Supplementary Methods

Two kinds of time in mammalian promoters: a refractory class that detects change faster per burst cycle

Matthew D. Schwartz and Martin Hemberg

This manuscript is preliminary. The first draft was written by Claude (Anthropic), and the text and computations are still under review by the authors.

Supplementary Methods. Exact likelihood of labeled and unlabeled RNA counts under a labeling window, its numerical validation, and the exact test

Related to Methods and Figures 1, 2, 3, 5 and 6.

This supplement contains the derivations and numerical details that the main text and the Methods summarize in words. Section 1 derives the labeling-window likelihood and Section 2 describes the fit, the exact test and the class rule. Section 3 introduces the one-parameter family of silent-period shapes, Section 4 reports the simulated detection power of a two-hour window at the depth of the data, Section 5 gives the allele-resolved models and Section 6 the closed forms for noise and detection delay used in Figure 6.

1 The labeling-window likelihood

Notation. Rates are per hour. A promoter copy has one active state, ON, left at rate $b \equiv k_{\text{off}}$, in which transcripts are made at rate $r \equiv k_{\text{syn}}$; each transcript is degraded at rate k_d. The silent period T_{off} has mean $1/a$ with $a \equiv k_{\text{on}}$, whatever its shape. The free parameters of a gene are $\theta = (a, b, r)$. The window has length T; σ indexes promoter states, o and n old and new counts, c cells, g genes and h depth bins with capture efficiencies β_h; q and q' are the label-misassignment probabilities of Section 1.5; ν is the uniformization rate and Λ always the likelihood-ratio statistic of Section 2.

1.1 Counts, genes and fixed quantities

The data are the deposited NASC-seq2 tables of reconstructed new and old molecules for primary mouse fibroblasts (adult tail) of an F1 hybrid cross (CAST/EiJ × C57BL/6J)[1]: two matrices of 37,249 genes by 8,916 cells, every cell exposed to 4-thiouridine (4sU) for $T = 2$ h before lysis. T-to-C conversions in the reads separate, per gene and cell, the molecules synthesized inside the window (new RNA) from those already present at its start and not degraded before its end (old RNA). We take the deposited reconstruction as given. Molecule numbers were rounded to the nearest integer, giving for gene g and cell c an old count o_{gc} and a new count n_{gc} with the two alleles pooled. The inclusion rule (Methods) keeps 23,285 genes, of which 39 with very large count ranges exceeded the largest feasible grid (Section 1.6), leaving 23,246 fitted (Table S1 lists all of them).

Two quantities, the mRNA degradation rate and the relative capture efficiency, are fixed rather than fitted. The degradation rate $k_{d,g}$ of each gene is set to the steady-state value the NASC-seq2 study tabulated for the same cells from the fraction of new molecules over the window, $k_{d,g} = -\ln(1 - f_{\text{new},g})/T$[1]; a gene absent from that table would receive the default $k_d = 0.065$ h^{-1}, but

every included gene has a tabulated rate. Sequencing depth enters through a relative capture efficiency. Let $t_c = \sum_g (o_{gc} + n_{gc})$ be the total molecule count of cell c and $d_c = t_c / \mathrm{median}_{c'}\, t_{c'}$ its relative depth. Cells are divided into $H = 20$ bins at the quantiles of d_c, and bin h receives the efficiency

$$\beta_h \;=\; \frac{\mathrm{median}_{c \in h}\, d_c}{\max_{h'}\ \mathrm{median}_{c \in h'}\, d_c}\,, \tag{1.1}$$

so that the deepest bin has $\beta_h = 1$ and every other bin is modeled as a binomial thinning of it. We write $h(c)$ for the bin of cell c and C_h for the number of cells in bin h.

1.2 Promoter models

A promoter occupies one of K states, exactly one of which is active. It is a continuous-time Markov chain with generator $\mathsf{W}(a, b)$, where $\mathsf{W}_{\sigma\sigma'}$ ($\sigma \neq \sigma'$) is the rate from σ to σ' and $\mathsf{W}_{\sigma\sigma} = -\sum_{\sigma' \neq \sigma} \mathsf{W}_{\sigma\sigma'}$. Every model we use is linear in the two promoter rates,

$$\mathsf{W}(a, b) \;=\; a\,\mathsf{W}_A + b\,\mathsf{W}_B\,, \tag{1.2}$$

with W_A supported on the silent rows and W_B on the active row. The two-state (telegraph) model has $K = 2$ and, in the state order (OFF, ON),

$$\mathsf{W}^{(1)}(a, b) \;=\; \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}, \tag{1.3}$$

so that the silent period is exponential with density $f_1(t) = a e^{-at}$. In the multi-step model with k steps the promoter returns from silence through k sequential first-order stages, ON $\xrightarrow{b}$ OFF_k $\xrightarrow{ka}$ $\cdots$ $\xrightarrow{ka}$ OFF_1 $\xrightarrow{ka}$ ON, each stage left at rate ka. Its silent period is the sum of k exponential stages, an Erlang-k variable with density

$$f_k(t) \;=\; \frac{(ka)^k\, t^{k-1} e^{-kat}}{(k-1)!}\,, \qquad \mathbb{E}[T_{\mathrm{off}}] = \frac{1}{a}\,, \qquad \mathrm{CV}^2[T_{\mathrm{off}}] = \frac{1}{k}\,, \tag{1.4}$$

so that $k = 1$ is the two-state model. This is the parameterization displayed in the Results. The genome-wide test uses $k = 2$, the smallest departure from a memoryless silence, whose generator in the order $(\mathrm{OFF}_1, \mathrm{OFF}_2, \mathrm{ON})$ is

$$\mathsf{W}^{(2)}(a, b) \;=\; \begin{pmatrix} -2a & 0 & 2a \\ 2a & -2a & 0 \\ 0 & b & -b \end{pmatrix}. \tag{1.5}$$

Nothing below uses more of W than Eq. (1.2) and a single active state, so any phase-type silent period (any distribution with a rational Laplace transform) is accommodated by a different pair $(\mathsf{W}_A, \mathsf{W}_B)$, and the models with unequal stages and parallel exits in Section 3 are built in this way.

The third model named in the main text, the clustered promoter, has a silent period with two timescales: on leaving ON the promoter enters a fast silent state OFF_f with probability w or a slow one OFF_s with probability $1 - w$, and returns to ON at rate a/x or a/y respectively. With states ordered $(\mathrm{OFF}_f, \mathrm{OFF}_s, \mathrm{ON})$,

$$\mathsf{W}^{\mathrm{C}}(a, b) \;=\; \begin{pmatrix} -a/x & 0 & a/x \\ 0 & -a/y & a/y \\ wb & (1-w)\,b & -b \end{pmatrix}, \tag{1.6}$$

where x and y are fixed by requiring a mean silent period $1/a$ and a squared coefficient of variation c (Section 3.1, Eq. (3.2)). The clustered class of the main text is defined by the equal-weight member $w = 1/2$ at $c = 2$, whose exit rates are $(2 \pm \sqrt{2})\, a$. The entry probability w and the dispersion c are part of the model, not fitted, so the clustered mixture $\mathrm{C}(w, c)$ has the same three free parameters as the other two models.

1.3 The joint chain over the window

Closed forms and error-controlled evaluations exist for the stationary count distribution of the two-state promoter and its extensions[2–4], and the window distribution of new counts under the two-state promoter has been evaluated and fitted gene by gene to these data[1]. The test used here requires the joint distribution of old and new counts under an arbitrary phase-type promoter, and that distribution depends in turn on the joint distribution of promoter state and transcript count at the moment the window opens.

Set $t = 0$ at the start of the window. Molecules present at $t = 0$ are old, molecules made in $(0, T]$ are new, and both kinds are degraded at rate k_d per molecule. The triple (σ_t, o_t, n_t) of promoter state, old count and new count is a Markov chain with transitions

$$\begin{aligned} (\sigma, o, n) &\to (\sigma', o, n) && \text{at rate } \mathsf{W}_{\sigma\sigma'}(a, b), \\ (\mathrm{ON}, o, n) &\to (\mathrm{ON}, o, n{+}1) && \text{at rate } r, \\ (\sigma, o, n) &\to (\sigma, o{-}1, n) && \text{at rate } k_d\, o, \\ (\sigma, o, n) &\to (\sigma, o, n{-}1) && \text{at rate } k_d\, n. \end{aligned} \tag{1.7}$$

Writing $p_t(\sigma, o, n)$ for its distribution, the master equation is

$$\begin{aligned} \frac{d}{dt}\, p_t(\sigma, o, n) \;=\; & \sum_{\sigma'} p_t(\sigma', o, n)\, \mathsf{W}_{\sigma'\sigma} \;+\; r\, \mathbf{1}_{\mathrm{ON}}(\sigma)\, [p_t(\sigma, o, n{-}1) - p_t(\sigma, o, n)] \\ & +\; k_d[(o{+}1)\, p_t(\sigma, o{+}1, n) - o\, p_t(\sigma, o, n)] \;+\; k_d[(n{+}1)\, p_t(\sigma, o, n{+}1) - n\, p_t(\sigma, o, n)], \end{aligned} \tag{1.8}$$

with $p_t(\sigma, o, -1) \equiv 0$. In matrix form $\dot{p}_t = p_t\, \mathsf{Q}$ with Q the generator of Eq. (1.7).

The initial condition expresses stationarity at the opening of the window. Let $\pi(\sigma, m)$ be the stationary joint distribution of promoter state and transcript count for the same rates, the normalized solution of

$$0 \;=\; \sum_{\sigma'} \pi(\sigma', m)\, \mathsf{W}_{\sigma'\sigma} \;+\; r\, \mathbf{1}_{\mathrm{ON}}(\sigma)\, [\pi(\sigma, m{-}1) - \pi(\sigma, m)] \;+\; k_d[(m{+}1)\, \pi(\sigma, m{+}1) - m\, \pi(\sigma, m)]. \tag{1.9}$$

Then

$$p_0(\sigma, o, n) \;=\; \pi(\sigma, o)\, \delta_{n,0}\,, \tag{1.10}$$

and the single-copy distribution of the window is the promoter-marginalized solution at $t = T$,

$$P_1(o, n \mid \theta;\, k_d, T) \;=\; \sum_{\sigma=1}^{K} p_T(\sigma, o, n)\,. \tag{1.11}$$

For every promoter model the old-count marginal of P_1 is a binomial thinning of the stationary count with survival probability $e^{-k_d T}$, and the two counts are correlated through the promoter state at $t = 0$.

The models are distinguished by the new count and by its joint distribution with the old count. The new count is accumulated from zero over an interval of known length comparable to the

promoter dwell times, and its distribution therefore depends on the shape of the silent-period distribution as well as on its mean, whereas a stationary snapshot constrains the silence mainly through its mean[5,6]. The shape of the silent-period distribution enters the distribution of the counts in three ways. First, for a stationary promoter silent at $t = 0$ the remaining silent time has the equilibrium residual-life distribution,

$$\Pr(\text{no return to ON in } (0,T] \mid \text{silent at } 0) \;=\; a\int_T^\infty [1 - F_{\text{off}}(u)]\, du, \tag{1.12}$$

equal to e^{-aT} for the exponential silence and $e^{-2aT}(1+aT)$ for the two-step silence of the same mean. The difference is largest near $aT \approx 1$ and vanishes as $aT \to 0$. Second, when the window spans several cycles the number of bursts inside it, and hence the spread of positive counts, is less variable under a regular silence. Third, the old count indicates how recently the promoter was active, and under a multi-step silence a promoter that has just burst is less likely to fire again soon. Old and new counts in one cell are therefore less positively associated than under a memoryless promoter with the same rates, and this third effect persists at small aT. We computed by simulation the power that these three effects give the test at the window length and depth of the data (Section 4).

1.4 Evaluation on a truncated grid by uniformization

The chain is evaluated on the finite grid $0 \le o \le N_o$, $0 \le n \le N_n$ (Section 1.6). Because old counts only decrease during the window, the window chain needs a boundary condition only on the new axis, where the birth out of $n = N_n$ is removed. In the stationary problem of Eq. (1.9) the birth out of $m = N_o$ is removed likewise. With these reflecting boundaries the truncated generator conserves probability exactly, and the distortion near the boundary is controlled by placing it far above the observed counts rather than by tracking an absorbed mass as in the finite state projection[7]. For the grids used in the fits (Section 1.6) we did not compute a separate bound on the probability mass at the boundary, because both distributions decay geometrically above the largest observed count.

