



Supplementary Information for

Multimodal transcriptional control of pattern formation in embryonic development

Nicholas C Lammers, Vahe Galstyan, Armando Reimer, Sean A Medin, Chris H Wiggins, and Hernan G Garcia

Hernan G Garcia

E-mail: hggarcia@berkeley.edu

This PDF file includes:

Supplementary text
Figs. S1 to S35
Tables S1 to S2
Captions for Movies S1 to S7
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Movies S1 to S7

Supplementary Figures

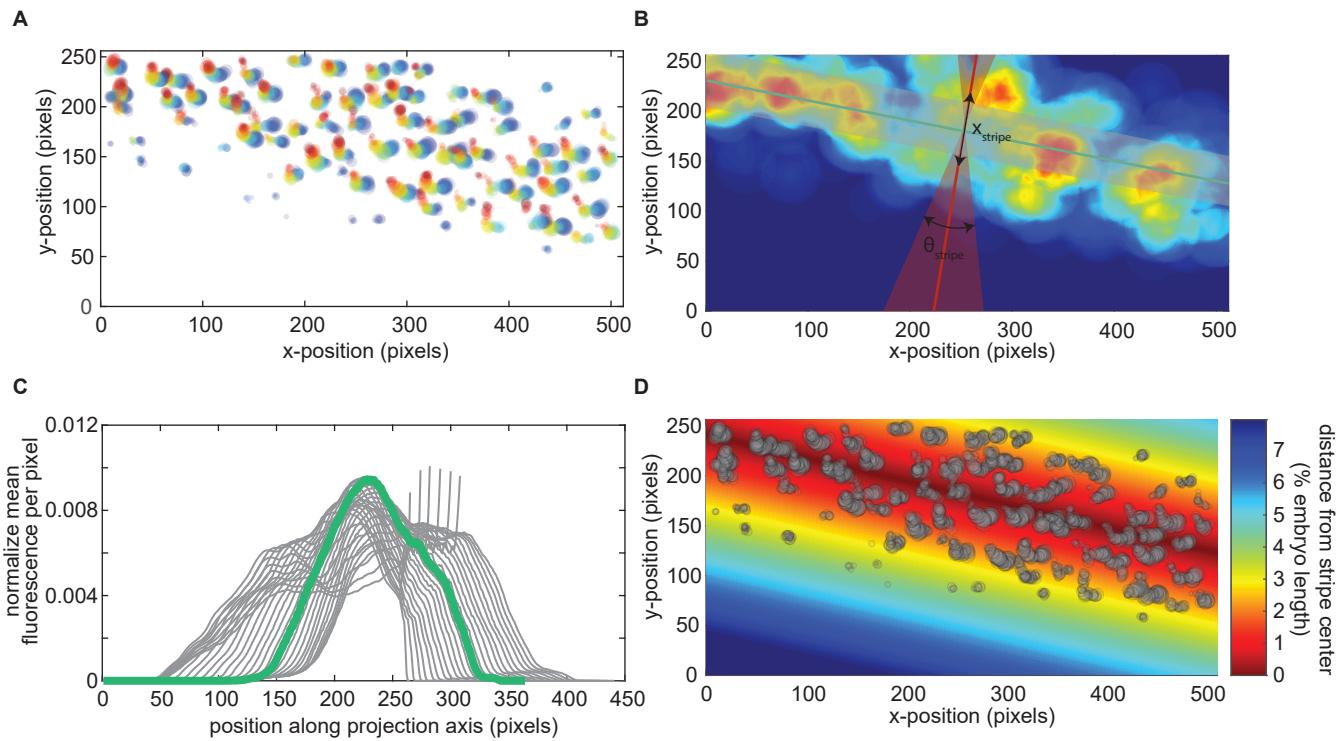


Fig. S1. Aligning stripes from multiple embryos. In order to minimize alignment errors when combining data from across multiple *Drosophila* embryos, an automated routine was employed to define a new experimental axis for each data set based upon the spatial distribution of transcriptional activity in the mature eve stripe 2 pattern. **(A)** Example of the spatiotemporal distribution of observed fluorescence for an individual embryo. Each circle corresponds to the fluorescence from a single locus at a single point in time. Only observations after 30 min into nuclear cycle 14 were used. Circle size indicates fluorescence intensity. Color indicates temporal ordering: 30 min (blue) to 47 min (red). **(B)** A Gaussian filter was convolved with the raw data points in (A). This filtering ameliorated stripe fitting artifacts that arose due to the relative sparsity of the raw data. The fitting procedure considered both a range of possible stripe orientations (θ_{stripe}) and, within each orientation, a range of possible positions of the stripe along the anterior-posterior axis (x_{stripe}) that, together constituted a set of possibilities for the new stripe center position and orientation. Here, the shaded red region indicates the range of values for θ_{stripe} that were considered. The red line indicates the best stripe axis inferred by the algorithm and the green line indicates the corresponding optimal stripe center. No constraints were placed on x_{stripe} , save for the limits of the experimental field of view. **(C)** For each proposed stripe orientation (θ_{stripe}), a projected stripe profile was generated by taking the average pixel intensity for each position, x_i , along the proposed stripe axis. To determine the optimal center location for each orientation, a sliding window with a width equal to 4% of the embryo length was used to determine the fraction of the total profile fluorescence that fell within 2% embryo length of the stripe center. For example, the gray shaded region in (B) illustrates what this range would be for the green stripe center line (B). This fraction of the total profile was used as a baseline for the comparison of potential stripe center positions. The θ_{stripe} and x_{stripe} that maximized this metric (green profile in (C)) were taken to define a new, empirically determined stripe center. **(D)** This inferred stripe position defined an experimental axis for each embryo that was used to aggregate observations from across embryos. Gray circles indicate experimental observations (size corresponds to intensity as in (A)) and shading indicates distance from inferred stripe center.

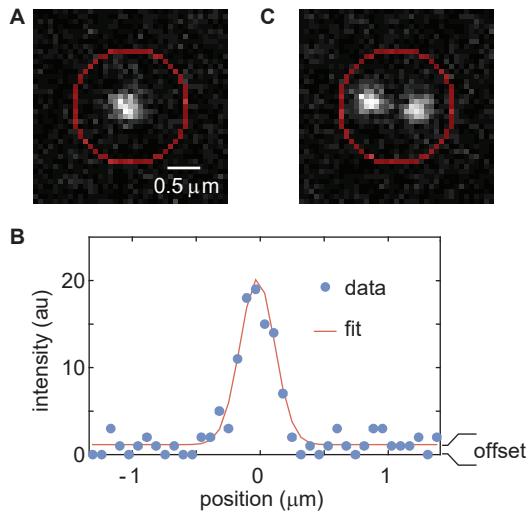


Fig. S2. Integrating MS2 Spots. (A) Sites of nascent transcription are identified in 3D using the Weka segmentation plugin for FIJI. Once identified, as described in (1), the Z-plane corresponding to the maximum fluorescence intensity is determined. On this Z-plane, fluorescence of a site is measured by integrating raw pixel intensities in a circular region around the fluorescent MS2 spot of a predefined area (indicated by the red circle) and then subtracting off the background intensity obtained by fitting a 2D Gaussian profile as outlined in (B). (B) X-Z projection of 2D Gaussian function fitted to MS2 spot shown in (A). Background intensity is estimated using the offset value for this Gaussian fit. The per-pixel offset is then multiplied by the area of the integration region. This background value is then subtracted from the fluorescence integrated across the area shown in (A). (C) The radius was chosen to be large enough to integrate the intensities from both sister chromatids, even when they are spatially separated and distinguishable .

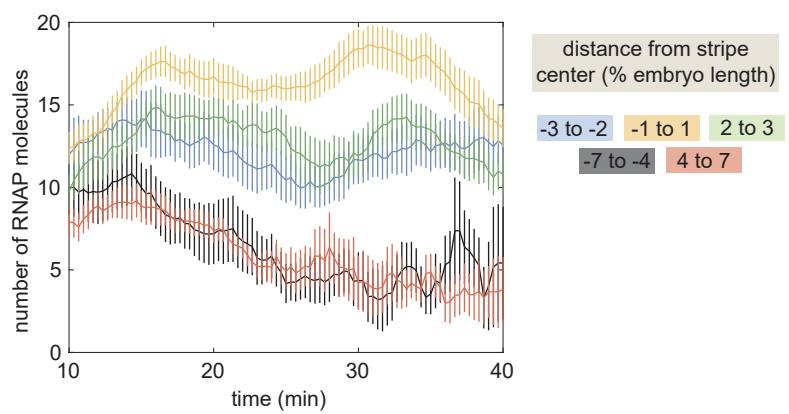


Fig. S3. Mean transcriptional activity. Mean transcriptional activity as a function of time for different positions along the stripe. (Average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

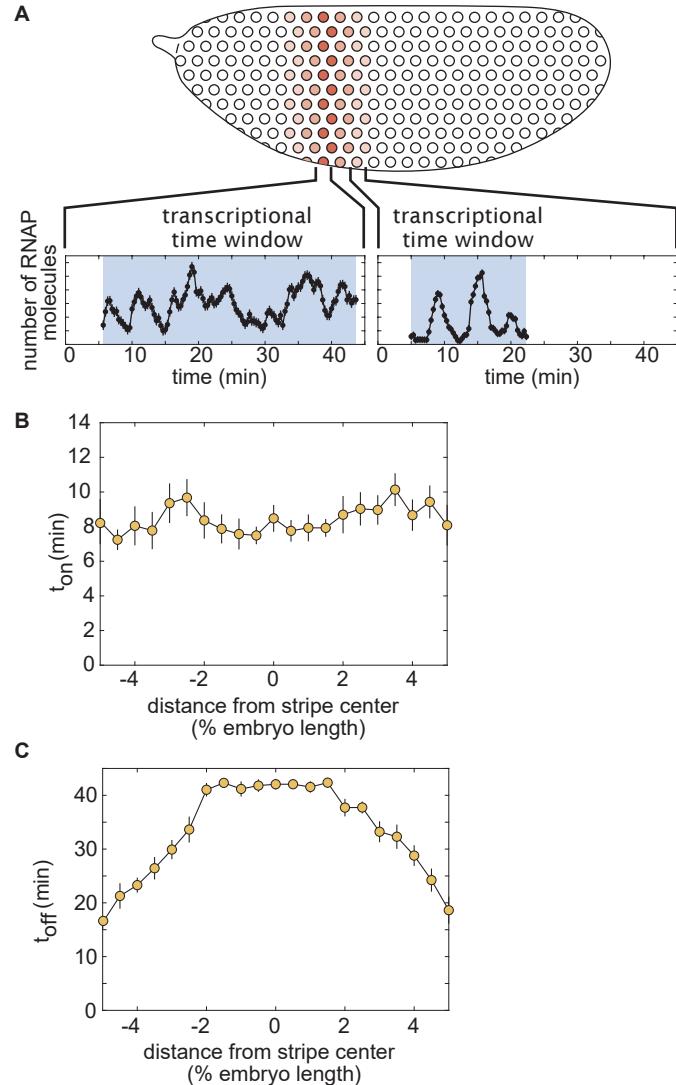


Fig. S4. Regulation of the transcriptional time window. (A) Single-nucleus measurements reveal that the duration of transcription is modulated along the stripe and that nuclei transcribe in a burst-like fashion. (B) Time for nuclei to activate transcription after mitosis, t_{on} , as a function of position along the stripe. (C) Time for nuclei to enter the quiescent transcriptional state, t_{off} . (B,C, average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

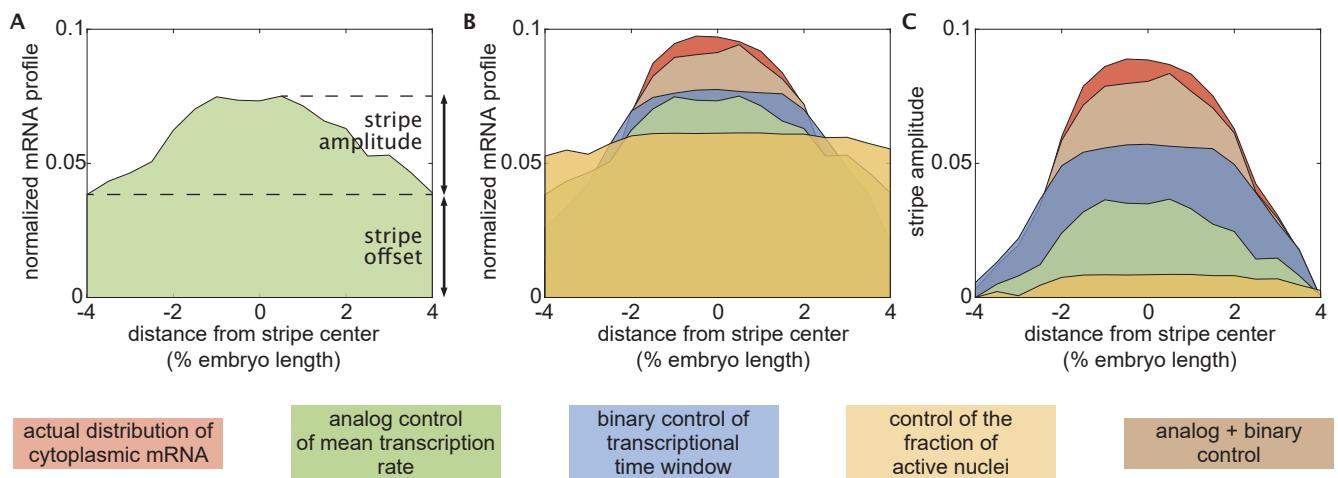


Fig. S5. Definition of stripe amplitude. (A) The normalized mRNA profile for the stripe can be separated into an offset and an amplitude. (B) Normalized mRNA profiles and (C) stripe amplitude for the cytoplasmic pattern of mRNA as well as for the contributions from the various regulatory strategies.

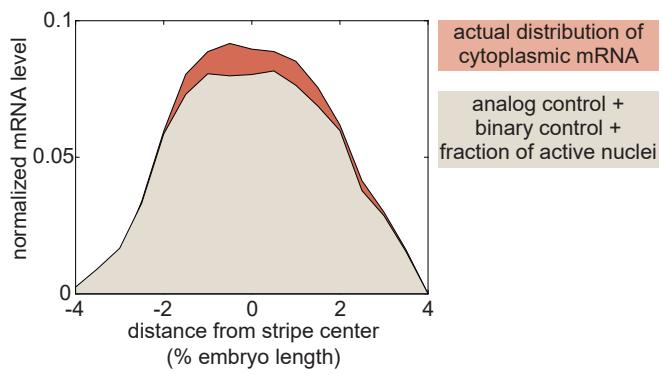


Fig. S6. Joint effect of mean rate, binary control, and fraction of active nuclei. Including of the predicted effect of anterior-posterior-dependent modulation of the fraction of active nuclei has little effect on the predicted cytoplasmic mRNA profile (compare brown profile in Figure 1G, gray profile above). The remaining difference between the full profile (red) and the gray profile can be attributed the effects of temporal variations in the mean rate of transcription.

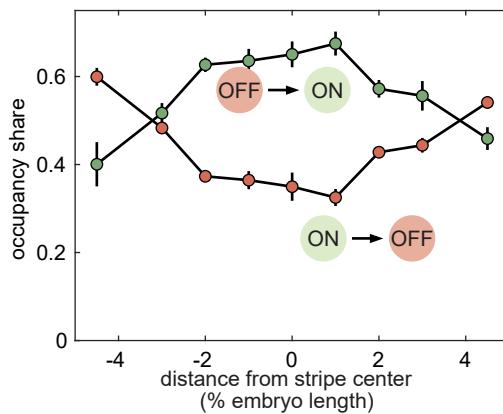


Fig. S7. Fraction of time spent in each transcriptional state. Fraction of time spent in the ON and OFF states as a function of the position along the stripe. (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

Supporting Information Text

Extended Materials and Methods

Data processing. Processed live-imaging movies were compiled from across 11 experiments (embryos) to form one master analysis set. While the position of *eve* stripe 2 along the anterior-posterior axis of the embryo was found to be consistent to within 1-2% of egg length, we sought to further reduce this embryo-to-embryo variation by defining new, “registered” AP axes for each experiment using the observed position and orientation of the mature stripe. To this end, an automated routine was developed to consistently establish the position and orientation of the *eve* stripe 2 center for each data set.

This routine, described graphically in [Figure S1](#), used observed spatial patterns of fluorescence measured from 30 minutes into nuclear cycle 14—the approximate time at which the mature stripe is first established ([2](#))—to the time of last observation (≥ 40 min) to find the natural position and orientation of the mature stripe. Generally, the *eve* stripes run roughly perpendicular to the anterior-posterior (AP) axis of the embryo; however, the approach allowed for the possibility that the true orientation of the *eve* 2 stripe deviated from the orientation implied by manual estimates of the anterior posterior axis. Thus, a variety of orientations for the natural stripe axis were considered, ranging between ± 15 degrees from the line perpendicular to the stripe with the manually specified anterior posterior axis. For each orientation, a sliding window of 4% embryo length in width was used to find the position along the proposed orientation that captured the largest fraction of the total fluorescence emitted by the mature stripe. The orientation and position that maximized the amount of fluorescence captured within this window defined a line through the field of view that was taken as the stripe center. All anterior-posterior positions used for subsequent analyses were defined relative to this center line.

Once the stripe centers for each set were established, fluorescence traces were interpolated to 20s resolution, with all times shifted to lie upon a common reference time grid. Traces near the edge of the field of view or that exhibited uncharacteristically large changes in fluorescence over a time step were flagged through a variety of automated and manual filtering steps. When necessary, these traces were removed from subsequent analyses to guard against the influence of non-biological artifacts.

cpHMM inference. To account for finite RNA polymerase (RNAP) elongation times, a compound state Markov formalism was developed in which the underlying two-promoter system—assumed to have three states (see [Figure 4E,F](#))—was transformed into a system with 3^w compound gene states, where w indicates the number of time steps needed for an RNAP molecule to traverse the full transcript (see [Figure S34](#)). These compound gene states played the role of the “hidden” states within the traditional HMM formalism. See *SI Appendix*, section D for details regarding the model’s architecture. Following this transformation from promoter states to compound gene states, it was possible to employ a standard version of the expectation-maximization (EM) algorithm, implemented using custom-written scripts in Matlab, to estimate bursting parameters from subsets of experimental traces (*SI Appendix*, section D). The scripts are available at the [GarciaLab/cpHMM](#) GitHub repository. Bootstrap sampling was used to estimate the standard error in our parameter estimates. Subsets of 8,000 data points were used to generate time-averaged parameter estimates. In order to accurately capture the time-averaged dynamics across the entirety of nuclear cycle 14, the full length of each experimental trace was used for time averaged inference. Sample sizes for windowed inference varied due to data set limitations. When possible, samples of 4,000 points were used. Only data points falling within a 15 minute window centered about the time point of interest were included in windowed inference runs. Inference was not conducted for spatiotemporal regions for which fewer than 1,250 time points were available. A minimum of 10 bootstrap samples were used to estimate each parameter value reported in this work. Reported values represent the median taken across bootstrap samples.

Input-output logistic regressions. The input-output analysis presented in [Figure 7](#) utilized input transcription factor data from immunostaining experiments presented in ([3](#)), as well as live measurements of a Bicoid-GFP fusion courtesy of Jonathan Liu and Elizabeth Eck. Logistic regression parameters were estimated in Matlab using the *fmincon* function. See *SI Appendix*, section H for further details.

Bootstrap error calculation. Bootstrap resampling was used frequently throughout this work to estimate the standard error in a variety of reported quantities, from trends estimated directly from raw experimental data in [Figure 1](#) to cpHMM inference results presented in [Figure 5](#) and [Figure 6](#). In this procedure, multiple bootstrap replicates, y_{boot}^i are generated by sampling with replacement from the pool of available experimental data, Y (see, e.g. ([4](#))). The parameter of interest (say, $t_{\text{on}}(x)$) is then calculated for each replicate and the mean of these estimates is taken as the bootstrap estimate of the parameter value, $\hat{t}_{\text{on}}(x)$, while the standard deviation across the pool of bootstrap parameter estimates is used to approximate the standard error in our estimate of $t_{\text{on}}(x)$. In our case, simply performing this procedure across the available pool of nuclei failed to account for biological variability that exists from embryo to embryo. To account for this, we introduced a hierarchical bootstrapping procedure. The first step in this procedure was to draw bootstrap samples from across the 11 embryos used in this study. Because these samples were taken with replacement, most bootstrap samples excluded some embryos out of the original set of 11 and included duplicates (or triplicates, etc.) of others. Each embryo-level bootstrap defined a subset of nuclei. The final set of nuclei used for parameter estimation was generated by performing another round of bootstrap sampling on this pool. Bootstrap averages and standard errors were then calculated as described above. This two-step procedure thus accounts for both embryo-to-embryo and nucleus-to-nucleus variability.

We note that the limited number of data points available for many spatiotemporal regions prevented us from performing this two-tiered bootstrap procedure in the case of our time-dependent cpHMM inference ([Figure 6D-F](#) and *SI Appendix*,

section G–Figure S28D–E). In these cases, we used all available sets (essentially skipping the first bootstrap resampling step) and took bootstrap samples from amongst available nuclei as in step two of the procedure described above.

Absolute calibration of MS2 signal. A major strength of the modelling techniques presented in this paper (*e.g.* cytoplasmic mRNA prediction, csHMM, logistic regressions) is that they can be applied directly to MS2 data, without a need to convert the signal into absolute counts of RNAP molecules. Because of this, none of the conclusions presented in this work depend upon an absolute AU to RNAP calibration. Still, it can be informative to view quantities in terms of biologically meaningful units.

Thus, in order to frame our results with respect to units with a clear physical interpretation, we calibrated our fluorescence measurements in terms of absolute numbers of mRNA molecules. This calibration was also used to inform our Poisson loading sensitivities (*SI Appendix*, section D). To calculate this calibration for our *eve* stripe 2 data, we relied on measurements reported by a previous study that utilized MS2 in conjunction with single molecule FISH to establish a calibration factor, α , between the integrated MS2 signal, F_{MS2} , and the number of mRNA molecules produced at a single transcriptional locus, N_{FISH} , (1) given by

$$\alpha = \frac{N_{\text{FISH}}}{F_{\text{MS2}}}. \quad [1]$$

This calibration factor can be used to estimate the average contribution of a single mRNA molecule to the observed (instantaneous) fluorescent signal. While the values for the parameters in 1 reported here pertain to the transcriptional output driven by the Bicoid activated P2 enhancer and promoter during nuclear cycle 13, the calibration should generalize to all measurements taken using the same microscope.

First, consider the total integrated fluorescence emitted by a single nascent mRNA while it is on the reporter gene,

$$F_1 = f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}}, \quad [2]$$

where f_{\max} denotes the instantaneous fluorescence emitted by a nascent mRNA that has transcribed the full complement of MS2 loops, L_I indicates the length of the MS2 loops, L_{II} indicates the distance between the end of the MS2 loop cassette and the 3' end of the gene, and v_{elong} indicates the elongation rate of RNAP molecules along the gene. We can solve for f_{\max} using α , namely,

$$F_1 = \frac{1}{\alpha} = f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}}, \quad [3]$$

such that

$$f_{\max} = \frac{v_{\text{elong}}}{\alpha} \frac{1}{\frac{1}{2}L_I + L_{II}}. \quad [4]$$

Here, we recognize that the cumulative fluorescence per RNAP molecule is simply the inverse of the number of molecules per unit fluorescence (α). Now we have the pieces necessary to derive an expression for the *instantaneous* fluorescence of a single RNAP molecule, that is,

$$\begin{aligned} F_{\text{RNAP}} &= \frac{1}{\tau_{\text{elong}}} f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}} \\ &= \frac{v_{\text{elong}}}{(L_I + L_{II})} f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}} \\ &= f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{(L_I + L_{II})} \\ &= \frac{v_{\text{elong}}}{\alpha} \frac{1}{(L_I + L_{II})}, \end{aligned} \quad [5]$$

resulting in

$$F_{\text{RNAP}} = \frac{v_{\text{elong}} F_{\text{MS2}}}{N_{\text{FISH}}} \frac{1}{(L_I + L_{II})}. \quad [6]$$

Measurements performed in (1) give N_{FISH} to be 220 (± 30) mRNA per nucleus and v_{elong} to be 1.5 (± 0.14) kb/min. Experimental measurements on the P2 enhancer (courtesy of Elizabeth Eck, Maryam Kazemzadeh-Atoufi and Jonathan Liu) indicate that the total fluorescence per nucleus, F_{MS2} , is 9,600 (± 320) AU minutes. For the reporter gene used to take these measurements, L_I and L_{II} are 1.275 kb and 4.021 kb, respectively. Thus, we obtain

$$\begin{aligned} F_{\text{RNAP}} &= \frac{1.5 \times 9610}{220} \frac{1}{(1.275 + 4.021)} \\ &= 13 \pm 1.7 \text{ AU/RNAP}. \end{aligned} \quad [7]$$

Though the error in our calibration is significant ($>13\%$), the conversion from arbitrary units to numbers of nascent mRNA nonetheless provides useful intuition for the implications of our inference results, and none of our core results depend upon having access to a precise calibration of the observed signal in terms of absolute numbers of RNAP molecules.

