Binary transcriptional control of pattern formation in development

Nicholas C Lammers[1†], **Vahe Galstyan**[2,3†], **Armando Reimer**[1], **Sean A Medin**[4], **Chris H Wiggins**[5,6,7,8*], **Hernan G Garcia**[1,4,9,10*]

***For correspondence:**
hggarcia@berkeley.edu (HGG);
chris.wiggins@columbia.edu (CHW)

[†]These authors contributed equally to this work

[1]Biophysics Graduate Group, University of California at Berkeley, Berkeley, California; [2]Biochemistry and Molecular Biophysics Option, California Institute of Technology, Pasadena, California; [3]Department of Physics, Columbia University, New York, New York; [4]Department of Physics, University of California at Berkeley, Berkeley, California; [5]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York; [6]Data Science Institute, Columbia University, New York, New York; [7]Department of Systems Biology, Columbia University, New York, New York; [8]Department of Statistics, Columbia University, New York, New York; [9]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California; [10]Institute for Quantitative Biosciences-QB3, University of California at Berkeley, Berkeley, California

Abstract During development, stochastic promoter switching between active and inactive states results in transcriptional bursts. We tested whether burst kinetics are sufficient to quantitatively recapitulate the formation of patterns of accumulated mRNA in *Drosophila* embryos by dissecting the transcriptional dynamics of *even-skipped* stripe 2. Using a novel memory-adjusted hidden Markov model, single-cell live imaging and theoretical modeling, we show that the regulation of bursting in space and time alone is insufficient to predict stripe formation. In addition to bursting, we discovered that the duration of the window of time over which genes transcribe is regulated, and that this binary (on/off) control of where and when gene expression occurs, not transcriptional bursting, is the main regulatory strategy governing stripe formation. Thus, a quantitative description of the regulation of both bursting and the transcriptional time window are necessary to capture the full complement of molecular rules governing the transcriptional control of pattern formation.

Introduction

During embryonic development, tightly choreographed patterns of gene expression—shallow gradients, sharp steps, narrow stripes—specify cell fates (***Gilbert, 2010***). These patterns arise from decisions made by individual cells to transcribe a particular gene (or not) in response to the nuclear concentrations of input activators and repressors, which are themselves regulated by other genes in the developmental network. In the last few years, a picture of how transcription is realized at individual loci has emerged in which promoters stochastically transition between transcriptional ON and OFF states (***Figure 1***A and ***Golding et al.*** (***2005***); ***Little et al.*** (***2013***)). In this scenario, RNA polymerase (RNAP) molecules are actively loaded only while the promoter is in the ON state, and thus promoter switching between OFF and ON states leads to punctuated bursts of transcriptional activity.

Transcriptional bursting in development has been proposed to constitute one of the main molecular forces behind the establishment of the gene expression patterns that dictate animal

body plans ***Little et al.*** (***2013***); ***Xu et al.*** (***2015***); ***Bothma et al.*** (***2014***); ***Fukaya et al.*** (***2016***); ***Zoller et al.*** (***2017***)). For example, the prevailing paradigm holds that a stripe of cytoplasmic mRNA within the embryo of the fruit fly *Drosophila melanogaster* arises because nuclei in the middle of the stripe transcribe with a higher burst frequency (controlled mainly by k_{on}) than nuclei on the boundaries of the pattern, as shown in ***Figure 1***B and C.

Here, we put the hypothesis that the spatial modulation of transcriptional bursting is sufficient to explain the formation of patterns of accumulated cytoplasmic mRNA to a stringent quantitative test in the context of the widely studied stripe 2 of the *even-skipped* (*eve*) gene in the developing fruit fly embryo (***Small et al., 1992***). Recently, single-cell live imaging was used to quantify the dynamics of transcriptional activity for this gene, revealing marked fluctuations in the number of actively transcribing RNAP molecules over time (***Garcia et al., 2013***; ***Bothma et al., 2014***). We develope a novel memory-adjusted hidden Markov model (mHMM) that captures bursting dynamics in individual nuclei as development progresses by inferring the instantaneous transcriptional state of the promoter at each time point (i.e., whether it is in the ON or OFF state) from these data. Using this model, we confirmed that bursting frequency, and not duration or intensity, is the main parameter under molecular control along the axis of the embryo. However, we discovered that, when confronted with estimates of accumulated transcript levels across *eve* stripe 2, transcriptional bursting fails to *quantitatively* recapitulate stripe formation. In contrast to the prevailing paradigm, we discovered that nuclei at the boundaries of *eve* stripe 2 undergo a coordinated transition into a transcriptionally quiescent state, and that this regulation of the timing with which promoters disengage (once and for all) from transcription is the main driver of pattern formation. Finally, by expanding our model to consider time-dependent transcriptional bursting, we detected significant temporal variations in the bursting parameters and explored hypotheses regarding the molecular mechanisms driving the onset of promoter quiescence. We conclude that a quantitative description of both the regulation of promoter bursting and the duration of the transcriptional time window needs to be adopted in order to reveal the molecular rules behind the transcriptional control of pattern formation and to reach a predictive understanding of development.

Results

A Quantitative Model of Pattern Formation by Transcriptional Bursting

Figure 1 presents a scenario in which the graded accumulation of cytoplasmic mRNA that leads to the formation of gene expression patterns is dictated by the modulation of burst frequency along the embryo. For example, in the case of the stripe shown in ***Figure 1***C, nuclei in the middle of the stripe transcribe with a higher bursting frequency than nuclei on the stripe boundary (compare ***Figure 1***D and E). In this section, we will turn this cartoon model into a precise mathematical statement in order to *quantitatively* predict how transcriptional bursting dictates pattern formation.

As a result of bursting, each promoter spends only a fraction of time in the ON state, given by $k_{\text{on}}(x,t)/(k_{\text{on}}(x,t)+k_{\text{off}}(x,t))$. When in this ON state, promoters transcribe at a rate $r(x,t)$. As a result, the mean rate of transcription is given by the product of the fraction of time spent in the ON state, and the transcription rate when in this active state (***Peccoud and Ycart, 1995***; ***Sanchez et al., 2011***), namely,

$$\langle \text{transcription rate} \rangle(x,t) = \underbrace{r(x,t)}_{\substack{\text{RNAP loading}\\ \text{rate}}} \times \underbrace{\frac{k_{\text{on}}(x,t)}{k_{\text{on}}(x,t)+k_{\text{off}}(x,t)}}_{\substack{\text{fraction of time}\\ \text{in ON state}}}. \tag{1}$$

Note that, in writing this equation, we have assumed a very general model that goes beyond the scenario put forth in ***Figure 1***D and E to allow for all bursting parameters to change both in space (x) and in time (t).

Modulation of the mean rate of transcription shown in ***Equation 1*** will lead to a pattern of accumulated cytoplasmic mRNA, which will ultimately result in the formation of a protein pattern

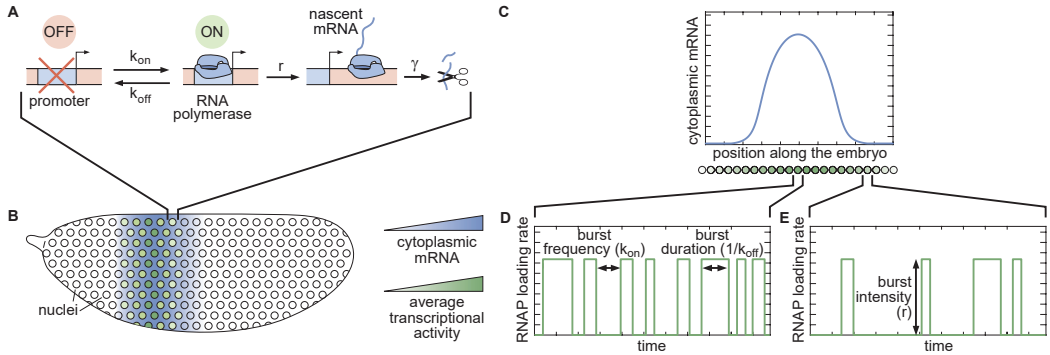


Figure 1. Model of pattern formation by transcriptional bursting. (**A**) Model of a promoter switching stochastically between transcriptionally active and inactive states that gives rise to bursts in transcription. (**B, C**) At different positions along the embryo, the rates defined in (**A**) dictate the average transcriptional activity per nucleus which, in turn, determines the cytoplasmic mRNA distribution. (**D, E**) These promoter switching parameters dictate the burst duration, burst frequency, and burst intensity. In particular, it has been proposed that the modulation of the burst frequency along the embryo is the main parameter dictating the spatial distribution of cytoplasmic mRNA.