Proposition 1 (Uniformization with an explicit tail)**.** *Let $\nu \ge \max_i |\mathsf{Q}_{ii}|$; on the truncated grid one may take ν equal to the largest promoter exit rate plus $r + (N_o + N_n)k_d$. Then $\mathsf{U} = \mathsf{I} + \mathsf{Q}/\nu$ is a stochastic matrix,*

$$p_T \;=\; \sum_{m=0}^{\infty} e^{-\nu T}\frac{(\nu T)^m}{m!}\, p_0\, \mathsf{U}^m, \tag{1.13}$$

and truncating the series at the first index M_ε at which the accumulated Poisson weight exceeds $1 - \varepsilon$ leaves an ℓ_1 error below ε:

$$\Big\| p_T - \sum_{m=0}^{M_\varepsilon} e^{-\nu T}\frac{(\nu T)^m}{m!}\, p_0 \mathsf{U}^m \Big\|_1 \;\le\; 1 - \sum_{m=0}^{M_\varepsilon} e^{-\nu T}\frac{(\nu T)^m}{m!} \;<\; \varepsilon. \tag{1.14}$$

The identity is the randomization of a continuous-time chain by a Poisson clock of rate ν[8–10]. The bound holds because every term $p_0\mathsf{U}^m$ is a probability vector with no negative entries, and the omitted terms therefore carry exactly the omitted Poisson weight. We use $\varepsilon = 10^{-13}$ and compute the achieved tail alongside every evaluated distribution. The number of terms is also capped, at $3\nu T + 60\sqrt{\nu T + 1} + 200$ and never more than 2,000 in the genome-wide fits (20,000 in the profile fits), and a penalty in the fit (Section 2.1) keeps trial parameters in the region where this cap is inactive.

The truncated stationary problem, Eq. (1.9), is block tridiagonal in m with $K \times K$ blocks: $\mathsf{L}\pi_{m-1} + \mathsf{D}_m\pi_m + \mathsf{B}_m\pi_{m+1} = 0$ with $\pi_m = (\pi(\sigma, m))_\sigma$, $\mathsf{L} = r\,\mathsf{E}$, $\mathsf{B}_m = (m{+}1)k_d\,\mathsf{I}$, $\mathsf{D}_m = \mathsf{W}^\top -$

$r\,\mathsf{E}\,\mathbf{1}[m < N_o] - mk_d\,\mathsf{I}$, and E the projector onto ON. Of our two implementations (Section 1.7), the reference implementation solves the assembled sparse system directly with one equation replaced by the normalization. The vectorized implementation eliminates from the top of the grid down: $\mathsf{G}_{N_o} = \mathsf{D}_{N_o}$, $\mathsf{T}_m = -\mathsf{G}_m^{-1}\mathsf{L}$, $\mathsf{G}_{m-1} = \mathsf{D}_{m-1} + \mathsf{B}_{m-1}\mathsf{T}_m$, then $\mathsf{G}_0\pi_0 = 0$ with the normalization and $\pi_m = \mathsf{T}_m\pi_{m-1}$. Eliminating in the opposite direction, from $m = 0$ upward, propagates rounding error into the decaying tail, where it is amplified relative to the true probabilities, and for slowly switching promoters the result can then be wrong at order one. The downward sweep agrees with the sparse direct solve to rounding error at every parameter point of Section 1.7. In the vectorized implementation p is held as a $K \times (N_o{+}1) \times (N_n{+}1)$ array and $p\,\mathsf{U}$ is applied as a sum of axis-local operations (a $K \times K$ contraction, a shift along the new axis in the ON slice, weighted shifts along each count axis), batched over parameter vectors that share a grid and a common ν. The reference implementation assembles Q as a sparse matrix and applies its exponential to p_0 by a truncated-Taylor algorithm[11].

1.5 Two alleles, label misassignment and capture

Three exact operations take P_1 to the distribution of the observed pair.

Two copies. The pooled counts of an autosomal gene are modeled as the sum over two independent copies with identical parameters,

$$P_2(o,n) \;=\; \sum_{o'=0}^{o}\sum_{n'=0}^{n} P_1(o',n')\,P_1(o-o',\,n-n')\,, \tag{1.15}$$

computed by fast Fourier transform on the grid $\{0,\dots,2N_o\}\times\{0,\dots,2N_n\}$, with negative rounding residues set to zero and the result renormalized. For an X-linked gene in a cell with one active X chromosome the count distribution is P_1 itself on the same grid, with the rest of the fit unchanged (Section 5.1). The one-copy and two-copy distributions are nearly degenerate in the bursty regime: when the duty fraction $\pi_{\mathrm{on}} = a/(a+b)$ is small the two copies are rarely active together, their superposed bursts resemble one copy that leaves silence twice as often, and

$$P_2(o,n;\,a,b,r) \;\approx\; P_1(o,n;\,2a,b,r)\,, \tag{1.16}$$

so that a two-copy fit to a one-copy gene gives about half the true a at nearly the right b and r. The two sides differ where both copies are active at once, which one copy cannot produce: the two-copy distribution has a heavier upper tail (simultaneous bursts) and, for the multi-step model, a less regular superposed silence than either copy alone. Through these differences the one-copy and two-copy models are distinguishable at the depth of the data, and the log-likelihood lost by fitting the wrong copy number is measured by simulation in Section 5.5.

Label misassignment. A new molecule none of whose reads carries a conversion is counted as old, and a sequencing or alignment error can make an old molecule look new. We model both errors per molecule and independently: a new molecule is read as old with probability $q = 0.10$ and an old molecule is read as new with probability $q' = 0.0208$. If $j \sim \mathrm{Bin}(o,q')$ old molecules and $i \sim \mathrm{Bin}(n,q)$ new molecules are misread, the apparent counts are $\tilde o = o - j + i$ and $\tilde n = n - i + j$ with the total conserved, and

$$P_3(\tilde o,\tilde n) \;=\; \sum_{o,n} P_2(o,n)\sum_{j=0}^{o}\sum_{i=0}^{n}\binom{o}{j}q'^{\,j}(1-q')^{o-j}\binom{n}{i}q^{\,i}(1-q)^{n-i}\;\mathbf{1}[\tilde o = o-j+i]\;\mathbf{1}[\tilde n = n-i+j]\,. \tag{1.17}$$

The sum is evaluated as diagonally shifted, binomially weighted copies of P_2. Indices whose weights fall below 10^{-16} everywhere are dropped, at most 60 values of j are used, and the result is renormalized (both implementations apply the same truncation). Both misassignment probabilities are fixed for all genes and cells, not fitted: q is the estimated probability that a labeled molecule acquires no T-to-C conversion at the per-position conversion rate reported for these cells [1], and q' is the median fraction of molecules called new in the 4sU-free control cells of the same study.

Capture. Each apparent molecule in a cell of depth bin h is captured independently with probability β_h of Eq. (1.1), irrespective of its label:

$$P_{\mathrm{obs}}^{(h)}(x,y) \;=\; \sum_{\tilde{o}\ge x}\sum_{\tilde{n}\ge y} \binom{\tilde{o}}{x}\beta_h^{x}(1-\beta_h)^{\tilde{o}-x}\binom{\tilde{n}}{y}\beta_h^{y}(1-\beta_h)^{\tilde{n}-y}\, P_3(\tilde{o},\tilde{n})\,, \tag{1.18}$$

a pair of matrix products, renormalized on the grid (the identity in the deepest bin). Misassignment and capture act independently on each molecule and the capture probability does not depend on the label, so the order of the two operations is immaterial. Capture variation finer than the H bins is not in the model, and its effect on the statistic is measured for each gene by the capture-jitter control of Section 2.3.

The log-likelihood of gene g under a promoter model with parameters θ is then

$$\ell_g(\theta) \;=\; \sum_c \log P_{\mathrm{obs}}^{(h(c))}\!\left(o_{gc},\, n_{gc}\,\middle|\,\theta;\, k_{d,g},\, T,\, q,\, q'\right), \tag{1.19}$$

with one evaluation of Eqs. (1.11) to (1.17) per parameter vector and one application of Eq. (1.18) per occupied depth bin. Probabilities are floored at 10^{-300} before the logarithm. The two-state, multi-step and clustered models differ only in the promoter block W that enters P_1.

1.6 Choice of the count grid

For gene g the single-copy grid is

$$N_o \;=\; \lfloor 1.6\, \max_c o_{gc} + 20 \rfloor, \qquad N_n \;=\; \lfloor 1.6\, \max_c n_{gc} + 15 \rfloor, \tag{1.20}$$

so that after the two-copy convolution every observed pair lies well inside the support. The margin keeps the probability mass at the reflecting boundaries negligible and includes the true counts above the observed ones over which Eqs. (1.17) and (1.18) sum. Genes evaluated in one batch share the elementwise maximum of their grids, rounded up to fixed multiples. Enlarging a grid adds only states of negligible probability. The cost of one evaluation grows as $K(N_o{+}1)(N_n{+}1)$ times the number of uniformization terms, itself proportional to $N_o + N_n$. Of the 23,285 genes passing inclusion, 33 were set aside before fitting because a lower bound on νT computed from their grid and degradation rate, $[(N_o + N_n)k_d + r_0]\,T$ with $r_0 = \max(0.05,\, k_d m_{\mathrm{tot}})$ and m_{tot} the mean total count per cell, exceeded 350. The remaining unfitted genes have count grids among the very largest and their fits did not complete. All 39 are listed in Table S1 with their grids.

1.7 Implementation and numerical validation

The vectorized likelihood, its exact first and second derivatives with respect to θ (forward-mode differentiation through the uniformization loop) and the optimizer of Section 2.1 are written in JAX in double-precision arithmetic and run unchanged on ordinary and graphics processors. Genes that share a count grid are fitted together. The reference implementation (sparse direct solve for

Eq. (1.9), Taylor action of the matrix exponential for Eq. (1.8), simplex optimization) shares no numerical component with the vectorized implementation.

We tested the numerical construction in four ways (Table S6 lists every parameter point with its outcome; Figure S1). (a) *Agreement of the two implementations.* $P_{\mathrm{obs}}^{(h)}$ of Eq. (1.18), with convolution, misassignment at $(q, q') = (0.10, 0.0208)$ and capture, was computed by both implementations at 64 parameter points (both models; eight combinations of the promoter rates spanning previously fitted fibroblast genes from typical to the most extreme; settings of (k_d, T, β) covering slow to fast turnover, two window lengths and two capture efficiencies) on grids with N_o from 74 to 1,167 and N_n from 29 to 186. At 64 of the 64 points the two implementations differed by less than the agreement threshold of 10^{-6}, the largest absolute difference over all entries of all distributions was 3.1×10^{-14}, and the achieved tail of Eq. (1.14) was below 9.8×10^{-14} everywhere. (b) *Agreement of the fitted statistic.* We simulated 40 synthetic genes from the two-state model at seven combinations of the promoter rates through the full observation model (120 cells per gene over depth bins of unequal capture) and fitted them under both models with both implementations and their respective optimizers. The statistic Λ of Eq. (2.1) agreed between the two implementations to within 0.023 for every gene (agreement criterion 0.05), and the remaining difference reflects where each optimizer terminates. (c) *Model-independent identities.* At the parameter point of (d), the old-count marginal of P_1 equals the binomially thinned stationary count and the mean old count of P_2 is twice that of P_1, both with error below 10^{-8}, and as $T \to 0$ the distribution reduces to π with no new molecules. (d) *Stochastic simulation.* The reference P_1 was compared with exact stochastic simulation of Eq. (1.7), 400,000 trajectories per model for the two-state and two-step promoters at one bursty parameter point with $T = 2$ h, each started from an exact draw of (σ, o) from π. The criterion is that more than 99% of grid entries with probability above 5×10^{-5} agree with the simulated frequencies within five Monte Carlo standard errors plus 5×10^{-5}, and the achieved fraction was 100% with a largest absolute deviation of 5.2×10^{-4}. Mass is conserved by construction: the truncated series sums to one minus the reported tail, and the renormalizations remove only rounding residue.

2 Fitting, the exact test, the controls and the class rule

2.1 Parameterization, optimizer and convergence

For one gene the data are the pairs (o_c, n_c) in the 8,916 cells, the depth bin $h(c)$ of each cell, and the gene's fixed k_d. The observed-count distribution $P_\theta^{(h)}$ is that of Eq. (1.18) and the log-likelihood $\ell(\theta)$ that of Eq. (1.19). The two-state and multi-step models differ only in the promoter block entering P_1. With $\hat{\ell}_{\mathrm{two}}$ and $\hat{\ell}_{\mathrm{multi}}$ the maximized log-likelihoods of the two-state model and of the multi-step model with $k = 2$, the statistic is

$$\Lambda \;=\; 2\,(\hat{\ell}_{\mathrm{multi}} - \hat{\ell}_{\mathrm{two}}), \tag{2.1}$$

positive when the multi-step model is favored. Both models have the free parameters (a, b, r). Neither is nested in the other, and Λ takes either sign.