Appendices

A. Theoretical model to predict cytoplasmic mRNA levels given from *in vivo* measurements of transcriptional activity.

Derivation details. Here we provide a more detailed treatment of the mathematical framework for connecting transcriptional activity in individual nuclei to levels of accumulated cytoplasmic mRNA. We begin with general expressions for the rate of mRNA production during the transcriptionally active and quiescent periods that dictate the transcriptional time window. When the promoter is actively transcribing ($t_{\text{on}} \leq t \leq t_{\text{off}}$), the net rate of mRNA production is

$$\frac{d\text{mRNA}}{dt}(x, t) = \underbrace{R(x, t)}_{\text{transcription rate}} - \underbrace{\gamma \text{mRNA}(x, t)}_{\text{degradation rate}}, \quad [8]$$

where γ is the mRNA degradation rate constant. For a promoter that has entered a transcriptionally quiescent state ($t > t_{\text{off}}$), we have

$$\frac{d\text{mRNA}}{dt}(x, t) = -\gamma \text{mRNA}(x, t), \quad [9]$$

such that degradation is now the only contribution to the change of mRNA concentration in time. Note that, in these two equations, we have ignored the contribution of mRNA diffusion. Previous measurements have estimated a diffusion coefficient of mRNA of $0.09 \mu\text{m}^2/\text{s}$ (5) and a typical mRNA degradation rate of 0.14 min^{-1} (6). Given these numbers, we expect an *eve* mRNA molecule to diffuse approximately $6 \mu\text{m}$, which corresponds to one nuclear diameter or 1% of the embryo length, before being degraded. Thus, given the overall width of the stripe mRNA profile of about 8% of the embryo length (Figure 3G), we expect diffusion to play a minimal role in stripe formation. Finally, note that we are also ignoring the delay between transcriptional initiation and the delivery of an mRNA molecule to the cytoplasm as a result of nuclear export. This delay would affect the *timing* of pattern formation, but would leave our conclusions about the relative role of transcriptional bursting and the regulation of the duration of the transcriptional time window unaffected.

To make progress, as in the main text, we make the simplifying assumption that the instantaneous rate of transcription can be well approximated by the time average at each position given by

$$R(x) \approx \langle R(x, t) \rangle_t. \quad [10]$$

We now consider the role of $t_{\text{on}}(x)$ in dictating pattern formation by envisioning a scenario where transcription begins at time $t_{\text{on}}(x)$, but does not cease. In this scenario, the accumulated mRNA is given by

$$\text{mRNA}_{\text{active}}(x, t) = \underbrace{R(x)}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma} (1 - e^{-\gamma(t-t_{\text{on}}(x))})}_{\text{time window}}. \quad [11]$$

Note that if the system evolves for a long amount of time, the second term in the parenthetical in Equation 11 becomes vanishingly small ($\gamma(t - t_{\text{on}}(x)) \gg 1$) such that all time dependence drops out of the expression and we recover the familiar expression for mRNA levels in steady state

$$\text{mRNA}_{\text{active}}(x, t) = \frac{R(x)}{\gamma}, \quad [12]$$

where mRNA production and degradation are balanced.

Next, consider the impact of regulating the timing with which nuclei cease transcriptional activity and become quiescent, t_{off} . Here, when $t > t_{\text{off}}(x)$, the amount of mRNA produced during the period of activity is subsumed within a decaying exponential envelope such that

$$\text{mRNA}_{\text{quiescent}}(x, t) = \underbrace{e^{-\gamma(t-t_{\text{off}}(x))}}_{\text{quiescent decay}} \left[\underbrace{R(x)}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma} (1 - e^{-\gamma(t_{\text{off}}(x)-t_{\text{on}}(x))})}_{\text{time window}} \right]. \quad [13]$$

Equation 13 represents a scenario in which the accumulation of cytoplasmic mRNA results from the interplay between two distinct regulatory strategies: the modulation of when the transcription starts and stops (binary control of the transcription time window) and the average rate with which transcription occurs within this time window (analog control of transcriptional bursting). We refactor Equation 13 to reflect this distinction and consider the case when $t > t_{\text{on}}$, giving

$$\text{mRNA}_{\text{full}}(x, t) = \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{e^{-\gamma(t-\min(t_{\text{off}}(x), t))} (1 - e^{-\gamma(\min(t_{\text{off}}(x), t) - t_{\text{on}}(x))}}_{\text{binary control}}, \quad [14]$$

which can be simplified slightly to yield

$$\text{mRNA}(x, t) = \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{(e^{-\gamma(t-\min(t_{\text{off}}(x), t))} - e^{-\gamma(t-t_{\text{on}}(x))})}_{\text{binary control}}. \quad [15]$$

Finally, we account for the fact that only some $p_{\text{active}}(x)$ fraction of nuclei within each region *ever* engage in transcription leading to

$$\text{mRNA}(x, t) = p_{\text{active}}(x) \times \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{(e^{-\gamma(t-\min(t_{\text{off}}(x), t))} - e^{-\gamma(t-t_{\text{on}}(x))})}_{\text{binary control}}. \quad [16]$$

This equation constitutes the basis of our theoretical dissection of pattern formation by transcriptional bursting and the control of the transcriptional time window.

Accounting for multiple transcriptional states. In the main text, [Equation 3](#) expresses the mean rate of mRNA production, $R(x)$, as a function of the bursting parameters k_{on} , k_{off} , and r . We can combine this equation with [Equation 16](#) to obtain an expression for the predicted amount of cytoplasmic mRNA that includes the burst parameters inferred by our cpHMM

$$\text{mRNA}(x, t) = p_{\text{active}}(x) \times \underbrace{\frac{r(x)}{\gamma} \frac{k_{\text{on}}(x)}{k_{\text{on}}(x) + k_{\text{off}}(x)}}_{\text{analog control}} \times \underbrace{(e^{-\gamma(t-\min(t_{\text{off}}(x), t))} - e^{-\gamma(t-t_{\text{on}}(x))})}_{\text{binary control}}. \quad [17]$$

While we present our results in terms of an effective two-state model in the main text, the presence of two transcriptional loci within each observed fluorescent spot suggests that the system is more naturally described using a three-state kinetic model. Here, we extend the framework presented in [Equation 17](#) to a scenario in which there are three distinct system states: 0 promoters on (0), 1 promoter on (1), and both promoters on (2) (see [Figure 4](#)). We begin with a general expression for this scenario that takes the contribution from the analog control term shown in [Equation 16](#) to be a sum over the output of each of the 3 activity states, namely,

$$\text{mRNA}(x, t) = p_{\text{active}}(x) \times \underbrace{\frac{1}{\gamma} \left(\sum_{i=0}^2 r_i(x) \pi_i(x) \right)}_{\text{analog control}} \times \underbrace{(e^{-\gamma(t-\min(t_{\text{off}}(x), t))} - e^{-\gamma(t-t_{\text{on}}(x))})}_{\text{binary control}}, \quad [18]$$

where $r_i(x)$ is the rate of RNAP loading for state i , and $\pi_i(x)$ indicates the fraction of time spent in state i . Note that the independent effect of the duration of the transcription time window and of mRNA decay on cytoplasmic mRNA levels remain unchanged in the multi-state case.

The fractional occupancies of the activity states ($\pi_i(x)$ terms in [Equation 18](#)) are a function of the rates with which the promoter switches between activity states. In general, the fractional occupancy of each activity state, π_i , may vary as a function of time; however we focus on their steady state values here, such that:

$$0 = \mathbf{R}(x)\boldsymbol{\pi}(x), \quad [19]$$

where $\mathbf{R}(x)$ is the transition rate matrix. Consistent with our inference results, we assume that no transitions are permitted between the high and low states (0 & 2). Thus, the transition rate matrix takes the following form:

$$\mathbf{R}(x) = \begin{bmatrix} -k_{01}(x) & k_{10}(x) & 0 \\ k_{01}(x) & -k_{10}(x) - k_{12}(x) & k_{21}(x) \\ 0 & k_{12}(x) & -k_{21}(x) \end{bmatrix}. \quad [20]$$

Together, [Equation 19](#) and [Equation 20](#) allow us to solve for the fractional occupancy of each activity state as a function of the transition rates that describe the system.

For the remainder of this derivation, we will drop the explicit x and t dependencies for ease of notation. Intuitively, the steady state (or stationary) distribution represents a limiting behavior of the system such that, upon reaching $\boldsymbol{\pi}$, no further shifts occur in the mean fraction of time spent in each activity state. [Equation 19](#) leads to a system of three equations:

$$0 = -\pi_0 k_{01} + \pi_1 k_{10} \quad [21]$$

$$0 = \pi_0 k_{01} - \pi_1 (k_{10} + k_{12}) + \pi_2 k_{21} \quad [22]$$

$$0 = \pi_1 k_{12} - \pi_2 k_{21}. \quad [23]$$

Before proceeding, we note that, since π is a probability distribution, we can eliminate one of our unknowns by enforcing normalization, that is,

$$1 = \pi_0 + \pi_1 + \pi_2. \quad [24]$$

With this in mind, we can solve [Equation 21](#) for π_1 to find

$$\pi_1 k_{10} = \pi_0 k_{01} \quad [25]$$

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}}. \quad [26]$$

Next, we use the normalization condition to eliminate π_2 from [Equation 23](#):

$$\begin{aligned} \pi_1 k_{12} &= \pi_2 k_{21} \\ &= (1 - \pi_0 - \pi_1) k_{21}. \end{aligned} \quad [27]$$

By combining this result with [Equation 26](#), we obtain

$$\pi_0 \frac{k_{01}}{k_{10}} k_{12} = (1 - \pi_0 - \pi_0 \frac{k_{01}}{k_{10}}) k_{21} \quad [28]$$

$$\pi_0 \frac{k_{01} k_{12}}{k_{10} k_{21}} = 1 - \pi_0 \frac{k_{10} + k_{01}}{k_{10}} \quad [29]$$

$$\pi_0 = \frac{k_{10} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \quad [30]$$

With [Equation 30](#) in hand, it is then straightforward to solve for the remaining π_i terms. First we obtain π_1 by substituting [Equation 30](#) into [Equation 26](#):

$$\begin{aligned} \pi_1 &= \pi_0 \frac{k_{01}}{k_{10}} \\ &= \frac{k_{01} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \end{aligned} \quad [31]$$

And finally π_2 :

$$\begin{aligned} \pi_2 &= 1 - \pi_0 - \pi_1 \\ &= \frac{k_{01} k_{12}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \end{aligned} \quad [32]$$

Thus, we arrived at the full expression for cytoplasmic mRNA levels in the 3-state case:

$$\text{mRNA}(x, t) = p_{\text{active}}(x) \underbrace{\frac{1}{\gamma} \left(r_1(x) \frac{k_{01}(x) k_{21}(x)}{\kappa(x)} + r_2(x) \frac{k_{01}(x) k_{12}(x)}{\kappa(x)} \right)}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t - \min(t_{\text{off}}(x), t))} - e^{-\gamma(t - t_{\text{on}}(x))} \right)}_{\text{binary control}}, \quad [33]$$

where, consistent with the 2-state case, we have taken $r_0(x)$ to be equal to zero and where $\kappa(x)$ denotes the denominator in [Equation 30](#), [Equation 31](#) and [Equation 32](#), namely,

$$\kappa = k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}. \quad [34]$$

Thus, from [Equation 33](#) we see that, while there are more terms comprising the analog control expression, the expression nonetheless takes on the same essential form as in [Equation 16](#).

Mapping the three-state model into an effective two-state model. Here we provide expressions relating the effective two-state parameters presented in the main text to parameters from the full three-state model. As we have done throughout this work, we take the transition rates between states (0) and (2) of the 3-state model to be negligible (consistent with inference results, see *SI Appendix*, section G). First, the on rate, $k_{\text{on}}^{\text{eff}}$ is directly equivalent to the transition rate between states (0) and (1), that is,

$$k_{\text{on}}^{\text{eff}} = k_{01}. \quad [35]$$

Similarly, since we do not observe from state (2) to state (0), $k_{\text{off}}^{\text{eff}}$ is equal to the transition rate from (1) to (0), weighted by the relative fraction of time the system spends in state (1) when it is in the effective ON state (1 or 2). Thus, we have:

$$k_{\text{off}}^{\text{eff}} = \frac{\pi_1 k_{10}}{\pi_1 + \pi_2} \quad [36]$$

$$= \frac{k_{01} k_{21} k_{10}}{k_{01} k_{21} + k_{01} k_{12}} \quad [37]$$

$$= \frac{k_{21}k_{10}}{k_{21} + k_{12}}. \quad [38]$$

Finally, r^{eff} is the occupancy-weighted average of the initiation rates for states (1) and (2), namely,

$$r^{\text{eff}} = \frac{\pi_1 r_1 + \pi_2 r_2}{\pi_1 + \pi_2} \quad [39]$$

$$= \frac{r_1 k_{01} k_{21} + r_2 k_{01} k_{12}}{k_{01} k_{21} + k_{01} k_{12}} \quad [40]$$

$$= \frac{r_1 k_{21} + r_2 k_{12}}{k_{21} + k_{12}}. \quad [41]$$

B. Measuring the amount of produced mRNA. Here, we outline our methodology for estimating rates of mRNA production depicted in [Figure 3A](#) and B, as well as the total cytoplasmic mRNA levels per nucleus shown in [Figure 3G](#).

Calculating rates of mRNA production. The observed fluorescence signal at transcriptional loci as a function of time, $F(t)$, is linearly related to the number of actively transcribing RNAP molecules. Thus, after a period equal to the amount of time needed for an RNAP molecule to transcribe the gene, τ_{elong} , the number of new mRNAs added to the cytoplasm will be proportional to $F(t)$ (2), that is,

$$F(t) \propto M(t + \tau_{\text{elong}}) - M(t), \quad [42]$$

where $M(t)$ indicates the total number of mRNA molecules that have been produced up to time t . We can relate this fluorescence signal to absolute numbers of RNAP molecules using the calibration procedure described in the Extended Materials and Methods. Drawing from the derivation provided in the SI Methods of (2), we take the rate of mRNA production at time t to be approximately equal to the observed fluorescence at time $t - \frac{\tau_{\text{elong}}}{2}$,

$$F\left(t - \frac{\tau_{\text{elong}}}{2}\right) \propto \frac{dM(t)}{dt}. \quad [43]$$

Here, the $\frac{\tau_{\text{elong}}}{2}$ term accounts for the time lag between the number of transcribing nascent mRNA molecules and their release into the cytoplasm. An alternative way to think about this is that [Equation 42](#) is essentially an expression for the time derivative of $M(t)$, centered at time $\frac{\tau_{\text{elong}}}{2}$. For ease of notation, we will ignore this offset factor for the remainder of this section. We will also treat the relationship in [Equation 43](#) as one of equality.

[Figure 3A](#) depicts the time-averaged rate of mRNA production for each nucleus within the experimental field of view for one of our 11 live imaging movies. For each nucleus, this quantity was obtained by averaging the fluorescence across all observed time points, from when the nucleus first turned on (t_{on}) through to the final time point where expression was detected (t_{off}), which is taken to be *either* the time at which the nucleus transitioned into a quiescent state *or* the time depicted in the figure (40 minutes into nuclear cycle 14)—whichever came first. Thus the average rate of mRNA production for nucleus i is obtained from

$$R_i = \frac{\alpha \sum_{t=t_{\text{on}}}^{t_{\text{off}}} F_i(t)}{\sum_{t=t_{\text{on}}}^{t_{\text{off}}} 1}, \quad [44]$$

where the denominator is the total number of time points over which the averaging is performed and α indicates a conversion factor with units of RNAP molecules per AU per unit time that accounts for two factors: (1) the conversion from fluorescence to absolute numbers of RNAP molecules and (2) the dwell time of RNAP molecules on the gene. From [Equation 7](#), we know that we have one RNAP molecule for every 13 AU. The second factor is analogous to the $v_{\text{elong}}/(L_I + L_{II})$ component of [Equation 5](#), but with the appropriate lengths and elongation rates for the *eve* stripe 2 reporter. See *SI Appendix*, section J for details about how we estimate the elongation time for our experimental system. Here, we will simply quote the results that $\alpha = 0.037$ RNAP molecules per AU per minute. Thus, the mean rate of mRNA production of a single nucleus is given by

$$R_i = \frac{0.037 \sum_{t=t_{\text{on}}}^{t_{\text{off}}} F_i(t)}{\sum_{t=t_{\text{on}}}^{t_{\text{off}}} 1}. \quad [45]$$

A similar procedure was performed to estimate the average rate of mRNA production for each region along the anterior-posterior that is depicted in [Figure 3B](#). This time, however, we summed over observed fluorescence values for *all* nuclei within the relevant region and time period and divided by the total number of time points such that

$$R_x = \frac{0.037 \sum_{i=1}^N \sum_{t=t_{\text{on}}}^{t_{\text{off}}} F_i(t)}{\sum_{i=1}^N \sum_{t=t_{\text{on}}}^{t_{\text{off}}} 1}, \quad [46]$$

where N indicates the total number of nuclei falling within anterior-posterior region x .

Calculating full mRNA profiles. In contrast to the production rates calculated above, determining the relative contributions to stripe formation from each regulatory strategy depicted in Figure 3G did not require an AU to RNAP calibration. Thus, we capture the calibration factor, along with all other proportionality constants, with a generic term β , with the expectation that β will drop out from all consequential stripe contribution calculations. For a given region along the axis of the embryo, the average observed fluorescence across all N nuclei (active, quiescent, and those that never engaged in transcription) within the region of interest was used as a proxy for the instantaneous rate of mRNA production per nucleus, given by

$$\begin{aligned}\frac{dM(x, t)}{dt} &= \frac{\beta}{N} \sum_{i=1}^N F_i(x, t) \\ &= \beta \langle F(t) \rangle_x.\end{aligned}\quad [47]$$

Here, $F_i(x, t)$ is the fluorescence of nucleus i at time t and position x . The x subscript in Equation 47 indicates that the average is taken over all nuclei falling within the same anterior-posterior region within the *eve* stripe 2 pattern.

Having obtained an expression for the rate of mRNA production as a function of space and time, we next sought to account for the degradation of mRNA over time. As indicated in the main text, we assumed a constant rate of mRNA decay, γ , over space and time. The next section in this appendix provides evidence for the validity of this assumption. For a constant mRNA decay rate, calculating the average concentration of mRNA amounts to taking a weighted sum over all preceding production rates for a position of interest, where the weight terms account for the effects of mRNA decay and are of the form $e^{-\gamma t}$. Thus, we summed over all time points for each region of interest to estimate the total amount of cytoplasmic mRNA present on average, yielding the quantity on the left-hand side of Equation 2, namely,

$$\text{mRNA}(x, t) = \beta \sum_{n=1}^T e^{-\gamma(t-n\Delta\tau)} \langle F(t - n\Delta\tau) \rangle. \quad [48]$$

Here $\Delta\tau$ is the experimental time resolution, and $T = \frac{t}{\Delta\tau}$ denotes the number of measurements taken through time t . The exponential term within the summand on the right-hand side captures the effects of mRNA decay (see SI Appendix, section A). Finally, to calculate the normalized mRNA profile shown in Figure 3G (red), the estimates for the total accumulated mRNA per nucleus found using Equation 48 must be divided by the sum across all spatial regions considered, namely

$$\text{mRNA}_{\text{norm}}(x_j, t) = \frac{\sum_{n=1}^T e^{-\gamma(t-n\Delta\tau)} \langle F(t - n\Delta\tau) \rangle_{x_j}}{\sum_{x_i \in X} \sum_{n=1}^T e^{-\gamma(t-n\Delta\tau)} \langle F(t - n\Delta\tau) \rangle_{x_i}}, \quad [49]$$

where X denotes the set of all regions along the anterior-posterior axis that were considered for the profile analysis and the subscripts i and j outside the angled brackets denote the spatial region over which the sum is taken. Note that the proportionality constant β cancels in the final expression for $\text{mRNA}_{\text{norm}}$. As a final step, we subtract the minimum across the anterior-posterior region considered to remove any basal offset such that

$$\text{mRNA}_{\text{full}}(x_j, t) = \text{mRNA}_{\text{norm}}(x_i, t) - \min_x (\text{mRNA}_{\text{norm}}(x_j, t)). \quad [50]$$

Validating the fluorescence model. We employed a stochastic simulation to test the validity of the relation proposed in Equation 47 and, more generally, of the approximate equality between time-lagged fluorescence and mRNA production asserted in Equation 42. Simulated traces were generated using the Gillespie Algorithm (7), adjusted to allow system parameters to vary in time. We assumed an effective off rate of 0.667 transitions per minute and an initiation rate of 16 RNAP molecules per minute. To generate a temporal trend, we varied the effective on rate from an initial value of 1.6 transitions per minute (0 to 12 minutes of our simulation) to a basal value of 0.5 transitions per minute (27.5 to 35 minutes of our simulation). Figure S8 compares the rate of mRNA production predicted using the time-lagged average of simulated fluorescence traces, to the true rate of mRNA production.

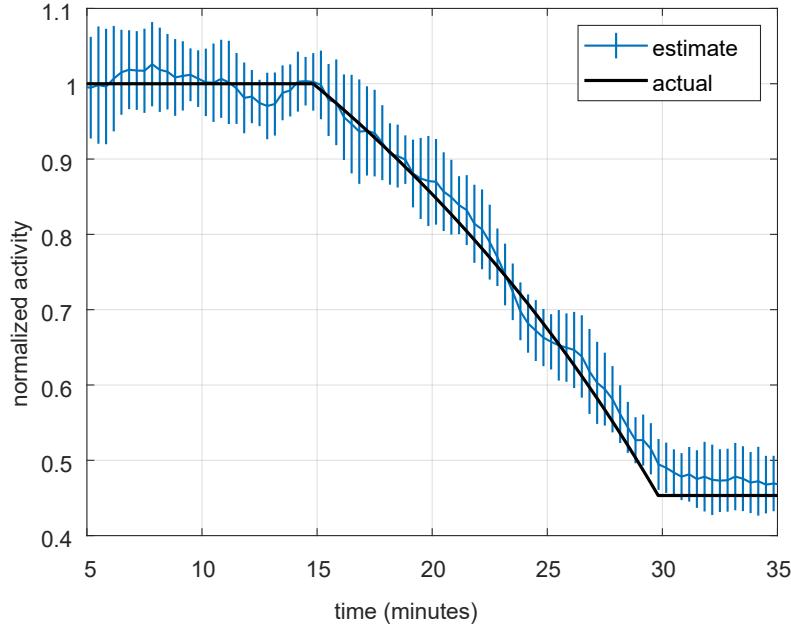


Fig. S8. Comparison of actual and estimated mRNA production rates. The black trend indicates the true rate of mRNA production as a function of time used for the simulation of transcription. The blue trend indicates the rate estimated from these simulated fluorescence traces. Error bars indicate standard deviation across 10 replicates containing 100 simulated traces each.

As expected, the approach faithfully recapitulates the true trend. Indicating that using spot fluorescence as a proxy for mRNA production should yield reliable results.

Calculating mRNA profiles due to the binary control of the transcriptional time window. The predicted profile due to binary control of the transcriptional time window alone (Figure 3G, blue) was calculated following the same procedure as for the full mRNA profile described above, save for the fact that, in this case, instantaneous fluorescence values for individual nuclei were converted to binary indicator variables ($f_i(t)$) that were set equal to 1 if $t < t_{\text{off}}^i$ and 0 otherwise. Additionally, only nuclei that were active at *some point* during nuclear cycle 14 were included in order to distinguish the effects of the transcriptional time window (Figure 1C) from the control of the fraction of active nuclei (Figure 1D). Thus, in this scenario, the “average rate” of mRNA production is equivalent to the fraction of nuclei engaged in transcriptional activity at a given point in time such that the rate of mRNA production is given by

$$\begin{aligned} \frac{d \text{mRNA}_{\text{binary}}(x, t)}{dt} &= \frac{1}{N(x)} \sum_{i=1}^{N(x,t)} f_i(t) \\ &= \langle f(x, t) \rangle \\ &= \frac{N_c(x, t)}{N(x)}, \end{aligned} \quad [51]$$

where $N_c(x, t)$ indicates the number of transcriptionally engaged nuclei at time t and position x , $N(x)$ indicates the *total* number of nuclei at position x that were transcriptionally competent at some point during nuclear cycle 14, and $\langle f(x, t) \rangle$ indicates the fraction of competent nuclei at position x and time t . The binary equivalent to Equation 48 takes the form of a time-weighted sum of the fraction of active nuclei within a region

$$\text{mRNA}_{\text{binary}}(x, t) = \sum_{n=1}^T e^{-\gamma(t-n\Delta t)} \frac{N_c(x, n\Delta t)}{N(x)}. \quad [52]$$

The steps for calculating the the normalized binary mRNA levels comprising the blue profile in Figure 3G from Equation 52 are identical to those shown for the full mRNA profile in Equation 49 and Equation 50 and are therefore not repeated here.

Comparison between predicted and measured cytoplasmic mRNA profiles. As a check for the validity of our approach to predicting levels of cytoplasmic mRNA from live imaging data (Equation 49 and Equation 50), we sought to compare our model's predictions to existing mRNA FISH data for the endogenous *eve* stripe 2 (8). For this comparison, we elected to use live imaging data for *eve* stripe 2 activity that was driven by a BAC containing the full *eve* locus (see (9) for details). This was done to minimize potential differences with the activity of the endogenous gene. Most notably, unlike the BAC reporter construct, the minimal reporter construct used for the majority of this work does not contain an enhancer sequence that is responsible for driving *eve* expression late in nuclear cycle 14 (10).

The researchers who generated the mRNA FISH data used the percent invagination of cellular membranes through cellularization as a means to break individual fixed embryos into rough temporal cohorts (11). We cross-referenced the invagination ranges for each temporal group in the FISH data from (11) with data provided by precise measurements of invagination for different time points in (3) to obtain estimates for the range of times encapsulated by each of these cohorts.

We elected to use the cohort comprised of embryos with ages ranging between 38 to 48 minutes into nuclear cycle 14 because this range was much narrower than the preceding cohort and because the stripe appeared to be relatively stable during this time period. We note that the authors of (11) measured invagination on the ventral surface of the embryo, while the authors in (3) used the dorsal surface. This difference could lead to inconsistencies, since invagination is known to proceed more rapidly on the ventral side of the embryo (11). However, the authors in (11) reported that this discrepancy is minimal up to the point where cell membrane extension has progressed to approximately 40% of its eventual full extent. The lower and upper bounds on the percent membrane invagination for the chosen cohort are 26% and 50% respectively. Thus, we expect the time estimate derived for the beginning of the period to be reasonably accurate, since dorsal and ventral membrane progression was reported to be comparable during this period. Moreover, to the degree that ventral invagination outpaces dorsal invagination at the end of our period of interest, this would result in an over-estimation of ending time. Thus, if anything, the true temporal window encompassed by the selected cohort may actually be tighter than 10 minutes, since the ending time might in fact be earlier than 48 minutes into nuclear cycle 14. Given the relative stability of the stripe profile during this period of development, we do not expect this potential discrepancy to have a material impact on our conclusions.