that feeds back into the developmental network. However, a quantitative model that connects single-cell transcriptional activity to the cytoplasmic accumulation of mRNA and the formation of macroscopic gene expression patterns needs to account not only for transcriptional bursting, but also for the decay of cytoplasmic mRNA given by the degradation rate γ. The net rate of mRNA production is the rate of mRNA synthesis minus the rate of mRNA degradation such that the amount of cytoplasmic mRNA is described by the differential equation

$$\frac{\mathrm{dmRNA}}{\mathrm{d}t}(x,t) = \underbrace{r(x,t)\frac{k_{\mathrm{on}}(x,t)}{k_{\mathrm{on}}(x,t)+k_{\mathrm{off}}(x,t)}}_{\langle\text{transcription rate}\rangle} - \underbrace{\gamma\,\mathrm{mRNA}(x,t)}_{\text{degradation rate}} \quad . \tag{2}$$

In this version of the model we are ignoring the effects of mRNA diffusion throughout the embryo (see ***Appendix 1*** for more details on this and other assumptions). Moreover, following both fixed-tissue and live-imaging studies of transcriptional bursting in development, we also assume that bursting is modulated along the axis of the embryo, but does not change during the nuclear cycle (***Pare et al., 2009***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Fukaya et al., 2016***; ***Desponds et al., 2016***; ***Zoller et al., 2017***). Following the initial rise in mRNA levels at the onset of transcription, this case is equivalent to demanding steady-state in ***Equation 2*** with $d\mathrm{mRNA}/dt = 0$, resulting in a steady-state distribution of mRNA given by

$$\mathrm{mRNA}(x) = \frac{1}{\gamma} r(x) \frac{k_{\mathrm{on}}(x)}{k_{\mathrm{on}}(x)+k_{\mathrm{off}}(x)}, \tag{3}$$

From this equation, we see that, in this steady-state scenario, spatial profiles of mRNA accumulation are created by the graded modulation of bursting parameters along the embryo which dictate the mean rate of transcription. For instance, one such scenario would be the modulation of burst frequency (through the modulation of k_{on}) along the embryo shown in ***Figure 1***C and D. More importantly, ***Equation 3*** provides a means to quantitatively test the current paradigm that transcriptional bursting can quantitatively recapitulate the formation of mRNA patterns in development. To determine whether burst frequency is the main molecular parameter under control in pattern formation and establish whether this regulation is sufficient to dictate pattern formation, it is necessary to infer bursting parameters and the amount of produced mRNA in embryos. Specifically, at each position x of the embryo we need to measure the values of k_{on}, k_{off} and r and use these values to calculate the predicted mRNA profile (left-hand side of ***Equation 3***), which will then be compared to direct measurements of the mRNA profile.

Revealing Transcriptional Bursting and the Mean Transcription Rate in Developing Embryos

In order to test the model of pattern formation by transcriptional bursting in ***Equation 3***, we first asked whether at least one of the promoter switching parameters (r, k_{on}, and k_{off}) is regulated throughout the stripe. In this scenario, ***Equation 1*** predicts that the mean rate of transcription across the stripe will also be modulated. We carried out this test by quantifying transcription of stripe 2 of *eve* in the fruit fly. This stripe is controlled by the combined action of two activators, Bicoid and Hunchback, and two repressors, Giant and Krüppel (***Frasch and Levine, 1987***; ***Small et al., 1992***). These activators initially direct a broad domain of transcriptional activity that is later refined by the repressors (***Small et al., 1992***; ***Bothma et al., 2014***).

To reveal how single-cell transcriptional dynamics underlie macroscopic patterns of gene expression (***Figure 1***), we imaged the transcription of an already established *eve* stripe 2 reporter using the MS2 system (***Garcia et al., 2013***; ***Bothma et al., 2014***). As shown in ***Figure 2***A, transcripts of a reporter gene driven by the *eve* stripe 2 enhancer and the *eve* promoter contain repeats of a DNA sequence that, when transcribed, form stem loops (***Bertrand et al., 1998***). These stem loops are recognized by the maternally provided MS2 coat protein fused to GFP (MCP-GFP) (***Figure 2***A). As a result, sites of nascent transcript formation appear as fluorescent puncta within individual nuclei (***Figure 2***B and Video 1). This fluorescence can be calibrated using single-molecule FISH in order to estimate the number of RNAP molecules actively transcribing the gene as a function of time (***Figure 2***C, see Materials and Methods and ***Garcia et al.*** (***2013***)).

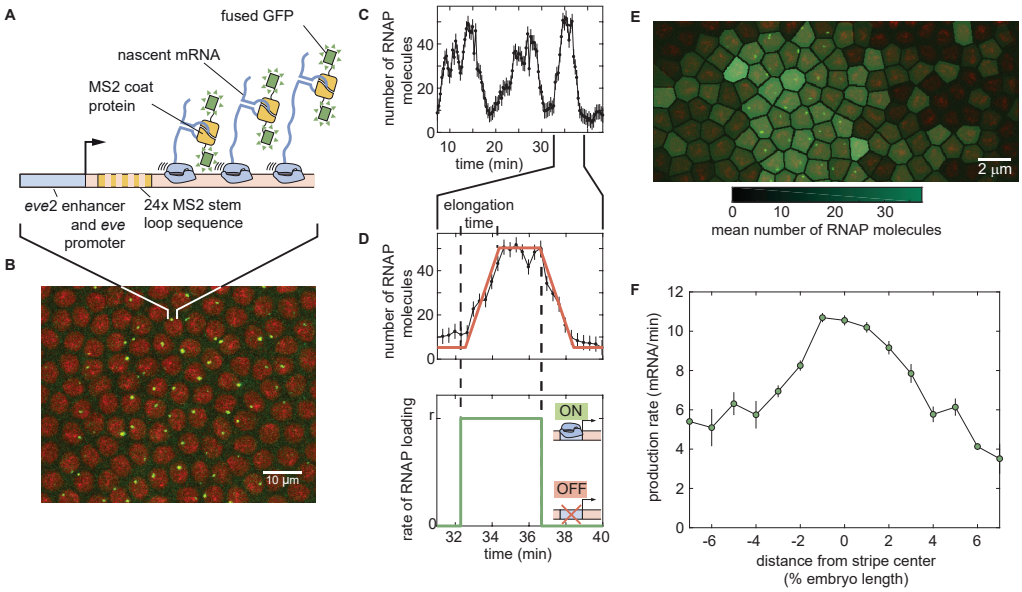


Figure 2. The mean transcription rate is modulated in a stripe-like fashion in *eve* stripe 2. (A) MS2 stem loops introduced in an *eve* stripe 2 reporter gene are bound by MS2 coat protein fused to GFP. **(B)** Sites of nascent transcript formation appear as green fluorescent puncta whose intensity reports on the number of actively transcribing RNAP molecules. Nuclei are visualized through a fusion of RFP to Histone. **(C)** The number of transcribing RNAP molecules fluctuates in a punctuated fashion over time. **(D)** We associate these peaks in the number of RNAP molecules with the transient switching of the promoter from the OFF to the ON state. **(E, F)** Mean transcriptional activity in (E) an individual embryo and (F) across 11 embryos along the stripe as a result of transcriptional bursting. (C,D error bars obtained from estimation background fluorescent fluctuations as described in Materials and Methods and ***Garcia et al.*** (***2013***); F, average over 11 embryos, error bars are generated via bootstrap re-sampling and approximate the standard error of the mean)

Figure 2–Figure supplement 1. Aligning stripes from multiple embryos.