Parameters are fitted in unconstrained coordinates $u \in \mathbb{R}^3$ mapped smoothly onto an open box,

$$\log \theta_j \;=\; l_j + (h_j - l_j)\,\varsigma(u_j), \qquad \varsigma(x) = \frac{1}{1 + e^{-x}}, \tag{2.2}$$

with bounds (e^{l_j}, e^{h_j}) of $(10^{-4}, 200)$ for a, $(10^{-3}, 2{,}000)$ for b and $(10^{-2}, 10^5)$ for r, all per hour. Inside the box the objective is exactly $-\ell$. A penalty $10^4 \max\{0, \nu_{\mathrm{u}} T/550 - 1\}$ on the uniformization load $\nu_{\mathrm{u}} T = [\max(a_{\mathrm{exit}}, b) + r + (N_o + N_n) k_d]\, T$, with $a_{\mathrm{exit}} = a$ for the two-state model and $2a$ for the

two-step model, keeps trial points out of the region where the series of Eq. (1.13) would need more than 2,000 terms. Wherever the penalty vanishes the cap is inactive and the likelihood is exact, which is the case at the fitted optima of every evaluable gene but one, whose endpoints lie on the soft boundary.

Every fit of a gene starts from one common point: a heuristic guess ($a = 0.02$, $b = 0.1$, $r = r_0$ per hour, with r_0 as in Section 1.6) moved by one Gauss–Newton step in $\log\theta$ that matches the logarithm of the depth-averaged model distribution to the logarithm of the empirical joint histogram of (o, n), weighted by the empirical frequencies. The fit then proceeds by 120 iterations of the Adam first-order method[12] in u, with the step size lowered in three stages from 0.10 to 0.01. These are followed by up to 8 damped Newton iterations $(\mathsf{H} + \lambda\mathsf{I})\,\delta = -\nabla$ on the exact gradient and Hessian in u, each clipped to $|\delta_j| \le 0.5$ and accepted only if the objective does not increase, with λ lowered after an accepted step and raised after a rejected one. A fit has converged when $\|\nabla\|_\infty < 10^{-3}$ at the endpoint or the last accepted Newton step has $\|\delta\|_\infty < 10^{-5}$.

A single start per model is unreliable for a likelihood-ratio statistic between non-nested models, because one model's fit can meet its convergence criterion in a local optimum far above the basin that the other model's fit reaches. In fits to counts simulated from the two-state model this happened for two synthetic genes, giving $\Lambda = 600.7$ and $\Lambda = -579.9$. Every gene is therefore fitted in four stages, applied unconditionally: (i) the two-state model from the common start; (ii) the multi-step model from the two-state optimum; (iii) the two-state model again from the multi-step optimum, keeping the better of its two endpoints; (iv) the multi-step model again from the best two-state endpoint, keeping the better of its two endpoints. The values entering Eq. (2.1) are the kept values, and convergence is assessed at the kept endpoints. A gene is evaluable when both kept fits converged (16,858 of 23,246; the breakdown of the rest is given in the Results and Table S1).

2.2 The exact test

The null hypothesis for each gene is its fitted two-state model, sampled at that gene's own cells and depths, and P is the probability under this null of a value of Λ at least as large as the one observed. Let $\hat{P}^{(h)}_{\text{two}}$ and $\hat{P}^{(h)}_{\text{multi}}$ be the observed-count distributions of Eq. (1.18) at the kept optima on the grid of the fit, and define the per-cell log-likelihood difference on the grid and its observed sum,

$$d_h(o,n) = \log \hat{P}^{(h)}_{\text{multi}}(o,n) - \log \hat{P}^{(h)}_{\text{two}}(o,n), \qquad S_{\text{obs}} = \sum_c d_{h(c)}(o_c, n_c) = \Lambda/2. \tag{2.3}$$

Under the fitted two-state model the cells are independent, the pair (O_c, N_c) of cell c is distributed as $\hat{P}^{(h(c))}_{\text{two}}$, and the null statistic $S_0 = \sum_c d_{h(c)}(O_c, N_c)$ is a sum of 8,916 independent discrete random variables with exactly known distributions, C_h of them identical in bin h. Its cumulant generating function is therefore the finite sum

$$\mathcal{K}(z) = \sum_{h=1}^{H} C_h \,\log \sum_{o,n} \hat{P}^{(h)}_{\text{two}}(o,n)\, e^{z\, d_h(o,n)}, \tag{2.4}$$

known without sampling or moment expansion for real and complex z alike (C_h is an integer, so that no branch of the complex logarithm has to be chosen).

Proposition 2 (Tail probability by contour inversion)**.** *Let $\hat{z}$ solve $\mathcal{K}'(\hat{z}) = S_{\text{obs}}$; $\mathcal{K}'$ is increasing, so the root is unique and has the sign of $S_{\text{obs}} - \mathcal{K}'(0)$. For $\hat{z} > 0$ the upper-tail probability of S_0 is*

$$P \;\equiv\; \Pr(S_0 \ge S_{\text{obs}}) \;=\; e^{\mathcal{K}(\hat{z}) - \hat{z} S_{\text{obs}}}\, \frac{1}{\pi} \int_0^\infty \operatorname{Re}\!\left[\frac{\exp\{\mathcal{K}(\hat{z}+iy) - \mathcal{K}(\hat{z}) - iy\, S_{\text{obs}}\}}{\hat{z}+iy}\right] dy, \tag{2.5}$$

and for $\hat{z} < 0$ the negative of the right side equals $\Pr(S_0 \le S_{\text{obs}})$, whose complement is P.

This is the Bromwich inversion of the moment generating function with the contour moved to the vertical line through the real saddle point, where the integrand decays fastest. No Gaussian approximation is made at the saddle. The integral is evaluated by the trapezoid rule with the step halved until the relative change is below 10^{-6}, in logarithmic space so that arbitrarily small values of P do not underflow. When $|\hat{z}|\sqrt{\mathcal{K}''(\hat{z})} < 0.05$, so that S_{obs} is near the null mean and P is moderate, the untilted inversion formula[13]

$$P = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Im}\left[e^{\mathcal{K}(iy) - iy\,S_{\mathrm{obs}}}\right]\frac{dy}{y} \tag{2.6}$$

is used instead. The prefactor of Eq. (2.5) alone is the Chernoff bound $P \le \exp\{-\sup_z[zS_{\mathrm{obs}} - \mathcal{K}(z)]\}$, reported as a rigorous upper limit where the contour integral did not converge on the largest count grids. The exact value was obtained for 16,844 of the 16,858 evaluable genes and the bound for 14.

Neither of the two standard approximations to the null distribution is used. Wilks' theorem does not apply: with k fixed at 2 the two models are distinct three-parameter families with no nesting, the two-state model lies at the boundary $\mathrm{CV}^2[T_{\mathrm{off}}] = 1$ of the Erlang family, and the finite-cell null distribution of Λ is neither χ^2_1 nor the half-χ^2_1 mixture. Because the generating function of Eq. (2.4) is exact, no asymptotic form is needed. The saddle point (Lugannani–Rice) approximation[14,15] is computed as a diagnostic but not used as the value, because the tilted null distribution mixes a smooth bulk with isolated atoms at rare grid points of large $|d|$, where the Gaussian expansion about the saddle is inaccurate.

The test holds both parameter vectors at their fitted values. We measured the variability that refitting adds under the null on a calibration cohort of 300 genes chosen before the genome-wide fit to span expression and grid size. For each gene, $B = 199$ datasets of 8,916 cells were drawn from $\hat{P}^{(h(c))}_{\mathrm{two}}$ with the real bin memberships and refitted under both models by the four-stage procedure, and the bootstrap P is $(1 + \#\{\Lambda^* \ge \Lambda\})/(199 + 1)$ with Λ^* the refitted statistics. Of the cohort, 206 genes have the complete set of replicates, and Table S6 pairs their exact and bootstrap P values (Figure S1). For 93% of the 206 cohort genes the two agree within a factor of two or the exact value lies below the resolution of the bootstrap, $1/(B + 1)$ (35 genes).

We controlled the false-discovery rate at 5% with the Benjamini–Hochberg procedure[16] over the 16,548 evaluable genes whose null distribution of S_0 is non-degenerate and whose tail probability came from the convergent integral. A gene with the Chernoff bound only, or whose two fitted models coincide so that every d_h vanishes, is not eligible for significance. On the test alone 3,737 genes are significant.

2.3 Gene-matched simulated controls and the class rule

A small P rejects the fitted two-state distribution as the source of the observed Λ but does not by itself identify an extra silent step as the only way to produce it. Two features of the data that the observation model omits could inflate Λ: capture variation finer than the H depth bins, and the doubling of gene copies in S and G2/M cells. We measure their effect gene by gene with a gene-matched simulated control (hereafter "control"): one synthetic dataset of 8,916 cells drawn from the gene's own fitted two-state model with the omitted feature added, passed through the same four-stage fit on the gene's grid from the gene's start, giving a null statistic specific to that gene's rates and depths.

(i) *Capture-jitter control.* Latent pooled counts are drawn from the gene's fitted single-copy two-state distribution $\hat{P}_{1,\mathrm{two}}$ convolved with itself and passed through the misassignment of Eq. (1.17), written $\mathcal{M}_{q,q'}(\hat{P}_{1,\mathrm{two}} * \hat{P}_{1,\mathrm{two}})$, and thinned per cell at $\beta_c = \min(1, \beta_{h(c)}J_c)$ with J_c lognormal of

Table M1: **Numbers of evaluable genes at each step of the class rule of Eq.** (2.7)**.** A control that could not be computed counts as failed.

Evaluable genes (both fits converged)	16,858
Significant on the exact test alone (5% FDR)	3,737
$\Lambda > 8.6$ (candidates)	1,231
of which significant on the exact test	1,229
capture-jitter control computed	1,168
passing the capture-jitter margin	358
copy-number control computed	1,182
passing the copy-number margin	356
passing both margins	304
surviving the G1 restriction (among candidates)	902
Refractory class (all conditions)	186
range of Λ in the class	12.9 to 461
Candidates failing at least one control (undetermined)	1,045
Two-state class ($\Lambda \leq -4.3$)	1,060
Undetermined	15,612

mean one and coefficient of variation 0.3. The refit gives Λ^{cap}. (ii) *Copy-number control.* Each simulated cell inherits the cell-cycle phase of the corresponding real cell (module scores of S and G2/M gene sets; Methods). Counts are drawn from the two-copy distribution for G1 cells, from its self-convolution (four copies) for G2/M cells and from either with probability one half for S cells, and thinned at $\beta_{h(c)}$ without jitter so that the copy-number effect is isolated. The refit gives Λ^{cn}. (iii) *G1 restriction.* Both models are refitted to the real counts of the G1 cells alone. Because Λ grows with the number of cells, the G1 value is compared with refits on two random subsets of all cells of the same size as the G1 set, and the gene survives when its G1 statistic is not lower than the smaller subset statistic by more than 4.3. Genes with fewer than 50 molecules in G1 cells cannot be tested, and a control that is missing or whose refit did not converge counts as failed.

A gene enters the refractory class when it is significant at 5% false-discovery rate on the exact test and

$$\Lambda > 8.6, \qquad \Lambda > 3\max(\Lambda^{\mathrm{cap}}, 4.3), \qquad \Lambda > 3\max(\Lambda^{\mathrm{cn}}, 4.3), \tag{2.7}$$

and it survives the G1 restriction. Because of the floor 4.3 inside the maxima, the smallest statistic with which a gene can enter is $3 \times 4.3 = 12.9$, irrespective of the control statistics. A gene enters the two-state class when $\Lambda \leq -4.3$. Evaluable genes meeting neither condition, including those above 8.6 that fail a control, are undetermined at this depth. The two thresholds are conventions fixed before the genome-wide fit, with $8.6 = 2 \times 4.3$, not quantiles of an asymptotic distribution (Section 2.2). We measured their operating characteristic instead: for data simulated from the two-state model at 24 grid points spanning silent period and burst size, the number of genes classified as refractory was 0, and for data simulated from the two-step model the power is the detection map of Section 4.1.

Table M1 gives the number of genes remaining after each condition of the rule. The control statistics are themselves often large (medians 11.2 and 16.4 over the candidates; 657 and 705 of them exceed 8.6), and for this reason entry requires a margin over the gene's own controls in addition to the fixed threshold. All 186 genes of the refractory class are significant on the exact test, so the class is the same whether or not that requirement is written into the rule. Membership under stricter and looser variants of Eq. (2.7) is reported in the Results, Figure S2 and Table S1.