SI Appendix, section B **Figure S9** summarizes the results of this comparison. To account for uncertainty regarding the precise dorsal-ventral (DV) orientation of embryos within our live-imaging set, we compared our model's predictions to mRNA measurements for a range of DV positions, encompassed by the green-shaded profile. We found a high degree of agreement between model predictions and reported levels of cytoplasmic mRNA. This conclusion is relatively insensitive to our assumptions regarding the average lifetime of *eve* mRNA as shown by the blue and red lines in the figure (predictions assuming mRNA lifetimes of 7 and 15 minutes, respectively). We thus concluded that the assumptions underlying our model for predicting cytoplasmic mRNA levels from *in vivo* single-cell transcriptional activity measurements are valid.

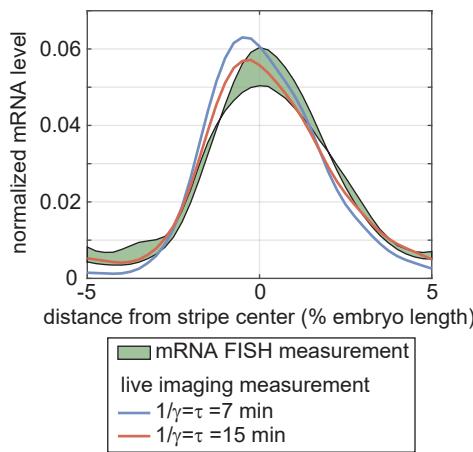


Fig. S9. Comparison of predicted cytoplasmic mRNA by live-imaging measurements to direct measurements by FISH. In an effort to check the validity of our modelling assumptions, we compared the predicted cytoplasmic mRNA profile stemming from live-imaging measurements of stripe 2 of an *eve* reporter from a BAC containing the full *eve* locus (9) to direct measurements of *eve* cytoplasmic mRNA levels using FISH (11). Here, the blue and red lines indicate our model's predictions under two different assumptions for the rate of mRNA degradation, and the shaded green profile indicates the range of reported mRNA levels for different ranges of dorsal-ventral positions. Comparisons indicate a high degree of agreement between prediction and measurement, indicating that our modelling assumptions are justified.

Sensitivity of results to mRNA lifetime assumption. In the main text we assume a degradation rate for *eve* of 0.14 min^{-1} (corresponding to a lifetime of roughly $\tau = 7 \text{ min}$). Since, to our knowledge, the decay rate of *eve* mRNA has not been measured directly, we follow (2) and base this estimate off of measurements for another of the pair rule genes, *fushi tarazu* (*ftz*) (12). In this section, we examine the degree to which the apparent contributions of each regulatory strategy (Figure 1) change under different assumptions for *eve* mRNA lifetime. Rather than conducting an exhaustive survey, we instead focus primarily on two limiting cases: rapid mRNA decay ($\tau = 1 \text{ min}$) and *no* mRNA decay ($\tau = \infty$).

Figure S10 summarizes the results of our analysis. We find that, regardless of the assumed mRNA lifetime, our model predicts that *eve* stripe 2 is formed almost entirely via the interplay between the binary control of the transcriptional time window and the analog modulation of the mean rate of transcription (compare brown and red profiles in Figure S10). However, we find that the relative importance of each factor depends, somewhat, on the assumed decay rate. In the case of rapid mRNA decay, as well as for the decay rate assumed in the main text, the time window (blue profile) is clearly the dominant factor in driving pattern formation (Figure S10A and B). If we assume the true mRNA lifetime is 15 minutes, slightly more than double our best guess of 7 minutes, we find that the time window is still predicted to contribute slightly more to stripe formation, but that the two contributions are now of order with one another (Figure S10C). Finally, in the limit where there is effectively no mRNA decay, the effects of the mean rate and time window are roughly equivalent (Figure S10D). This result can be explained

by the fact that the mean rate strategy is insensitive to the decay rate, whereas the effect of the time window is enhanced by the action of mRNA decay.

Thus, overall, we found that our model's prediction that the control of the transcriptional time window plays a primary role in stripe formation holds for mRNA lifetimes less than or equal to 15 minutes, which is more than double the measured life time of *ftz* mRNA (6). Perhaps more importantly, both factors are found to play a significant role, *irrespective* of mRNA decay rate, indicating that our central finding is robust to our assumption regarding mRNA decay dynamics.

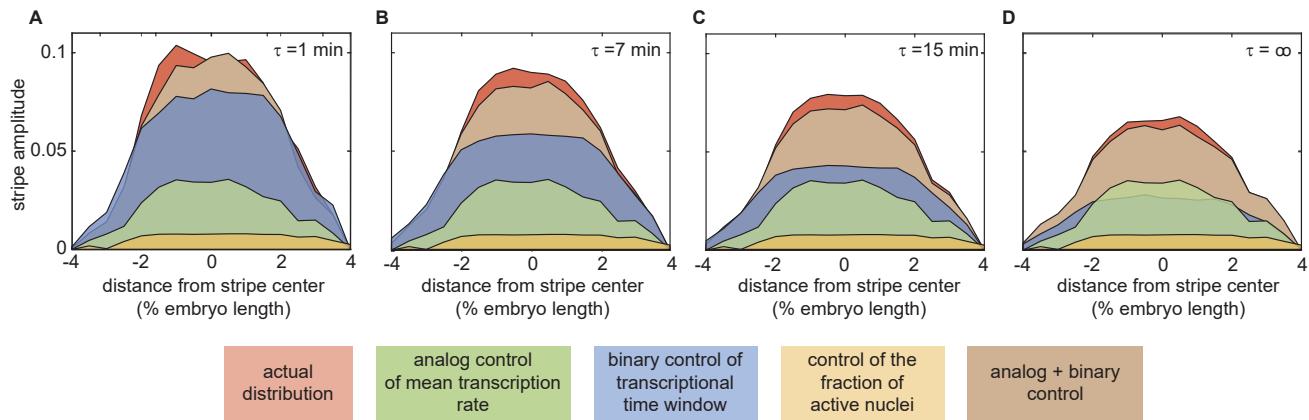


Fig. S10. Sensitivity of regulatory strategy contribution to assumed mRNA lifetime. The average lifetime of *eve* mRNA is a significant assumed parameter in our model. This figure compares the predicted contributions of each regulatory strategy for the mRNA lifetime assumed in the main text ($\tau = 7$ min) to limiting cases in which mRNA is assumed to decay almost instantaneously ($\tau = 1$ min) on the one hand, and infinitely slowly on the other ($\tau = \infty$). Even at these extremes, the central conclusion that the stripe is formed via the joint action of mean rate modulation (green profile) and the time window (blue profile) remains intact. As expected, the relative contribution of the time window is sensitive to the assumed τ , yet even in the limit of no significant mRNA decay, its impact is still of order with the effect of mean rate modulation.

Control strategy contributions for eve BAC. A key question regarding the results in the main text is whether and to what degree the relative contributions of the regulatory control strategies we identified in Figure 1 and Figure 3 for the reporter containing only the *eve* stripe 2 enhancer hold true for the formation of the stripe in the endogenous context. While we cannot directly query activity at the endogenous *eve* locus, we were able to examine the dynamics of stripe formation for an *eve* BAC used in the companion paper to this manuscript (9). Since this BAC contains the full *eve* regulatory locus, it likely provides a better proxy for stripe formation in the endogenous context than the isolated *eve* 2 reporter. Figure S11 shows the results of this analysis. As with the reporter construct used in the main text (Figure S11A), we find that in the endogenous context (Figure S11B) the stripe is formed primarily through the interplay between two regulatory strategies: the modulation of the average rate of production (green) and of the duration of transcriptional activity (blue). As with the reporter, the binary control of the transcriptional time window is the dominant driver of stripe formation (compare with Figure 1G). Interestingly, unlike the reporter construct, the full predicted profile (red profile) that accounts for the interplay between mRNA decay temporal fluctuations in the mean rate of mRNA production differs substantially from the simpler model (brown profile) that approximates mRNA production as constant over time. We speculate that this difference is attributable to the influence of the “late enhancer”—which is present in the *eve* BAC but *not* in the reporter—that takes over control of *eve* activity late in nuclear cycle 14. Further work will be necessary to fully elucidate the regulatory impact of this late element on the formation of the mature *eve* stripe pattern.

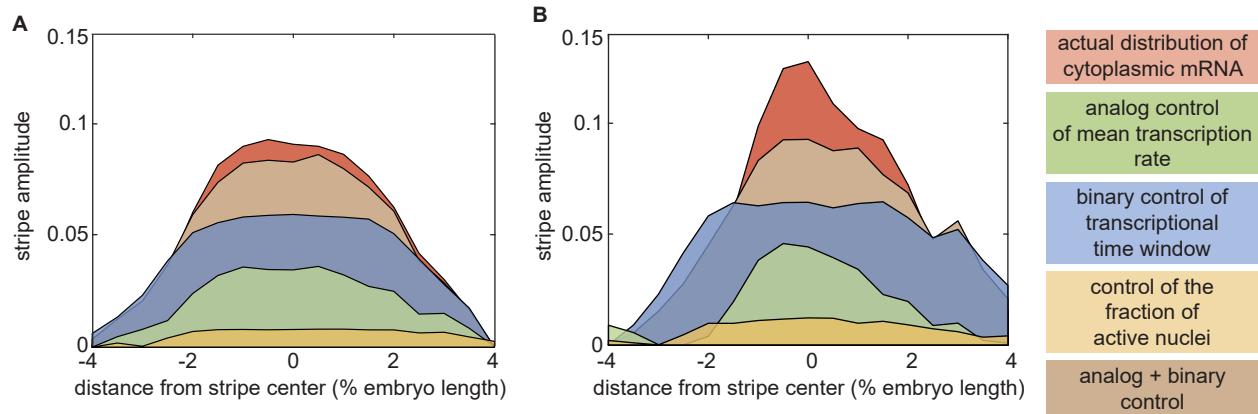


Fig. S11. Regulatory strategy contributions to eve stripe 2 formation in endogenous context. (A) Control strategy contributions for the reporter *eve* stripe 2 reporter construct (reproduced from Figure 3G). (B) Predicted contributions for *eve* BAC. As with the reporter construct, the formation of *eve* stripe 2 in the context of the full *eve* regulatory locus is dominated by the interplay between mean rate modulation (green) and control of the time window of transcriptional activity (blue).

C. Accounting for effects of experimental detection threshold. A number of analyses in this work rely (directly or indirectly) on the estimation of when gene loci first enter into a transcriptionally engaged state (t_{on}) as well as when they stop transcribing (t_{off}). These quantities are estimated using our live imaging data. Because live imaging experiments are subject to a detection limit (below which dim loci will not register as being active), there is a potential for bias in our estimates of t_{on} and t_{off} , as well as other metrics that derive in part from these quantities (duration of transcriptional activity, for instance). Similarly, there is a potential for bias in the measurements of the fraction of active nuclei as well as the transcriptional time window (Figure 3C-F) and the contributions of these regulatory strategies to pattern formation depicted in Figure 3G. These quantities were calculated assuming that undetected loci produced *no* mRNA when, in fact, it is possible that they produce at low levels periodically or even throughout the whole nuclear cycle. To estimate the nature and severity of these potential biases, we first estimated the detection limit for our live imaging experiments. Where appropriate, we used this limit to determine its potential effect on our conclusions.

Estimating the detection limit. We followed a methodology that was laid out in a previous work that employed the MS2 system in the fly to estimate the detection limit of our live imaging experiments (1). Specifically, we calculated the minimum observed fluorescence value for each gene locus in our dataset. We then fit a Gaussian distribution to this set of minimum values to estimate the detection limit for our dataset. As shown in Figure S12, this procedure returned an estimated detection limit of $54 \text{ AU} \pm 1$ which, according to our estimate of the absolute calibration, corresponds to approximately 4 RNAP molecules.

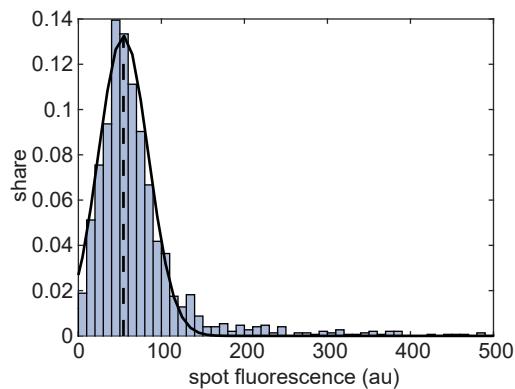


Fig. S12. Estimating the fluorescence detection limit. Distribution of minimum observed fluorescence values for each fluorescence trace in our data set ($N=1484$). The solid black line indicates the Gaussian probability density function that best fits the empirical data. The vertical dashed line indicates the inferred detection limit of $54 \text{ AU} \pm 1$.

Estimating detection threshold artifacts for t_{on} and t_{off} . As illustrated in Figure S13A, the presence of a detection threshold will generally lead to estimated on-times that are too late and estimated off-times that are too early. To gain quantitative estimates for these biases, we used the results of our burst parameter inference shown in Figure 5 to simulate 50 activity trajectories for each region along the anterior-posterior axis that accurately recapitulated observed position-specific burst dynamics. For each simulated trace, we imposed a detection threshold of 54 AU and examined how far the resulting threshold-impacted estimates for t_{on} and t_{off} diverged from the true on and off times. Averaging these effects across 50 traces for each anterior-posterior region indicated that the detection limit did lead to biases. Specifically, we found that it would result in a 30-60 second *overestimate* (too late) of t_{on} (Figure S13B) and a 20-50 second *underestimation* (too early) of t_{off} (Figure S13C). These errors compound for our estimate of the duration of the transcriptional time window, leading to underestimates of between 50 and 100 seconds (Figure S13D). As shown in the figure, while statistically significant, comparisons between raw and threshold-adjusted estimates for these three quantities reveal that these biases are small compared to the quantities of interest, and thus have a minimal effect on the observed trends.

In addition to the estimation of average on and off times as a function of AP position, analyses presented in Figure 6C (fraction of quiescent nuclei over time) and Figure 7A,D, and E (input/output inference for the transcriptional time window time window) of the main text, as well as Figure S32 in this SI, required us to distinguish between transcriptionally engaged and quiescent nuclei. In these cases we employed a simpler definition of t_{off} , taking it to be equal to the last time at which fluorescence was observed for a given locus. According to the results for our t_{off} analysis in Figure S13B, we would expect the impact of this simpler (but more empirical) approach to be minimal, leading to a slight overestimation of t_{off} (1-2 minutes).

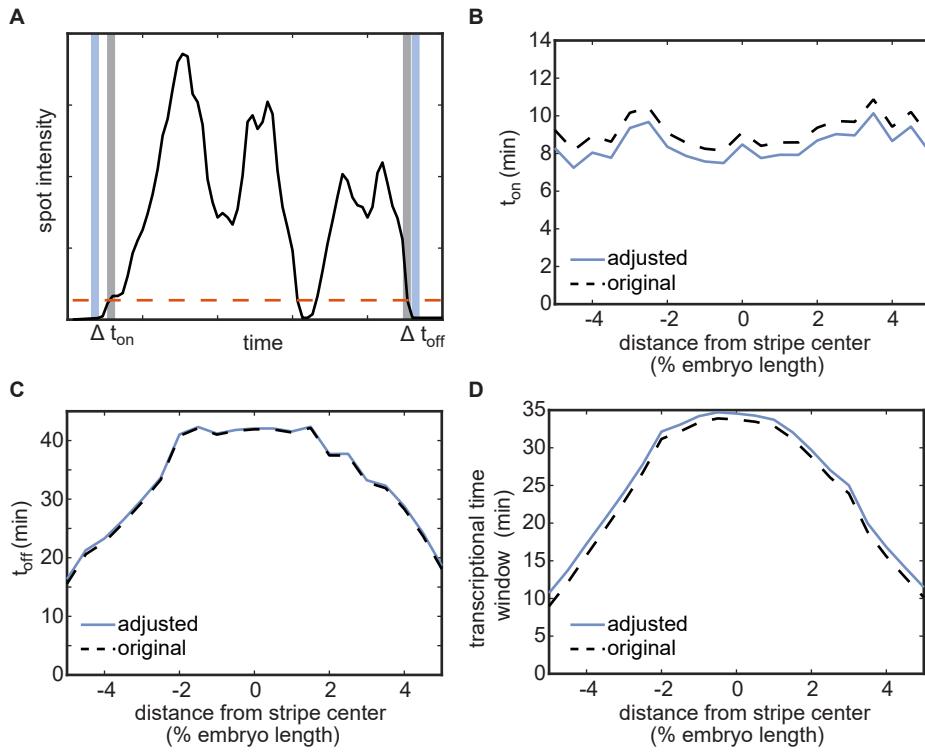


Fig. S13. Accounting for effects of detection limit on off and on time calculations. (A) Example of a simulated fluorescence trace. Red dashed line indicates the estimated detection limit of 54 AU. Blue lines indicate the true on and off times and gray lines indicate the apparent on and off times due to the effects of the detection threshold. (B-D) Plots of raw and adjusted anterior-posterior trends for (B) t_{on} , (C) t_{off} , and (D) the duration of the time window ($t_{off} - t_{on}$).

Possibility of basal expression does not impact conclusions about regulatory strategy contributions. This work invokes two regulatory strategies, the control of the transcriptional time window and of the fraction of active nuclei (Figure 1C,D), that assume that undetected loci are transcriptionally silent; that is, that they do not produce any mRNA. Yet, due to the detection limit of our experiment, we cannot completely rule out the possibility that some undetected loci are actually expressing at a basal level. We thus examined what effects, if any, the existence of such basal expression would have on the conclusions presented in this work.

First, we examined how the predicted contributions of each regulatory control strategy in Figure 3G would change in the presence of basal expression. We first examined what would happen to our predictions if all nuclei that we classified as never having turned on during the whole nuclear cycle actually expressed at a basal level equal to our detection limit of 54 AU for the duration of the nuclear cycle. Figure S14A and B compare the predicted contributions to stripe formation in the absence and presence of this basal transcriptional activity amongst this population. For the analysis in Figure S14B, we maintained the distinction between active and inactive loci, but assumed that all transcriptionally inactive loci (those for which transcription was never detected) were actually expressing at a rate equal to the detection limit throughout nuclear cycle 14. Thus, instead of emitting *no* fluorescence (or, equivalently, producing no mRNA), they were assumed to emit at a constant fluorescence of 54 AU. As a result, the influence of the control of the fraction of active nuclei in pattern formation was reduced, but not eliminated. The size of the effect depends on the relative magnitudes of the detection threshold and the average expression level amongst active gene loci. For our data, the detection limit is approximately 10% of the average expression level. As shown in the figure, this is small enough that the effects of this potential basal expression are predicted to be minuscule. Indeed, it is difficult to distinguish between Figure S14A and B by eye. This is because (1) the fraction of active nuclei contributes negligibly to stripe formation in the first place, and (2) the detection threshold is quite small compared to the average transcription rate amongst transcriptionally engaged loci.

We next examined an even more extreme case wherein *all* undetected loci—both those that were never detected as transcriptionally active and those that shut off early as a part of the control of the transcriptional time window—expressed at the basal rate throughout the nuclear cycle (Figure S14C). As expected, incorporating basal activity at all undetected loci has the effect of decreasing the overall prominence of the stripe pattern, since the expression floor is effectively raised from 0 to 54 AU. However, despite this change, the core conclusion that the analog control of the mean rate and binary control of the transcriptional time window jointly drive pattern formation (Figure 3) remains valid. Thus, we conclude that these results are robust to the possible existence of basal activity amongst quiescent nuclei.

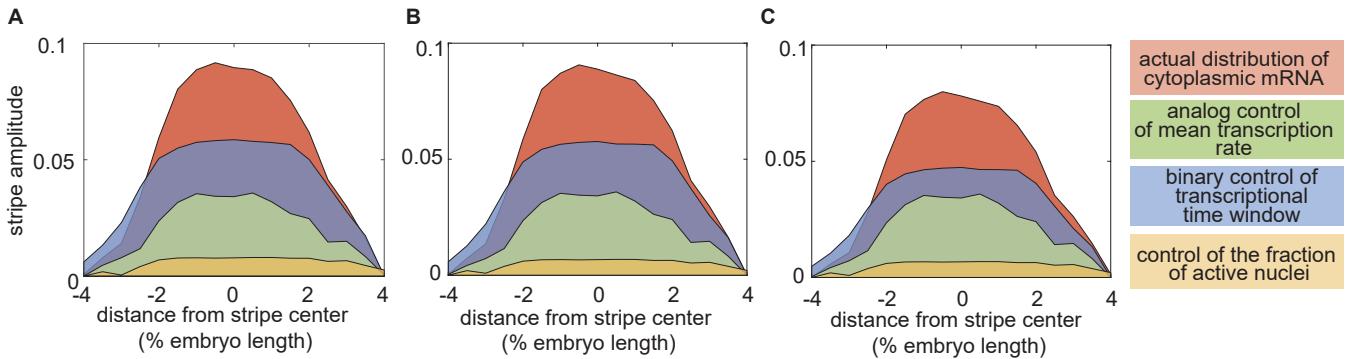


Fig. S14. Impact of possible basal expression on control strategy contributions to stripe formation. Predicted contributions of the regulation of the mean rate (green), time window (blue), and fraction of active nuclei (yellow) to pattern formation. **(A)** Predicted contributions assuming no basal activity. **(B)** Predicted contributions with all nuclei that never turn on assumed to express at detection limit of 54 AU. **(C)** Predicted contributions with both all nuclei that never turn on and all nuclei that become quiescent at some point during nuclear cycle 14 assumed to express at detection limit.

Any undetected basal expression would have radically different burst dynamics. To remain undetected a reporter gene must never have more than four nascent RNAP molecules actively transcribing along its length. In the main text, we invoke the two-state model of bursting to describe the highly stochastic expression patterns exhibited by *eve* gene loci. A fundamental feature of this model is that gene loci transition between an active and an inactive state (see Figure 4). If basally expressing loci also adhered to this model, the burst parameters (k_{on} , k_{off} , and r) would need to be such that the bursts were either small enough or rare enough such that basal loci never contain more than four actively transcribing RNAP molecules at any time point. With this in mind, let us consider the consequences of tuning each burst parameter (see Figure 6A) to satisfy this requirement.

Figure S15A features a simulated fluorescence trace generated using the bursting parameters inferred at the anterior flank of the stripe (anterior-posterior position -4 in Figure 5D and E). How would these bursting dynamics have to change in order to seemingly silent nuclei to be transcribing at a low, basal rate?

First, if the difference between basally expressing and detectable loci stems from a reduction in k_{on} , it would still be the case that bursts, when they occurred, would last about 90 s ($1/k_{off}$ in Figure 5E) and load RNAP molecules at a rate of about $1/3 \text{ s}^{-1}$ (Figure 5D), resulting in the loading of approximately 32 RNAP molecules over the duration of the burst. This number of RNAP molecules on the gene could not escape detection, as it is significantly over our detection limit of four RNAP molecules. Thus, as illustrated in Figure S15B, the only way for loci to avoid detection in this scenario is for k_{on} to be so low that it is unlikely for a single burst to occur within the approximately 40 minute time window of our observation. If we require that the odds of observing a single burst during the 40 minute window of observation be at most 90%, then we have

$$p_{\text{burst}} \leq 0.10 \quad [53]$$

$$1 - e^{-40k_{on}} \leq 0.10 \quad [54]$$

$$-k_{on} \geq \frac{\ln 0.9}{40} \quad [55]$$

$$k_{on} \leq 0.003 \text{ min}^{-1}, \quad [56]$$

Where p_{burst} is the probability of a burst occurring. We see that this requirement results in a k_{on} value that is *two orders of magnitude* lower than what is observed in detected loci (Table S1, low k_{on}). Such a locus would be active a mere 0.4% of the time (compared to around 50% among observed loci). As a result, even if these loci did transcribe with bursting dynamics, these dynamics would be so different from those considered in this work that basally expressing loci would constitute a qualitatively distinct population from detected, high-expressing loci, and not a mere population of active loci that happened to present slightly lower fluorescence values.

Second, if basal expression is realized by increasing k_{off} (decreasing burst duration), the decrease must be sufficiently large that the probability of loading more than 4 RNAP molecules in a burst is low. At a loading rate of 20.5 RNAP per minute, this means that it must be rare for bursts to last for more than $4 \text{ RNAP}/(20.5 \text{ RNAP}/\text{min}) \approx 0.2 \text{ min}$. If we demand that the probability of observing such a burst is no greater than 10%, this yields

$$p_{\text{long}} \leq 0.10 \quad [57]$$

$$e^{-0.2k_{off}} \leq 0.10 \quad [58]$$

$$k_{off} \geq -\frac{\ln 0.1}{0.2} \quad [59]$$

$$k_{off} \geq 11.5 \text{ min}^{-1}. \quad [60]$$

Here, p_{long} denotes the likelihood of a burst longer than 0.2 minutes. The k_{off} value consistent with these constraints is already an order of magnitude larger than the k_{off} inferred for transcriptionally engaged loci. However we have not yet accounted for the

fact that are typically multiple bursts over the course of the nuclear cycle and we require that it be unlikely for *any* single burst to crest the detection limit. This means that the basal k_{off} would need to be larger still. If we assume (conservatively) that there will be approximately 10 bursts per 40 minute time period, then we have

$$(1 - p_{\text{long}})^{10} \geq 0.9 \quad [61]$$

$$e^{-0.2k_{\text{off}}} \leq 1 - 0.9^{0.1} \quad [62]$$

$$k_{\text{off}} \geq -\frac{\ln(0.011)}{0.2} \quad [63]$$

$$k_{\text{off}} \geq 23 \text{ min}^{-1}. \quad [64]$$

Thus we see that this additional requirement implies a k_{off} of at least 23 min^{-1} , ~ 30 -times higher than the values observed for this magnitude throughout this work. As illustrated in [Figure S15C](#), this would lead to a dramatically different kind of activity (compare to [Figure S15A](#)). In such a scenario, bursts would only last 3 seconds on average, with only a single RNAP loaded per burst. Thus, as with k_{on} , we conclude that the dramatic modulation of k_{off} necessary to explain the presence of basally expressing loci would demand that these loci behave qualitatively different from active ones.