The peaks and troughs in the number of active RNAP molecules (***Figure 2***C) have been related to the rate of RNAP loading at the *eve* promoter by assuming that promoter loading is "burst-like", with the promoter loading RNAP molecules onto the gene at a constant rate over discrete periods

of time (***Figure 2***D and ***Garcia et al.*** (***2013***); ***Bothma et al.*** (***2014***)). This and other evidence from live imaging (***Bothma et al., 2014***; ***Fukaya et al., 2016***; ***Desponds et al., 2016***), as well as data from fixed-tissue approaches (***Pare et al., 2009***; ***Bothma et al., 2014***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Zoller et al., 2017***), support the promoter-switching model in ***Figure 1***A.

Using MS2, we measured the mean transcriptional rate of our *eve* reporter construct over multiple embryos as described in ***Appendix 5*** and ***Figure 2–Figure Supplement 1***, and found that this rate is modulated along the embryo's axis (***Figure 2***E and F, Video 2 and Materials and Methods). Thus, we conclude that at least one of the promoter switching parameters (r, k_{on}, and k_{off}) is under regulatory control across the stripe. However, data such as shown in ***Figure 2***C cannot reveal *which* parameter is actually subject to control by the input transcription factors to *eve* stripe 2.

The Presence of Sister Chromatids Suggests an Effective Three-State Model of Transcription

A fundamental assumption in deriving the prediction in ***Equation 3*** is that transcription of *eve* stripe 2 can be described by a 2-state model of promoter switching (***Figure 1***A). However, close examination of our data revealed that individual fluorescent puncta from our *eve* reporter often transiently separated into two puncta (***Figure 3***A and Video 3). We and others have hypothesized that these two puncta correspond to sister chromatids that spend most of the time localized within the same diffraction-limited spot (***Little et al., 2011***). This hypothesis is supported by the fact that, in the early development of *D. melanogaster*, the genome is rapidly replicated at the beginning of each nuclear cycle (***Rabinowitz, 1941***; ***Shermoen et al., 2010***), well before transcription becomes detectable after anaphase (***Shermoen and O'Farrell, 1991***; ***Garcia et al., 2013***). Sister chromatids resulting from this replication event stay in close proximity (***Little et al., 2013***; ***Senaratne et al., 2016***).

If each fluorescent punctum contains two promoters, then it is necessary to revisit the widely used 2-state model. In this revised scenario, each promoter on one of the sister chromatids undergoes fast ON/OFF switching. Therefore each punctum can be in one of 3-states: (0) both promoters OFF, (1) one promoter ON and the other OFF, and (2) both promoters ON (***Figure 3***B). States (1) and (2) are expected to exhibit different rates of RNAP loading, r_1 and r_2, respectively. Previous studies on this and other reporter constructs have posited the existence of multiple transcriptional states, each equipped with its unique rate of RNAP loading (***Bothma et al., 2014***; ***Corrigan et al., 2016***; ***Desponds et al., 2016***).

Our interpretation that fluorescent puncta contain two active promoters suggests three constraints on the bursting parameters in the model in ***Figure 3***B. First, the probability of both promoters transitioning simultaneously should be negligible; we expect no transitions between states (0) and (2) such that $k_{02} = k_{20} = 0$. Second, if these two promoters transcribe independently, then state (2) will have double the loading rate of state (1) such that $r_2 = 2r_1$. Finally, if the promoters switch between their states in an independent manner, then there will be an extra constraint on their transitions 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, $k_{10} = 2k_{21}$ (see ***Appendix 2*** for further details). While this independence of sister chromatids is supported by recent single-molecule FISH experiments (***Little et al., 2011***; ***Zoller et al., 2017***), classic electron microscopy work suggests a scenario in which sister chromatids are tightly correlated in their transcriptional activity (***McKnight and Miller, 1977***). In light of this uncertainty regarding chromatid independence, we elected to employ a general 3-state model that makes no assumptions about the nature and strength of sister chromatid interactions.

In light of these considerations, we revised the framework presented in ***Equation 3*** to account for multiple states (***Figure 3***B). In this revised model, the steady-state distribution of mRNA is given

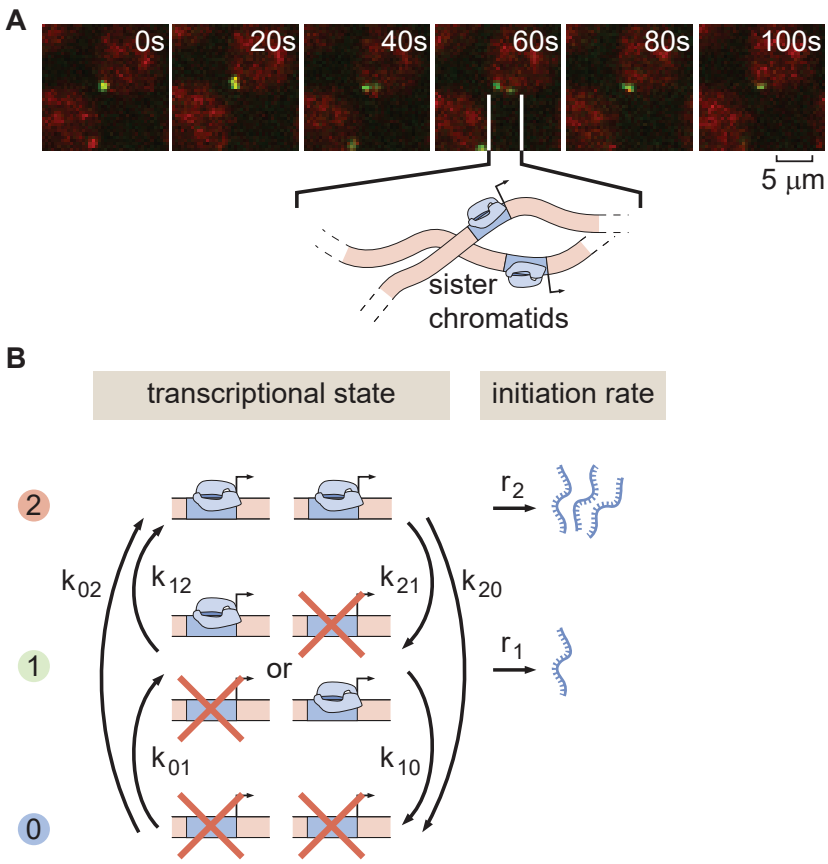


Figure 3. Fluorescent puncta report on the combined transcription of sister chromatids. (A) DNA is rapidly replicated in each cycle of early embryonic development such that each fluorescent punctum observed is actually composed of two distinct transcriptional loci within a diffraction-limited spot, each one corresponding to a sister chromatid. **(B)** Revised three-state model of promoter switching within a fluorescent punctum that accounts for the combined action of both sister chromatids.

by

$$\text{mRNA}(x) = \frac{1}{\gamma}\left(r_1 \underbrace{\frac{k_{01}k_{21}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}}}_{\substack{\text{probability of} \\ \text{being in state (1)}}} + r_2 \underbrace{\frac{k_{01}k_{21}}{k_{10}k_{12} + k_{01}k_{21} + k_{01}k_{12}}}_{\substack{\text{probability of} \\ \text{being in state (2)}}} \right). \tag{4}$$

Here, the amount of cytoplasmic mRNA now depends on the probability of the punctum being in one of the effective states multiplied by the rate of RNAP loading when in that state. Note that, for clarity, we have omitted the spatial dependence of all rates in ***Equation 4*** (see ***Appendix 1*** for details of this derivation). In addition, we have assumed that $k_{02} = k_{20} = 0$ as a means to further simplify this expression. This assumption will be tested in detail below. Regardless, of these molecular details, ***Equation 4*** provides a revised prediction for how promoter switching parameters in the 3-state model dictate the formation of the stripe.