For the clustered class we used the same likelihood, exact test and controls, applied inside the two-state class only, with $\mathrm{C}(1/2, 2)$ of Eq. (1.6) in the role of the multi-step model. The statistic is twice the difference between the maximized log-likelihoods of the clustered and two-state models, positive when the clustered model is favored. Both refit controls are regenerated with the clustered model as the alternative, and Eq. (2.7) with the G1 restriction is applied unchanged. Every gene with a converged two-state fit was refitted under the clustered model on its own grid except 135 whose grids exceed the cap used for the controls. Of the 1,060 two-state genes 1,037 were fitted, and the 23 that were not fitted count as not clustered. Genes whose comparison between the two-state and multi-step models is undetermined are not considered for the clustered class, whatever the outcome of their refits, and the clustered class is therefore a subset of the two-state class and disjoint from the refractory class.

3 The shape of the silent period

The two-state and multi-step models differ only in the distribution of the silent period T_{off}. Here both are embedded in a one-parameter family indexed by the squared coefficient of variation $s \equiv \mathrm{Var}[T_{\mathrm{off}}]/\mathbb{E}[T_{\mathrm{off}}]^2$ at fixed mean $1/a$, with everything downstream of the promoter unchanged. The two-state model has $s = 1$, the k-step model $s = 1/k$ and the clustered model $s = c > 1$. Every member has the same three free parameters (a, b, r) per gene, so maximized log-likelihoods at different s are compared directly.

3.1 The clustered mixture

In the clustered mixture $\mathrm{C}(w, c)$ of Eq. (1.6) the silent period is hyperexponential with mean dwell times $x < y$ of the fast and slow states in units of $1/a$, so that $\mathbb{E}[T_{\mathrm{off}}] = [wx + (1-w)y]/a$ and $\mathbb{E}[T_{\mathrm{off}}^2] = 2[wx^2 + (1-w)y^2]/a^2$. Requiring $\mathbb{E}[T_{\mathrm{off}}] = 1/a$ and $s = a^2\,\mathbb{E}[T_{\mathrm{off}}^2] - 1 = c$ gives

$$wx + (1-w)y = 1, \qquad wx^2 + (1-w)y^2 = \frac{1+c}{2}, \tag{3.1}$$

and eliminating y yields a quadratic in x whose root with $x < y$ is

$$x = 1 - \sqrt{\frac{(1-w)(c-1)}{2w}}, \qquad y = 1 + \sqrt{\frac{w(c-1)}{2(1-w)}}. \tag{3.2}$$

The other root exchanges the labels of the two states (weight w on the slow state is the model with fast weight $1-w$). A positive fast rate requires $x > 0$, that is

$$c < \frac{1+w}{1-w} \quad \Longleftrightarrow \quad w > \frac{c-1}{c+1}, \tag{3.3}$$

together with $c \geq 1$ ($c = 1$ gives both rates equal to a, the two-state model), and the bound of Eq. (3.3) is attained as the fast rate diverges. At $c = 2$ the fast weight must therefore exceed $1/3$. This bound excludes the symmetric pair of weights $(1/4, 3/4)$ as perturbations about the equal weight, because a fast weight of one quarter would cap s at $5/3$ and a slow-component weight of one quarter is the model $w = 3/4$ relabeled. Of that pair only $w = 3/4$ is thus a distinct admissible model, and the low-side perturbation of Section 3.5 is $w = 0.4$ instead. Both ends of the admissible range reduce observationally to the two-state model (instant returns from the fast state as $w \to 1/3$; a vanishing slow component as $w \to 1$). Table M2 lists the members fitted genome-wide.

Table M2: **Exit-rate multipliers of the clustered mixture** $\mathrm{C}(w, c)$**.** From Eq. (3.2). In every row the mean silent period is $1/a$ and its squared coefficient of variation is c; the last column is the share of silent time spent in the fast state, wx. The second row is the model that defines the clustered class, and the other rows are the sensitivity analyses of Section 3.5. Admissible fast weights at $c = 2$, 3, 4 are $w > 1/3$, $1/2$, $3/5$.

w	c	fast rate $/a = 1/x$	slow rate $/a = 1/y$	wx
0.40	2	$4 + 2\sqrt{3} \approx 7.464$	$(3 - \sqrt{3})/2 \approx 0.634$	0.054
0.50	2	$2 + \sqrt{2} \approx 3.414$	$2 - \sqrt{2} \approx 0.586$	0.146
0.75	2	$(1 - 1/\sqrt{6})^{-1} \approx 1.690$	$(1 + \sqrt{3/2})^{-1} \approx 0.449$	0.444
0.75	3	$(3 + \sqrt{3})/2 \approx 2.366$	$(\sqrt{3} - 1)/2 \approx 0.366$	0.317
0.75	4	$2 + \sqrt{2} \approx 3.414$	$(1 + 3/\sqrt{2})^{-1} \approx 0.320$	0.220

The equal-weight member at $c = 2$ has $x = 1 - 1/\sqrt{2}$, exit rates $(2 \pm \sqrt{2})\, a$, and a slow component that accounts for $(1-w)y = 85\%$ of the silent time. In live-cell recordings of a promoter with clustered bursts the fast component lasts minutes and the squared coefficient of variation of the silent period lies between 2.3 and 4.3[17,18], so $c = 2$ falls below the measured range and is conservative. Because the equal-weight member is admissible only for $c \leq 3$ (Eq. (3.3)), the shape perturbations of Section 3.5 use $w = 3/4$.

3.2 Two unequal stages and k equal stages

For $1/2 \leq s < 1$ the silent period is the sum of two exponential stages of unequal rate: on leaving ON at rate b the promoter enters the first stage, moves to the second at rate $m_1 a$ and returns to ON at rate $m_2 a$, with

$$m_1 = \frac{2}{1 + \sqrt{2s - 1}}, \qquad m_2 = \frac{2}{1 - \sqrt{2s - 1}}. \tag{3.4}$$

Then $\mathbb{E}[T_{\mathrm{off}}] = (1/m_1 + 1/m_2)/a = 1/a$ and $\mathrm{Var}[T_{\mathrm{off}}] = (1/m_1^2 + 1/m_2^2)/a^2 = s/a^2$. At $s = 1/2$ the stages are equal ($m_1 = m_2 = 2$) and the model is the two-step model of Eq. (1.5) exactly. As $s \to 1$ the second stage becomes instantaneous and the model approaches the two-state model. We denote the member with two unequal stages by $\mathrm{G}(s)$. For $s < 1/2$ the silent period is the chain of k equal stages of Eq. (1.4), with $s = 1/k$.

Each value of s then corresponds to exactly one model: the k-step model at $s = 1/k$ for integer $k \geq 3$, $\mathrm{G}(s)$ for $1/2 \leq s < 1$, the two-state model at $s = 1$, and $\mathrm{C}(1/2, s)$ for $1 < s < 3$ (admissible below 3 by Eq. (3.3)). The single coordinate s thus orders the models from bursty ($s > 1$) through memoryless ($s = 1$) to regular ($s < 1$), which is the classification of renewal kinetic schemes by the dispersion of the interval[19,20]. The main grid was $s \in \{0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 1, 1.3, 1.6, 2, 2.5\}$ and the extension toward regular silences was $s \in \{1/3, 1/4, 1/6, 1/8, 1/12, 1/20\}$, that is $k \in \{3, 4, 6, 8, 12, 20\}$ (the grid specified before analysis, Table S3, had 0.97 in place of 0.95). For every member the joint distribution of old and new counts at $t = T$ was computed by the uniformization of Section 1.4 with the promoter block of the generator replaced and ν taken from the largest total outflow of the full generator (near $s = 1$, $m_2 a$ can exceed b). On three parameter sets spanning slow to fast silences, $\mathrm{G}(1/2)$ reproduced the two-step distribution and $\mathrm{C}(1/2, 2)$ the independently coded clustered distribution to rounding error, $\mathrm{G}(s)$ approached the two-state distribution as $s \to 1$, and every distribution summed to one.

3.3 Profile likelihood over the shape

For gene g and each grid value of s we computed the profile log-likelihood

$$\ell_g(s) = \max_{a,\,b,\,r} \ \ell_g(a, b, r;\ s), \tag{3.5}$$

where ℓ_g is the exact log-likelihood of Eq. (1.19) under the full observation model of Section 1.5: two independent copies, misassignment probabilities q and q', binomial capture at the efficiency of each cell's depth bin, and the gene's fixed k_d. At each grid value we ran the optimizer of Section 2.1 with a shorter first-order schedule from two starts (the gene's two-state and two-step estimates mapped into the family; on the extension grid, the optimum at $s = 1/2$ and the two-step estimate) and kept the better endpoint. The shape estimate is $\hat{s}_g = \arg\max_s \ell_g(s)$ on the grid, and the likelihood interval is

$$\mathcal{I}_g = \left\{ s : 2\left[\ell_g(\hat{s}_g) - \ell_g(s)\right] < \chi^2_{1,\,0.95} = 3.84 \right\}, \tag{3.6}$$

the grid values within 1.92 log-likelihood units of the maximum. We call $\mathcal{I}_g$ entirely regular if all its elements lie below 1, entirely bursty if all lie above 1, and undecided if it contains 1. It is a descriptive likelihood interval on a grid with no coverage statement attached, for the reasons that exclude Wilks' theorem in Section 2.2.

The gene set was the 1,246 classified genes together with unclassified genes with converged fits drawn at random, to a total of 2,000. The extension grid was fitted for the refractory genes and a contrast set of 186 two-state genes. Complete main-grid profiles were obtained for 1,658 genes (184 refractory, 1,055 two-state, 419 unclassified). The rest failed to converge at one or more grid points and are marked in Table S6, which lists $\ell_g(s)$ for every profiled gene.

3.4 Shape estimates by class

Over the full grid all 184 of the 184 refractory intervals are entirely regular (Figure S2). The maximum is at the most regular shape on the grid, $\hat{s} = 1/20$, for 168 of them, with the likelihood still rising toward smaller s, and no refractory gene has $\hat{s} \geq 1/2$. The median log-likelihood gain at $\hat{s}$ is 27 units over the two-state model and 14.0 units over the two-step model (natural-log units, not doubled). Because most of the class has its maximum at this end of the grid, $\hat{s}$ for a refractory gene is an upper bound on its dispersion and $1/\hat{s}$ a lower bound on its number of steps, not an estimate: the likelihood increases toward silences at least as regular as 20 equal stages, and a nearly deterministic refractory delay is not excluded. For this reason, and for a second one given in Section 4.2, we describe the class only as "more regular than a memoryless promoter allows, with the fitted number of steps a lower bound".

In the two-state class the shape estimates lie mostly on the bursty side: of 1,055 genes, 1,047 have intervals entirely above 1, 4 entirely below 1 and 4 that contain 1, with a median gain over $s = 1$ of 7.5 units. The distribution of $\hat{s}$ over the bursty grid values is shown in Figure S2. Among the 419 unclassified genes the profiles are nearly flat (median gain 0.36) and 348 intervals contain 1, as expected for genes whose two-model comparison is itself undetermined.

3.5 Sensitivity of the clustered class to the mixture weight and shape

To test the sensitivity to the weight, we refitted the 1,033 two-state genes whose count grids were within the cap used for the controls (27 of the 1,060 exceed it) under C$(0.4, 2)$ and C$(0.75, 2)$ by the same procedure on the same grids. We then re-applied the rule of Eq. (2.7) at each weight with the equal-weight control statistics and G1 outcomes (an approximation, since both are weight-specific; a non-converged refit counts as lost). Of the 329 clustered genes, 156 remain at $w = 0.4$ and 260

at $w = 0.75$ (118 at all three weights). Of the 731 telegraph genes, 9 enter at $w = 0.4$ and 202 at $w = 0.75$. The likelihood is highest at 0.75 for 157 of the 329, at the equal weight for 134 and at 0.4 for 38.

To test the sensitivity to the shape, we refitted the same genes under C(0.75, 3) and C(0.75, 4), both inside the live-cell range, at the weight where $c = 4$ is admissible. Of the 329 clustered genes, 326 remain at $c = 3$ and 321 at $c = 4$. Over $c \in \{2, 3, 4\}$ at this weight the likelihood peaks at 4 for 169 of the 329, at 3 for 154 and at 2 for 6, with median gains over $c = 2$ of 12 and 15 log-likelihood units. Because a two-state gene refitted under a more dispersed mixture also gains likelihood, counting the telegraph genes that would enter at these shapes would require controls regenerated at each shape, which we did not compute. The refits at other weights and shapes show, first, that defining the clustered class by the equal-weight, $c = 2$ member is conservative in both the weight and the shape, because the likelihood is higher for a heavier fast component and for a more dispersed silence, under either of which the class would be larger. Second, a weight or shape per gene is not identifiable from one two-hour window at this depth, and fitting one would require a second window length or deeper sequencing per cell.