Finally, if we assume that basal expression is realized by decreasing the rate of RNAP loading during transcriptional bursts, similar arguments to those considered above for k_{off} indicate that an undetected gene locus must load RNAP molecules at a rate no faster than $0.5/\text{min}$ to remain undetected ([Table S1](#)); a rate that is 40 times higher than that of active loci.

All of the above scenarios strain the bursting model to the breaking point. If k_{on} is modulated in basally expressing loci, then this would imply that basal gene loci are active a mere 0.4% of the time ([Figure S15A](#)). If k_{off} or r are modulated, then “bursts” would consist, on average, of a single polymerase loading event ([Figure S15C](#) and D). Further studies employing single-molecule techniques will be needed to establish the presence or absence of low-level expression. In the meantime, we conclude that while we cannot rule out the existence of basal expression, we *can* confidently state that, if it exists, such activity must be radically different in character from the burst dynamics observed amongst the “transcriptionally engaged” loci we identify in the main text.

Table S1. Comparing burst characteristics of three possible basal expression schemes to “normal” loci expressing above the detection limit.

	normal	low k_{on}	high k_{off}	low r
burst frequency	0.5 min^{-1}	0.003 min^{-1}	0.5 min^{-1}	0.5 min^{-1}
burst duration	1.5 min	1.5 min	0.05 min	1.5 min
initiation rate	$20.5 \text{ RNAP min}^{-1}$	$20.5 \text{ RNAP min}^{-1}$	$20.5 \text{ RNAP min}^{-1}$	$0.5 \text{ RNAP min}^{-1}$
burst size	31.5 RNAP	31.5 RNAP	0.9 RNAP	0.9 RNAP
% of time bursting	41.1 %	0.4%	1.8 %	41.1 %

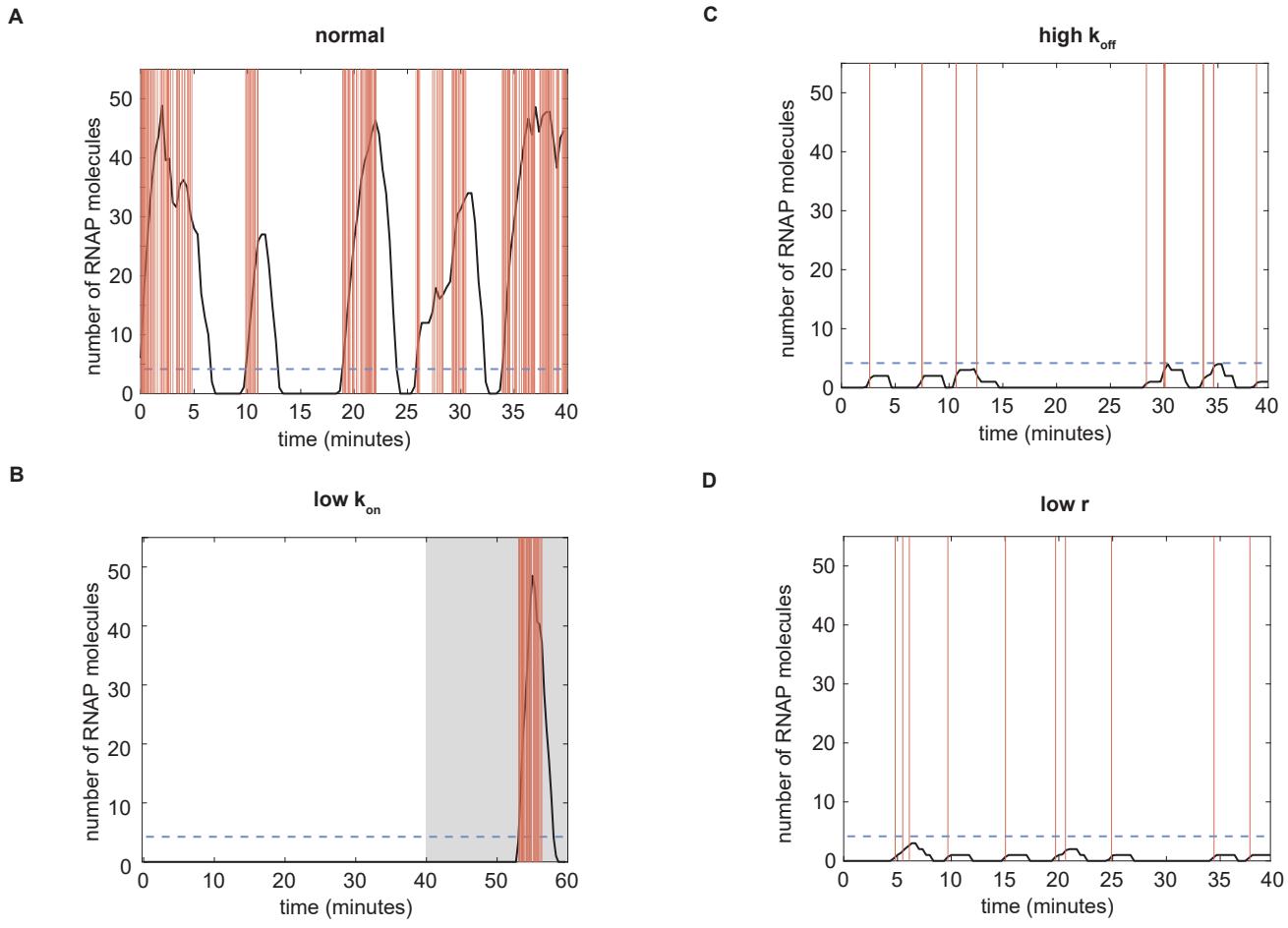


Fig. S15. Possible basal burst dynamics scenarios. Simulated activity traces illustrating bursting dynamics for basally expressing gene loci. Vertical red lines indicate RNAP initiation events. Dashed blue line indicates detection limit. **(A)** illustrative trace generated using burst parameters from anterior stripe flank. **(B-D)** hypothetical basal traces with **(B)** reduced k_{on} (0.003 min^{-1}), **(C)** elevated k_{off} (23.7 min^{-1}), and **(D)** reduced RNAP loading rate ($0.5 \text{ RNAP min}^{-1}$). Gray region in **(B)** indicates the time period coming after 40 minutes of observation.

D. The compound-state hidden Markov model.

Model introduction. To model the dynamics of an observed fluorescence series, $\mathbf{y} = \{y_1, y_2, \dots, y_T\}$, where T is the number of data points in a trace, we assume that, at each time step, the sister promoters can be in one of K effective states. In the analysis of *eve* stripe 2 data, we use a simple model with the number of effective states equal to three ($K = 3$). The method, however, allows for more complex transcription architectures with higher numbers of states. Transitions between the effective promoter states are assumed to be Markovian, meaning that the hidden promoter state z_t at time step t is conditionally dependent only on the state in the previous time step. This dependency is modeled through a $K \times K$ transition probability matrix $\mathbf{A} = p(z_t|z_{t-1})$, where A_{kl} is the probability of transitioning from the l^{th} state into the k^{th} state in the time interval $\Delta\tau$, where $\Delta\tau$ is the data sampling resolution. We assign a characteristic RNAP initiation rate, $r(k)$, with units of RNAP per minute, to each effective promoter state, $z(k)$, $1 \leq k \leq K$. Thus, the number of polymerases initiated between time steps $t - 1$ and t will be $r(z_t)\Delta\tau$. Because the fluorescence intensity contributed by each polymerase depends on the number of transcribed MS2 stem loops, the contribution will vary with the position of the polymerase on the gene. In our transcription model we assume that polymerase elongation takes place at a constant rate. Therefore, if τ_{MS2} is the time it takes to transcribe the MS2 loops, the fluorescence contribution of an RNAP molecule will initially grow linearly ($\tau \leq \tau_{\text{MS2}}$) and will then stay constant for the remainder of transcription ($\tau_{\text{MS2}} \leq \tau \leq \tau_{\text{elong}}$). Given this time dependence, we define a maximum fluorescence emission per time step for each state as $v(k) = F_{\text{RNAP}}r(k)\Delta\tau$, $1 \leq k \leq K$, where F_{RNAP} is the fluorescence calibration factor determined using smFISH experiments (see Materials and Methods).

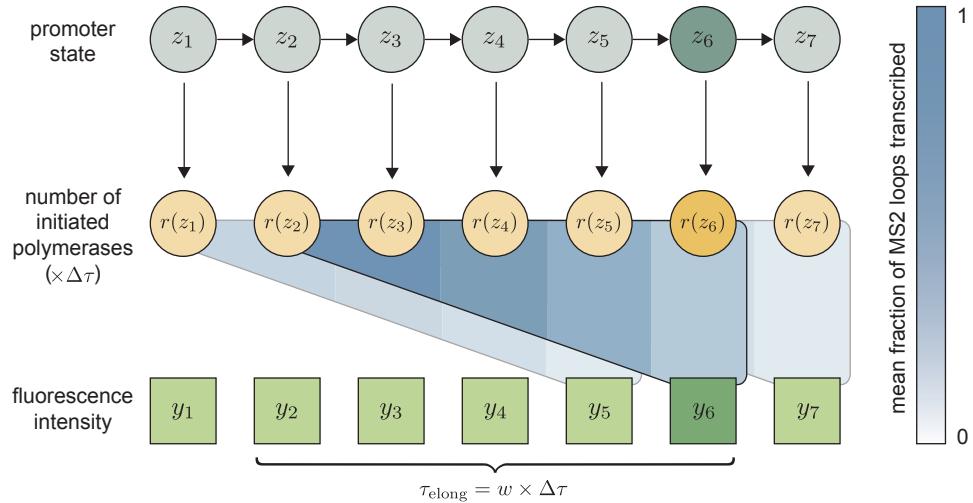


Fig. S16. Schematic overview of the cpHMM architecture. The sister promoters are modeled as undergoing a series of Markovian transitions between effective transcriptional states (z_t). Each promoter state uniquely determines the number of polymerases initiated in a single time step ($r(z_t)\Delta\tau$). Fluorescence emissions from polymerases initiated in the most recent w steps combine to produce the observed fluorescence intensity (y_t). The color bar indicates the mean fraction of MS2 loops that have been transcribed and contribute fluorescence at the moment of observation. The color corresponding to the more recently initiated polymerases is therefore lighter (fewer loops transcribed) than that corresponding to polymerases initiated at earlier times (more loops transcribed).

The instantaneous fluorescence intensity is the cumulative contribution from polymerases initiated in the previous w time steps, where $w = \tau_{\text{elong}}/\Delta\tau$ is the system-dependent integer memory. Here $\Delta\tau$ indicates the observational time resolution, a quantity set by experimental conditions. The time required for an RNAP molecule to transcribe our reporter gene (τ_{elong}) is *a priori* unknown. We developed an autocorrelation-based method to estimate τ_{elong} directly from our experimental data (see *SI Appendix*, section J and (13)). The observation y_t at time step t conditionally depends not only on the hidden promoter state z_t , but also on the hidden states in the previous w time steps, $\{z_t, z_{t-1}, \dots, z_{t-w+1}\}$. To be able to describe the observed system dynamics through a hidden Markov model, the observation at time step t needs to be conditionally independent from the states at earlier time steps. We therefore introduce the concept of a compound state, $s_t = \{z_t, z_{t-1}, \dots, z_{t-w+1}\}$, which, together with the set of model parameters, θ , is sufficient to define the probability distribution of the observation y_t , thereby satisfying the Markov condition. Since $z_t \in \{1, \dots, K\}$, each compound state can take one of K^w different values, $s_t \in \{1, \dots, K^w\}$. While the number of possible compound states is K^w , only K different transitions are allowed between them, since the most recent $w-1$ promoter states are deterministically passed from one compound state to the next, i.e. the last $w-1$ elements in $s_{t+1} = \{z_{t+1}, z_t, \dots, z_{t-w+2}\}$ are present in s_t as well. The schematic overview of the cpHMM architecture is shown in [Figure S16](#).

We model the fluorescence emission probabilities corresponding to each hidden compound state as Gaussian distributions with a standard deviation σ , which we learn during inference. The joint probability distribution $p(\mathbf{y}, \mathbf{s}|\theta)$ of the series of hidden compound states, $\mathbf{s} = \{s_1, s_2, \dots, s_T\}$, and fluorescence values, $\mathbf{y} = \{y_1, y_2, \dots, y_T\}$, is given by

$$p(\mathbf{y}, \mathbf{s}|\theta) = p(s_1|\pi) \prod_{t=1}^T p(y_t|s_t, \mathbf{v}, \sigma) \prod_{t=2}^T p(s_t|s_{t-1}, \mathbf{A}). \quad [65]$$

Here π is a K -element vector, with π_k being the probability that the trace starts at the k^{th} effective promoter state, and \mathbf{v} is a K -element vector of fluorescence emission values per time step.

Our goal is to find an estimate of the model parameters, $\hat{\theta} = \{\hat{\pi}, \hat{\mathbf{v}}, \hat{\mathbf{A}}, \hat{\sigma}\}$, which maximizes the likelihood $p(\mathbf{y}|\theta)$ of observing the fluorescence data, namely,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} p(\mathbf{y}|\theta). \quad [66]$$

The likelihood can be obtained by marginalizing the joint probability distribution, $p(\mathbf{y}, \mathbf{s}|\theta)$, over the hidden compound states, that is,

$$p(\mathbf{y}|\theta) = \sum_{\mathbf{s}=\{s_1, s_2, \dots, s_T\}} p(\mathbf{y}, \mathbf{s}|\theta). \quad [67]$$

Note that the summation is performed over all possible choices of \mathbf{s} — a vector of T elements, each of which can take K^w possible values. The total number of terms in the sum is thus equal to K^{wT} , which grows exponentially with the number of time points. To make the estimation of the model parameters tractable, we use an approximate inference method, the expectation-maximization (EM) algorithm.

We note that the notion of a compound state was also introduced in an earlier work (14) to account for the memory effect in hidden Markov modeling of *actin* transcription and then an EM methodology was applied to learn the kinetic parameters from MS2-based transcription data. Unlike their approach, however, we do not explicitly model the recruitment of individual RNAP molecules, but instead, assign a continuous RNAP initiation rate to each promoter state. Additionally, our model estimates the magnitude of the background noise present in the experimentally measured fluorescence signal, whereas the model presented in (14) takes this quantity as an input, requiring that it be estimated separately. We believe that these differences serve to make our model more flexible. Moreover, by eliminating the need for absolute calibration and noise estimation, we hoped to facilitate the use of our model in a wide variety of experimental contexts, for which one or the other quantity may not be readily obtainable. In the "Continuous vs. Poisson promoter loading" section of *SI Appendix*, section E we demonstrate that relaxing the continuous RNAP loading assumption when generating synthetic data does not significantly affect the accuracy of the cpHMM inference.

Expectation-maximization (EM) algorithm. Consistent with standard EM approaches (cf. Bishop (15), Chapter 13), at each iteration we maximize the lower bound of the logarithm of the likelihood using the current estimate of the model parameters, namely,

$$\hat{\boldsymbol{\theta}}_{k+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \mathcal{L}(\boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k), \quad [68]$$

$$\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{\mathbf{s}=\{s_1, s_2, \dots, s_T\}} p(\mathbf{s} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) \log p(\mathbf{y}, \mathbf{s} | \boldsymbol{\theta}) \leq \log p(\mathbf{y} | \boldsymbol{\theta}). \quad [69]$$

Here $\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k)$ is the objective function, $\hat{\boldsymbol{\theta}}_k$ is the estimate of the model parameters in the k^{th} expectation step of the EM algorithm. Since we model the transitions between the effective sister promoter states as a Markov process, the logarithm of the joint probability distribution, $\log p(\mathbf{y}, \mathbf{s} | \boldsymbol{\theta})$, can be written as

$$\log p(\mathbf{y}, \mathbf{s} | \boldsymbol{\theta}) = \log p(s_1 | \boldsymbol{\pi}) + \sum_{t=1}^T \log p(y_t | s_t, \mathbf{v}, \sigma) + \sum_{t=2}^T \log p(s_t | s_{t-1}, \mathbf{A}). \quad [70]$$

Now, we introduce several notations: $s_t^i := 1$ if and only if $s_t = i$; $\Delta(s_t, d) :=$ the d^{th} digit of the promoter state sequence $s_t = \{z_t, z_{t-1}, \dots, z_{t-(w-1)}\}$, starting from the left end; $C_{zs} = 1$ if and only if $\Delta(s, 1) = z$; $B_{s',s} = 1$ if and only if the transition $s \rightarrow s'$ between the compound states s and s' is allowed, which happens when the latest $(w-1)$ promoter states in the compound state s match the earliest $(w-1)$ promoter states of the compound state s' . With these notations in hand, the terms in Equation 70 can be rewritten as

$$\log p(s_1 | \boldsymbol{\pi}) = \sum_{i=1}^{K^w} \sum_{k=1}^K s_1^i C_{ki} \log \pi_k, \quad [71]$$

$$\log p(y_t | s_t, \mathbf{v}, \sigma) = \frac{1}{2} \sum_{i=1}^{K^w} s_t^i (\log \lambda - \log(2\pi) - \lambda(y_t - V_i(\mathbf{v}))^2), \quad [72]$$

$$\log p(s_t | s_{t-1}, \mathbf{A}) = \sum_{i,j=1}^{K^w} \sum_{k,l=1}^K B_{ij} s_t^i s_{t-1}^j C_{ki} C_{lj} \log A_{kl}. \quad [73]$$

Here $\lambda = 1/\sigma^2$ is the Gaussian precision parameter, and $V_i(\mathbf{v})$ is the aggregate fluorescence produced in the w consecutive promoter states of the i^{th} compound state.

Because of the finite time τ_{MS2} it takes a single polymerase to transcribe the MS2 sequence, the fluorescence contribution of polymerases is weighted at different positions in the window of w time steps. If we define $n_{\text{MS2}} = \tau_{\text{MS2}}/\Delta\tau$ as the number of time steps (not necessarily an integer) necessary for transcribing the MS2 sequence, the mean fraction of the full MS2 sequence transcribed by a polymerase at the d^{th} time step of the elongation window will be given by

$$\kappa(d) = \begin{cases} 1, & \text{if } \lceil n_{\text{MS2}} \rceil < d \leq w \\ d - n_{\text{MS2}} + \frac{n_{\text{MS2}}^2 - (d-1)^2}{2n_{\text{MS2}}}, & \text{if } \lfloor n_{\text{MS2}} \rfloor < d \leq \lceil n_{\text{MS2}} \rceil \\ \frac{d-1/2}{n_{\text{MS2}}}, & \text{if } 1 \leq d \leq \lfloor n_{\text{MS2}} \rfloor \end{cases}$$

where $\lceil n_{\text{MS2}} \rceil$ and $\lfloor n_{\text{MS2}} \rfloor$ are the ceiling and the floor of n_{MS2} , respectively. The dependence of the weighting function $\kappa(d)$ on the position for a specific choice of parameters is illustrated in Figure S17.

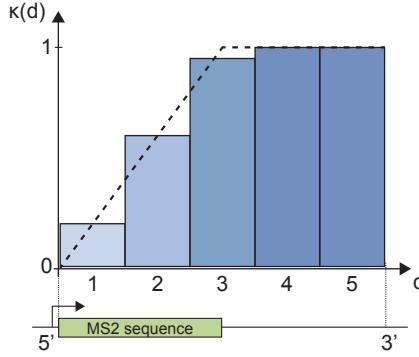


Fig. S17. The weighting function $\kappa(d)$ evaluated at different positions along the genome. The dashed line represents the fraction of the MS2 loops transcribed at a given position. Parameters used for plotting: $\tau_{\text{elong}} = 100 \text{ sec}$, $\tau_{\text{MS2}} = 50 \text{ sec}$, $\Delta\tau = 20 \text{ sec}$, $w = \tau_{\text{elong}}/\Delta\tau = 5$, $n_{\text{MS2}} = \tau_{\text{MS2}}/\Delta\tau = 2.5$.

Accounting for the weighted fluorescence contribution of polymerases, the aggregate fluorescence $V_i(\mathbf{v})$ becomes

$$V_i(\mathbf{v}) = F_{i,:}\mathbf{v}, \quad [74]$$

where the i^{th} row of the $K^w \times K$ matrix \mathbf{F} is the number of times each promoter state is present in the i^{th} compound state, weighted by the position-dependent function $\kappa(d)$. For example, if we consider a promoter with $K = 3$ states and memory $w = 5$, then the row of \mathbf{F} corresponding to the compound state $s = \{1, 1, 3, 2, 3\}$ will be $[\kappa(1) + \kappa(2), \kappa(4), \kappa(3) + \kappa(5)]$.

Having all the pieces of the logarithm of the joint probability distribution, $\log p(\mathbf{y}, \mathbf{s}|\boldsymbol{\theta})$, we obtain a final expression for the objective function, namely,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) &= \sum_{i=1}^{K^w} \sum_{k=1}^K \langle s_1^i \rangle C_{ki} \log \pi_k \\ &+ \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^{K^w} \langle s_t^i \rangle (\log \lambda - \log(2\pi) - \lambda(y_t - F_{i,:}\mathbf{v})^2) \\ &+ \sum_{t=1}^T \sum_{i,j=1}^{K^w} \sum_{k,l=1}^K B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{lj} \log A_{kl}. \end{aligned} \quad [75]$$

Here $\langle s_t^i \rangle$ and $\langle s_t^i s_{t-1}^j \rangle$ are the expectation coefficients at the k^{th} step of the EM algorithm defined as

$$\langle s_t^i \rangle = \sum_{s=\{s_1, s_2, \dots, s_T\}} s_t^i p(s|\mathbf{y}, \hat{\boldsymbol{\theta}}_k), \quad [76]$$

$$\langle s_t^i s_{t-1}^j \rangle = \sum_{s=\{s_1, s_2, \dots, s_T\}} s_t^i s_{t-1}^j p(s|\mathbf{y}, \hat{\boldsymbol{\theta}}_k). \quad [77]$$

Using the current estimate of the model parameters, $\hat{\boldsymbol{\theta}}_k$, the expectation coefficients $\langle s_t^i \rangle$ and $\langle s_t^i s_{t-1}^j \rangle$ are calculated using the forward-backward algorithm. From the definitions in Equation 76 and Equation 77, we obtain

$$\langle s_t^i \rangle = \sum_{s_1, s_2, \dots, s_T} s_t^i p(s_1, s_2, \dots, s_T | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t} s_t^i p(s_t | \mathbf{y}, \hat{\boldsymbol{\theta}}_k), \quad [78]$$

$$\langle s_t^i s_{t-1}^j \rangle = \sum_{s_1, s_2, \dots, s_T} s_t^i s_{t-1}^j p(s_1, s_2, \dots, s_T | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t, s_{t-1}} s_t^i s_{t-1}^j p(s_t, s_{t-1} | \mathbf{y}, \hat{\boldsymbol{\theta}}_k). \quad [79]$$

Following the conventional implementation of the forward-backward algorithm (cf. Bishop (15), Chapter 13), we use the Markov property of the promoter state dynamics, together with the sum and products rules of probability, to write

$$p(s_t | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) = \frac{\alpha_t(s_t) \beta_t(s_t)}{p(\mathbf{y} | \hat{\boldsymbol{\theta}}_k)}, \quad [80]$$

$$p(s_{t-1}, s_t | \mathbf{y}, \hat{\boldsymbol{\theta}}_k) = \frac{\alpha_{t-1}(s_{t-1}) p(y_t | s_t, \hat{\boldsymbol{\theta}}_k) p(s_t | s_{t-1}, \hat{\boldsymbol{\theta}}_k) \beta_t(s_t)}{p(\mathbf{y} | \hat{\boldsymbol{\theta}}_k)}, \quad [81]$$

$$\alpha_t(i) = p(y_1, \dots, y_t, s_t = i | \hat{\boldsymbol{\theta}}_k), \quad [82]$$

$$\beta_t(i) = p(y_{t+1}, \dots, y_T | s_t = i, \hat{\boldsymbol{\theta}}_k). \quad [83]$$

Here $\alpha_t(i)$ is the joint probability of observing the fluorescence emission values in the first t steps and being at the i^{th} compound state at step t ; while $\beta_t(i)$ is the conditional probability of observing fluorescence values from the time point $(t+1)$ till the end of the series, given that the compound state at time t is i . Note that α and β can be treated as $K^w \times T$ matrices, where each column is a vector of length K^w , accounting for the K^w possible values of i in [Equation 82](#) and [Equation 83](#). We evaluate the elements of α and β matrices recursively as

$$\alpha_t(i) = p(y_t | s_t = i, \hat{\theta}_k) \sum_{j=1}^{K^w} \alpha_{t-1}(j) p(s_t = i | s_{t-1} = j, \hat{\theta}_k), \quad [84]$$

$$\beta_t(i) = \sum_{j=1}^{K^w} \beta_{t+1}(j) p(y_{t+1} | s_{t+1} = j, \hat{\theta}_k) p(s_{t+1} = j | s_t = i, \hat{\theta}_k). \quad [85]$$

The boundary values for $\alpha_1(i)$ and $\beta_T(i)$ at the first and last columns of α and β matrices, respectively, are given by

$$\alpha_1(i) = p(y_1 | s_1 = i, \hat{\theta}_k) p(s_1 = i | \hat{\theta}_k), \quad [86]$$

$$\beta_T(i) = 1, \quad [87]$$

where the first follows from the definition of $\alpha_t(i)$, and the second is obtained from [Equation 80](#) by setting $t = T$. Having evaluated the α and β matrices, the likelihood $p(\mathbf{y} | \hat{\theta}_k)$ that appears in the denominator of [Equation 80](#) and [Equation 81](#) can be found by setting $t = T$ in [Equation 80](#) and summing over s_T , namely,

$$\left(\sum_{s_T=1}^{K^w} p(s_T | \mathbf{y}, \hat{\theta}_k) \right) p(\mathbf{y} | \hat{\theta}_k) \equiv p(\mathbf{y} | \hat{\theta}_k) = \sum_{s_T=1}^{K^w} \alpha_T(s_T). \quad [88]$$

With the probabilities $p(s_t | \mathbf{y}, \hat{\theta}_k)$ and $p(s_{t-1}, s_t | \mathbf{y}, \hat{\theta}_k)$ known, the expectation coefficients follow directly from [Equation 78](#) and [Equation 79](#).