A Memory-Adjusted Hidden Markov Model Infers Bursting Parameters from Live Imaging Data

In order to test the revised prediction for how transcriptional bursting in the 3-state model determines spatial mRNA profiles in ***Equation 4***, we moved on to extracting the kinetic parameters governing this model as a function of the nucleus' position along the stripe. Typically, the *in vivo* molecular mechanism of transcription factor action (whether the transcription factor acts on k_{on}, k_{off}, or r in a two-state model of promoter switching) is inferred from measurements of transcriptional

noise obtained through snapshots of dead and fixed embryos or cells using theoretical models (***Zenklusen et al., 2008***; ***So et al., 2011***; ***Little et al., 2013***; ***Jones et al., 2014***; ***Senecal et al., 2014***; ***Xu et al., 2015***; ***Zoller et al., 2017***). In contrast, MS2-based live imaging can directly inform on the dynamics of promoter switching and bursting in real time. The MS2 approach, however, reports on the total number of actively transcribing RNAP molecules and not on the instantaneous rate of RNAP loading at the promoter, which is necessary to directly read out k_{on}, k_{off}, and r (***Figure 1***D,E and ***Figure 2***D). To date, approaches for extracting bursting parameters from such data have mainly relied on correlative approaches (***Larson et al., 2011***; ***Coulon et al., 2014***; ***Desponds et al., 2016***) or the manual analysis of single-nucleus transcriptional dynamics (***Bothma et al., 2014***; ***Fukaya et al., 2016***). A computational method for inferring the hidden rates of RNAP loading (***Figure 4***A, bottom) from the total number of actively transcribing RNAP molecules (***Figure 4***A, top) is thus needed to obtain the promoter-switching parameters.

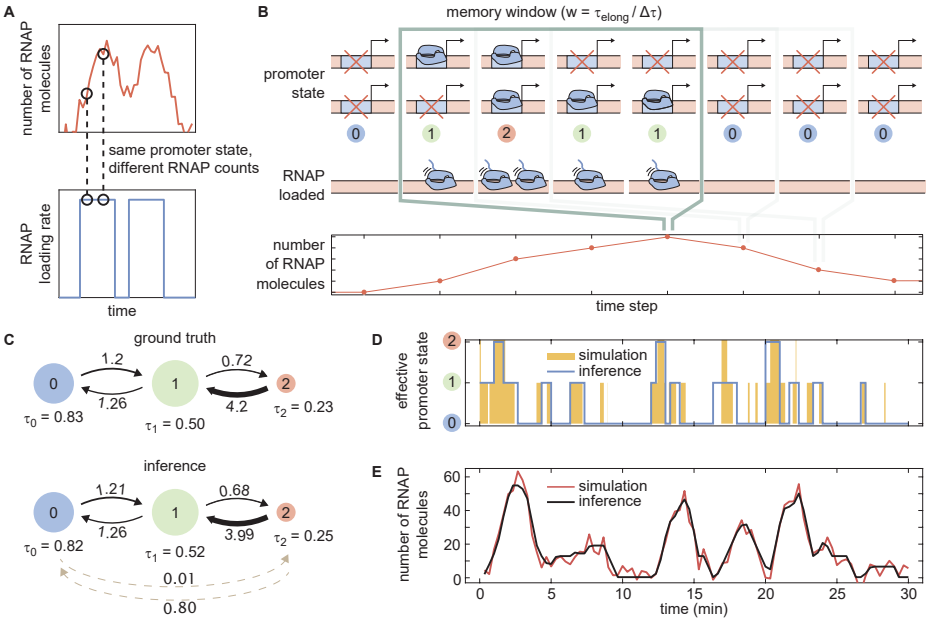


Figure 4. Statistical validation of mHMM. (A) The same hidden rate of RNA polymerase loading can correspond to different observable numbers of RNA polymerase molecules on the gene, such that standard HMM approaches cannot be used to decode the hidden promoter state. **(B)** In our mHMM architecture, the trajectory of effective promoter states over the memory time window dictates the number of RNA polymerase loaded onto the gene. **(C)** 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 4–Figure Supplement 1***. **(D)** Sample simulated promoter activity trace (yellow) generated using the parameters in C, overlaid with the best fitted trace (blue) obtained using the Viterbi algorithm (see ***Appendix 3*** for further details). **(E)** Simulated and best fitted observable number of RNA polymerase molecules corresponding to the promoter trajectory shown in D.

Figure 4–Figure supplement 1. Inference statistics for the mHMM validation.

Hidden Markov models (HMMs) are widely used to uncover the dynamics of a system as it transitions through states that are not directly accessible to the observer (***Bronson et al., 2009***). For example, this approach is often applied to study ion channels by inferring their opening and closing dynamics from single-molecule patch-clamp experiments (***Qin et al., 1997***, ***2000***). However, there is a significant difference between ion-channel data and our transcriptional data. In ion channels, the observable (current) relates directly to the instantaneous underlying molecular state. In contrast, our observable (the MS2 signal) does not correspond to the hidden variable of interest (promoter state) in a one-to-one fashion (***Figure 4***A). Instead, the observable MS2 signal reflects

the net effect of promoter switching over a period equal to the time that an RNAP molecule takes to transcribe the whole gene, τ_{elong} (***Figure 2***D and ***Figure 4***B). Thus, unlike ion-channel currents, instantaneous fluorescence in our transcription dynamics experiments does not just depend on the current promoter state; it exhibits a dependence on how active the promoter has been over a preceding window of time, which effectively constitutes a memory for recent promoter states. Classic HMM approaches cannot account for this kind of system memory.

In order to model the process of transcription and extract the kinetic parameters of promoter switching, we augmented classic HMMs to account for memory. A similar approach was recently introduced to study the transcriptional dynamics of the *actin* promoter in cell culture (***Corrigan et al., 2016***). Our memory-adjusted hidden Markov model (mHMM) assumes that each site of nascent transcript formation may exist in the three distinct activity states (***Figure 3***B). We assign a continuous rate of RNAP initiation to each effective state, without explicitly modeling the recruitment of individual RNAP molecules. We verified that this continuous arrival model is capable of accurately recovering parameters even for simulated data that feature discrete, stochastic RNA polymerase loading statistics (see ***Appendix 3*** for more details). In its initial version, this mHMM assumes that the promoter switching rates do not vary in time as in ***Equation 4***. This condition will be relaxed later on to allow for time-dependence of rates as in ***Appendix 1***.

The instantaneous count of actively transcribing RNAP molecules is the cumulative number of RNAP molecules initiated in the previous $w = \tau_{\text{elong}}/\Delta\tau$ time steps, where $\Delta\tau$ is the data sampling resolution. Our mHMM relates the hidden activity of effective promoter states in the previous w steps to the observed cumulative fluorescence contributed by actively transcribing RNAP molecules (***Figure 4***B).

In order to validate this method, we simulated fluorescence traces using biologically plausible parameters (***Appendix 3–Table 1***) and inferred back the model parameters with mHMM. ***Figure 4***C-E demonstrate the performance of mHMM. ***Figure 4***C shows the parameters used to simulate the promoter trajectory as it switches through the multiple possible states in ***Figure 4***D (yellow). This promoter trajectory leads to the simulated trajectory of the number of RNAP molecules actively transcribing the gene in ***Figure 4***E (red). Using mHMM, we find the best fitted path for our observable (***Figure 4***E, black) and the corresponding most likely promoter state trajectory (***Figure 4***D, blue). Comparison of the simulated and inferred parameters (***Figure 4***C) 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 (***Figure 4–Figure Supplement 1***). Additional studies on the dependence of inference accuracy on sampling resolution are presented in ***Appendix 3***. Thus, we conclude that our newly established mHMM reliably extracts the kinetic parameters of transcriptional bursting from live-imaging data, providing an ideal tool for testing the predictions from ***Equation 4***.

Transcription Factors Regulate the Fraction of Time in the ON State

We used the quantitative power afforded by our mHMM to infer the promoter-switching parameters from our 3-state model (***Figure 3***B) from our real-time transcriptional data (***Figure 2***). To make this possible, we independently estimated the time it takes for an individual RNAP molecule to terminate transcription, τ_{elong}, building upon an established autocorrelation approach (see ***Appendix 4*** and ***Coulon and Larson*** (***2016***)).