3.6 Cell-to-cell rate heterogeneity on the bursty side

If a varies across cells, the pooled silent periods are a mixture of exponentials, hyperexponential with $s > 1$ even when every cell is memoryless. A profile favoring $s > 1$ therefore does not by itself separate burst clustering within a cell from rate heterogeneity between cells, and $\hat{s}$ on the bursty side is an upper bound on the within-cell dispersion. Because mixing over cells can only add dispersion, this confounding acts in one direction: the likelihood interval of a gene whose promoter is memoryless in every cell cannot be shifted by rate heterogeneity entirely into the regular range $s < 1$, where the refractory class lies.

3.7 The twenty-gene grid over the number of steps and the two observation models

Separately from the genome-wide profile, we fitted twenty genes with small count ranges (twelve refractory and eight two-state, at evenly spaced ranks of Λ within each class) with chains of $k = 1, 2, 3, 5, 8$ equal steps and with non-integer values of k between them (a gamma-distributed silence of shape k). We extended the chains to $k = 16$ and 32 for eleven refractory genes, and profiled the two-state genes over the dispersion $s \in [1, 2.9]$ of the equal-weight mixture. Figure 1G is drawn from these fits, and Table S6 lists them under each of the two observation models we used. The refractory gains (twice the log-likelihood gain at $k = 5$ from 17.6 to 88.0, median 30.0, as in the Results; flat beyond $k = 8$ for nine of the eleven extended genes, beyond 16 for one, still rising at 32 for one) were computed under the full observation model of Section 1.5. The curves in Figure 1G and the two-state deficits quoted beside them (-0.04 to -17.2 at $k = 2$; -0.2 to -34.9 at $k = 5$, median -7.6) come from a dispersion-tolerant variant that integrates the likelihood over lognormal per-cell capture, as in the capture-jitter control. Under this variant the refractory gains at $k = 8$ range from 8.2 to 78.6, and the ordering of the models is the same for every gene under both observation models. The dispersion profiles of the two-state genes peak above the exponential value for seven of the eight ($\hat{s}$ from 1.5 to 1.8) and exclude $s = 1$ by Eq. (3.6) for five. A further twelve clustered genes, chosen by the same count-range rule at evenly spaced ranks of the clustered statistic, were fitted with the same five chains under the same variant. Their deficits at $k = 5$ (median 12.1, versus 7.4 for the seven telegraph genes of the panel) were compared with those of the telegraph genes by a one-sided Mann–Whitney test specified before analysis with $\alpha = 0.05$ ($P = 0.07$).

4 Detection power of a two-hour window at this depth

Every gene is classified by one statistic, Λ, measured in one experimental design: 8,916 cells, a window of $T = 2$ h, and full-length plate chemistry at the depth bins of Section 1.1. Throughout this section a simulated gene is a set of (o, n) pairs, one per real cell, drawn from a stated model with known parameters through the full observation model of Section 1.5 at that cell's own depth bin and refitted by the four-stage procedure exactly as real counts are. A simulated multi-step gene counts as detected when its Λ exceeds 8.6. We did not simulate the margins over the controls that the full rule of Eq. (2.7) also requires, and every detection rate below is therefore an upper bound for that rule.

4.1 Detection map at the real depth

We simulated three sets of genes with known kinetics. (i) Two-state genes were re-simulated as two-step genes at their own fitted (a, b, r, k_d), stratified by fitted mean silent period in octaves from under one hour to over sixteen hours (up to twenty per stratum, 108 in all; few two-state genes have sub-hour silences). (ii) A design grid of two-step parameter values was simulated, with mean silent period $1/a \in \{1, 2, 4, 8, 16, 32\}$ h by burst size $r/b \in \{2, 3.4, 6, 12\}$ by mRNA half-life $\ln 2/k_d \in \{3, 10\}$ h at an active period of $1/b = 1$ h (48 points), together with the 24 short-half-life points re-simulated from the two-state model as the reference for false positives. (iii) The refractory genes were re-simulated at their own fitted two-step rates (Section 4.2).

In set (i) no gene with a silent period under four hours was detected (0 of 48); 6 of 40 were detected between four and sixteen hours and 11 of 20 above sixteen hours. In set (ii) no point with burst size 2 or 3.4 was detected at any silent period or half-life (0 of 24; largest $\Lambda = 7.1$), whereas 11 of the 16 points with silent periods of four hours or more and burst sizes of six or more were. Among the 16 points with silent periods of one or two hours, 1 exceeded the threshold, with $\Lambda = 1{,}491$, whereas all seven of its grid neighbors were below the floor of Eq. (2.7). Its fit converged by the criterion of Section 2.1, and we have no explanation for this isolated value. Of the 24 points simulated from the two-state model, the number exceeding 8.6 was 0 (largest 8.2). The detection map is shown in Figure S1 and every simulated point is listed in Table S6. A promoter that recovers within a few hours, or whose bursts are of ordinary size, is invisible to the test however regular its silence, and about one gene in seven with a silent period between four and sixteen hours is detected. The long silences and large bursts of the detected class are therefore partly a selection effect of the design, and the statement that most classified genes are memoryless holds only at this resolution.

4.2 Re-simulation of the refractory genes and goodness of fit

Set (iii) was designed as a positive control, but only 3 of the 40 refractory genes re-simulated as two-step promoters at their own fitted rates and depths were re-detected, with a median refitted statistic of 2.0 (largest 15.1) compared with the median of 27.2 observed for the class. For three genes examined individually, re-simulation with more steps raised the statistic toward the observed value by different amounts, and refitting their real counts with q' halved or doubled left them far above the threshold, which excludes a misspecified conversion background as the source of the large statistics. The real counts of refractory genes thus have a feature that the two-step model accommodates better than the two-state model but that neither generates at the fitted rates.

To locate the feature that the fitted models miss, we examined the marginal count histograms of the 186 refractory genes and their expression-matched two-state partners. We compared observed and expected numbers of cells with a given count under each gene's fitted class model (two-step for

refractory, two-state for the partner) through the standardized residual

$$z_x = \frac{O_x - E_x}{\sqrt{E_x}}, \tag{4.1}$$

where O_x and E_x are the observed and expected numbers of cells with count x, evaluated at $x = 1$ and pooled over $x \geq 4$, and contrasted the two classes by paired Wilcoxon tests. An excess of cells with exactly one molecule is common to both classes (median z_1 for new counts 8.6 in refractory genes, 6.9 in their partners; $P = 0.051$). A light upper tail, in contrast, is specific to the refractory class: the median $z_{\geq 4}$ for new counts is -3.4 in refractory genes versus 1.0 in their partners ($P = 5 \times 10^{-28}$), and the refractory residual is the higher of the pair in only 10% of pairs. The old counts show the same difference (-2.9 versus $+0.5$), and within the refractory class Λ is correlated with the tail deficit (Spearman $\rho = -0.65$ with $z_{\geq 4}$).

We considered two technical explanations for the residual pattern of the refractory genes, and the data exclude both. Ambient single-molecule contamination predicts that a cell's odds of carrying exactly one molecule of a gene rise with the number of high-expressing cells of that gene on its plate. In the data, however, the per-gene odds ratios are close to one in both classes (medians 0.99 and 1.04 over the 14 pairs where estimable; $P = 0.39$). A stable expressing subpopulation predicts that molecules concentrate in few cells. The share of a gene's molecules carried by its top percentile of cells is indeed higher in refractory genes (median 0.116 versus 0.054; $P = 8 \times 10^{-28}$), as their long silences require, but within the class it is anticorrelated with Λ (Spearman $\rho = -0.25$), the opposite of what a stable expressing subpopulation would produce.

We interpret the low re-detection rate and the light upper tail as showing that the statistic is sensitive to silences more regular, and count distributions tighter, than two equal steps produce. Models with more steps fit the refractory genes better (168 of 184 profiles still rising at 20 steps, Section 3.4), although the data do not fix the number of steps. Sub-geometric burst sizes or a regular active period would also produce a light upper tail at fixed mean, and the window does not separate these from a regular silence. The class is accordingly defined by "more regular than a memoryless promoter allows", the fitted number of steps is a lower bound, and the two-step model is the alternative that detects the class, not a description that reproduces its counts.

4.3 Pulse length

Imaging-based metabolic labeling reaches more cells at lower depth per cell with pulses of an hour or less[21]. To determine the pulse length required to detect the class on such a platform, we simulated a single pulse of length $T_{\mathrm{p}} \in \{0.5, 1, 2, 4, 8, 20\}$ h with no chase over the 19,603 cells of that design, with kinetics taken from the fitted two-step rates (for power) or two-state rates (for false positives) of each of 15 expression-stratified genes. The simulated mean silent periods span 0.4 to 426 h. In these simulations the decision threshold was the conventional $|\Lambda| > \chi^2_{1,\,0.95} = 3.84$ rather than 8.6. With the higher threshold every fraction below would be lower still. For genes simulated from the two-step model the fraction classified as two-step was 0% with a two-hour pulse, 7% with four, 27% with eight and 73% with 20 hours, with 20% and 13% classified as two-state at the two longest pulses. For genes simulated from the two-state model none was falsely classified as two-step at four hours or less or at the longest pulse, and 33% were at eight hours. Power thus rose monotonically with pulse length, and only with a label of about a day were most genes classified correctly with no false positives. On the imaging platform's own counts (a one-hour pulse) bursting is favored over constitutive synthesis for 94% of 758 testable genes, but the two-state and multi-step models are not distinguishable, as the simulations predict for a one-hour pulse irrespective of the detection chemistry.

4.4 Depth

To determine the sequencing depth per cell required to detect the class, we thinned the real counts. Each captured molecule was kept with probability f, cells were subsampled to 4,000, and the thinned counts were refitted with every bin's efficiency set to $f\beta_h$, which is exact within the model because binomial thinnings compose,

$$Y \mid X \sim \mathrm{Bin}(X, \beta), \quad Z \mid Y \sim \mathrm{Bin}(Y, f) \implies Z \mid X \sim \mathrm{Bin}(X, f\beta). \tag{4.2}$$

Thinning to $f = 0.15$ leaves about 2,200 new molecules per median cell in place of 15,000, the depth of combinatorial-indexing labeling[22]; $f = 0.03$ leaves about 440, the depth of droplet assays[23]. At $f = 0.15$, 7 of 186 refractory genes kept $\Lambda > 8.6$. The median statistic fell from 27.2 to 0.68, a median per-gene ratio of 0.024 compared with the 0.067 expected if Λ scaled with the number of molecules (f times 4,000/8,916), and 1 of 1,024 two-state genes kept $\Lambda < -8.6$. At $f = 0.03$ no refractory gene survived (median $\Lambda = 0.015$) and none of 564 two-state genes kept its classification. The statistic thus decreases faster than linearly with the number of molecules. At the depth of droplet or combinatorial-indexing assays only a handful of the strongest refractory genes would be detected, and classifying the genes of another cell type requires full-length chemistry at a comparable depth in thousands of cells.

To our knowledge the only other NASC-seq2 dataset at this depth is of human K562 cells: 613 cells treated with 4sU, 13,600 genes passing the inclusion rule of Section 1.1, median 10,330 new molecules per cell[1]. Per-cell depth is comparable to the fibroblast data, but Λ is a sum over cells and 613 cells are too few to classify genes in either direction. For this reason we did not classify K562 genes (Limitations).

4.5 Capture noise and model-free observables

Model-free summaries of the real data (turnover, the probability of no new molecule given old molecules present, the log odds of persistence across the window) form one unimodal cloud, which might be read as evidence against discrete kinetic classes. When two discrete classes are simulated through the same cells and observation model (the refractory genes from their two-step fits and every other gene from its two-state fit, 3,246 genes in all), the summaries are as unimodal as the data (no dip test[24] rejects; smallest $P = 0.73$). When every gene is instead simulated from its two-state fit, so that no gene has memory, the summaries remain unimodal. Refractory genes also separate from expression-matched two-state genes equally well whether or not memory is simulated (area under the receiver operating characteristic curve for the persistence odds 0.92 with memory, 0.93 without, 0.87 in the data). Unimodality of these summaries is therefore uninformative about discreteness, and the position of the refractory genes in the cloud is set by their operating point rather than by memory. The difference between the models appears only in the full joint distribution of old and new counts that the likelihood uses.

5 Allele-resolved models

New molecules, but not old ones, are assigned to a parental chromosome of the F1 hybrid by strain-specific variants, and every allele-level distribution below is therefore a distribution of new counts, marginal over old RNA. With capture β_h and a new molecule read as old with probability q, allele-resolved new counts are thinned at the effective efficiency $e_c = \beta_{h(c)}(1 - q)$. The genome-wide fits pool the two alleles, and allele-resolved counts are used only in this section.

5.1 One active copy

For X-linked genes, which have one active copy, the count distribution is P_1 of Eq. (1.11) evaluated on the two-copy grid, with misassignment and capture applied as in Section 1.5 and the fitting procedure of Section 2.1 unchanged. Its parameters are related to those of a two-copy fit by the near-degeneracy of Eq. (1.16). The one-copy and two-copy fits of the X-linked refractory genes are compared with simulation in Section 5.5 (Figure 5H).