The optimal model parameters in the $(k+1)^{\text{th}}$ step of the EM algorithm are obtained by maximizing the objective function $\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}, \hat{\theta}_k)$ in [Equation 75](#) with respect to $\{\boldsymbol{\pi}, \mathbf{v}, \lambda, \mathbf{A}\}$, subject to the probability constraints $\sum_{k=1}^K \pi_k = 1$ and $\sum_{k=1}^K A_{kl} = 1$, $1 \leq l \leq K$. The update equations for the model parameters are found as

$$\text{initial state probabilities: } \hat{\pi}_m = \frac{\sum_{i=1}^{K^w} \langle s_1^i \rangle C_{mi}}{\sum_{k=1}^K \sum_{i=1}^{K^w} \langle s_1^i \rangle C_{ki}}, \quad [89]$$

$$\text{fluorescence emission rates: } \hat{\mathbf{v}} = \mathbf{M}^{-1} \mathbf{b}, \text{ where} \quad [90]$$

$$M_{mn} = \sum_{t=1}^T \sum_{i=1}^{K^w} \langle s_t^i \rangle F_{in} F_{im}, \quad [91]$$

$$b_m = \sum_{t=1}^T \sum_{i=1}^{K^w} \langle s_t^i \rangle y_t F_{im}, \quad [92]$$

$$\text{noise: } \frac{1}{\hat{\lambda}} = \hat{\sigma}^2 = \frac{\sum_{t=1}^T \sum_{i=1}^{K^w} \langle s_t^i \rangle (y_t - F_{i,:} \hat{\mathbf{v}})^2}{\sum_{t=1}^T \sum_{i=1}^{K^w} \langle s_t^i \rangle}, \quad [93]$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{t=1}^T \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{mi} C_{nj}}{\sum_{k=1}^K \sum_{t=1}^T \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{nj}}. \quad [94]$$

We note that in the update steps we impose no constraints on the inferred emission rates for the generality of treatment and therefore, expect the effective OFF state to have a nonzero but small inferred emission rate compared with that of the ON states.

Pooled inference on multiple traces. Since the information available in a single MS2 fluorescence trace is not sufficient for the accurate inference of underlying model parameters, we perform a pooled EM inference assuming that the traces are statistically independent and governed by the same parameters. If $\mathbf{y}_{1:N}$ are N different fluorescence traces with corresponding trace lengths $T_{1:N}$, and $\mathbf{s}_{1:N}$ are the hidden compound state sequences corresponding to each trace, we obtain

$$p(\mathbf{y}_{1:N}, \mathbf{s}_{1:N} | \boldsymbol{\theta}) = \prod_{n=1}^N p(\mathbf{y}_n, \mathbf{s}_n | \boldsymbol{\theta}), \quad [95]$$

$$p(\mathbf{s}_n | \mathbf{y}_{1:N}, \hat{\theta}_k) = p(\mathbf{s}_n | \mathbf{y}_n, \hat{\theta}_k), \quad 1 \leq n \leq N. \quad [96]$$

Therefore, the objective function $\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k)$ maximized at each EM iterations takes the form

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) &= \sum_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N} p(\mathbf{s}_{1:N} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\mathbf{y}_{1:N}, \mathbf{s}_{1:N} | \boldsymbol{\theta}) \\
&= \sum_{n=1}^N \sum_{\mathbf{s}_n} p(\mathbf{s}_n | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\mathbf{y}_n, \mathbf{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^N \sum_{\mathbf{s}_n} p(\mathbf{s}_n | \mathbf{y}_n, \hat{\boldsymbol{\theta}}_k) \log p(\mathbf{y}_n, \mathbf{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^N \mathcal{L}_n(\boldsymbol{\theta} | \mathbf{y}_n, \hat{\boldsymbol{\theta}}_k).
\end{aligned} \tag{97}$$

From the above equation, we recognize that the objective function for the pooled inference is the sum of objective functions written for each individual trace. Using the expression for the single-trace objective function obtained earlier (Equation 75), we find

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) &= \sum_{n=1}^N \sum_{i=1}^{K^w} \sum_{k=1}^K \langle s_1^i(n) \rangle C_{ki} \log \pi_k \\
&\quad + \frac{1}{2} \sum_{n=1}^N \sum_{t=1}^{T_n} \sum_{i=1}^{K^w} \langle s_t^i(n) \rangle (\log \lambda - \log(2\pi) - \lambda(y_t(n) - F_{i,:} \mathbf{v})^2) \\
&\quad + \sum_{n=1}^N \sum_{t=1}^{T_n} \sum_{i,j=1}^{K^w} \sum_{k,l=1}^K B_{ij} \langle s_t^i(n) s_{t-1}^j(n) \rangle C_{ki} C_{lj} \log A_{kl},
\end{aligned} \tag{98}$$

where $\langle s_t^i(n) \rangle$ and $\langle s_t^i(n) s_{t-1}^j(n) \rangle$ are now the expectation coefficients obtained for the n^{th} fluorescence trace via the forward-backward algorithm, and $y_t(n)$ is the fluorescence at time step t in the n^{th} trace. The update equations are then derived analogous to the single-trace case, with an additional summation performed over all traces, namely,

$$\text{initial state probabilities: } \hat{\pi}_m = \frac{\sum_{h=1}^N \sum_{i=1}^{K^w} \langle s_1^i(h) \rangle C_{mi}}{\sum_{k=1}^K \sum_{h=1}^N \sum_{i=1}^{K^w} \langle s_1^i(h) \rangle C_{ki}}, \tag{99}$$

$$\text{fluorescence emission rates: } \hat{\mathbf{v}} = \mathbf{M}^{-1} \mathbf{b}, \quad \text{where} \tag{100}$$

$$M_{mn} = \sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i=1}^{K^w} \langle s_t^i(h) \rangle F_{in} F_{im}, \tag{101}$$

$$b_m = \sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i=1}^{K^w} \langle s_t^i(h) \rangle y_t(h) F_{im}, \tag{102}$$

$$\text{noise: } \frac{1}{\lambda} = \hat{\sigma}^2 = \frac{\sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i=1}^{K^w} \langle s_t^i(h) \rangle (y_t(h) - F_{i,:} \hat{\mathbf{v}})^2}{\sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i=1}^{K^w} \langle s_t^i(h) \rangle}, \tag{103}$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i(h) s_{t-1}^j(h) \rangle C_{mi} C_{nj}}{\sum_{k=1}^K \sum_{h=1}^N \sum_{t=1}^{T_h} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i(h) s_{t-1}^j(h) \rangle C_{ki} C_{nj}}. \tag{104}$$

Execution of the cPHMM method. Execution of the cPHMM method starts by initializing the model parameters. $\boldsymbol{\pi}$ and each column of \mathbf{A} , both of which are vectors of size K , are initialized by randomly sampling from a Dirichlet distribution given by

$$f(\mathbf{x}) \sim \frac{\Gamma(\sum_{k=1}^K u_k)}{\prod_{k=1}^K \Gamma(u_k)} \prod_{k=1}^K x_k^{u_k-1}. \tag{105}$$

The Dirichlet distribution parameters u_k are all set equal to one, which makes each initial promoter state equally likely to be occupied, and equally likely to be transitioned into.

To initialize the fluorescence emission rates, \mathbf{r} , and the Gaussian precision parameter, $\lambda = 1/\sigma^2$, we first treat the fluorescence data $\mathbf{y}_{1:N}$ as identical and independently distributed (i.i.d.) and use a simplified time-independent EM algorithm to find their optimal values (cf. Bishop (15), Chapter 13). We initialize the highest emission rate by randomly choosing a value between 70% and 130% of the highest emission rate inferred by the i.i.d. approach. The lowest emission rate is initialized to 0 because of the apparent silent periods in the activity traces. The remaining ($K - 2$) emission rates are initialized by choosing random

values between 0 and the highest emission rate. Finally, we initialize the Gaussian noise σ by randomly choosing a value between 50% and 200% of the noise inferred by the i.i.d. approach.

After initializing the model parameters, we iterate between the expectation and maximization steps of the EM algorithm until the relative changes in the Euclidean norms of the model parameters after consecutive iterations become smaller than $\varepsilon = 10^{-4}$ or the number of iterations exceeds 500. Because EM approaches typically infer locally optimal parameter values, the algorithm is run on the same dataset using multiple randomly chosen initial parameters (10-20 in our implementations), and the globally optimal set of values is chosen in the end. In the Matlab implementation of the EM algorithm, the variables are all stored in logarithmic forms to avoid overflow and underflow issues, which could occur when recursively evaluating the elements of the α and β matrices. Also, special care is taken when accounting for time points less than the elongation time, i.e. $t < w$, in which case the compound state is a collection of not w , but t promoter states, i.e. $s_t = \{z_t, z_{t-1}, \dots, z_1\}$.

Because of the exponential scaling of the model complexity with the integer memory window ($w = 7$ for the *eve* construct with $\Delta\tau = 20$ sec data sampling resolution), significant computational resources were used when conducting inference on simulated and experimental data. It took approximately 2 hours to conduct 25 cpHMM inferences with different initialization conditions on a machine with 24 CPU cores. Users of the cpHMM method are advised to have this metric as a reference when estimating the computational cost of their inference.

Windowed cpHMM. To investigate temporal trends in bursting parameters, we extended the cpHMM method to allow for a sliding window inference approach. From a technical perspective, this required a revision of the inference formalism to be compatible with fragments of fluorescent traces in which the beginning of the trace (initial rise in y_t from $t = 1$) was not included.

To that end, we modified the first term in [Equation 70](#) to allow for all possible promoter state sequences that could lead to the observation of the first fluorescence measurement in the chosen time window ($[T_1, T_2]$), namely,

$$\log p(\mathbf{y}_{T_1:T_2}, s_{T_1:T_2} | \boldsymbol{\theta}) = \log p(s_{T_1} | \boldsymbol{\pi}^{(T_1-w+1)}, \mathbf{A}) + \sum_{t=T_1}^{T_2} \log p(y_t | s_t, \mathbf{r}, \sigma) + \sum_{t=T_1}^{T_2} \log p(s_t | s_{t-1}, \mathbf{A}), \quad [106]$$

$$\begin{aligned} \log p(s_{T_1} | \boldsymbol{\pi}^{(T_1-w+1)}, \mathbf{A}) &= \log \left(p(z_{T_1-w+1} | \boldsymbol{\pi}^{(T_1-w+1)}) \prod_{t=T_1-w+2}^{T_1} p(z_t | z_{t-1}, \mathbf{A}) \right) \\ &= \sum_{i=1}^{K^w} \sum_{n=1}^K s_{T_1}^i D_{ni}^w \log \pi_n^{(T_1-w+1)} + \sum_{i=1}^{K^w} \sum_{d=2}^w \sum_{k,l=1}^K s_{T_1}^i D_{ki}^{d-1} D_{li}^d \log A_{kl}. \end{aligned} \quad [107]$$

Here $\boldsymbol{\pi}^{(T_1-w+1)}$ is the probability distribution of the earliest promoter state that still has an impact on the observation of the first measurement in the sliding window, and D_{ni}^d is an indicator variable which takes the value 1 only if the promoter state in the d^{th} position of the i^{th} compound state is n .

The modified expression for the joint probability distribution does not change the functional form of the equations used for calculating the expectation coefficients. Maximization equations for the emission rates and the noise also remain intact. Only the maximization equation for the transition probabilities is revised from [Equation 94](#) into

$$\hat{A}_{mn} = \frac{\sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{mi} C_{nj} + \sum_{i=1}^{K^w} \sum_{d=2}^w \langle s_{T_1}^i \rangle D_{mi}^{d-1} D_{ni}^d \log A_{mn}}{\sum_{k=1}^K \sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{nj} + \sum_{k=1}^K \sum_{i=1}^{K^w} \sum_{d=2}^w \langle s_{T_1}^i \rangle D_{ki}^{d-1} D_{ni}^d \log A_{kn}}. \quad [108]$$

We make a steady-state assumption within the sliding window and choose $\boldsymbol{\pi}^{(T_1-w+1)}$ to be the stationary distribution of the current transition probability matrix, i.e. $\mathbf{A} \boldsymbol{\pi}^{(T_1-w+1)} = \boldsymbol{\pi}^{(T_1-w+1)}$. We therefore use the current estimate of \mathbf{A} to evaluate $\boldsymbol{\pi}^{(T_1-w+1)}$ at each EM iteration, instead of performing a maximization step.

E. Statistical validation of cpHMM. We validated cpHMM for the three-state ($K = 3$) architecture schematically illustrated in [Figure S18A](#) by generating synthetic trajectories of effective promoter states using the Gillespie algorithm [\(7\)](#) and adding Gaussian noise to the resulting activity traces. Parameters in [Table S2](#) were used for data generation. Pooled inferences were conducted on 20 independent datasets, each containing 9,000 data points, representative of the number of experimental data points in a central stripe region. The top panel of [Figure S18B](#) shows the kinetic architecture used to simulate the promoter trajectory in [Figure S18C](#) (yellow) as it switches through the multiple possible states. This promoter trajectory leads to the simulated trace of [Figure S18D](#) (red). Using cpHMM, we found the best fitted path for our observable ([Figure S18D](#), black) and the corresponding most likely promoter state trajectory ([Figure S18C](#), blue).

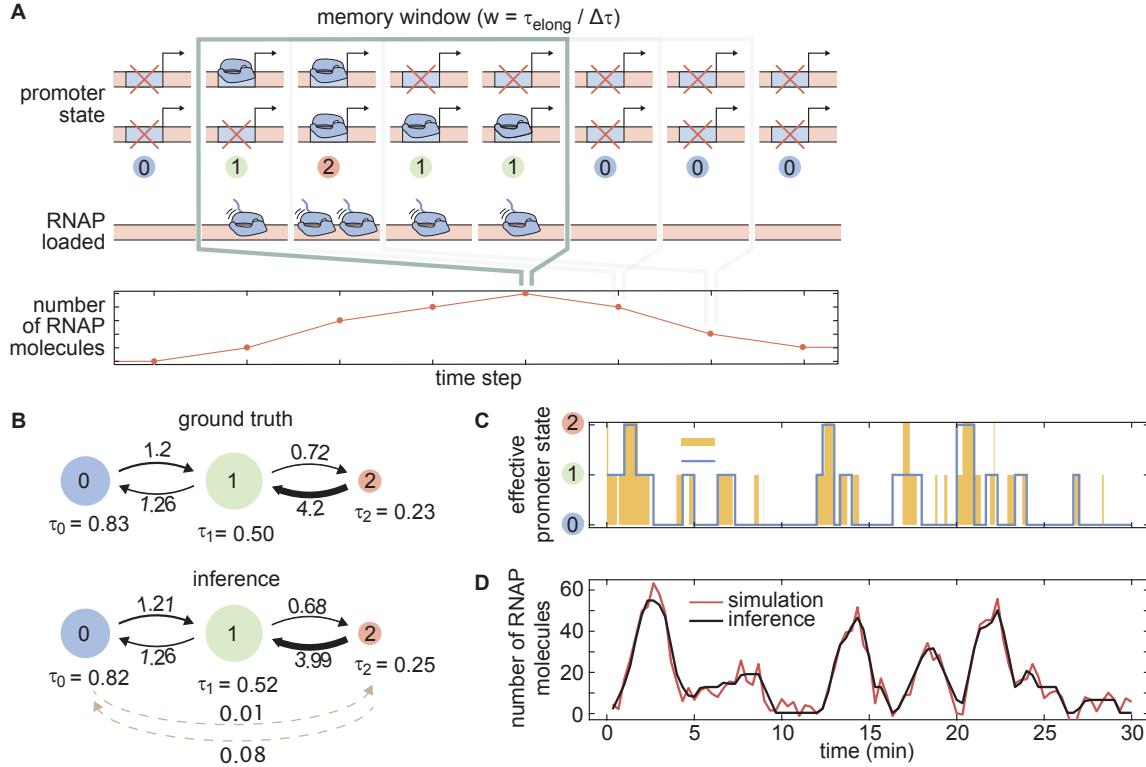


Fig. S18. Statistical validation of cpHMM. (A) Three-state cpHMM architecture where ON and OFF promoter states on each sister chromatid result in an effective three-state model. The trajectory of effective promoter states over the memory time window given by the elongation time dictates the number of RNAp molecules loaded onto the gene. (B) Flow diagrams of promoter states and transition rates for the true parameters used to simulate trajectories (top) and corresponding average inference results obtained from 20 independent datasets (bottom). The area of each state circle is proportional to the relative state occupancy, and the thickness of the arrows is proportional to the transition rates. Dashed lines correspond to inferred transitions with very slow rates that were absent in the simulation. Rates are in min^{-1} and dwell times are in min. Error bars for the mean inferred parameters are shown in Figure S19. (C) Sample simulated promoter activity trace (yellow) generated using the parameters in (B), overlaid with the best fitted trace (blue) obtained using the Viterbi algorithm (16). (D) Simulated and best fitted observable number of RNAp molecules corresponding to the promoter trajectory shown in (C).

Table S2. Parameter values used for generating synthetic datasets in the statistical validation of the model. In order to perform this validation, we chose parameters that approximated those obtained through the cpHMM inference on experimental data shown in Figure 5.

Parameter	Value
Promoter switching rates ($k_{01}, k_{10}, k_{12}, k_{21}$)	$(1.2, 1.26, 0.72, 4.2) \text{ min}^{-1}$
RNAp initiation rates (r_0, r_1, r_2)	$(0, 18.5, 46) \text{ RNAp/min}$
Measurement noise (σ)	4.5 RNAp
RNAp elongation time (τ_{elong})	140 sec
Data sampling resolution ($\Delta\tau$)	20 sec
Memory window ($w = \tau_{\text{elong}} / \Delta\tau$)	7
MS2 loop transcription time (τ_{MS2})	30 sec
Duration of each trace	30 min
Number of time points per dataset	9,000
Number of traces per dataset	100
Number of independent datasets	20

As shown in Figure S18B and Figure S19, comparison of the simulated and inferred parameters indicates that we reliably recovered the parameters used to generate our simulated data with high precision. We accurately inferred transition rates, dwell times, fraction of time spent in each state, and the rates of RNAp loading over 20 independent datasets of simulated traces.

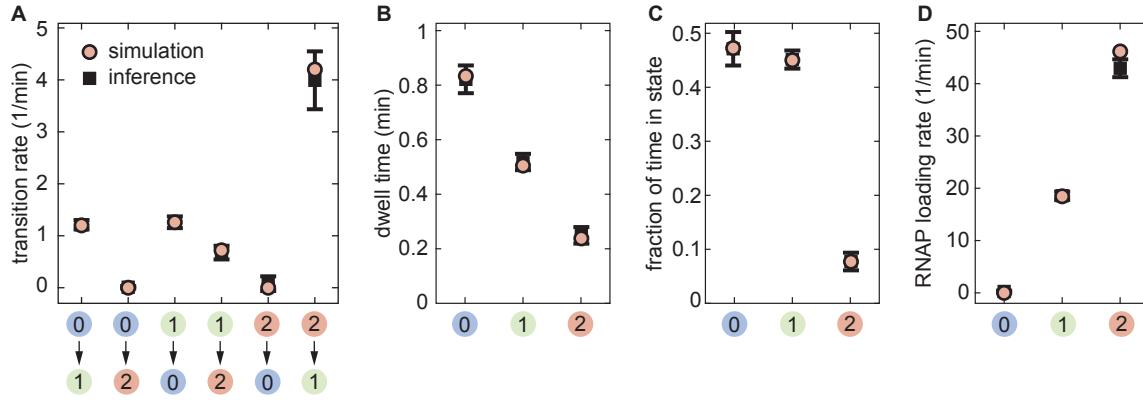


Fig. S19. Inference statistics for the cpHMM validation. The true and inferred values of (A) transition rates, (B) dwell times in states, (C) state occupancies, and (D) RNAP loading rates are compared. Statistics on the inferred values are obtained from 20 independently generated datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates).

Validation details. We used the relation between the transition rate matrix, \mathbf{R} , and the inferred transition probability matrix, \mathbf{A} , defined in *SI Appendix*, section D to obtain estimates of the transition rates, namely,

$$\mathbf{A} = e^{\mathbf{R}\Delta\tau}, \quad [109]$$

$$R_{ij} = \left(\frac{1}{\Delta\tau} \log \mathbf{A} \right)_{ij}. \quad [110]$$

Here, the exponential and logarithm operations act on matrices $\mathbf{R}\Delta\tau$ and \mathbf{A} , respectively. Occasionally, taking the matrix logarithm of the transition probability matrix \mathbf{A} yielded small negative values for transition rates between states (0) and (2), which were originally zero during data generation. In those cases, we assigned them a 0 value to keep them physically admissible.

Continuous vs. Poisson promoter loading. To demonstrate the validity of our choice to use continuous RNAP initiation rates in the transcription model (*SI Appendix*, section D), we repurposed our simulation to, instead of considering a constant rate of RNAP loading, explicitly account for individual RNAP loading events when generating the traces. We assumed that individual polymerase molecules traverse at a constant elongation rate ($v_{\text{elong}} = 46 \text{ bp/sec}$, *SI Appendix*, section J) and that their arrival to the promoter region has a Poisson waiting time distribution, provided that the promoter is cleared from the previous polymerase molecule which has a finite footprint size of $l_{\text{RNAP}} = 50 \text{ bp}$ (17). This led to a two-step model for the process of RNAP initiation, with Poisson-distributed wait times for the recruitment of RNAP to the promoter followed by a finite wait period as the RNAP cleared the promoter—a process taken to be approximately deterministic. With this information in hand, we expressed the mean loading time of RNAP at a single promoter (r_1^{-1}) as the sum of the mean time of polymerase arrival at an empty promoter, $\langle \tau_{\text{arrival}} \rangle$, and the time required to clear it after arrival, $\frac{l_{\text{RNAP}}}{v_{\text{elong}}}$, that is,

$$\frac{1}{r_1} = \langle \tau_{\text{arrival}} \rangle + \frac{l_{\text{RNAP}}}{v_{\text{elong}}}. \quad [111]$$

Having the values of r_1 , l_{RNAP} , and v_{elong} , we found $\langle \tau_{\text{arrival}} \rangle$ and used it in simulating the arrival events of individual polymerases.

We performed inference on these simulated traces using cpHMM with the objective of determining whether a Poisson loading rate had an effect on the obtained parameters. As shown in Figure S20, when the data is generated using Poisson RNAP loading, cpHMM slightly overestimates the high transition rate, but otherwise manages to accurately recover the model parameters. This therefore justifies our modeling approach of assigning continuous RNAP initiation rates to each promoter state, instead of explicitly modeling the recruitment of individual polymerases.

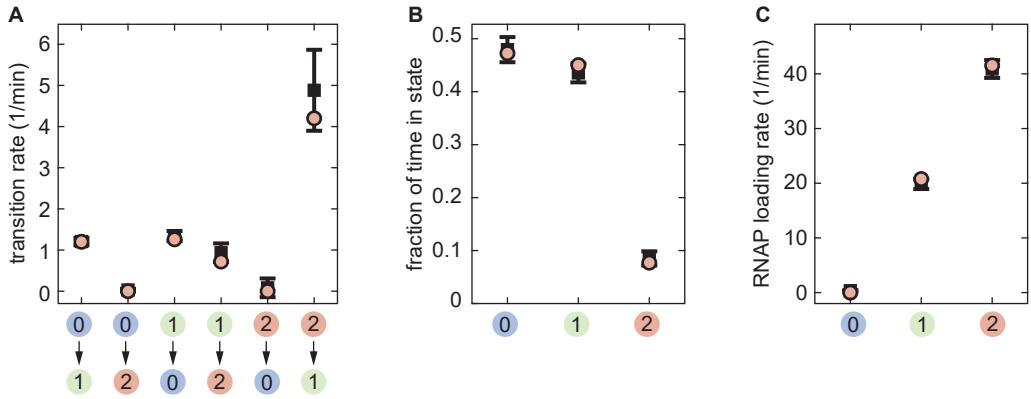


Fig. S20. Validation of cpHMM on Poisson RNAP loading data. (A) Transition rates, (B) state occupancies and (C) RNAP loading rates inferred from 15 independently generated datasets assuming Poisson loading of RNAP. (Error bars represent one standard deviation calculated across these 15 independent replicates.)

Sensitivity of cpHMM to data sampling resolution. In our cpHMM framework, we modeled the stochastic transitions between effective promoter states using a discrete time Markov chain model which assumes that the state of the promoter remains constant during the experimental time step ($\Delta\tau$), and that transitions to the next promoter state can occur only at the end of each step. This means that, if the fastest promoter switching rate is greater than the data sampling rate ($1/\Delta\tau$), our model might be unable to capture all those transitions. To study this possible limitation of cpHMM, we conducted inference on synthetic activity traces generated with varying sampling rates. Since the system memory ($w = \tau_{\text{elong}}/\Delta\tau$) needs to be an integer, we varied w in the [3, 7] range, correspondingly changing the sampling resolution from low ($\tau_{\text{elong}}/3 \approx 46\text{s}$) to high ($\tau_{\text{elong}}/7 = 20\text{s}$). We used the values in Table S2 for the remaining model parameters.

Figure S21 summarizes the findings of this study. As expected, the accuracy of inference improves with increasing data sampling rate, and inference results get very close to the ground truth values when the highest sampling rate (1/20 sec = 0.05s^{-1}) becomes comparable to the fastest transition rate (0.07s^{-1}). Except for the fastest transition rate, all other rates are inferred accurately for the whole spectrum of sampling resolutions (Figure S21A). The accuracy of inferred state occupancies is also remarkably high, making it robust to variations in the data sampling rate (Figure S21B). The high RNAP loading rate tends to be underestimated for slower sampling resolutions, which is reasonable since the chances of promoter leaving state (2) during a single time step become greater, effectively reducing the net rate of loaded RNAP molecules per time step (Figure S21C). Generally, we find the inference of model parameters to be reasonably accurate for the entire spectrum of experimentally realizable data sampling rates, and highly accurate when the timescale of the fastest transition and data sampling are comparable.