Figure 5A contains a typical experimental trace for a nucleus in the core of the stripe together with its best fit, which corresponds to the mHMM-inferred promoter trajectory in ***Figure 5***B (details about implementation of the mHMM method are given in ***Appendix 3***). Our ability to infer the instantaneous promoter state in individual nuclei throughout development is further illustrated in ***Figure 5***C and Video 4. The snapshots revealed that, as development progresses and the stripe sharpens, fluorescent puncta continuously fluctuate among their three transcriptional states on a time scale of of approximately 1-2 minutes.

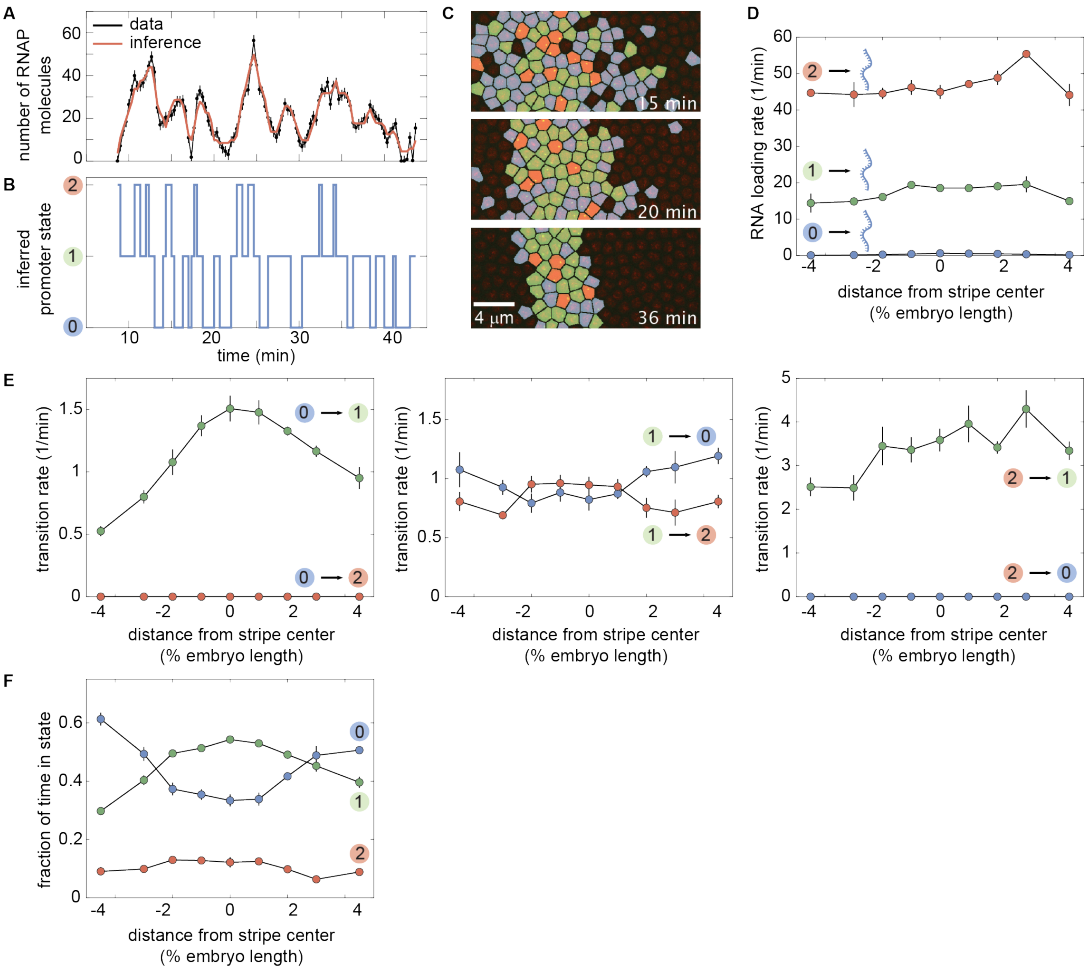


Figure 5. Mean promoter switching parameter inference using mHMM. (A) Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (color code as in B). **(D)** The rate of initiation for each transcriptional state is not significantly modulated along the embryo. **(E)** mHMM reveals that mainly the transition rates between states (0) and (1), and between states (1) and (2) are up-regulated in the stripe center and that there are no transitions between states (0) and (2). **(F)** This modulation of rates increases the share of time the promoter spends in the active states in the stripe center. (A, error bars obtained from estimation background fluorescent fluctuations as described in Materials and Methods and ***Garcia et al.*** (***2013***); D-F, error bars indicate bootstrap estimates of the standard error in mHMM inference as described in ***Appendix 3***).

We used all traces from the same region along the anterior-posterior axis to perform a time-averaged inference of bursting parameters. The rates of RNAP loading, r_1 and r_2, remained constant throughout the stripe for all promoter states (***Figure 5***D), suggesting that none of the transcription factors regulating *eve* stripe 2 act on this kinetic parameter. In contrast, a subset of the transition rates between transcriptional states change along the embryo axis (***Figure 5***E). Specifically, the transitions from state (0) to (1) and from state (1) to (2) are up-regulated in the stripe core; we also inferred a slight down-regulation of the transition from state (1) to (0) in this same stripe region. These observations suggest that transcription factors act primarily on the rate of promoter turning on, consistent with previous results (***Xu et al., 2015***; ***Fukaya et al., 2016***). This regulation effectively increases the fraction of time spent in transcriptionally active states in loci near the stripe center (***Figure 5***F; ***Zoller et al.*** (***2017***)).

Finally, our inferred rates support the hypothesis that each fluorescent punctum contains two sister chromatids, each with its own promoter capable of transcribing in bursts: we detected no transitions between states (0) and (2) (***Figure 5***E) such that $k_{02} = k_{20} = 0$, as assumed for the

derivation of ***Equation 4***. However, our inference also suggests that these two promoters do not act independently. Note that the rate of transcription of state (2) is not twice the rate corresponding to state (1) ($r_2 \neq 2r_1$ in ***Figure 5***D). Further, the transition rates between states are inconsistent with promoters switching between ON and OFF states independently as detailed in ***Appendix 2***. With our mHMM-mediated inference of bursting parameters, we are now in a position to test ***Equation 4***.

Transcriptional Bursting Does Not Dictate *Eve* Stripe 2 Formation

Having established the magnitude of the bursting parameters that dictate transcriptional bursting in *eve* stripe 2, we next sought to determine whether promoter switching can quantitatively explain pattern formation as predicted by ***Equation 4***. The left-hand side of ***Equation 4*** corresponds to the total amount of mRNA produced, which can be measured by integrating our raw fluorescence traces (***Appendix 5*** and ***Garcia et al.*** (***2013***); ***Bothma et al.*** (***2014***)). This calculation yields the profile of cytoplasmic mRNA as a function of the position and time along the embryo (distribution in ***Figure 6***, green). The contribution of transcriptional bursting to pattern formation can be calculated by substituting the inferred bursting parameters (***Figure 5***D,E) into the the right-hand side of ***Equation 4***. Surprisingly, as shown in ***Figure 6***, transcriptional bursting cannot quantitatively reproduce the pattern of accumulated mRNA. Bursting parameters are regulated along the stripe as inferred in ***Figure 5***, but this modulation is not sufficient to quantitatively explain how the pattern of accumulated mRNA is formed. While ***Figure 6*** shows mRNA levels at 40 minutes into nc14, once the stripe has matured (***Bothma et al., 2014***), our conclusions hold for any time point in the second half on nc14 as shown in ***Figure 6–Figure Supplement 1***.

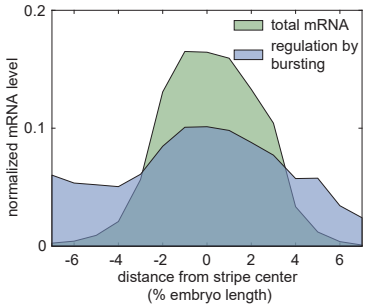


Figure 6. Contributions of transcriptional bursting to *eve* stripe 2 formation. Cytoplasmic mRNA distribution resulting from integrating the raw fluorescence data compared to the distribution obtained from the regulation of transcriptional bursting along the embryo following ***Equation 4***.