5.2 Between-allele correlation by bivariate mixing

Conditional on the promoter path, an allele's new count at unit capture is Poisson with mean equal to the capture-free window activity

$$\vartheta = \int_0^T r\,\mathbf{1}_{\mathrm{ON}}(t)\,e^{-k_d(T-t)}\,dt, \tag{5.1}$$

so the observed count in a cell of efficiency e is mixed Poisson, $P(n) = \int \mathrm{Poi}(n \mid e\vartheta)\,\phi(\vartheta)\,d\vartheta$, with ϕ fixed by the fitted kinetics; P is the per-copy new-count distribution of Section 1 at the gene's class model and fitted rates, marginalized over old RNA and thinned at e. Since $e\vartheta\,\mathrm{Poi}(n \mid e\vartheta) = (n+1)\mathrm{Poi}(n+1 \mid e\vartheta)$,

$$\int e\vartheta\,\mathrm{Poi}(n \mid e\vartheta)\,\phi(\vartheta)\,d\vartheta = (n+1)\,P(n+1) \tag{5.2}$$

for any mixing density. Now give the two alleles of one gene in one cell the joint mixing density

$$\phi(\vartheta_1,\vartheta_2) = \phi_1(\vartheta_1)\,\phi_2(\vartheta_2)\,\big[1+\omega(\vartheta_1-\mu_1)(\vartheta_2-\mu_2)\big], \tag{5.3}$$

with μ_i and ς_i^2 the mean and variance of ϑ_i, the simplest bivariate family with given marginals and a linear correlation term[25], for which $\mathrm{corr}(\vartheta_1,\vartheta_2) = \omega\varsigma_1\varsigma_2$. Conditional independence of the two counts given $(\vartheta_1,\vartheta_2)$ and Eq. (5.2) then give the joint count distribution in closed form,

$$P(n_1,n_2) = P_1(n_1)\,P_2(n_2)\,\big[1+\omega\,\gamma_1(n_1)\,\gamma_2(n_2)\big], \qquad \gamma_i(n) = \frac{(n+1)\,P_i(n+1)}{e\,P_i(n)} - \mu_i, \tag{5.4}$$

which needs only each allele's marginal distribution in the cell's depth bin; μ_i and $\varsigma_i^2 = \mathbb{E}[n(n-1)]/e^2 - \mu_i^2$ are its factorial moments at unit capture, and $\vartheta_{i,\max} = r(1-e^{-k_dT})/k_d$ bounds the activity. The density of Eq. (5.3) is nonnegative on $[0,\vartheta_{1,\max}]\times[0,\vartheta_{2,\max}]$ only for

$$-\big[\max\{(\vartheta_{1,\max}-\mu_1)(\vartheta_{2,\max}-\mu_2),\,\mu_1\mu_2\}\big]^{-1} \le \omega \le \big[\max\{\mu_1(\vartheta_{2,\max}-\mu_2),\,\mu_2(\vartheta_{1,\max}-\mu_1)\}\big]^{-1}, \tag{5.5}$$

which caps the correlation the family can express, and Table S4 marks the estimates that lie at one of these bounds.

The allele-resolved tables report new-RNA read units rather than reconstructed molecules (a median of 3.7 units per molecule), and each allele is therefore assigned a unit yield $\kappa_i = \sum_c n_{ic}/(\mu_i\sum_c e_c)$ estimated from its own mean. We computed three estimators of the correlation $\varrho = \omega\varsigma_1\varsigma_2$ per gene over the 8,912 cells with allele-resolved counts. The model-moment estimator

$$\hat{\varrho} = \frac{\sum_c (n_{1c}-e_c\kappa_1\mu_1)(n_{2c}-e_c\kappa_2\mu_2)}{\kappa_1\kappa_2\,\varsigma_1\varsigma_2\sum_c e_c^2} \tag{5.6}$$

is unbiased for ϱ under any unit-generation noise independent between alleles given $(\vartheta_1,\vartheta_2)$. Its empirical-moment variant replaces $\varsigma_1\varsigma_2$ by the observed excess variance of the depth-normalized units

over their Poisson expectation. The likelihood estimator maximizes $\sum_c \log[1+\omega\gamma_1(n_{1c})\gamma_2(n_{2c})]$ over the interval of Eq. (5.5), with γ_i from the distribution thinned at $e_c\kappa_i$, and its value is $\hat{\omega}\varsigma_1\varsigma_2$. The null distribution is obtained by permuting one allele's counts across cells within depth bins (200 permutations per gene). Three reference values set the scale of $\hat{\varrho}$: a cell-level floor, a negative control and a positive control. The cell-level floor is the same statistic between the reference allele of a refractory gene and the reference allele of its expression-matched two-state partner in the same cells, which share every cell-wide input and no promoter. The negative control is the set of X-linked genes, whose one active copy cannot correlate with a silent one (median $\hat{\varrho} = -0.07$ over 12 genes). The positive control is the set of annotated cell-cycle-periodic genes, both of whose copies follow one strong cell-level input (median 0.52, compared with 0.15 for all other classified genes; 12 genes, one-sided depth-adjusted $P = 4 \times 10^{-6}$).

On simulated pairs drawn from Eq. (5.4) the three estimators recover ϱ, with a wider spread at refractory than at two-state rates because at refractory rates most cells carry no new molecule on either allele (Figure S5). On the data the three estimators agree on the two facts stated in the Results. First, the two copies of one gene are correlated in either class (same-gene medians 0.14, 0.19 and 0.18 for refractory genes and 0.32, 0.18 and 0.13 for two-state genes by the moment, likelihood and empirical estimators; 86% and 96% of genes at permutation $P < 0.05$). Second, the same-gene correlation is an order of magnitude above the cross-gene floor (0.03 over 324 pairs by the moment estimator, 0.02 by the empirical one; same-gene minus cross-gene positive in 85% and 97% of genes). They disagree on the class contrast, however: over the 149 autosomal matched pairs the moment estimate is higher in the two-state member ($P = 4 \times 10^{-12}$), the empirical estimate higher in the refractory member ($P = 8 \times 10^{-4}$) and the likelihood estimate equal ($P = 0.41$), because the three normalize by different variances and the refractory per-copy distribution is concentrated at zero. We therefore use this construction only for the correlation level and the gene-versus-cell contrast, and not for a comparison between the classes (Table S4).

5.3 Chromosome-resolved co-transcription of neighboring genes

For each autosomal gene with at least 50 positive cells on each allele (12,557 genes) we formed new-RNA indicators I_g^R, I_g^A (allele count above zero) on the reference and alternative chromosome over all 8,916 cells. For a same-chromosome pair (i, j) the statistic is

$$X_{ij} = \tfrac{1}{2}[\rho(I_i^R, I_j^R) + \rho(I_i^A, I_j^A)] - \tfrac{1}{2}[\rho(I_i^R, I_j^A) + \rho(I_i^A, I_j^R)], \tag{5.7}$$

with ρ the Pearson correlation across cells. A cell-level covariate (depth, cycle phase, a shared upstream signal) enters both brackets identically and cancels; a state shared along one chromosome copy enters only the first. Allele assignment thins both equally, so X keeps its sign but not its scale and is reported as a fraction of the opposite-copy correlation. On 20,000 random different-chromosome pairs $X = -0.00007 \pm 0.00007$. Pairs were binned by promoter distance with overlapping gene bodies excluded. Intervals are pair-bootstrap percentiles or, in the largest bins, normal.

Computed over all cells, $X > 0$ at every separation (0.0014 even beyond 5 Mb), which is the signature of whole-chromosome allelic imbalance in a minority of cells rather than of local co-transcription. We therefore screened each cell chromosome by chromosome and marked a cell as imbalanced on a chromosome when it had at least 20 allele-assigned new reads there and a parental fraction above 80% (595 cells, 6.7%, marked at least once; 16 to 92 per autosome), and we masked that cell's counts on that chromosome. Masking lowered the distance-independent level from 0.0014 to 0.0004 and left the short-range component above it in place (0.0028 before, 0.0026 after, within 10 kb). The excesses quoted in the Results and Figure 5F are computed after

this masking, whereas the different-chromosome baseline and the contact-domain comparison are computed without it. After masking, $X = 0.0030$ within 10 kb (1,776 pairs), a same-copy excess of 8% of the opposite-copy correlation. Head-to-head pairs within 10 kb, which cannot share a read-through transcript, give the same value (0.0030; 1,399 pairs). Imprinted domains, which are transcribed from only one parental chromosome, provide a positive control: among all pairs within 200 kb (median 0.001) the largest X belong to the imprinted pairs (*Meg3–Rian* 0.23, *Dio3os–Dio3* 0.18, *Ndn–Peg12* 0.14). Sharing a contact domain, defined from fibroblast Hi-C insulation (3,060 boundaries at 25-kb resolution; Methods), has no additional effect at matched distance: at 50 to 200 kb the same-domain minus different-domain excess is 0.0001 (-0.0003 to 0.0004), compared with a distance effect of 0.0023 in the same bin.

5.4 Inter-allele overdispersion and the expression-matched contrast

A second, model-free statistic is based on how the new units in each cell are split between the two alleles. For each cell with $t = n_1 + n_2 \geq 1$ allele-resolved units of gene g, n_1 given t is taken as beta-binomial,

$$P(n_1 \mid t) = \binom{t}{n_1} \frac{\mathrm{B}(n_1 + \alpha_1,\, t - n_1 + \alpha_2)}{\mathrm{B}(\alpha_1, \alpha_2)}, \qquad \psi_g = \frac{1}{\alpha_1 + \alpha_2 + 1}, \tag{5.8}$$

with B the beta function and ψ_g the intra-class correlation, fitted per gene by maximum likelihood. Larger ψ_g means splits more extreme than binomial and hence alleles less often active in the same window. The number of read units per molecule inflates ψ_g (units from one molecule fall on one allele) and differs between the classes (medians 4.22 in refractory and 1.67 in matched two-state genes), which by itself would raise ψ_g in the refractory class, opposite to the observed difference. The absolute values are therefore not interpretable as correlation levels, and we compare the classes only through the expression-matched contrast: median $\psi = 0.59$ in refractory versus 0.73 in two-state genes over 156 matched pairs (Mann–Whitney $P = 3 \times 10^{-7}$, Cliff's $\delta = -0.34$), lower meaning more correlated, with the same sign in each expression tercile (Table S4).

5.5 Log-likelihood lost under a misspecified copy number

For the X-chromosome test of the Results we refitted 45 of the 46 confidently classified X-linked genes (one was excluded for grid size) under the one-copy model, with the same starts and convergence criteria as the genome-wide fits and with two-copy fits from the same code for comparison. A gene retained its refractory classification when its one-copy statistic exceeded 4.3. To understand why two of the ten refractory X-linked genes were not retained, we measured the log-likelihood lost by misspecifying the copy number on counts simulated with known parameters at those two genes' fitted rates. Genes simulated with two copies lose 123 to 187 log-likelihood units when fitted under the one-copy model, and genes simulated with one copy gain 11 to 43 under the one-copy model relative to the two-copy fit. The real counts of the two genes that were not retained fit 520 and 900 units worse under one copy than under two, far outside the simulated two-copy band, and therefore reject the one-copy assumption itself rather than the refractory classification (Figure 5H). All ten refractory X genes converged under both copy numbers. Conversely, two of the two-state X genes acquire a nominal multi-step statistic above 4.3 under one copy. Neither converged under one copy, and the one-copy model does not apply to one of them, which is expressed from both alleles in these cells (an escapee of X inactivation). Over all 45 genes the one-copy fits converged for 68% compared with 96% for the two-copy fits. As Eq. (1.16) predicts, the one-copy estimate of a is about twice the two-copy value for every refractory X gene, with b and r nearly unchanged (Table S4), which means

that two-copy fits underestimate the burst-on rates of X-linked genes while leaving the classification unchanged for eight of the ten refractory genes.

6 Noise, information and detection delay in closed form

In this section a promoter has one active state, left at rate b, in which transcripts are made at rate k_{syn} (the r of the inference sections), and each transcript is degraded at rate k_d. The silent period T_{off} is a positive random variable with mean $1/a$ and Laplace transform $\hat{f}(\lambda) = \langle e^{-\lambda T_{\mathrm{off}}} \rangle$. Three distributions are used: exponential with rate a (the two-state promoter, $k = 1$); Erlang with k stages each left at rate ka (the multi-step promoter; the fitted refractory model is $k = 2$); and the clustered mixture C$(1/2, 2)$, an equal mixture of exponentials with rates $(2 \pm \sqrt{2})a$. All share the duty fraction $\pi_{\mathrm{on}} = a/(a+b)$, cycle time $\tau_c = 1/a + 1/b$, stationary mean $k_{\mathrm{syn}}\pi_{\mathrm{on}}/k_d$ and burst size $B = k_{\mathrm{syn}}/b$. Logarithms are natural (nats) unless bits are stated.