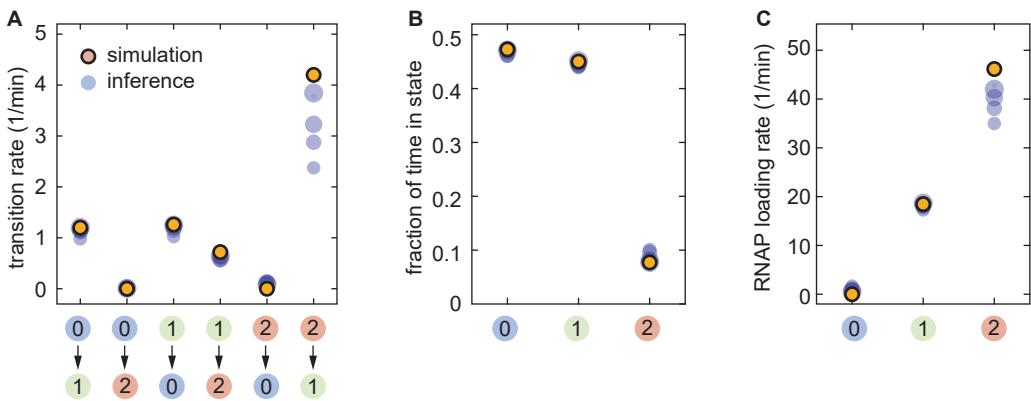


Fig. S21. Sensitivity of cpHMM to data sampling resolution. (A) Transition rates, (B) state occupancies and (C) RNAP loading rates inferred from datasets generated with varying time resolutions. Transparent circles represent averages over 20 independently generated samples. The increasing size of the blue circles corresponds to higher data sampling resolutions (largest: 20s, smallest: 46s).

Performance of cpHMM in different kinetic regimes. Thus far, the validation of cpHMM was performed on datasets that were generated using parameters similar to those inferred for the *eve* promoter. These parameters have characteristic low ON rates (k_{01}, k_{12}) and a high OFF rate (k_{21}), where “low” and “high” are relative to the data sampling frequency, which for our experimental setup is 3/min. To assess the utility of our inference method for a generic choice of model parameters, we performed additional inference studies in three different parameter regimes: low ON rates and low OFF rates (E–Figure S22A-C), high ON rates and low OFF rates (Figure S22D-F), and high ON rates and high OFF rates (Figure S22G-I).

As expected, the inference is the most accurate when the data sampling frequency is greater than the transition rates (Figure S22A-C), in which case multiple transitions within a single time frame occur only rarely, making our discrete Markovian

representation of the state dynamics a valid approximation. The largest deviations of the inferred model parameters from their ground truth values occur when the ON rates are high and the OFF rates are low (Figure S22D-F). Since the promoter rarely remains in the lower initiation states (0 or 1) for the entire duration of a frame and tends to rapidly transfer into a higher initiation state (1 or 2, respectively), the rates of RNAP loading for states 0 and 1 are significantly overestimated (Figure S22F). Despite the inaccuracies in estimating the RNAP loading rates, all transition rates, with the exception of k_{10} , are inferred with a high accuracy (Figure S22D). Remarkably, the deviations caused by the high ON rates get substantially suppressed when the OFF rates are also made comparably high (Figure S22G-I). This can be thought of as a consequence of an effective counterbalancing between unwanted ON and OFF transitions within a single time frame.

Overall, these additional studies, together with the statistical validation studies discussed earlier (Figure S19), elucidate the domain of applicability of cpHMM: the method performs accurate inference when the ON/OFF transition rates are respectively slow/slow, slow/high, or high/high; and is not successful in accurately inferring some of the model parameters when the ON rates are high, but the OFF rates are low. We hope that these characteristics of the method will be useful in informing the design of promoter architectures and new experiments.

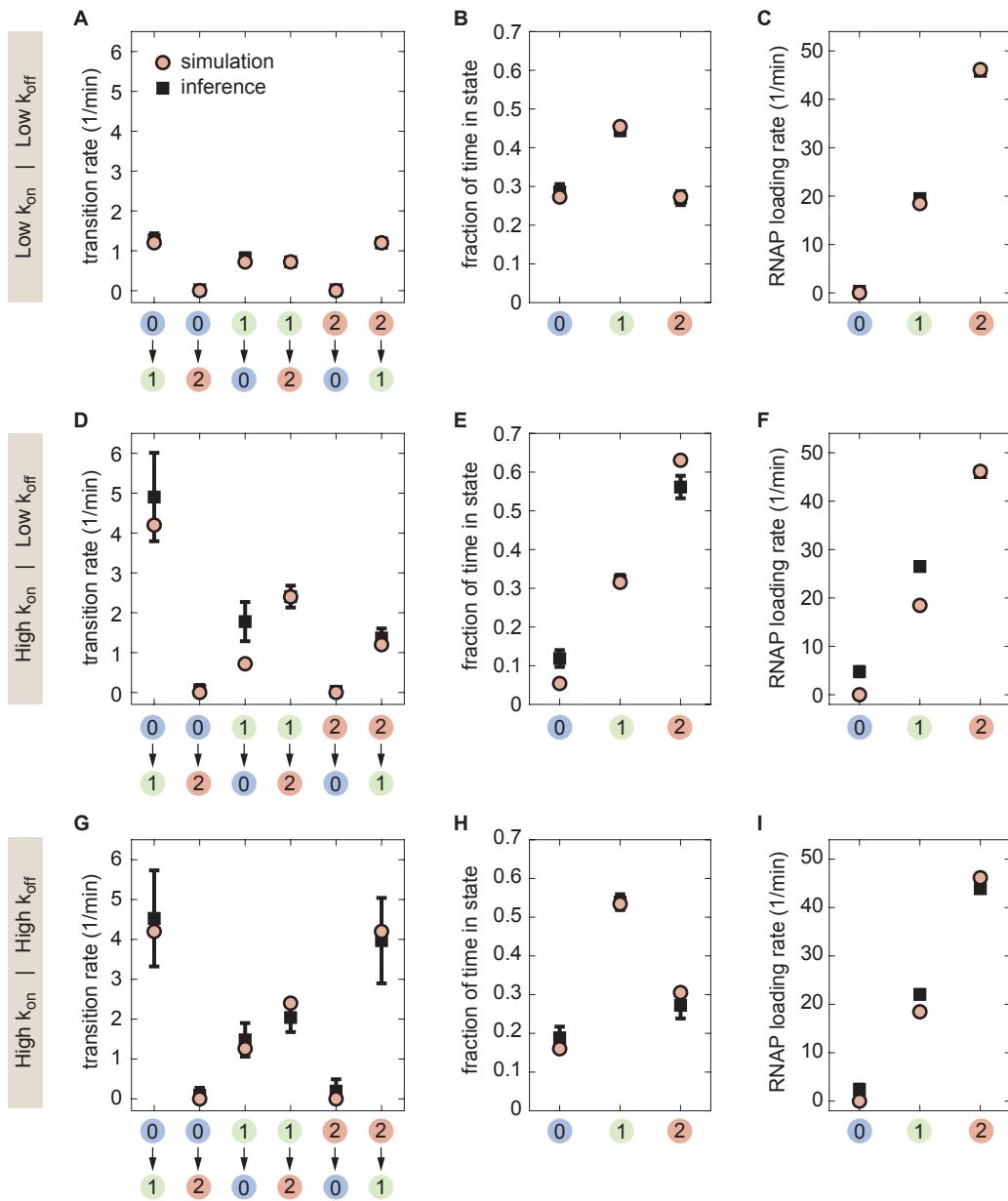


Fig. S22. Study of cpHMM performance for different choices of the ON/OFF transition rates Comparison of inference performance for different ON/OFF rates using a data sampling frequency of 3/min. (A-C) low/low, (D-F) high/low, (G-I) high/high. The statistics of inferred model parameter values is obtained from 20 independent datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates.)

Windowed cpHMM. To check that our windowed cpHMM was capable of fitting time-varying data, we conducted statistical validation using simulated traces exhibiting various time-dependent trends in the bursting parameters. We studied three

scenarios that mimicked ways in which bursting parameters could, in principle, be modulated to drive the onset of transcriptional quiescence: a decrease in k_{on} over time, an increase in k_{off} and a decrease in r . We also studied the case of increasing k_{on} , as this was the strongest temporal trend observed in our experimental data. [Figure S23](#) summarizes the results for these validation tests.

For each test, 100 simulated traces, 40 minutes in length, were generated ($\Delta\tau = 20$ s) that exhibited the desired parameter trends. Consistent with our approach to the experimental data, a sliding window of 15 minutes was used for inference, meaning that for each inference time, τ_{inf} , all data points within 7.5 minutes of τ_{inf} were included in the inference. This led to inference groups consisting of 4500 data points, with the exception of the first and last time points, each of which had 3700 data points (first and last $w+1$ points were excluded from inference). Transition and initiation rates shown in [Figure S23](#) are associated with state (1) of the three-state model ($k_{on} = k_{01}/2$, $k_{off} = k_{10}$ and $r = r_1$ in [Figure S23A](#)), as these were found to provide the most faithful indication of underlying system trends.

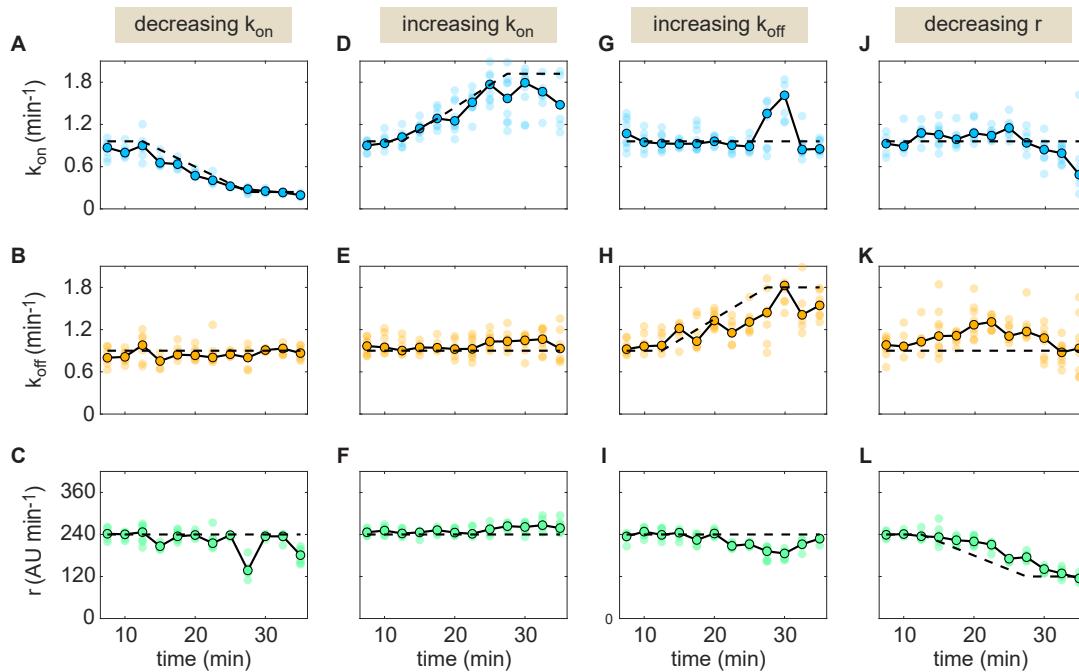


Fig. S23. Validation of windowed cpHMM inference. The method's accuracy was tested for four distinct sets of parameter time trends. Results for each scenario are organized by column. In each plot, the black dashed line indicates the true parameter value as a function of time. Connected points (outlined in black) indicate the median inferred parameter value at each time point across 10 distinct replicates. Translucent points indicate inference values from individual replicates. Thus, the dispersion of these replicates at a given time point indicates the precision of the inference.

For each scenario, we assessed whether and to what degree the windowed cpHMM method could accurately recover the temporal profiles. In general, the method was found to perform quite well within the parameter regimes that were tested. For both the increasing and decreasing k_{on} scenarios ([Figure S23A-C,D-E](#)), windowed cpHMM inference accurately captured the modulation in k_{on} with no significant variation evident in the r and k_{off} trends. In the case of increasing k_{off} ([Figure S23G-I](#)), we observed deviations in k_{on} and r from their true values at the inflection point of the k_{off} curve (around 30 min). However, the deviation in r is relatively mild and the “blip” in k_{on} , while of larger magnitude, is comprised of only two time points and so would likely not be mistaken for a legitimate indication of underlying system behavior. In the case of a decrease in the initiation rate ([Figure S23J-L](#)) we observe a ~ 5 min delay in the model response. We attribute this delay to the finite dwell time of RNAP molecules on the gene (in this case $\tau_{elong} = 140$ sec, although further studies will be needed to determine why the observed lag appears larger than the elongation time). In addition, we note a degradation in the precision of the inference of k_{on} and k_{off} at low r (RHS of [Figure S23J, K](#)).

Overall, we conclude that the windowed cpHMM method is capable of accurately inferring time-resolved parameter values. An important caveat to these results is that the size of the sliding window (15 min in this case) places an inherent limit on the time scales of the parameter trends the model is capable of inferring. Changes that occur on shorter time scales will be registered, but the temporal averaging introduced by the sliding window will lead to underestimates of the rate of the parameter changes in the underlying system.

F. Sister chromatids.

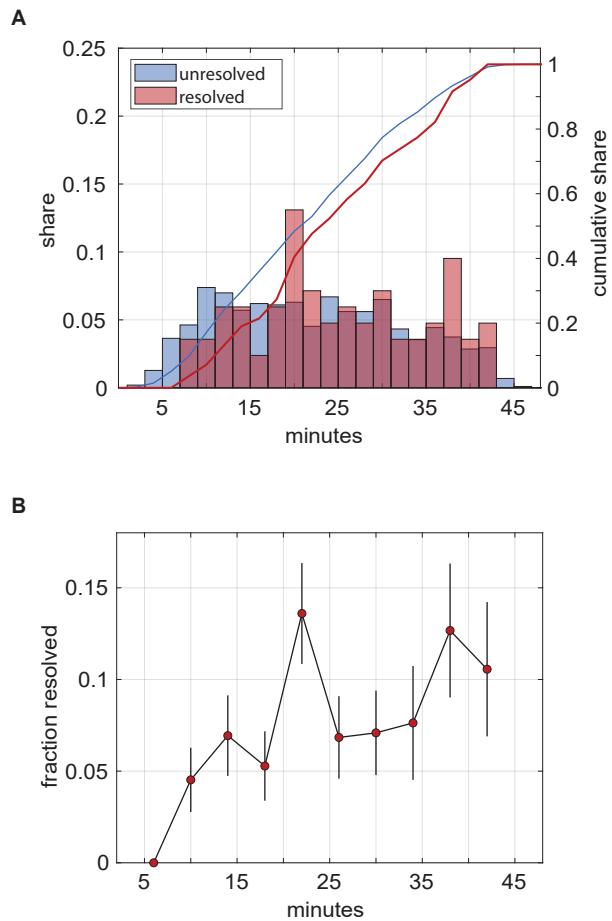


Fig. S24. Live imaging data indicate timing of sister chromatid appearance. (A) Distribution of observation times for frames in which chromatids were resolvable (red) and diffraction-limited (blue). Bars indicate empirical probability distribution function. Lines indicate cumulative density function. Data indicate the presence of chromatids by no later than 7-8 minutes into nuclear cycle 14. (B) Fraction of frames featuring resolved chromatids as a function of time. Trend suggests replication of relevant portion of genome across all observed nuclei is completed by approximately 10 minutes into nuclear cycle 14. Initial lag is likely attributable—at least in part—to stochastic turn-on times between sister *eve* loci and lower fluorescence levels early on in the nuclear cycle.

Detection of sister chromatid appearance. Previous studies have indicated that the *D. melanogaster* genome is quickly replicated at the beginning of each nuclear cycle in early development (18, 19), suggesting that each diffraction-limited spot in our imaging data likely contains two distinct *eve* loci. We sought to use our live imaging data to verify whether genome replication occurred early enough in the nuclear cycle such that the presence of the replicated promoters would have to be taken into account. While the two *eve* loci are located within a diffraction-limited spot for the majority of frames in our data, there are a subset of frames in which two distinct puncta can be clearly observed due to fluctuations in the separation between chromatids (see Figure 4D). We reasoned that, by tracking the frequency of frames with resolved puncta over time, we could ascertain how the timing of genome replication compares to the onset of transcription. If replication precedes the onset of transcription, then the fraction of resolved frames should be relatively stable over the duration of *eve* expression in nuclear cycle 14. If, on the other hand, replication happens after the onset of transcription, we should see a significant increase in the frequency of resolved sister chromatids over time as development progresses.

To pursue this question, we randomly selected snapshots of transcriptional loci in 100 different nuclei for each of the 11 embryos used in this study. We then determined the fraction of these sampled snapshots in which two distinct puncta were clearly visible by eye and observed how these instances of resolved chromatids were distributed in time. As indicated in Figure S24A, we see evidence for resolved puncta by around 7 minutes into nuclear cycle 14. This is well within the average range for turn-on times observed throughout the stripe (see Figure S4B). Our results indicate that, at the very least, the genomic region containing our *eve* stripe 2 reporter is replicated within *some* nuclei by 6-8 minutes into nuclear cycle 14. Figure S24B tracks the share of total observations for which we detected resolved puncta as a function of time. A systematic delay in DNA replication would be expected to result in a progressive increase in this metric over time. However, such a trend is not evident. While we see *no* resolved sister loci between 4 and 8 minutes (first point in the plot in Figure S24B), this absence could be attributed to other factors at play early on in nuclear cycle 14. For example, part of this apparent lag could

be attributable to the fact that loci are, on average, dimmer early on in the nuclear cycle, which could mask the presence of two *eve* loci by reducing the probability of both producing observable amounts of fluorescence at the same time. It is also possible that the precise timing of locus turn-on varies for each sister locus, as it does for loci in different nuclei. Regardless, even if the initial rise between 6 and 10 minutes in Figure S24B is reflective of the replication of the locus during this period of time, the relative stability of the frequency of resolved loci from 10 minutes onward indicates that this process is restricted to the first few minutes of transcription. Additional experiments are needed to further elucidate the interplay between DNA replication and the onset of transcription. Regardless, the examination of our live imaging data supports the conclusion that the majority of our data consist of diffraction limited spots containing two distinct *eve* loci.

Probing for interactions between sister chromatids.

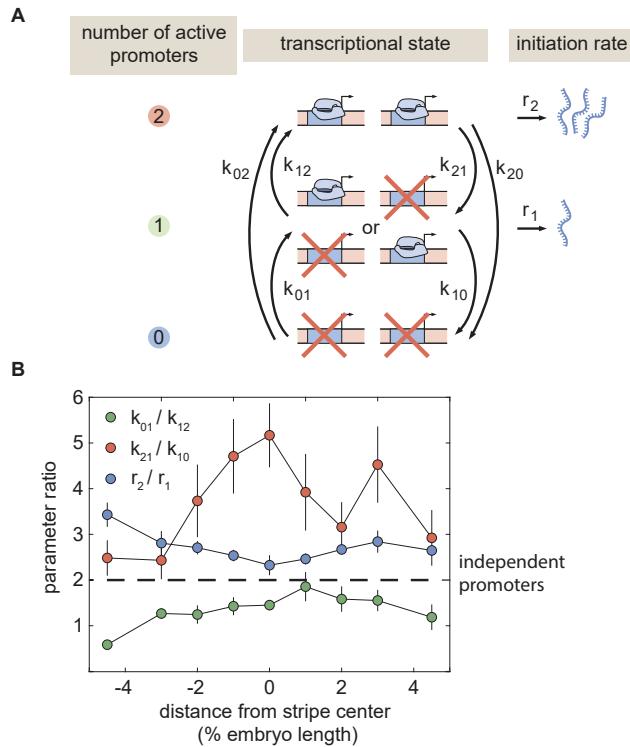


Fig. S25. Probing combined transcription of sister chromatids. (A) Revised three-state model of promoter switching within a fluorescent punctum that accounts for the combined action of both sister chromatids. (B) Summary of bursting parameter ratios. All three bursting parameter ratios deviate from their expected values under the independence assumption given by the horizontal dashed line. (Error bars indicate magnitude of difference between first and third quartiles of cpHMM inference results for bootstrap samples of experimental data over multiple embryos. See Materials and Methods for details.)

If each fluorescent punctum contains two promoters (Figure 4D), then it is necessary to revisit the widely used two-state model of transcriptional bursting. In this revised scenario, each promoter on one of the sister chromatids undergoes fast ON/OFF switching. Therefore, each spot (encompassing two identical loci) can be in one of three states: (0) both promoters OFF, (1) one promoter ON and the other OFF, and (2) both promoters ON (Figure S25B). States (1) and (2) are expected to exhibit different rates of RNAP loading, r_1 and r_2 , respectively. See *SI Appendix*, section A and *SI Appendix*, section D and for details regarding the implementation of this three-state model.

The presence of two transcriptional loci within each fluorescent punctum suggests three constraints on the relationship between bursting parameters in the model shown in Figure S25A. First, if these two promoters transcribe independently, then state (2) will have double the loading rate of state (1) such that $r_2 = 2r_1$. Second, the probability of both promoters transitioning simultaneously should be negligible; we expect no transitions between states (0) and (2) such that $k_{02} = k_{20} = 0$. Finally, if the promoters switch between their states in an independent manner, then there will be an extra constraint on their transition rates. For example, there are two paths to transition from (0) to (1) as either promoter can turn on in this case. However, there is only one possible trajectory from (1) to (2) because only one promoter has to turn on. This condition sets the constraint $k_{01} = 2k_{12}$. Similarly, $k_{10} = k_{21}/2$.

While the independence of sister chromatids is supported by recent single-molecule FISH experiments (20, 21), classic electron microscopy work suggests a scenario in which sister chromatids are tightly correlated in their transcriptional activity (22, 23). Given this uncertainty regarding chromatid independence, we elected to employ a general three-state model that makes no assumptions about the nature and strength of sister chromatid interactions. In addition to permitting greater flexibility, this agnostic approach also meant that the structure of the kinetic model returned by cpHMM inference provided clues regarding the nature of the coupling between sister loci. Specifically, we examined the ratios between the high and low on rates (k_{01} and k_{12}), off rates (k_{21} and k_{10}), and initiation rates (r_2 and r_1). A deviation from these expectations would

indicate either that the two sister loci do not initiate RNAP independently (first constraint), or that they do not transition between activity states independently (second and third constraint).

Overall, our results suggest that the two loci are coupled to a nontrivial degree. We observe that the rate of initiation for the high state, $r_2(x)$, (corresponding to two active promoters) is consistently greater than twice the middle state, $r_1(x)$ ([Figure S25B](#), blue). This trend suggests some sort of synergy in the RNAP initiation dynamics of the sister promoters. Even more strikingly, we observe that the rate of switching from (2) to (1), k_{21} , is *much* higher than twice the rate of switching from (1) to (0), k_{10} , ([Figure S25C](#), red). This indicates that each promoter is more likely to switch off when its sister locus is also active. This anti-correlation is consistent with some form of competition between the loci, a scenario that could arise, for instance, if local concentrations of activating TFs are limiting. In addition, we observe substantial variation in the relationship between the high and low on rates (k_{01} and k_{12} , respectively), ranging from one of near equality in the anterior flank to nearly the 2-to-1 ratio that would be expected of independent loci in the stripe center and posterior ([Figure S25C](#), green). Finally, as shown in [Figure S26](#), we observe no transitions between the (0) and (2) states, lending support to the hypothesis that, despite their correlation, our spots do contain two promoters.

Further experiments in which the sister chromatids are labeled in an orthogonal manner are needed to confirm and elaborate upon these results. One important consideration to address is the fact that the spatial proximity of the two loci appears to fluctuate significantly over time. Thus, if (as seems plausible) the strength of the coupling between loci depends in some way upon the radial separation of the loci, then the results reported here are effectively an average of time-varying system behavior. Valuable information may be obscured as a result of this averaging.

G. cpHMM inference sensitivities.

Full three-state inference results. For the sake of simplicity, we presented our inference results in the main text using an effective two-state model in which two distinct active transcriptional states were combined into a single effective ON state (see [Figure 4E](#) and F). Here, for completeness, we include time-averaged and time-resolved inference results for the full three-state model where, as shown in [Figure S25](#), (0) corresponds to the state where both promoters are in the OFF state, (1) indicates the state where either promoter is in the ON state, and (2) represents the states where both promoters are in the ON state.

As indicated in the main text, the full three-state results ([Figure S26](#)) exhibited the same trends as were evident in the effective two-state plots ([Figure 5](#)). In agreement with the effective two-state model, the rate of transcript initiation is not modulated to a significant degree across the stripe ([Figure S26D](#)). Moreover, we once again see that activation rates, and specifically the rate of switching from OFF to the middle ON rate (states 0 and 1 in [Figure S26E](#)) are strongly elevated in the stripe center.

Like the time-averaged results, time-resolved inference trends for the full three-state model agree closely with effective two-state results shown in main text (compare [Figure S27](#) to [Figure 6D-F](#)). Due to a lack of statistics for state (2), we show only transition rates into and out of the first active state (middle state in [Figure 4E](#)).

Two-state inference results. Although the presence of sister chromatids indicated that the three-state model was most appropriate for the *eve* stripe 2 system, we wanted to check that our conclusions were robust to this assumption. To do this, we conducted time-averaged and windowed inference assuming a simpler, two-state model (*see, e.g.* [Figure 4B](#)). Note that this approach is distinct from the *effective* two-state results presented in the main text. There, as outlined in [Figure 4D-F](#), a three-state model was specified for inference and the results for the two active (ON) states were aggregated after the fact to simplify the presentation of the results. Conversely, here, we explicitly conducted inference using a two-state model.