Figure 6–Figure supplement 1. Contributions of transcriptional bursting to *eve* stripe 2 formation over time.

Binary Control of the Transcriptional Time Window Is the Main Driver of Pattern Formation

The failure of ***Equation 4*** to explain how the stripe of accumulated mRNA is formed through transcriptional bursting led us to closely examine our data. This analysis, shown in ***Figure 7***A, revealed that, in addition to the bursting parameters, the duration of the window of time over which promoters engage in the transcription process is also modulated along the embryo.

Whereas the time at which each nucleus becomes transcriptionally active, $t_{\mathrm{on}}(x)$, was constant across the stripe, with all nuclei becoming active 9 ± 4 min after the previous anaphase (***Figure 7***B), the time at which nuclei stop transcribing and become quiescent, $t_{\mathrm{off}}(x)$, showed a strong modulation along the embryo's axis as shown in ***Figure 7***C. As a result, the time window over which each punctum is engaged in the transcription process, $\Delta t = t_{\mathrm{off}} - t_{\mathrm{on}}$ is sharply modulated along the stripe as shown in ***Figure 7***D,E and Video 5.

In order to explore the effect of the the modulation of the transcriptional time window on stripe formation, we revised our theoretical model of pattern formation by single-cell transcriptional

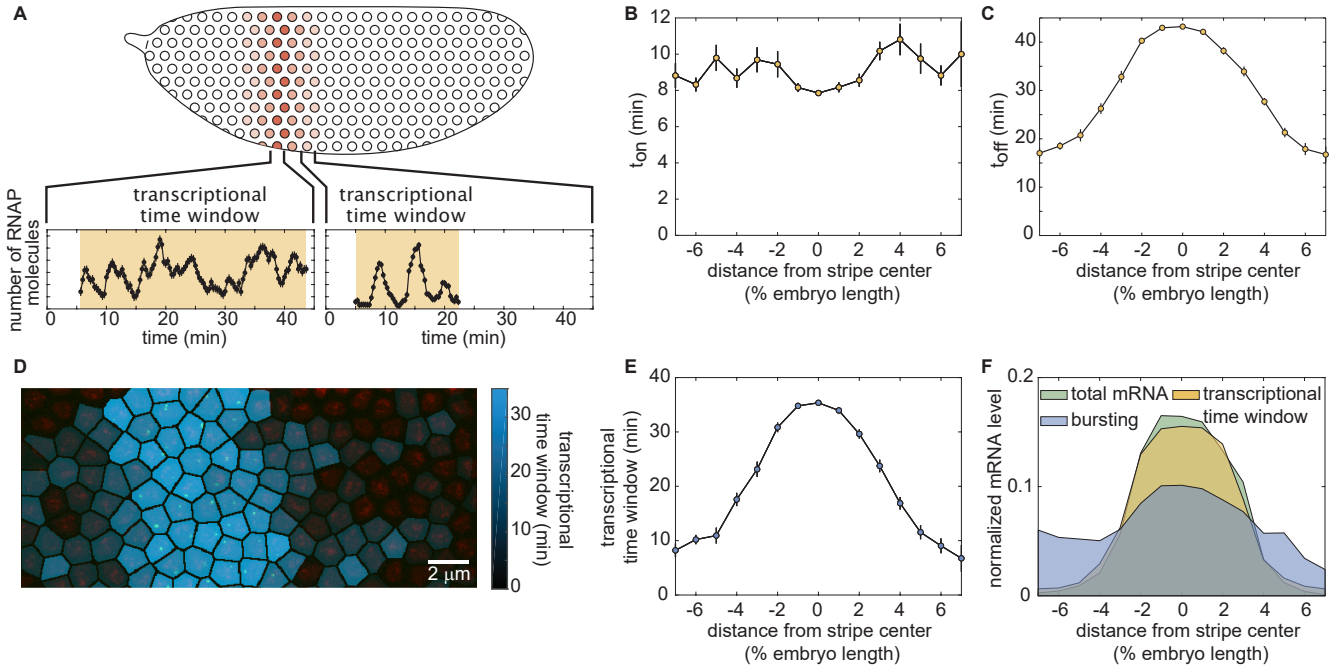


Figure 7. The role of the transcriptional time window in dictating stripe formation. (A) Single-nucleus measurements reveal that both transcriptional bursting dynamics and the transcriptional time window are modulated along the stripe. **(B)** Time for nuclei to activate transcription after mitosis, t_{on} as a function of position along the stripe. **(C)** Time for nuclei to enter the quiescent transcriptional state, t_{off}. **(D, E)** Duration of the transcriptional time window along the stripe. **(F)** mRNA distribution resulting from integrating the raw fluorescence data compared to the contributions of transcriptional bursting ("analogue control" in ***Equation 7***) and the transcriptional time window ("binary control" in ***Equation 7***) to stripe formation. (B,C,E, average over 11 embryos, error bars are generated via bootstrap re-sampling and calculating the standard error of the mean).

Figure 7–Figure supplement 1. Relative contributions to stripe formation as a function of time

Figure 7–Figure supplement 2. Contributions of inactive nuclei to stripe formation.

activity. When the nucleus is actively transcribing and bursting ($t_{\text{on}} \leq t \leq t_{\text{off}}$; shaded regions in ***Figure 7***A), the net rate of mRNA production is the rate of mRNA synthesis minus the rate of mRNA degradation such that

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = \underbrace{r_1 \frac{k_{01}k_{21}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}} + r_2 \frac{k_{01}k_{21}}{k_{10}k_{12} + k_{01}k_{21} + k_{01}k_{12}}}_{\langle \text{transcription rate} \rangle} - \underbrace{\gamma\, \text{mRNA}(x,t)}_{\text{degradation rate}} . \tag{5}$$

For $t > t_{\text{off}}$, once transcription ceases and the promoter enters the quiescent state, the mean rate of production drops to zero (white region in ***Figure 7***A), and only the mRNA decay term remains, resulting in

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = -\gamma\, \text{mRNA}(x,t). \tag{6}$$

Equation 5 and ***Equation 6*** can be solved in steady-state to obtain an expression for the amount of mRNA available at a position x and time point t taking into account transcriptional bursting as well as the transcriptional time window. This calculation, which is presented in detail in ***Appendix 1*** leads to

$$\text{mRNA}(x,t) = \underbrace{\frac{r_1}{\gamma} \frac{k_{01}k_{21}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}} + \frac{r_2}{\gamma} \frac{k_{01}k_{21}}{k_{10}k_{12} + k_{01}k_{21} + k_{01}k_{12}}}_{\text{analogue control}} \times \underbrace{\left(e^{-\gamma(t - \min(t_{\text{off}}(x), t))} - e^{-\gamma(t - t_{\text{on}}(x))}\right)}_{\text{binary control}} . \tag{7}$$

Here, we have once again assumed that the transition rates can vary in space and not in time, an assumption that we will test in detail below. Two distinct regulatory strategies for pattern formation emerge from ***Equation 7***. First, if promoters were to never enter quiescence and the system was allowed to reach steady state, only the first factor in the equation would remain, leading to ***Equation 4***. This first factor describes how transcriptional bursting parameters can be systematically varied across the embryo in order to control the mean rate of gene expression in a graded fashion. As a result, we identify this regulatory strategy with the *analogue control* of gene expression. Second, even if bursting parameters were constant throughout the embryo, gene expression patterns could still be realized through the spatial modulation of the timing of the onset of transcription and of the entry into the quiescent state given by $t_{\text{on}}(x)$ and $t_{\text{off}}(x)$, respectively. This effect of the transcriptional time window on pattern formation is captured by the second factor in ***Equation 7***. We identify this regulatory strategy—akin to an on/off switch—with the *binary control* of gene expression. Thus, our revised model predicts how pattern formation arises from the interplay between two distinct gene expression strategies: the analogue regulation of the mean transcriptional rate the and binary regulation of the transcriptional time window.