6.1 Stationary Fano factor for an arbitrary silent-period distribution

Let σ_t be the promoter state with generator Q and stationary distribution π, n_t the count, and $y_j = \mathbb{E}[n\,\mathbf{1}\{\sigma = j\}]$. Births occur at rate k_{syn} in the active state only and deaths at rate $k_d n$, so the stationary moment equations close at every order. The first-order equations $(k_d \mathsf{I} - \mathsf{Q}^{\mathsf{T}})y = k_{\mathrm{syn}}\pi_{\mathrm{on}} e_{\mathrm{on}}$ sum to $\mathbb{E}[n] = k_{\mathrm{syn}}\pi_{\mathrm{on}}/k_d$, the second factorial-moment equations sum to $\mathbb{E}[n(n-1)] = k_{\mathrm{syn}} y_{\mathrm{on}}/k_d$, and therefore

$$
\begin{aligned}
F \equiv \frac{\operatorname{Var} n}{\mathbb{E}\, n} &= 1 + \frac{y_{\mathrm{on}}}{\pi_{\mathrm{on}}} - \mathbb{E}[n], \qquad y_{\mathrm{on}} = k_{\mathrm{syn}}\pi_{\mathrm{on}}\mathcal{R}(k_d), \\
\mathcal{R}(\lambda) &= \left[(\lambda \mathsf{I} - \mathsf{Q})^{-1}\right]_{\mathrm{on,on}} = \int_0^\infty e^{-\lambda t} \Pr(\sigma_t = \mathrm{ON} \mid \sigma_0 = \mathrm{ON})\, dt.
\end{aligned}
\tag{6.1}
$$

The process renews at each entry into the active state: from there the promoter stays an exponential time of rate b, makes one silent sojourn, and is active again. Hence $\mathcal{R} = 1/(b+\lambda) + [b/(b+\lambda)]\hat{f}(\lambda)\mathcal{R}$, that is $\mathcal{R}(k_d) = 1/(k_d + b\varphi)$ with $\varphi = 1 - \hat{f}(k_d)$, and

$$
F = 1 + \frac{k_{\mathrm{syn}}}{k_d + b\varphi} - \frac{k_{\mathrm{syn}} a}{k_d(a+b)} = 1 + \frac{k_{\mathrm{syn}}\, b\, (k_d - a\varphi)}{k_d\, (a+b)\, (k_d + b\varphi)}, \qquad \varphi = 1 - \langle e^{-k_d T_{\mathrm{off}}} \rangle. \tag{6.2}
$$

The variance decomposition of the doubly stochastic Poisson count, $\operatorname{Var} n = \mathbb{E}\, n + \operatorname{Var} \mu$ with $\mu_t = k_{\mathrm{syn}} \int_{-\infty}^{t} e^{-k_d(t-u)} \mathbf{1}_{\mathrm{ON}}(u)\, du$, closed by the same renewal, gives the same result. By Jensen's inequality $\hat{f}(k_d) \geq e^{-k_d/a} > 1 - k_d/a$, hence $F > 1$ for every silent-period distribution. The Fano factor depends on the silence only through the convex functional $\langle e^{-k_d T_{\mathrm{off}}} \rangle$, and since $\partial F/\partial \varphi < 0$ a more regular silence of the same mean lowers F and a more dispersed one raises it. The transforms are $\hat{f}_1 = a/(a+k_d)$, which reduces Eq. (6.2) to the two-state result $F_1 = 1 + k_{\mathrm{syn}} b/[(a+b)(a+b+k_d)]^2$; $\hat{f}_k = [ka/(ka+k_d)]^k$; and $\hat{f}_{\mathrm{C}} = 2a(a+k_d)/(2a^2 + 4ak_d + k_d^2)$ for the clustered mixture, for which $F_{\mathrm{C}} - F_1 = k_{\mathrm{syn}} ab/[(a+b+k_d)(2a^2 + 2ab + 4ak_d + bk_d + k_d^2)] > 0$. Since $(1 + x/k)^k$ increases strictly in k (put $x = k_d/a$), $\hat{f}_k$ decreases and φ_k increases in k, so

$$
F_1 > F_2 > \cdots > F_\infty > 1
$$

at any positive rates, where F_∞ is Eq. (6.2) with $\varphi = 1 - e^{-k_d/a}$, the value for a deterministic silence. In the operating-point variables, with $a + b = 1/[\tau_c \pi_{\mathrm{on}}(1 - \pi_{\mathrm{on}})]$,

$$
F = 1 + B(1 - \pi_{\mathrm{on}})^2\, \frac{(a+b)(k_d - a\varphi)}{k_d(k_d + b\varphi)}, \qquad F_1 = 1 + \frac{B(1 - \pi_{\mathrm{on}})^2}{1 + k_d \tau_c \pi_{\mathrm{on}}(1 - \pi_{\mathrm{on}})}, \tag{6.3}
$$

linear in burst size at fixed duty fraction, cycle time and k_d. Equation (6.2) agrees with the exact stationary moments from the numerically solved moment equations of the joint promoter–RNA chain at the 1,246 fitted rate sets, for $k = 1, 2, 3, 5, 7$ and the clustered mixture, to a largest relative difference of 8×10^{-12}. The median Fano factor over the fitted sets falls from 2.51 at $k = 1$ to 2.28 at $k = 2$ and 2.09 at $k = 7$, with limit 2.01 (Figure 6D): a multi-step silence at the same rates lowers stationary noise slightly and never raises it.

6.2 Information per interval and detection delay

Consider a step $a \to a'$ in the silent-exit rate at an unknown time, b and k_{syn} unchanged. When bursts are resolved ($k_{\text{syn}} \gg k_d \gg a, b$) the alternating silent and active periods are visible in the count path, and each promoter cycle is then one observation. Cycles are independent and identically distributed before and after the change, and only T_{off} is informative about a. For $T_{\text{off}} \sim \text{Gamma}(k, ka)$ the log-likelihood ratio of an interval of length t is $k \ln(a'/a) - k(a' - a)t$; its mean under the post-change distribution, where $\mathbb{E}\, T_{\text{off}} = 1/a'$, is the Kullback–Leibler information per interval,

$$D_k = k\Big[\ln \frac{a'}{a} + \frac{a}{a'} - 1\Big] = k\, D_1. \tag{6.4}$$

Equation (6.4) holds because an Erlang-k interval is the sum of k independent exponential stages whose rates the step multiplies by a common factor, and information adds over independent stages. A k-step silence therefore carries k times the information of an exponential one of the same mean about a change in its own rate[26]. By Lorden's theorem and its refinements[27–30], among stopping rules whose mean number of pre-change intervals to a false alarm is at least γ_c the least worst-case mean delay is $(\ln \gamma_c)/D_k$ intervals to first order as $\gamma_c \to \infty$, attained by the cumulative-sum rule[31] and to the same order by the Shiryaev–Roberts rule. The alarm is a stopping time on the cycle filtration, so by Wald's identity the delay in time is the delay in cycles times the post-change mean cycle:

$$\mathbb{E}[\text{delay}] \simeq \frac{\ln \gamma}{I_k}, \qquad I_k = \frac{D_k}{1/a' + 1/b} = \frac{k\,[\ln(a'/a) + a/a' - 1]}{1/a' + 1/b}, \tag{6.5}$$

with $\gamma = \gamma_c \tau_c$ the mean time to false alarm (the conversion shifts $\ln \gamma$ by a constant). A Monte Carlo estimate of the log-likelihood-ratio growth rate over simulated post-change cycles agrees with I_k within one standard error, which confirms that the post-change mean cycle is the correct denominator. For a step in b or in k_{syn} the corresponding rates involve only the active period or the burst-size distribution and do not depend on k, so that a regular silence shortens detection only for changes in the burst-on rate. For the clustered mixture D_{C} depends on a'/a alone (the family is a scale family), has no elementary closed form, is bounded above by D_1 by convexity of the divergence, and by quadrature is $0.59\, D_1$ at $a'/a = 1.5$. When $a, b \gg k_d$ the births tend to a Poisson process of rate $k_{\text{syn}} \pi_{\text{on}}$ under every distribution and the advantage vanishes.

At practical false-alarm budgets the first-order delay $(\ln \gamma_c)/D_k$ overstates the advantage of a regular silence, because the threshold that yields γ_c lies well below $\ln \gamma_c$ when $D_k \ll 1$. With per-cycle Shiryaev–Roberts detectors on independent silent periods ($a'/a = 1.5$, threshold set by bisection to a pre-change mean run length of $\gamma_c = 100$ cycles, change at cycle zero, 20,000 paths) the simulated delay is the fraction η_k of $(\ln \gamma_c)/D_k$ with

$$\eta_1 = 0.39, \quad \eta_2 = 0.51, \quad \eta_3 = 0.58, \quad \eta_5 = 0.67, \quad \eta_7 = 0.73, \quad \eta_{\text{C}} = 0.31.$$

The delay of a k-step promoter relative to the two-state promoter of the same mean at this budget

is therefore not $1/k$ but

$$\frac{\text{delay}_k}{\text{delay}_1} = \frac{\eta_k}{\eta_1} \cdot \frac{1}{k} : \qquad \frac{0.51}{0.39} \cdot \frac{1}{2} = 0.65 \ (k = 2), \qquad \frac{0.58}{0.39} \cdot \frac{1}{3} = 0.50 \ (k = 3), \tag{6.6}$$

and for the clustered mixture $(\eta_C/\eta_1)/0.59$ in place of $1/0.59$. The per-gene delays in hours of Section 6.4 and Table M3 include these factors.

Detector and calibration. The Shiryaev–Roberts statistic is $R_n = (1 + R_{n-1})L_n$, $R_0 = 0$, with L_n the likelihood ratio of the nth observation under the post-change relative to the pre-change distribution. It is Shiryaev's Bayesian detector in the limit of a vanishing change prior[32,33]. An alarm is raised when $R_n > e^u$, with u set by bisection on simulated pre-change sequences to the required mean time to false alarm. The change occurs at a geometric time (mean 10 intervals). The delay is averaged over the realizations in which the alarm followed the change, and their fraction is reported beside every delay (Figures 6A–6C). The operating points of Figure 6 are the measured interburst times of *TFF1* at half-maximal and saturating estradiol[17] (185 and 66 or 86 min, steps of 2.8- and 2.2-fold) and a 24-fold step at the same scale, at mean times to false alarm of 100 and 1,000 pre-change intervals: 44 settings of number of steps, budget and operating point at 100,000 realizations each (Table S5).

When bursts are not resolved the same detector is applied to mRNA counts with $L_n = e^{\zeta_n}$, where ζ_n is the difference of the one-step predictive log-probabilities of two exact forward filters over the hidden promoter. The per-frame information rate $\mathbb{E}_1[\zeta_n]$ then reproduces the factor k of Eq. (6.4) at slow switching. The ratios of the delays are smaller than this factor, however, because $\mathbb{E}_0[e^{\zeta_n} \mid y_{<n}] = 1$ makes $\mathbb{E}_0 R_n$ grow linearly ($\mathbb{E}_0 R_n = n$) whether or not information accrues, and at budgets of a few pre-change cycles R_n crosses the threshold through this linear growth alone in a minority of settings (Figure S6).

6.3 The matched-mean comparison across architectures

For the comparison across architectures (Figure S6A; Table S5) we applied the interval-level detector to exponential, Erlang-2, Erlang-5, Erlang-7 and clustered interval distributions with identical pre- and post-change means. The clustered distributions are two-component mixtures with slow-component weight 0.05, 0.15 or 0.30 and slow-to-fast mean ratio 10 or 180 so as to include the live-cell operating point. A 2.8-fold and a 24-fold rate step were applied under two conventions, to every component's mean (whole-mean) or to the fast component alone (single-component; the mean interval then changes 1.01- to 2.7-fold), at budgets of 100 and 1,000 pre-change mean intervals: $3 \times 2 \times 2 \times 2 \times 2 = 48$ settings at 40,000 realizations each. Detection is faster with Erlang-5 intervals than with exponential intervals of the same mean at all 48 settings. The exponential delay is 2.0 to 3.8 times longer (median 3.1) under the whole-mean convention and 1.02 to 3.7 times (median 2.4) under the single-component convention. For Erlang-2 the factors are 1.6 to 1.8 and 1.01 to 1.75. For the clustered distributions the whole-mean step is a common rescaling of both components, and joint convexity of the divergence then bounds their information per interval by D_1, as for $\mathrm{C}(1/2, 2)$ in Section 6.2. Their simulated delays are listed with the others in Table S5 and drawn in Figure S6.