Most of our findings remained unchanged in the context of the two-state model. Consistent with the three-state case, the two-state time-averaged cpHMM inference indicated that the fraction of time spent in an active state, rather than the rate of RNAP initiation, drives the difference in mRNA production rates across the stripe ([Figure S28A-C](#)). Moreover, as with the three-state case, two-state results indicated that the bulk of this variation stem from modulation in k_{on} ([Figure S28C](#), green). Interestingly, whereas we did see a degree of spatial dependence in k_{off} for 3-states, we observed no such trend for 2-states ([Figure S28C](#), red). In general, this is not surprising, as our use of a simpler model likely means that multiple switching rates are being projected onto the k_{off} parameter. Specifically, if the *eve* stripe 2 system is indeed a true three-state system, then we would expect the two-state k_{off} estimate to reflect the joint action of the k_{10} , k_{21} , and k_{12} rates from the three-state model. As a result, the spatial dependence of each one of these rates would get averaged out when combined onto k_{off} .

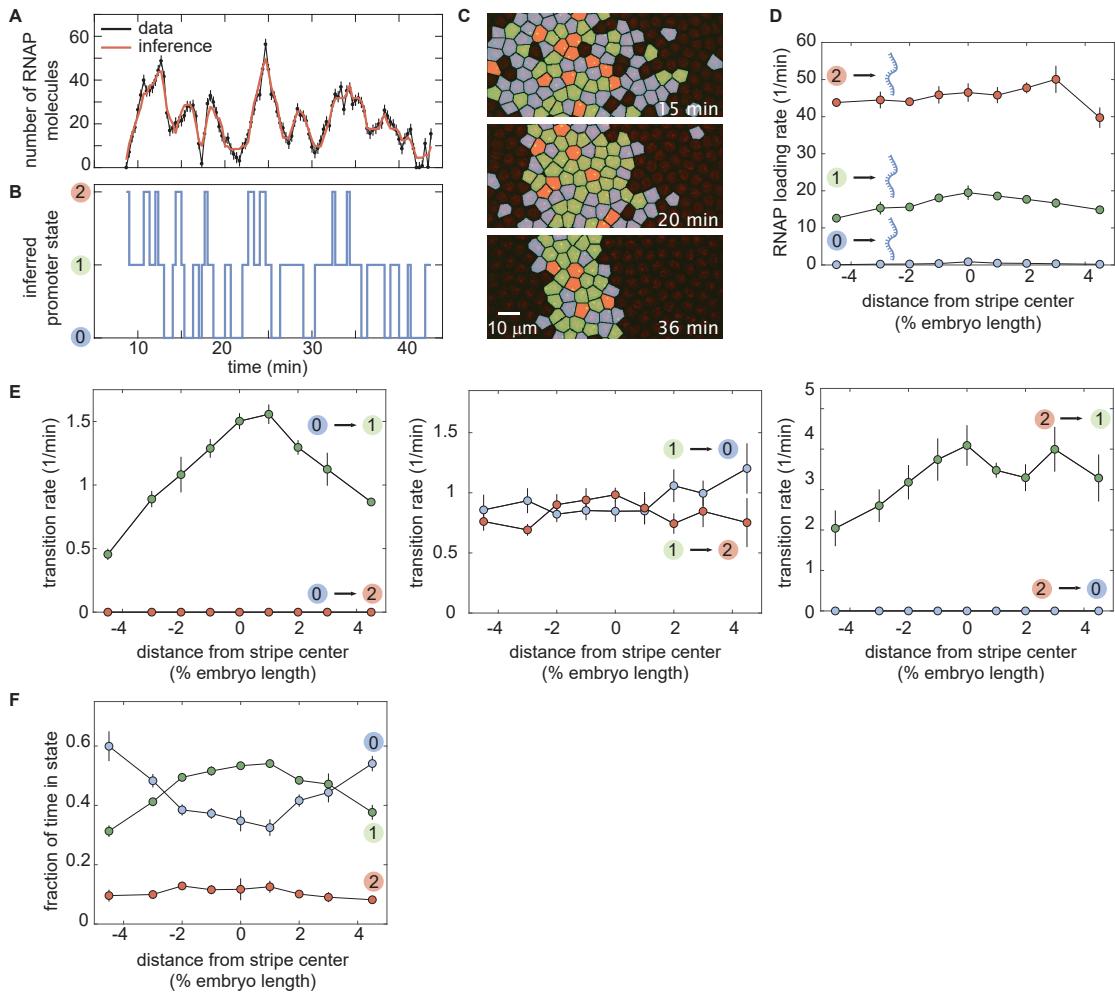


Fig. S26. Full three-state results for time-averaged cpHMM inference. (A) Representative experimental trace along with its best fit and (B) its most likely corresponding promoter state trajectory. (C) Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (colors as in B). (D) The rate of initiation for each transcriptional state is not significantly modulated along the embryo. (E) Our cpHMM revealed that the transition rate between the OFF (0) and middle ON state (1) is up-regulated in the stripe center. In contrast, the rates of switching out of the middle and high ON states show little to no significant AP-dependent modulation. (F) The modulation of the rate of switching from 0 to 1 acts to increase the fraction of time the promoter spends in the active states in the stripe center. (A, error bars obtained from estimation of background fluorescent fluctuations, as described in Materials and Methods and (1); D, E, and F, error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

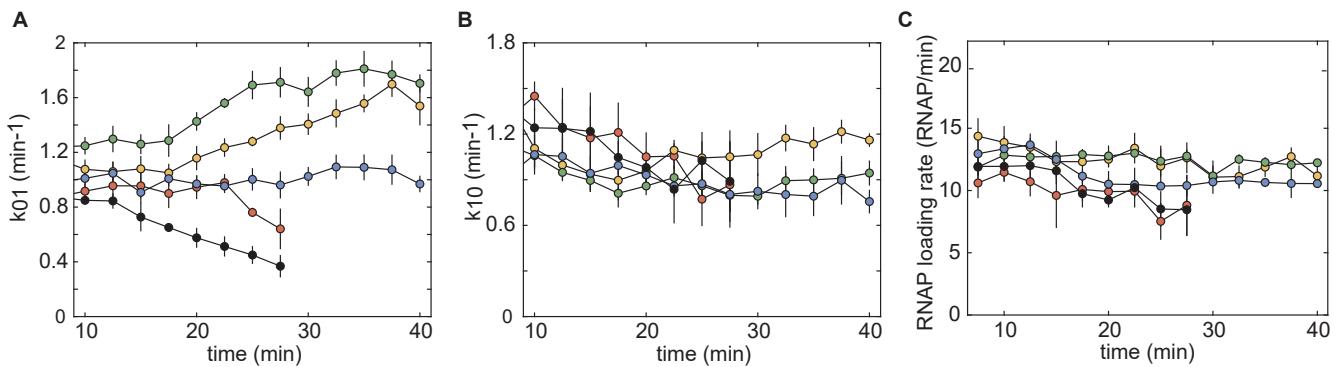


Fig. S27. Full three-state results for time-dependent cpHMM inference. (A) Transition rate from transcriptionally inactive state (0) to the first active state (1). Same trends evident as for effective 2 state model. (B) Transition rate from first on state (1) to OFF state (0). (C) Rate of transcript initiation in first on state (1) as a function of time. (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

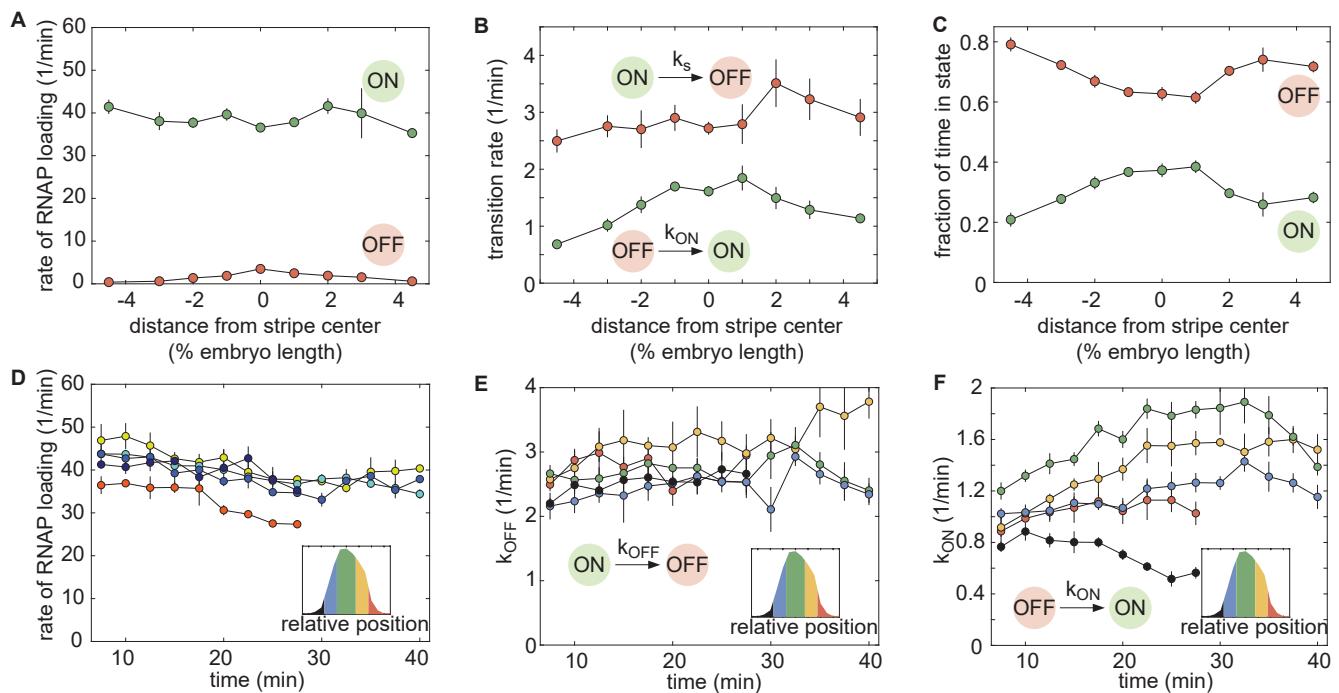


Fig. S28. Two-state cpHMM inference. (A-C) Time-averaged 2-state inference results. (A) Consistent with three-state inference results, we observed no significant modulation in the rate of initiation along the axis of the embryo. Moreover, we found that k_{on} (green plot in (B)) was modulated along the anterior-posterior axis to vary the amount of time the promoter spent in the ON state (green curve in (C)). In a departure from the three-state case, we observed no significant spatial trend in k_{off} , though we noted a spike in k_{off} at 3% of the stripe center. (D-F) Time-resolved (windowed) two-state cpHMM results. (D) Consistent with the 3-state inference, we saw little to no modulation in the rate of RNAP loading r over time, although we noted a mild downward trend across all AP bins that was most pronounced in the posterior flank (red curve). (E) Two-state inference indicated no significant temporal trends in k_{off} . (F) k_{on} time trends largely agreed with the three-state case, although we noted that the decrease in k_{on} in the posterior flank that was apparent in the three-state results was not observable in this two-state context (Figure 6E, red). (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrapped samples of experimental data. See Materials and Methods for details.)

As with the time-averaged case, we found that results for two-state windowed cpHMM were generally consistent with three-state trends. A notable exception to this rule was the absence of any significant decrease in k_{on} in the posterior stripe flank (Figure S28F, red). This is not entirely surprising, as the trend returned by the three-state inference was relatively mild (Figure 6E, red), encompassing only the final two time points for which there was sufficient data to conduct inference. It is possible that the added complexity of the three-state model allowed it to register a subtle shift in the activation rate that was convolved with countervailing features in the two-state case. Future work will seek to elucidate the source of this discrepancy and further test the validity of the trend uncovered in the three-state case.

Comparing true and effective two-state inference results. Here, for completeness, we provide direct comparisons between the time-averaged inference for the effective two-state results presented in the main text and the true two-state results presented in the previous section.

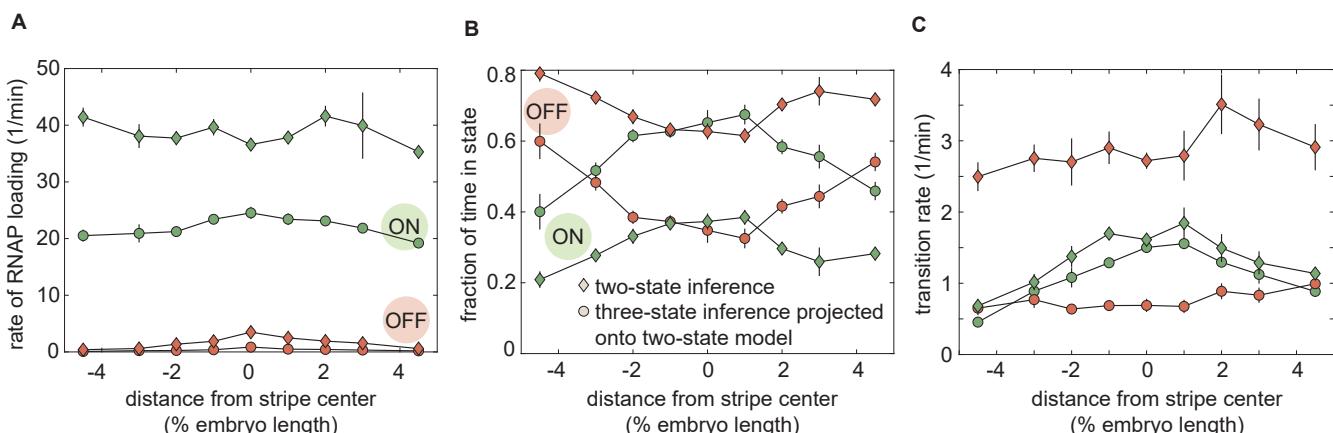


Fig. S29. Comparing two- and three-state cpHMM inference results. Three-state inference results can be presented in terms of a two-state model in which states (1) and (2) are aggregated into a single ON state (see Figure 4E and F). Here, color schemes are consistent with those employed in Figure S28A-C. Squares indicate true two-state results (presented in the previous section) and circles indicate effective two-state trends derived from the three-state results presented in Figure 5. **(A)** Anterior-posterior-dependent trends in the rate of RNAP initiation are nearly identical between the true and effective initiation rates, however the initiation rate returned by two-state cpHMM inference (green squares) is roughly twice as large as that implied by the three state results (green circles). **(B)** As with the initiation rates, we observe similar trends between the true and effective cases, but substantial differences in magnitude. The effective two-state model recovers an ON state occupancy that is roughly double that returned by two state cpHMM inference. **(C)** While the ON rate trends and magnitudes are nearly identical, the OFF rate returned by two-state cpHMM inference is roughly triple that implied by three-state inference. Thus it is clear that this difference in OFF rate underlies the observed departures in both state occupancies (B) and state initiation rates (A). (Error bars indicate magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

As Figure S29 makes clear, while anterior-posterior-dependent parameter trends are by and large consistent between the true and effective two state models, we do observe substantial differences in the absolute magnitudes of parameter values. These differences originate (directly or indirectly) from the three-fold difference in the value of k_{off} between the true and effective models (Figure S29C, red squares and circles, respectively). The k_{off} value for the effective two-state model is defined as

$$k_{\text{off}} = \frac{k_{10}k_{21}}{k_{21} + k_{12}}. \quad [112]$$

See *SI Appendix*, section A for expressions for all three effective two-state bursting parameters (k_{on} , k_{off} , and r) in terms of these three-state transition rates. This value represents the inverse of the mean amount of time the system, upon switching out of state (0), spends in one of the active states before returning to (0), and we can see that it is necessarily less than or equal to k_{10} .

Thus, the two- and three-state results imply that the systems switch out of the active state(s) on substantially different timescales. On the other hand, the ON rates are strikingly similar across the two models. As a result, the effective two-state model implies that the system is in one of the active states for between 40 and 70% of time, whereas two-state cpHMM inference implies significantly lower shares falling between 20 to and 40%. Since both models must reproduce the same mean production rate—this is an inherent feature of the experimental traces—we see that the two-state cpHMM inference returns an estimated initiation rate that is consistently twice as large as the initiation rate implied by the effective two-state model.

Thus, while most of the conclusions featured in this paper are robust to our choice of model architecture, this decision does, nonetheless hold important implications for how we understand the underlying system. Further work is needed elucidate the root cause of this discrepancy and move towards a more concrete understanding of the correspondence between the structure of the model and that of the physical system.

H. Input-Output analysis details. In this appendix, we provide additional information about data sources, inference methodology, and inference sensitivities related to the input-output analysis presented in the main text.

Data sources. The input-output analysis presented in the main text made use of previously published data sets for the spatiotemporal concentration profiles of the gap genes Hunchback, Krüppel, and Giant (Figure S30A, C and D). These data derive from elegant experiments in which individual embryos were co-immunostained for transcription factors of interest and precisely staged by measuring progressive cellularization over the course of nuclear cycle 14 to generate a time series of protein concentration profiles spanning the course of this period of development (3). The Bicoid concentration data used for this analysis derives from live imaging experiments using a Bicoid-GFP fusion established by (24). These data come courtesy of Jonathan Liu and Elizabeth Eck (Figure S30B).

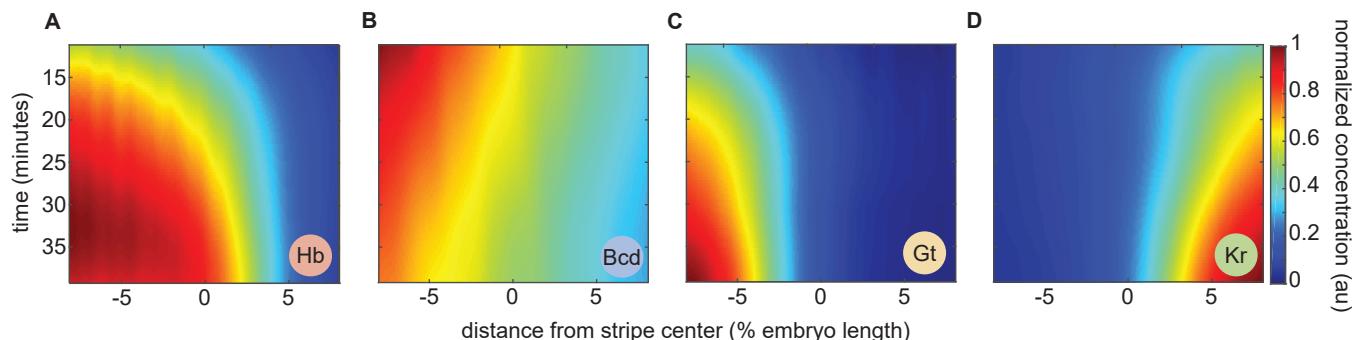


Fig. S30. Spatiotemporal transcription factor concentration maps. Heatmaps indicate normalized concentration profiles for the eve stripe 2 regulators **(A)** Hunchback, **(B)** Bicoid, **(C)** Giant, and **(D)** Krüppel as a function of space and time. In each case, levels were normalized relative to the maximum concentration observed within the spatiotemporal window of interest.

Data processing. To prepare the Krüppel, Giant, and Hunchback profiles for use in our logistic regression analysis, we adopted an approach similar to that described in (3). Dorso-ventral orientation of embryos was found to have negligible effect on calculated intensity profiles and was ignored (i.e. all embryos were included, regardless of orientation). For each time point in nuclear

cycle 14, a weighted temporal average was calculated using a sliding Gaussian kernel with $\sigma_t = 5$ min. For each time point, the minimum observed value across all anterior-posterior positions was then calculated and subtracted in order to remove background fluorescence. Normalized profiles were then calculated using the formula

$$I_{\text{norm}} = \frac{I_{\text{raw}}}{\max(I_{\text{raw}}) - \min(I_{\text{raw}})} \quad [113]$$

An identical procedure was followed for processing the Bicoid-GFP data, with the addition of a spatial averaging step using a sliding Gaussian window of $\sigma_{\text{AP}} = .5$ % embryo length. This step was necessitated by the fact that, because individual embryos were imaged for the duration of nuclear cycle 14, multiple experiments contributed concentration data along the anterior-posterior axis for each time point. Thus averages in both space and time were needed in order to effectively aggregate these data into a single average spatiotemporal profile.

Finally, we discovered that the anterior-posterior axes in our live imaging data (both for *eve* stripe 2 and Bicoid-GFP) were inconsistent with the axes employed by the fixed data reported by the authors in (3). We addressed this issue by using *eve* stripe 2 as a fiduciary mark to register the positions of the fixed and live data sets. Specifically, we aligned the mRNA peak predicted by our model at 40 minutes into nuclear cycle 14 with the peak in second stripe of the *eve* protein profile at 40 minutes, as reported in (25).

Logistic regression framework. The binomial logistic regression is a widely used statistical method for assessing the relationship between a set of predictor variables and a response variable of interest that is constrained to take on one of only two possible outcomes. In the context of our analysis, the predictor variables were the normalized transcription factor concentration profiles and the response variables were (i) the overall transcriptional state given by the transcriptional time window (active or silent?) and (ii) the bursting state amongst transcriptionally active loci (ON or OFF?). Inference was conducted at the level of individual gene loci. fmincon, a standard matlab function for constrained optimization, was used to fit all models discussed both in the main text and in this appendix.

To prevent overfitting at the stripe centers, the selection of data sets for input-output inference were weighted to ensure equal representation of data points from across all regions of space and time included in the analysis. The data were divided into cells of size 1% of the embryo length in width and 1 minute in duration for the purpose of calculating and assigning these weights. The number of data points in adjacent regions were factored into each region's weight score using a 2D Gaussian averaging kernel. Regions with fewer than 25 total data points were not included in the inference.

Inference details: transcriptional time window. For the time window input-output analysis, we considered only loci that were transcriptionally active for one or more time steps in nuclear cycle 14. Loci were classified as transcriptionally active for all time points between the first and last time points for which they exhibited detectable levels of transcriptional activity and silent for all time points following their final shut-off for which their nuclei were still present in the experimental field of view. Time points preceding the onset of activity were discarded. Figure S31A illustrates how this quantity varies over space and time in our experimental data. We considered a class of logistic regression models in which each transcription factor was permitted to appear at most once, thus requiring that each factor act on *eve2* in a uniform manner through space and time; i.e., the same protein could not activate expression on one stripe flank and repress on the other.

Inference details: transcriptional bursting. The bursting input-output analysis focused exclusively on transcriptionally engaged loci. The Viterbi algorithm was used to infer the instantaneous activity state (ON vs. OFF) for all loci. This activity state was taken as the response variable in our regression analysis. In all other respects, the inference procedure was identical to that conducted for the time window.

Results of unconstrained inference: time window. For the input-output inference results presented in the main text (Figure 7), we used prior knowledge about the regulatory function of each input transcription factor to constrain its range of permissible values in our inference. Specifically, we constrained the activators Bicoid and Hunchback to play activating roles in our model and, likewise, required that the repressors Krüppel and Giant played repressing roles. In several cases, this constrained inference led to models in which one or more transcription factors played no significant regulatory role (Bicoid and Hunchback for the time window and Bicoid for transcriptional bursting). In this section, we tested the sensitivity of the conclusions presented in the main text to our use of functional constraints by conducting unconstrained input-output inference runs.

The results of our unconstrained input-output inference for the transcriptional time window are identical to those presented in the main text. Despite the fact that no limitations were imposed on the regulatory function of each factor, we nonetheless recovered a model in which the two repressors, Giant and Krüppel, are necessary and sufficient to explain the onset of transcriptional quiescence in the stripe flanks. In agreement with the constrained case, we found that the addition of Hunchback and Bicoid to this two-repressor model had no qualitative effect on the output profile predicted by the model (Figure S31B). A quantitative comparison of model fit scores confirmed that the addition of Hunchback and Bicoid did nothing to improve model fit (Figure S31C). Thus, we conclude that our finding that the transcriptional time window can be explained entirely by the joint repressive action of Krüppel and Giant is insensitive to our choice to impose functional constraints.

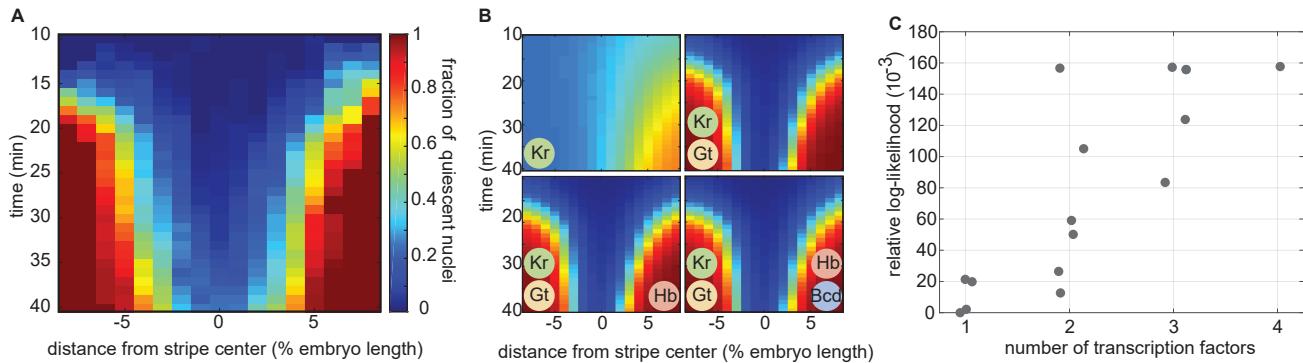


Fig. S31. Unconstrained inference results for the transcriptional time window. (A) Observed fraction of quiescent nuclei as a function of space and time. Identical data to that presented in Figure 7A. (B) Relaxing constraints on the functional nature of each transcription factor had no appreciable effect on the inference results. Profiles shown here are indistinguishable from those shown in Figure 7D. Once again, we find that the joint action of the repressors Giant and Krüppel is sufficient to explain the progressive onset of transcriptional quiescence in the stripe flanks. (C) A quantitative comparison of model fits reinforces the qualitative conclusions drawn from (B). Models including 3 and 4 transcription factors cannot improve on the fit achieved by the simpler double repressor model. Here blue dots indicate models for which only Giant and Krüppel make significant contributions to the model fit. This indicates that, while the 3 and 4 transcription factor models include additional parameters, these do not contribute appreciably to overall model fit, emphasizing the fact that these models behave, effectively, as double repressor models.