In ***Figure 7***F we compare the degree to which each regulatory strategy is sufficient to generate the observed mRNA profile as predicted by ***Equation 7***. To make this possible, we assume a degradation rate of *eve* mRNA of 0.14 min^{-1} (***Edgar et al., 1987***). The figure reveals that the regulation of the duration of the time window is sufficient to recapitulate the formation of the stripe of accumulated mRNA from the single-cell transcriptional dynamics underlying it. Thus, our results show that the main strategy the embryo uses to generate *eve* stripe 2 is not the *analogue* modulation of gene expression through transcriptional bursting. Instead, the *binary* control of the transcriptional time window across the nascent *eve* stripe 2 pattern plays a dominant role in driving the formation of the mature stripe pattern (see also ***Figure 7–Figure Supplement 1*** and ***Figure 7–Figure Supplement 2*** for a discussion of other contributions to stripe formation). Thus, the decisive metric for stripe formation is not the rate at which nuclei produce mRNA while actively transcribing, but the timing with which nuclei transition into a state of transcriptional quiescence.

Uncovering the Molecular Origins of the Transcriptional Time Window

Our discovery that the temporal control of the onset of transcriptional quiescence plays a dominant role in the formation of the mature *eve* stripe 2 pattern motivated us to search for the mechanisms

governing the transition into quiescence. So far, our model, embodied in ***Equation 7***, assumed that the rates of promoter switching do not change in time. This setup made the implicit assumption that transcriptional quiescence results from the promoter transitioning into an extra, silent state (***Figure 8***A, top). This silent state could, for example, be linked to irreversible chromatin modifications. According to this hypothesis, the onset of quiescence reflects a fundamental change to the molecular character of the transcriptional locus such that the bursting framework no longer applies.

Alternatively, if we abandon the widespread assumption that the rates of promoter switching are not modulated in time, quiescence could be explained without the need to invoke a silent state that lies outside of our model. In this scenario, one or multiple promoter switching rates would change over time in order to progressively reduce the frequency, intensity, and/or duration of transcriptional bursts. Such modulation could be achieved by downregulating k_{on}, downregulating r, and/or upregulating k_{off} (***Figure 8***A, bottom). Any of these effects, if sufficiently strong, could abolish all activity at a transcriptional locus for the remainder of the nuclear cycle, leading to quiescence.

If quiescence can be explained by the bursting model, then one or more bursting parameters must be modulated in time (***Figure 8***A, bottom). As a result, in order to discriminate between these two possible scenarios, we sought to determine whether the bursting dynamics varied over time. To probe for time-dependence in the 3-state model parameters, we split the stripe into the five regions shown in ***Figure 8***B and analyzed the single-cell trajectories returned by our original mHMM inference (***Figure 5***B). As a first pass, we examined the average time spent in the (0) state as a function of time. As shown in ***Figure 8***C, this analysis revealed significant temporal trends in this (0) state dwell time. These and other trends found in our data (***Figure 8–Figure Supplement 1***) suggested that the rate of promoter turning on, k_{on}, was being regulated over time.

In order to investigate these apparent temporal trends in k_{on} further and probe for trends in other switching parameters, we extended our mHMM method to obtain promoter bursting parameters over discrete periods of time by performing inference on our MS2 traces using a sliding window (see ***Appendix 3*** for details). Consistent with our initial findings, the transition rate between states (0) and (1), k_{01}, presented a strong spatiotemporal modulation (***Figure 8***D). Specifically, nuclei in both the anterior and posterior stripe boundaries (black and red regions in ***Figure 8***B) transcribe with a value of k_{01} that decreases as development progresses. In addition, the rate of RNAP loading when in state (1), r_1, also decreased slightly at the stripe's flanks (***Figure 8***E). This coincidence of the decrease in k_{01} and r in flank nuclei with the onset of transcriptional quiescence (***Figure 8***F) supports the hypothesis that quiescence in the stripe flanks is driven, at least in part, by the temporal modulation of bursting parameters (***Figure 8***A, bottom).

However, unlike the stripe flanks, the center stripe regions exhibited no strong correlations between bursting parameter trends and quiescence. Indeed, although 60% and 40% of nuclei in the regions directly anterior and posterior of the stripe center (blue and yellow regions in ***Figure 8***B) are quiescent by 40 min into the nuclear cycle (***Figure 8***F), we detected no corresponding decrease in k_{01}. In fact, k_{01} actually *increased* in some inner regions of the stripe (***Figure 8***D). In addition, the transition rate between (1) and (0), k_{10}, decreased slightly in the anterior stripe flank.

The fact that, as one subpopulation of nuclei becomes quiescent, the rate of transcription in the remaining nuclei remains constant or even increases runs counter to the hypothesis that quiescence is exclusively driven by the temporal modulation of the promoter switching parameters. There are two (potentially complementary) explanations that could reconcile this observation with the promoter switching hypothesis. First, temporal changes in bursting parameters associated with the onset of quiescence might be too fast to be captured by our mHMM as described in ***Figure 8–Figure Supplement 2*** and in ***Appendix 3***. Second, even if the transition into quiescence happens on a slow enough timescale to be detectable in principle, the gradual nature of the onset of quiescence in the stripe center (see ***Figure 8***F yellow , green, and blue) implies that only a small fraction of nuclei are undergoing the transition at any given point in time. Thus, it is possible that the subset of nuclei for which the onset of quiescence is imminent *do* experience a change in their bursting parameters, but that the activity of the remaining, actively bursting population washes out any signature of the

