 2001.

[17] Ole E. Barndorff-Nielsen, Sidney I. Resnick, and Thomas Mikosch, editors. *Lévy Processes*. Birkhäuser Boston, Boston, MA, 2001.

[18] Gennady Gorin and Lior Pachter. Special function methods for bursty models of transcription. *Physical Review E*, 102(2):022409, August 2020.

[19] J. A. Bondy and U. S. R. Murty. *Graph Theory*, volume 244 of *Graduate Texts in Mathematics*. Springer London, London, 2008.

[20] Paul J. Gans. Open First-Order Stochastic Processes. *The Journal of Chemical Physics*, 33(3):691–694, September 1960.

[21] Donald A. McQuarrie. Kinetics of Small Systems. I. *The Journal of Chemical Physics*, 38(2):433–436, January 1963.

[22] Chetan Gadgil, Chang Hyeong Lee, and Hans G. Othmer. A stochastic analysis of first-order reaction networks. *Bulletin of Mathematical Biology*, 67(5):901–946, September 2005.

[23] Matthias Reis, Justus A. Kromer, and Edda Klipp. General solution of the chemical master equation and modality of marginal distributions for hierarchic first-order reaction networks. *Journal of Mathematical Biology*, 77(2):377–419, August 2018.

[24] Ravi P. Agarwal and Donal O'Regan. *Ordinary and partial differential equations: with special functions, Fourier series, and boundary value problems*. Springer, New York, NY, 2009. OCLC: ocn227032618.

[25] Lucy Ham, Rowan D. Brackston, and Michael P. H. Stumpf. Extrinsic Noise and Heavy-Tailed Laws in Gene Expression. *Physical Review Letters*, 124(10):108101, March 2020.

[26] Ole. E. Barndorff-Nielsen and Steen Thorbjørnsen. Self-Decomposability and Lévy Processes in Free Probability. *Bernoulli*, 8(3):323–366, 2002.

[27] Makoto Yamazato. Unimodality of Infinitely Divisible Distribution Functions of Class L. *The Annals of Probability*, 6(4):523–531, 1978. Publisher: Institute of Mathematical Statistics.

[28] Nicola Cufaro Petroni and Piergiacomo Sabino. Gamma Related Ornstein-Uhlenbeck Processes and their Simulation. Preprint, arXiv: 2003.08810, March 2020.

[29] Dimitris Karlis and Evdokia Xekalaki. Mixed Poisson Distributions. *International Statistical Review / Revue Internationale de Statistique*, 73(1):35–58, 2005.

[30] Pavol Bokes, John R. King, Andrew T. A. Wood, and Matthew Loose. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *Journal of Mathematical Biology*, 64(5):829–854, April 2012.

[31] Gennady Gorin and Lior Pachter. Length Biases in Single-Cell RNA Sequencing of pre-mRNA. Preprint, bioRxiv: 10.1101/2021.07.30.454514, July 2021.

[32] Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson, and Jason H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, April 2017.

[33] O Kessler, Y Jiang, and L A Chasin. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Molecular and Cellular Biology*, 13(10):6211–6222, October 1993.

[34] Heather L. Drexler, Karine Choquet, and L. Stirling Churchman. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Molecular Cell*, 77(5):985–998.e8, March 2020.

[35] Steffen Heber, Max Alekseyev, Sing-Hoi Sze, Haixu Tang, and Pavel A. Pevzner. Splicing graphs and EST assembly problem. *Bioinformatics*, 18(suppl_1):S181–S188, July 2002.

[36] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018.

[37] Vanessa M Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C Wilson, Renee Moore, Terrill K McClanahan, Svetlana Sadekova, and Joel A Klappenbach. Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology*, 35(10):936–939, October 2017.

[38] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, September 2017.

[39] Luyi Tian, Jafar S. Jabbari, Rachel Thijssen, Quentin Gouil, Shanika L. Amarasinghe, Hasaru Kariyawasam, Shian Su, Xueyi Dong, Charity W. Law, Alexis Lucattini, Jin D. Chung, Timur Naim, Audrey Chan, Chi Hai Ly, Gordon S. Lynch, James G. Ryall, Casey J.A. Anttila, Hongke Peng, Mary Ann Anderson, Andrew W. Roberts, David C.S. Huang, Michael B. Clark, and Matthew E. Ritchie. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. preprint, Genomics, August 2020.

[40] Michael Hagemann-Jensen, Christoph Ziegenhain, Ping Chen, Daniel Ramsköld, Gert-Jan Hendriks, Anton J. M. Larsson, Omid R. Faridani, and Rickard Sandberg. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology*, 38(6):708–714, June 2020.

[41] Michael Hagemann-Jensen, Christoph Ziegenhain, and Rickard Sandberg. Scalable full-transcript coverage single cell RNA sequencing with Smart-seq3xpress. Preprint, bioRxiv: 2021.07.10.451889, July 2021.

[42] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, October 2002.

[43] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J, 3rd ed edition, 2003.

[44] A. G. Asuero, A. Sayago, and A. G. González. The Correlation Coefficient: An Overview. *Critical Reviews in Analytical Chemistry*, 36(1):41–59, January 2006.

[45] Norman MacDonald. *Time Lags in Biological Models*, volume 27 of *Lecture Notes in Biomathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1978.

[46] Kevin Burrage, Pamela Burrage, Andre Leier, and Tatiana T. Marquez-Lago. A Review of Stochastic and Delay Simulation Approaches in Both Time and Space in Computational Cell Biology. In David Holcman, editor, *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*. Springer International Publishing, Cham, 2017.

[47] Tomáš Gedeon and Pavol Bokes. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. *Biophysical Journal*, 103(3):377–385, August 2012.

[48] Andre Leier and Tatiana T. Marquez-Lago. Delay chemical master equation: direct and closed-form solutions. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150049, July 2015.

[49] Jacek Miekisz, Jan Poleszczuk, Marek Bodnar, and Urszula Foryś. Stochastic Models of Gene Expression with Delayed Degradation. *Bulletin of Mathematical Biology*, 73(9):2231–2247, September 2011.

[50] Farzad Fatehi, Yuliya N. Kyrychko, and Konstantin B. Blyuss. A new approach to simulating stochastic delayed systems. *Mathematical Biosciences*, 322:108327, April 2020.

[51] Manuel Barrio, Kevin Burrage, Andre Leier, and Tianhai Tian. Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation. *PLOS Computational Biology*, 2(9):e117, September 2006.

[52] L F Lafuerza and R Toral. Exact solution of a stochastic protein dynamics model with delayed degradation. *Physical Review E*, 84:051121, November 2011.

[53] L F Lafuerza and R Toral. Role of delay in the stochastic creation process. *Physical Review E*, 84:021128, August 2011.

[54] Tao Jia and Rahul V Kulkarni. Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting. *Physical Review Letters*, 106:058102, February 2011.

[55] Qingchao Jiang, Xiaoming Fu, Shifu Yan, Runlai Li, Wenli Du, Zhixing Cao, Feng Qian, and Ramon Grima. Neural network aided approximation and parameter inference of non-Markovian models of gene expression. *Nature Communications*, 12(1):2618, December 2021.