6.4 Stationary noise and detection delay across the genome

At fixed rates, adding steps to the silent period lowers F and raises I_k, so that for a single gene there is no trade-off between stationary noise and detection speed. Across genes both quantities

Table M3: **Class medians of stationary noise and detection delay at the fitted rates.** Fano factor from Eq. (6.2); delay for a 1.5-fold step in burst-on rate at a false-alarm budget of 100 cycles, in own cycles to first order (Eq. (6.7)) and in hours with the finite-budget factor of Eq. (6.6). Telegraph and clustered genes are evaluated under their own silent-period distributions and refractory genes at $k = 2$; the refractory delays are upper bounds, and their advantage over the other classes a lower bound, because $k = 2$ is a lower bound on the number of steps.

	telegraph (731)	clustered (329)	refractory (186)
Fano factor F	2.1	3.0	6.3
cycle time τ_c (h)	5.5	4.5	103
delay, own cycles (first order)	50	79	22
delay, hours (finite budget)	110	115	1,100

vary with the operating point $(a, b, k_{\text{syn}}, k_d)$. From Eq. (6.5),

$$\text{delay} = \ln \gamma_c \, \frac{1/a' + 1/b}{D_k(a'/a)}, \qquad \frac{\text{delay}}{\tau_c} = \ln \gamma_c \, \frac{(1/a' + 1/b)/(1/a + 1/b)}{D_k(a'/a)}, \tag{6.7}$$

the first proportional to cycle time and independent of burst size, the second constant within an architecture up to a duty-fraction factor in $(a/a', 1)$, while Eq. (6.3) is linear in burst size. We evaluated both at the fitted rates of the 1,246 classified genes (telegraph and clustered genes under their own distributions, refractory genes at $k = 2$) for $a'/a = 1.5$ and $\gamma_c = 100$, the delay in hours both to first order and with the factor η of Eq. (6.6) (class medians in Table M3). The Spearman correlation of F with the delay in hours is 0.74 (0.69 with the finite-budget factors) and with the delay in cycle units -0.10. Detection is slower in hours for noisier genes because burst size and cycle time rise together across the genome, whereas in units of each gene's own cycle detection is fastest for the refractory class and slowest for the clustered class.

6.5 Information rate about a fluctuating environment

To quantify the information that a promoter's transcript count carries about a fluctuating input, we added a two-state environment ϵ_t, switching at rate $1/\tau$ in each direction, as a hidden coordinate of one chain (environment, promoter, count) propagated over each sampling interval (frame) of length Δ by the kernel $\exp(\mathsf{Q}\Delta)$. In the high state one parameter (synthesis rate, burst-on rate or active-exit rate) differs so that the stationary mean rises by a factor χ from m. We compared two-state promoters and multi-step promoters with three sequential silent steps at matched duty fractions, with a constitutive promoter of the same means as reference, over 36 combinations of architecture, duty fraction, m, χ and τ, and descriptively at the class-median duty fractions. The information rate is

$$I_{\text{env}} = \frac{1}{\Delta \ln 2} \, \mathbb{E}\big[\ln p(y_t \mid y_{<t}, \epsilon_{\le t}) - \ln p(y_t \mid y_{<t}) \big] \tag{6.8}$$

bits per mRNA lifetime, both terms from exact forward filters (the first conditioned on the realized environment transitions), averaged over realizations of a readout population of 150 cells. Because the environment enters at frame boundaries, Eq. (6.8) is a lower bound on the information rate about the full environment path, and the values below are lower bounds in the same sense. Modulation of the synthesis rate (the burst-size channel) is the most informative at 31 of the 36 combinations and modulation of the burst-on rate (the burst-frequency channel) at the other five. On the burst-frequency channel the information rate of the multi-step promoter exceeds that of the two-state promoter by more than two standard errors at 16 of 18 matched points, by a median factor of 1.6,

and its Fano factor in the low state is the lower of the two at all 18 points. At one representative point the rates are 0.0085 and 0.0033 bits per lifetime and the Fano factors 5.29 and 5.50 for the multi-step and two-state promoters, respectively. No switching promoter exceeds the constitutive one in information rate by more than two standard errors in any of 120 comparisons, and at the class-median duty fractions the refractory operating point is both the less informative and the noisier (Table S5). The two-standard-error criterion was chosen after the calculation and is labeled post hoc wherever it appears.

6.6 Burst-train information rate

The information a burst train carries about its own promoter is the mutual-information rate i between the promoter path and the point process of transcription events, whose intensity is $\lambda_t = k_{\mathrm{syn}}\mathbf{1}\{\sigma_t = \mathrm{ON}\}$. For a point-process channel $i = \mathbb{E}[\varphi(\lambda_t)] - \mathbb{E}[\varphi(\hat{\lambda}_t)]$ with $\varphi(v) = v \ln v$ and $\hat{\lambda}_t$ the causal estimate of the intensity from past events[34,35]. The first term is $r_e \ln k_{\mathrm{syn}}$ with $r_e = k_{\mathrm{syn}}\pi_{\mathrm{on}}$ the event rate. Since only one state emits, every event resets the filter to the active state, and between events the unnormalized filter obeys the linear flow

$$\dot{\mathbf{q}} = (\mathsf{Q}^{\mathsf{T}} - \mathsf{R})\,\mathbf{q}, \qquad \mathbf{q}(0) = e_{\mathrm{on}}, \qquad \mathsf{R} = \mathrm{diag}(k_{\mathrm{syn}}\mathbf{1}_{\mathrm{ON}}), \tag{6.9}$$

so $\mathcal{S}(t) = \mathbf{1}^{\mathsf{T}}\mathbf{q}(t)$ is the survival function of the inter-event interval, $f = -\mathcal{S}'$ its density, $\hat{\lambda} = f/\mathcal{S}$ at the current age, and events form a renewal process with stationary age density $r_e\mathcal{S}$. Using $\int_0^\infty f \ln \mathcal{S}\, dt = -1$,

$$i = r_e \ln k_{\mathrm{syn}} - r_e \int_0^\infty f \ln \frac{f}{\mathcal{S}}\, dt = r_e\,[h(f) - (1 - \ln k_{\mathrm{syn}})], \qquad h(f) = -\int_0^\infty f \ln f\, dt, \tag{6.10}$$

the event rate times the excess of the interval's differential entropy over that of an exponential interval of rate k_{syn}. We compute $h(f)$ by quadrature on the eigen-expansion of Eq. (6.9) (for the two-state promoter $h(f)$ has a closed form in Gauss hypergeometric functions), and a Monte Carlo estimate from the causal filter agrees with it. As a worked example, a promoter with two unequal silent steps left at rates 0.9 and 0.6, an active state left at rate 1.4 and $k_{\mathrm{syn}} = 2.3$ (arbitrary time units) has $i = 0.5536$ nats per unit time, compared with 0.5409 for the two-state promoter with the same mean active and silent periods, a ratio of 1.02. At matched dwell times a regular silence changes the information rate of the burst train by only a few percent. Because this rate does not distinguish the classes, we compare them in Figure 6 by detection delay instead.

References

[1] Daniel Ramsköld, Gert-Jan Hendriks, Anton J. M. Larsson, Juliane V. Mayr, Christoph Ziegenhain, Michael Hagemann-Jensen, Leonard Hartmanis, and Rickard Sandberg. Single-cell new RNA sequencing reveals principles of transcription at the resolution of individual bursts. *Nature Cell Biology*, 26(10):1725–1733, 2024. doi: 10.1038/s41556-024-01486-9.

[2] J. Peccoud and B. Ycart. Markovian modeling of gene-product synthesis. *Theoretical Population Biology*, 48:222–234, 1995.

[3] Zhixing Cao and Ramon Grima. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proceedings of the National Academy of Sciences of the United States of America*, 117:4682–4692, 2020. doi: 10.1073/pnas.1910888117.

[4] Matthew D. Schwartz and Martin Hemberg. Controlled inference of transcriptional bursting from single-cell RNA counts. Manuscript, bootloops.ai, 2026.

[5] A. R. Stinchcombe, C. S. Peskin, and D. Tranchina. Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression. *Physical Review E*, 85(6):061919, 2012. doi: 10.1103/PhysRevE.85.061919.

[6] Andrew G. Nicoll, Juraj Szavits-Nossan, Martin R. Evans, and Ramon Grima. Transient power-law behaviour following induction distinguishes between competing models of stochastic gene expression. *Nature Communications*, 16:2833, 2025. doi: 10.1038/s41467-025-58127-4.

[7] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124:044104, 2006. doi: 10.1063/1.2145882.

[8] Arne Jensen. Markoff chains as an aid in the study of Markoff processes. *Skandinavisk Aktuarietidskrift*, 36:87–91, 1953. doi: 10.1080/03461238.1953.10419459.

[9] Winfried K. Grassmann. Transient solutions in Markovian queueing systems. *Computers & Operations Research*, 4:47–53, 1977. doi: 10.1016/0305-0548(77)90007-7.

[10] Andrew Reibman and Kishor Trivedi. Numerical transient analysis of Markov models. *Computers & Operations Research*, 15:19–36, 1988. doi: 10.1016/0305-0548(88)90026-3.

[11] Awad H. Al-Mohy and Nicholas J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM Journal on Scientific Computing*, 33: 488–511, 2011. doi: 10.1137/100788860.

[12] Diederik P. Kingma and Jimmy Ba. Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations (ICLR 2015), 2015. arXiv:1412.6980.

[13] J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38:481–482, 1951. doi: 10.2307/2332598.

[14] H. E. Daniels. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25:631–650, 1954. doi: 10.1214/aoms/1177728652.

[15] Robert Lugannani and Stephen Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12:475–490, 1980. doi: 10.2307/1426607.

[16] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57: 289–300, 1995.

[17] Joseph Rodriguez, Gang Ren, Christopher R. Day, Keji Zhao, Carson C. Chow, and Daniel R. Larson. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell*, 176:213–226.e18, 2019.

[18] J. Tünnermann, G. Roth, J. Cramard, and L. Giorgetti. Enhancer control of promoter activity and variability via frequency modulation of clustered transcriptional bursts. *Nature Genetics*, 58(8):1967–1975, 2026. doi: 10.1038/s41588-026-02676-x.

[19] Juan M. Pedraza and Johan Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861):339–343, 2008. doi: 10.1126/science.1144331.

[20] Benjamin Zoller, Damien Nicolas, Nacho Molina, and Felix Naef. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology*, 11: 823, 2015.

[21] Jingyi Ren, Haowen Zhou, Hu Zeng, Connie Kangni Wang, et al. Spatiotemporally resolved transcriptomics reveals the subcellular RNA kinetic landscape. *Nature Methods*, 20:695–705, 2023.

[22] Junyue Cao, Wei Zhou, Frank Steemers, Cole Trapnell, and Jay Shendure. Sci-fate characterizes

the dynamics of gene expression in single cells. *Nature Biotechnology*, 38(8):980–988, 2020. doi: 10.1038/s41587-020-0480-9.

[23] Qi Qiu, Peng Hu, Xiaojie Qiu, Kiya W. Govek, Pablo G. Cámara, and Hao Wu. Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nature Methods*, 17 (10):991–1001, 2020. doi: 10.1038/s41592-020-0935-4.

[24] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The Annals of Statistics*, 13: 70–84, 1985. doi: 10.1214/aos/1176346577.

[25] Mei-Ling Ting Lee. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics – Theory and Methods*, 25:1207–1222, 1996. doi: 10.1080/03610929608831759.

[26] S. Koyama and L. Kostal. The effect of interspike interval statistics on the information gain under the rate coding hypothesis. *Mathematical Biosciences and Engineering*, 11(1):63–80, 2014. doi: 10.3934/mbe.2014.11.63.

[27] G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6):1897–1908, 1971. doi: 10.1214/aoms/1177693055.

[28] Moshe Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, 13: 206–227, 1985. doi: 10.1214/aos/1176346587.

[29] George V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14:1379–1387, 1986. doi: 10.1214/aos/1176350164.

[30] Tze Leung Lai. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, 44:2917–2929, 1998. doi: 10.1109/18.737522.

[31] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954. doi: 10.2307/2333009.

[32] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 8:22–46, 1963.

[33] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8:411–430, 1966.

[34] M. Sinzger, M. Gehri, and H. Koeppl. Poisson channel with binary Markov input and average sojourn time constraint. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2873–2878, 2020. doi: 10.1109/ISIT44484.2020.9174360.

[35] M. Gehri, N. Engelmann, and H. Koeppl. Mutual information of a class of Poisson-type channels using Markov renewal theory. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 1931–1936, 2024. doi: 10.1109/ISIT57864.2024.10619620.