Results of unconstrained inference: transcriptional bursting. In the context of the transcriptional bursting input-output analysis, the removal of functional constraints led to a significantly more complex landscape of inferred regulatory models. While the functional roles of Krüppel, Giant, and Hunchback were consistent with the constrained case (repressing, repressing, and activating, respectively), Bicoid was consistently inferred to play a repressive role. Despite this complication, the three-factor Krüppel-Giant-Hunchback model favored by the constrained inference remained the best-fitting three-factor model (Figure S32C, red circle). While the addition of Bicoid as a repressor to create a model dependent on all four input transcription factors led to a small improvement in model fit (Figure S32C), comparison of this four-factor model's predicted activity profile with that of the Krüppel-Giant-Hunchback model revealed no material improvement in the model's agreement with the experimental data (Figure S32B, bottom left vs. bottom right). Moreover, there is (to our knowledge) no experimental evidence for Bicoid playing a repressive role in the regulation of *eve* stripe 2. Indeed, there is strong evidence that Bicoid is necessary for *eve* stripe 2 activity (26). We thus conclude the Krüppel-Giant-Hunchback model remains the most plausible option in the unconstrained case.

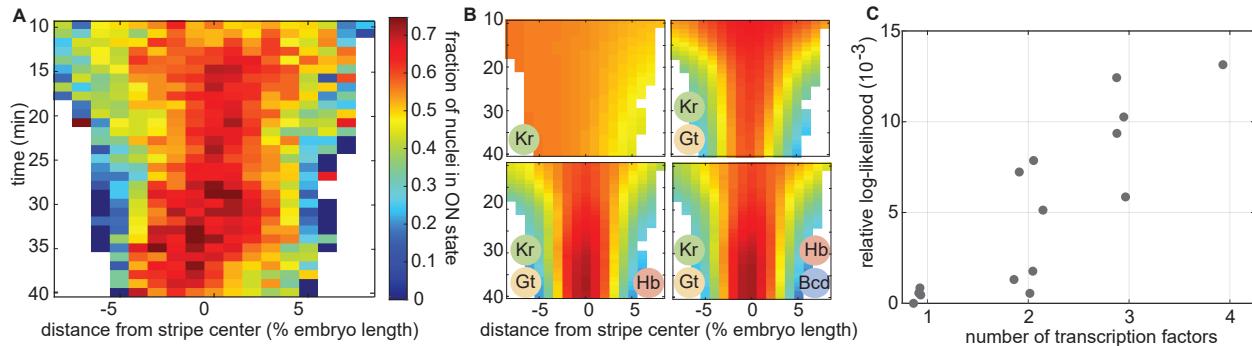


Fig. S32. Unconstrained inference results for transcriptional bursting. (A) Observed fraction of transcriptionally active nuclei in the ON (bursting) state. Identical data to that presented in Figure 7B. (B) As with time window, relaxing the constraints on the functional nature of each transcription factor did little to alter the inference results presented in the main text (compare to Figure 7E). As with the constrained results, the joint action of Giant, Krüppel, and Hunchback appears sufficient to explain the spatiotemporal activity pattern revealed by cpHMM inference. (C) A quantitative comparison of model fits.

I. Inherent limits of burst parameter inference. By definition, the onset of transcriptional quiescence coincides with the cessation of observable bursting activity. If this cessation is driven by changes in the bursting parameters as in scenario (ii) in Figure 6A, there is an inherent limit to the timescale of such changes that could be detected: changes that unfold over time scales of the same order or faster than the characteristic timescale of the process of transcriptional bursting itself (1-3 min) cannot be detected. Notably, this is not a limit of the cpHMM method, but, rather is inherent to the system—in order to infer bursting parameters, we must observe bursts and, in order to infer a change in parameters, we must have access to bursting activity that reflects this change. Thus, the characteristic frequency of bursts sets an insurmountable resolution limit for any kind of bursting parameter inference. To illustrate this limitation, we simulated three scenarios in which k_{on} decreases to 0 over periods 15, 5, and 1 min in length. We then sought to recover the trend in k_{on} . To emphasize that the limitations are not specific

to cpHMM, but rather, are an inevitable consequence of the structure of the system, we used the *true* promoter trajectories (those used to generate the simulated data) to estimate k_{on} . These estimates thus represent the absolute best-case scenario for parameter inference, in which we recover the underlying behavior of the system exactly.

Figure S33G-H indicate that, even with perfect knowledge of the bursting state at each gene locus, it is not possible to recover a change in the on rate that happens within the span of one minute. These results show that—even under ideal circumstances—there exists a time scale below which time-dependent burst parameter inference will fail to detect shifts in burst parameter values. The absence of bursts following the transition means that, not only are we unable to accurately recover the true trend, but we are also unable even to determine whether any decrease in k_{on} occurred (on any time scale). Thus, in this scenario, it would be impossible to determine that a modulation in the bursting parameters—as opposed to a transition into some alternative, silent state—drives the onset of quiescence.

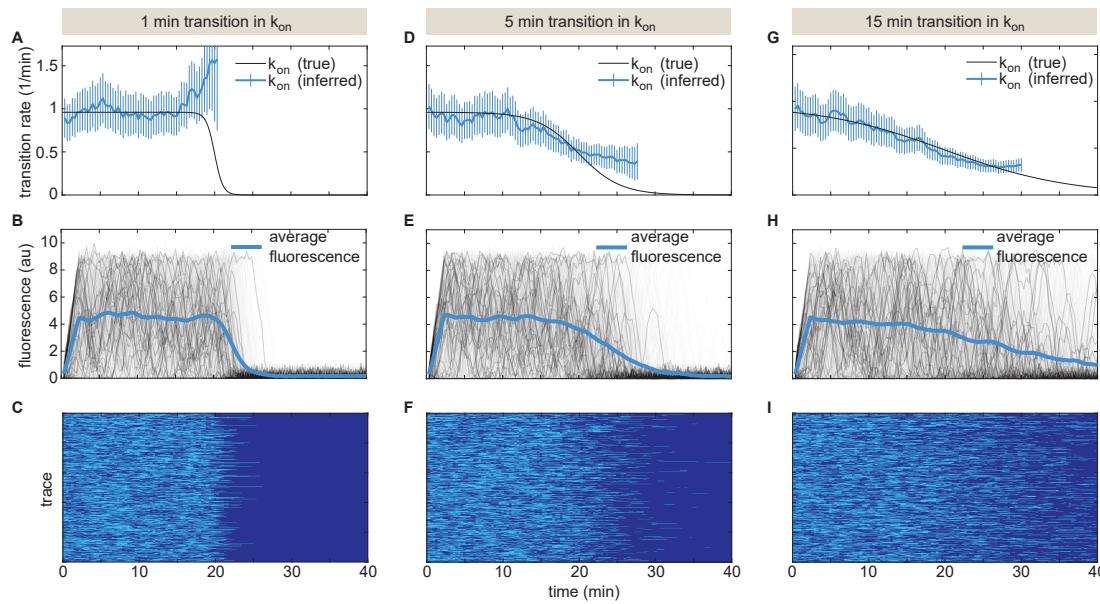


Fig. S33. Limitations of burst parameter inference. (A-C) 15-min transition. (A) Black curve indicates true k_{on} value as a function of time and blue curve indicates inferred value. Because the change unfolds on a time scale that is much slower than the bursting timescale, it is possible to accurately recover the underlying k_{on} trend from the fluorescent traces. (B) The temporal trend in the average fluorescence across simulated traces (blue curve) reflects this gradual decrease in k_{on} . Note that variation in simulated traces (gray) unfolds on a significantly faster timescale than the change in the mean. (C) Visualization of promoter switching. Light blue indicates ON periods and dark blue indicates OFF periods. The observation that bursts of activity are interspersed throughout the k_{on} transition makes it possible to recover the temporal trend. (D-F) 5 min transition. (D) We are able to recover first half of k_{on} trend, but due to the speed of transition, insufficient active traces remain to permit the accurate recovery of the full profile. (E, F) The onset of quiescence is much starker than in the 15 min case. Because the transition happens faster than in (A-C), there are fewer bursts that unfold during the transition and, hence, we have fewer reference points with which to infer the underlying trend. (G-I) 1 min transition. Here the k_{on} transition occurs on the timescale of a single burst. As a result, we are unable to recover the temporal trend. (H-I) The period of observation is divided in a nearly binary fashion. (A,D,G, error bars indicate 95% confidence interval of exponential fits used to estimate k_{on}).

J. Determining the RNAP dwell time using autocorrelation. In order to conduct cpHMM inference, it is necessary to specify the number of time steps w required for an RNAP molecule to traverse the reporter gene,

$$w = \frac{\tau_{\text{elong}}}{\Delta\tau}, \quad [114]$$

where $\Delta\tau$ is set by the temporal resolution of our data acquisition and τ_{elong} is the elongation time which is unknown *a priori*. Past studies have estimated elongation rates for other systems involved in early patterning in the *Drosophila* embryo, but there is substantial disparity between the reported values. A live imaging study of transcriptional activity driven by the *hunchback* P2 enhancer reported an elongation rate of $1.4 - 1.7 \text{ kb min}^{-1}$ (1). However, a recent study of the same regulatory element reported elongation rates of $2.4 - 3.0 \text{ kb min}^{-1}$ —nearly twice as fast (27). These results suggested that RNAP elongation rates measured for other systems might not apply to our *eve* stripe 2 reporter. Thus, in order to ensure the validity of our inference, we developed an approach that uses the mean autocorrelation function of experimental fluorescence traces to estimate the elongation time directly from our data.

The autocorrelation function $R_F(\tau)$ quantifies the degree to which a signal, $F(t)$, is correlated with a lagged version of itself, $F(t - \tau)$, and is given as a function of the time delay, τ , between the two signal copies being compared such that

$$R_F(\tau) = \frac{E[(F(t) - \mu_f)(F(t - \tau) - \mu_f)]}{\sigma_f^2}, \quad [115]$$

where μ_f is the average observed fluorescence, σ_f is the standard deviation of the fluorescence and E denotes the expectation value operator. As illustrated in [Figure S34A](#), the fact that it takes RNAP molecules some finite amount of time to traverse the gene implies that the observed fluorescence at a transcriptional locus at some time t , $F(t)$, will be correlated with preceding fluorescence values $F(t - \tau)$ so long as $\tau < \tau_{\text{elong}}$ because the two time points will share a subset of the same elongating RNAP molecules. As τ increases, the correlation between $F(t)$ and $F(t - \tau)$ due to these shared RNAP molecules will decay in a linear fashion until it reaches zero when $\tau = \tau_{\text{elong}}$ ([Figure S34B](#), blue curve).

The dramatic change in the slope of the autocorrelation function that occurs at $\tau = \tau_{\text{elong}}$ can be used to estimate the elongation time of the system; however, it is not the only feature present in [Equation 115](#). Because the time series of promoter states constitutes a Markov chain, the instantaneous promoter state and, therefore, the instantaneous rate of RNAP loading, exhibits a nontrivial, positive autocorrelation due to the promoter switching dynamics of the system. For instance, if it takes the promoter an average of 1 minute to switch states, then it is clear that promoter activity for $\tau < 1$ min will be strongly correlated with itself. Thus, we see that the rates of promoter switching dictate the speed with which this “dynamic” autocorrelation decreases with increasing τ . More precisely, the dynamics autocorrelation will take the form of a decaying exponential in τ , with the time scale set, approximately, by the second largest eigenvalue of the Markov chain’s transition rate matrix ([Figure S34B](#), red curve)

$$R_P(\tau) \sim e^{-\lambda_2 \tau}. \quad [116]$$

Thus, the observed autocorrelation function contains, at a minimum, signatures of both the finite RNAP dwell time (τ_{elong}) and of promoter switching dynamics. As a result, inferring elongation times from the change in slope in the mean autocorrelation is often relatively subtle in practice.

A theoretical analysis of $R_F(\tau)$ indicated that the second derivative of the mean autocorrelation function reliably exhibits a peak that can be used to read out the value of τ_{elong} . [Figure S34C](#) shows the analytic prediction for the autocorrelation and second derivative when τ_{elong} is equal to 7 time steps ($w = 7$). We confirmed that the same second derivative approach works in the context of stochastic simulations using realistic parameters for the *eve* stripe 2 system ([Figure S34D](#)). Having confirmed the efficacy of the autocorrelation method for simulated data, we next applied the same technique to uncover τ_{elong} for our experimental traces.

The black profile in [Figure S34E](#) indicates the form of the autocorrelation second derivative for the set of traces used for cpHMM inference. We observed that, while there is a definite inflection point, the peak for the experimental data is much broader than for simulated traces. The most likely cause of this feature is the existence of variability in τ_{elong} (see below). From comparisons of the position of the second derivative peak for experimental traces with simulated profiles, we concluded that an elongation time of $w = 7$ ($\tau_{\text{elong}} = 140$ s) best characterized our data ([Figure S34E](#), green curve). This implies that

$$\begin{aligned} v_{\text{elong}} &= \frac{6444 \text{ bp}}{140 \text{ s}} \\ &= 46 \text{ bp s}^{-1} \\ &= 2.8 \text{ kb min}^{-1}, \end{aligned} \quad [117]$$

where the length used represents the distance from the start of the MS2 step loop sequence to the end of the 3' end of the construct. Interestingly, this elongation rate falls within the $2.4 - 3.0 \text{ kb min}^{-1}$ range reported in [\(27\)](#).

[Figure S34F](#) shows how a simple adjustment to our simulation approach, wherein the elongation time steps w for individual RNAP molecules were drawn from a Gaussian distribution with mean $\mu_w = 7$ and standard deviation $\sigma_w = 2.5$ time steps can qualitatively reproduce the wider profile observed in experimental data, indicating that our observations are indeed consistent with the presence of variability in RNAP elongation times. Additional experimental and theoretical work will be necessary to uncover the biological source of this variability.

In light of the ambiguity introduced by the broad second derivative peak exhibited by our experimental data, we also verified that our inference was robust to the choice of τ_{elong} , testing cases where $\tau_{\text{elong}} = 120$ s and $\tau_{\text{elong}} = 160$ s (see below).

cpHMM inference is insensitive to small changes in RNAP dwell time. Due to the uncertainty in our estimate of τ_{elong} , we conducted sensitivity estimates to ensure that our inference results were robust to our input assumption for w . As shown in [Figure S35](#), we conducted cpHMM inference on our experimental data assuming different values of w . Based upon our autocorrelation analysis, w values of 6, 7 and 8 seemed the most plausible candidates for the average system elongation time (see [Figure S34E](#)). While small quantitative difference are apparent across these three cases (with a median coefficient of variation of 11%), the results for different values of w showed a constant offset throughout the embryo, such that qualitative trends were largely robust to the assumed w value.

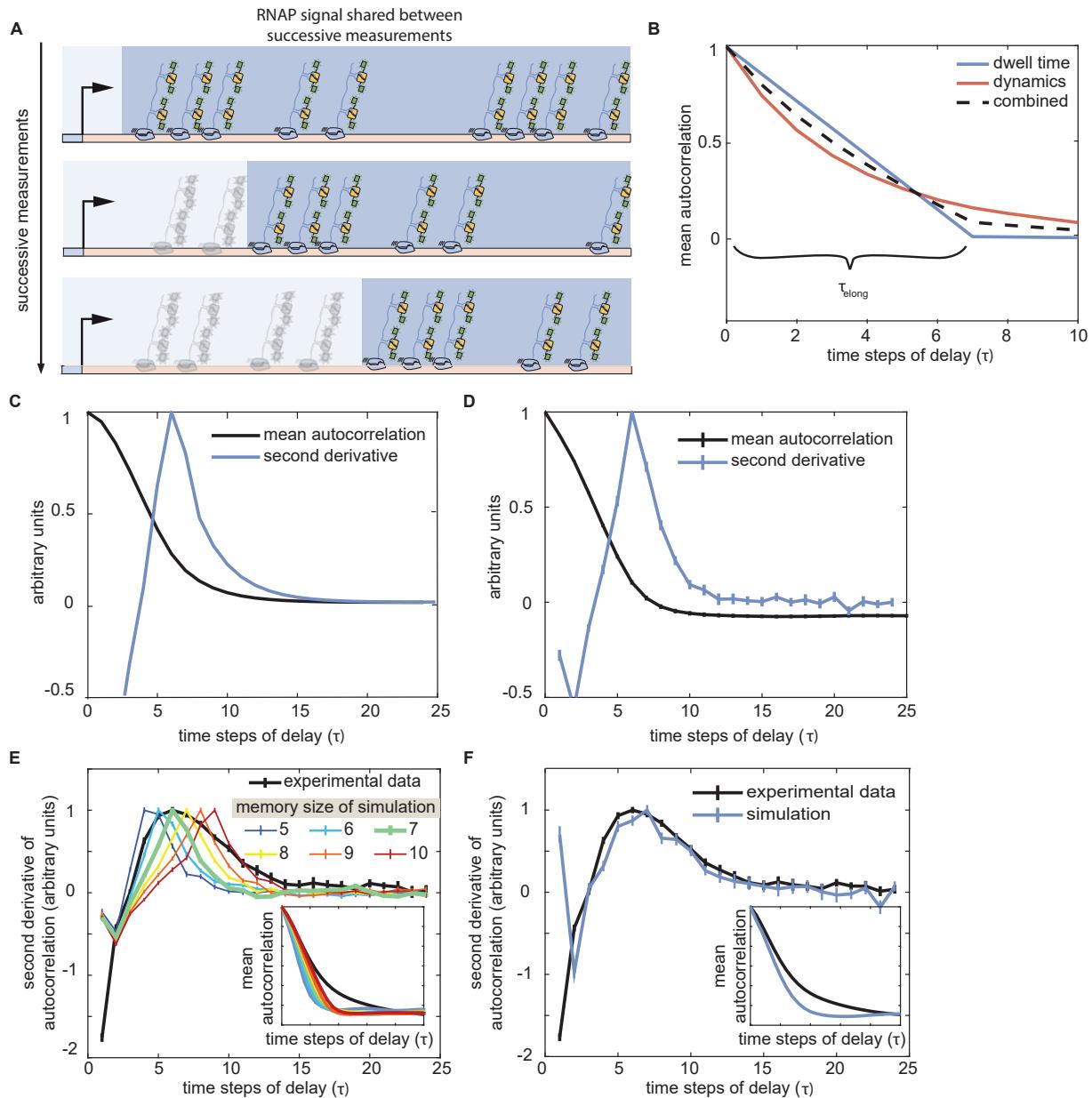


Fig. S34. Using the autocorrelation of the fluorescence signal to estimate RNAP dwell time. (A) It takes a finite amount of time for RNAP molecules to transcribe the full length of the reporter gene. As a result, successive fluorescence measurements will contain some of the same GFP-tagged RNAP molecules. Dark blue-shaded regions indicate the subset of RNAP molecules that are present on the gene for successive measurements. (B) This overlap causes successive measurements to be correlated, and the degree of correlation due to the overlap decays linearly, reaching zero when the separation between measurements is equal to the elongation time, τ_{elong} (blue curve). However, the trace autocorrelation function contains other signatures that can obscure the inflection induced by RNAP elongation dynamics. For instance, successive time points also exhibit correlation due to the promoter switching dynamics (red curve). (C) Theoretical analysis of the autocorrelation function and (D) stochastic simulations indicate that the second derivative of the mean autocorrelation function (dark blue curves) can be used to find the structural break in the function (black curves) that corresponds to τ_{elong} . Here, a peak at 6 time steps of delay indicates an elongation time of 7 time steps (140 s). (E) Simulated traces with elongation time of 7 time steps (green curve) exhibit a peak in the second derivative that coincides with the maximum of the experimental curve. Inset plots show corresponding mean autocorrelation curves for experimental data and simulations. (F) Stochastic simulations in which we allow for variation in elongation times distributed around a mean of 7 time steps qualitatively recapitulates the observed curve. (C-F, second derivative profiles depicted here are normalized relative to their maximum value for ease of depiction.)

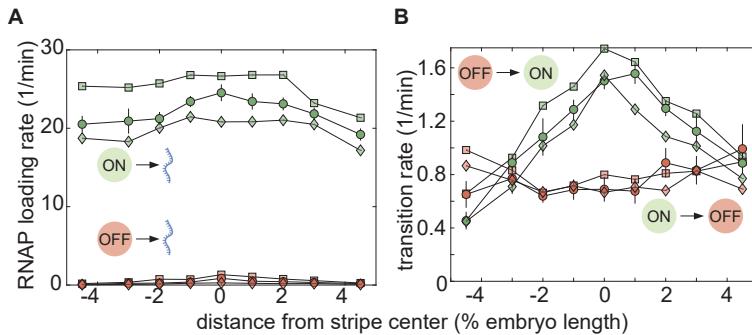


Fig. S35. Elongation time sensitivities. Square, circle, and diamond symbols denote inference results for memory time window values w of 6, 7, and 8, respectively. $w = 7$ plots are bolded. Bootstrap errors are shown for $w = 7$ case. (A) Initiation rates are robust to w assumption. (B) Transition rates also exhibit high degree of robustness to the w used for inference. (Error bars indicate magnitude of difference between first and third quartiles of cpHMM inference results for bootstrap samples of experimental data).

Movies

Movie S1. [Video 1. Transcriptional activity of eve stripe 2 reported by MS2.](#) Raw MS2 signal where fluorescent puncta report on the number of actively transcribing RNAP molecules.

Movie S2. [Video 2. Mean rate of transcription of eve stripe 2 reported by MS2.](#) Nuclei false colored by their time-averaged transcriptional activity (up to the depicted time point).

Movie S3. [Video 3. Transcriptional time window.](#) Nuclei along the stripe false colored after the duration of their transcriptional time window.

Movie S4. [Video 4. Fraction of active nuclei.](#) Nuclei along the stripe false colored according to whether they engaged in transcription at any time point during the nuclear cycle.

Movie S5. [Video 5. Fluorescent puncta contain sister chromatids.](#) Fluorescent puncta transiently separate to reveal the presence of sister chromatids as shown by the white circles throughout the movie.

Movie S6. [Video 6. Real-time inferred promoter states.](#) Real-time inference of effective promoter ON (green) and OFF (red) state in individual nuclei.

Movie S7. [Video 7. Average embryo containing all inputs and the output.](#) Average concentrations of Bicoid (blue), Hunchback (red), Krüppel (green) and Giant (yellow) combined with the average transcriptional activity of the *eve* reporter (purple). (Hunchback, Krüppel and Giant data obtained from (3)).

References

1. HG Garcia, M Tikhonov, A Lin, T Gregor, Quantitative imaging of transcription in living drosophila embryos links polymerase activity to patterning. *Curr Biol* **23**, 2140–5 (2013).
2. JP Bothma, et al., Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living drosophila embryos. *Proc Natl Acad Sci U S A* **111**, 10598–10603 (2014).
3. JO Dubuis, R Samanta, T Gregor, Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Syst Biol* **9**, 639 (2013).
4. B Efron, T Hastie, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. (Cambridge University Press, New York, NY, USA), 1st edition, (2016).
5. JM Halstead, et al., Translation. an rna biosensor for imaging the first round of translation from single cells to living animals. *Science* **347**, 1367–671 (2015).
6. BA Edgar, GM Odell, G Schubiger, Cytoarchitecture and the patterning of fushi tarazu expression in the drosophila blastoderm. *Genes Dev* **1**, 1226–37 (1987).
7. DT Gillespie, General method for numerically simulating stochastic time evolution of coupled chemical-reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
8. CC Fowlkes, et al., A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. *Cell* **133**, 364–74 (2008).
9. A Berrocal, NC Lammers, HG Garcia, MB Eisen, Kinetic sculpting of the seven stripes of the drosophila even-skipped gene. *bioRxiv* (2018).

10. J Jiang, T Hoey, M Levine, Autoregulation of a segmentation gene in drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. *Genes Dev* **5**, 265–77 (1991).
11. CL Luengo Hendriks, et al., Three-dimensional morphology and gene expression in the drosophila blastoderm at cellular resolution i: data acquisition pipeline. *Genome Biol.* **7**, R123 (2006).
12. BA Edgar, G Schubiger, Parameters controlling transcriptional activation during early drosophila development. *Cell* **44**, 871–7 (1986).
13. A Coulon, D Larson, Fluctuation analysis: Dissecting transcriptional kinetics with signal theory. *Methods Enzymol.* (2016).
14. AM Corrigan, E Tunnacliffe, D Cannon, JR Chubb, A continuum model of transcriptional bursting. *Elife* **5** (2016).
15. C Bishop, *Pattern recognition and machine learning*. (Springer, New York), (2006).
16. AJ Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inform. Theory* **13**, 260–269 (1967).
17. GA Rice, MJ Chamberlin, CM Kane, Contacts between mammalian rna polymerase ii and the template dna in a ternary elongation complex. *Nucleic Acids Res.* **21**, 113–118 (1993).
18. M Rabinowitz, Studies on the cytology and early embryology of the egg of drosophila melanogaster. *J. Morphol.* **69**, 1–49 (1941).
19. AW Sherman, ML McCleland, PH O'Farrell, Developmental control of late replication and s phase length. *Curr Biol* **20**, 2067–77 (2010).
20. SC Little, G Tkacik, TB Kneeland, EF Wieschaus, T Gregor, The formation of the bicoid morphogen gradient requires protein movement from anteriorly localized mrna. *PLoS Biol* **9**, e1000596 (2011).
21. B Zoller, SC Little, T Gregor, Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell* **175**, 835–847 e25 (2018).
22. SL McKnight, J Miller, O. L., Electron microscopic analysis of chromatin replication in the cellular blastoderm drosophila melanogaster embryo. *Cell* **12**, 795–804 (1977).
23. SL McKnight, J Miller, O. L., Post-replicative nonribosomal transcription units in d. melanogaster embryos. *Cell* **17**, 551–63 (1979).
24. T Gregor, EF Wieschaus, AP McGregor, W Bialek, DW Tank, Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141–52 (2007).
25. MD Petkova, G Tkacik, W Bialek, EF Wieschaus, T Gregor, Optimal decoding of cellular identities in a genetic network. *Cell* **176**, 844–855 e15 (2019).
26. S Small, A Blair, M Levine, Regulation of even-skipped stripe 2 in the drosophila embryo. *EMBO J* **11**, 4047–57 (1992).
27. T Fukaya, B Lim, M Levine, Rapid rates of pol ii elongation in the drosophila embryo. *Curr Biol* **27**, 1387–1391 (2017).