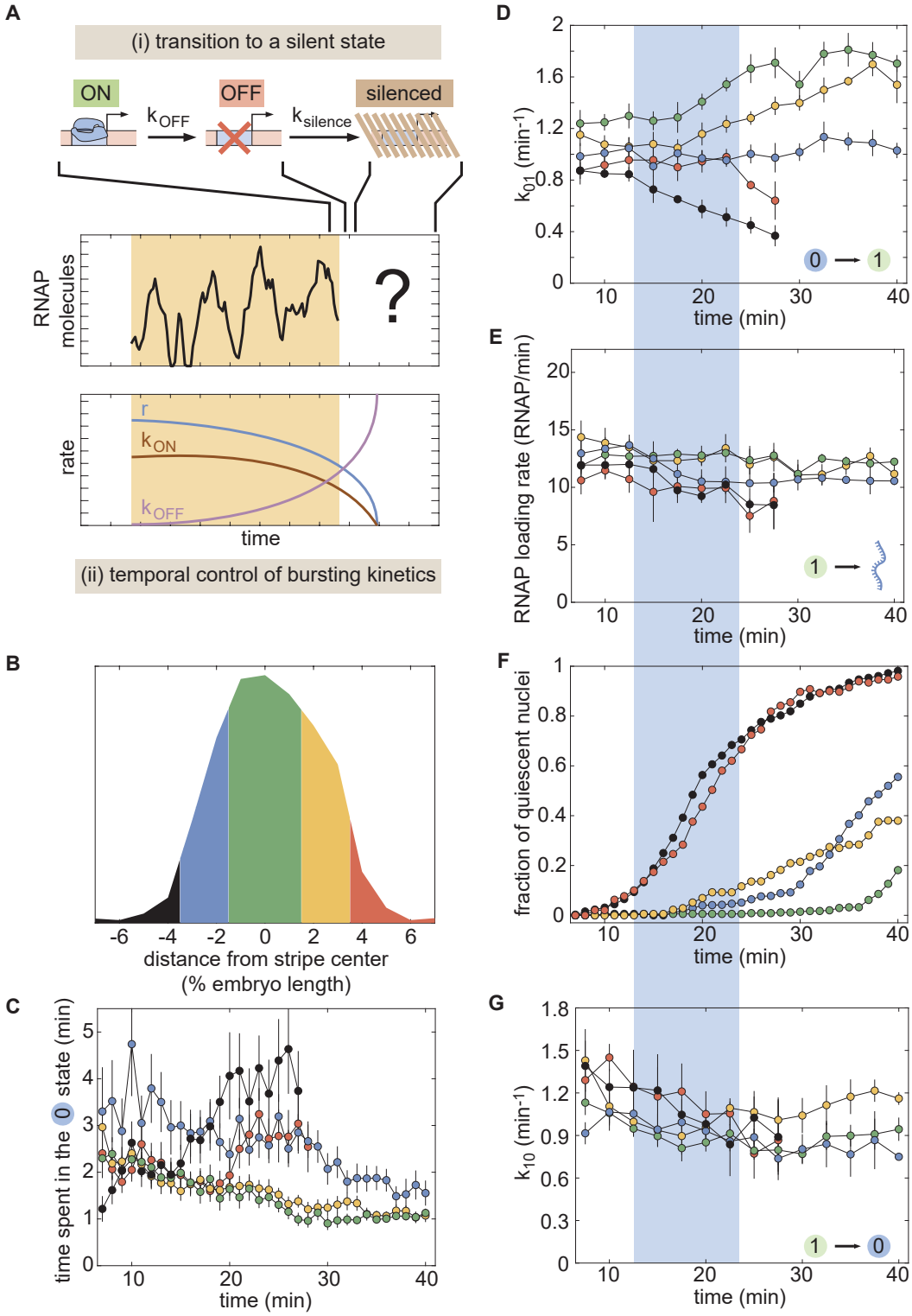


Figure 8. Investigating the molecular character of quiescence. (A) Two possible hypotheses explaining promoter quiescence onset by (i) an irreversible transition into an alternative, transcriptionally silent state and (ii) the modulation of one or more bursting parameters over time. **(B)** Splitting of the stripe into five regions for our analysis. **(C)** Time spent in the (0) state, **(D)** transition rate between the (0) and (1) states, **(E)** rate of RNAP loading off of state (1), **(F)** fraction of quiescent nuclei, and **(G)** transition rate between the (1) and (0) states as a function of time and position along the stripe. (Error bars indicate the bootstrap estimate of the standard error.)

Figure 8–Figure supplement 1. Temporal regulation of bursting dynamics.

Figure 8–Figure supplement 2. Limits to temporal inference.

changes in our inference results. As a result, the molecular nature of transcriptional quiescence and, in particular, the relationship between quiescence and transcriptional bursting remains an open question. Additional experiments and improved data-analysis pipelines will be necessary to definitively elucidate the impact of the temporal modulation of promoter switching rates on quiescence, as detailed in the Discussion and ***Appendix 3***.

Discussion

The realization that many genes in embryonic development are transcribed in a burst-like fashion has led researchers to ask how bursting parameters are modulated along an embryo in order to realize gene expression patterns (***Lionnet et al., 2011***; ***Bothma et al., 2014***; ***Xu et al., 2015***; ***Fukaya et al., 2016***; ***Zoller et al., 2017***). Despite the appeal of this scenario, to our knowledge, no stringent *quantitative* test of this hypothesis had been carried out.

To close this gap, we derived a simple theoretical model connecting pattern formation to transcriptional bursting and tested its predictions experimentally. This theoretical model predicted how transcriptional bursting parameters dictate the cytoplasmic accumulation of mRNA and the subsequent formation of a gene expression pattern (***Equation 4***). In order to test our model's predictions, we used the MS2 system to capture the formation of the widely studied stripe 2 of the *eve* gene in fly embryos at the single-cell level. However, this MS2 technique cannot directly report on the instantaneous state of the promoter. We developed a memory-adjusted Hidden Markov Model (mHMM) that is capable of inferring the instantaneous promoter state in an automated and statistically robust manner. Using mHMM we directly obtained, for the first time in a multicellular organism, promoter-switching parameters across a pattern of gene expression (***Figure 5***) by visualizing a transcriptional process in real time. In agreement with previous measurements on different gene expression patterns (***Xu et al., 2015***; ***Fukaya et al., 2016***; ***Zoller et al., 2017***), our results revealed that the main bursting parameter regulated by the input transcription factors to *eve* stripe 2 is the bursting frequency, which is controlled by k_{on}.

It is important to note that our mHMM algorithm is not limited to the *eve* stripe 2 system and should prove useful to infer the underlying promoter state of any gene that is tagged using approaches such as the MS2 or PP7 systems in any organism (***Larson et al., 2011***; ***Hocine et al., 2012***; ***Fukaya et al., 2016***). Further, mHMM could be used to infer the state of the ribosome as mRNA is being translated into protein in novel single-molecule *in vivo* translation assays (***Morisaki et al., 2016***; ***Wang et al., 2016***; ***Yan et al., 2016***; ***Wu et al., 2016***). Thus, we envision that our mHMM approach will serve as a useful tool for the broader biophysical analysis of *in vivo* cellular processes at the single-molecule level.

To our surprise, confronting our measurements of the total amount of mRNA produced over a nuclear cycle with the bursting parameters inferred by our mHMM revealed that the analogue modulation of the mean rate of transcription afforded by transcriptional bursting cannot quantitatively recapitulate the formation of a sharp gene expression stripe across the embryo as predicted by ***Equation 4*** (***Figure 6***). We discovered that, in addition to transcriptional bursting, the duration of the window of time over which promoters engage in transcription is also regulated, and that this binary regulation of the transcriptional time window mediated by promoters entering into a quiescent state constituted the main regulatory contribution to the formation of the stripe (***Figure 7***F). Thus, our results suggest that, in order to make progress toward a quantitative and predictive picture of how the stripe is formed, it is necessary to go beyond the widespread steady-state, static picture of pattern formation in development put forward by previous single-cell transcriptional activity studies that focused on how gene expression patterns are formed by transcriptional bursting (***Pare et al., 2009***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Fukaya et al., 2016***; ***Desponds et al., 2016***; ***Zoller et al., 2017***).

The realization of the importance of the the regulation of the duration of the transcriptional time window in pattern formation led us to ask how this entry into transcriptional quiescence unfolds at the molecular level (***Figure 8***A). To answer this question, we expanded our mHMM to go beyond

time-independent models of promoter switching and to infer the regulation of these rates both in space and time. We conclude that, while the temporal modulation of the promoter switching rates may explain quiescence in certain regions of the pattern, there is substantial evidence indicating that an additional silent state may need to be invoked to explain entry into quiescence in other regions. However, in order to move toward sharper molecular hypotheses, such as whether repressors only act when the promoter is transiently in the OFF state (***Bothma et al., 2014***; ***Fukaya et al., 2016***), it will be necessary to expand mHMM to correlate input transcription factor concentration dynamics with output transcriptional activity. Experimentally, we recently measured inputs and output simultaneously using novel fluorescent labeling technology (***Bothma et al., 2018***). Thus, there is a clear experimental and computational path to uncover the detailed mechanisms behind the molecular control of transcriptional bursting and quiescence in development.

Our mHMM inference also suggests that the two sister promoters contained within our fluorescent puncta neither behave independently from each other (***Little et al., 2013***; ***Zoller et al., 2017***) nor in a perfectly correlated fashion (***McKnight and Miller, 1977***). Thus, we speculate that sister chromatids are capable of interacting and coordinating their transcriptional activities, perhaps as a result of shared resources in their immediate environment. This speculation is consistent with recent observations of transcriptional coordination among promoters that share an enhancer (***Fukaya et al., 2016***) and among alleles engaged in transvection (***Lim et al., 2018***), and with reports of spatially resolved domains of high concentration of transcription factors in the nucleus (***Crocker and Ilsley, 2017***; ***Mir et al., 2017***). In order to shed light on the molecular processes underlying sister chromatids coordination, it will be necessary to develop approaches to label each promoter in an orthogonal manner.

Regardless, the main conclusion from this study does not depend on the molecular details of how entry into quiescence is realized or whether a 2-state or 3-state model accounting for sister chromatids is considered: the formation of gene expression patterns in development, which are what the embryo ultimately utilizes in order to drive its developmental program (***Dubuis et al., 2013***), might be effectively independent of the details of transcriptional bursting, and, instead, arise predominantly from the spatial regulation of the duration of the transcriptional time window. In other words, the experiment/theory discourse deployed in this work suggests that the binary control of *whether* a promoter transcribes could be more relevant to patterns of gene expression than the analogue control of *how* this promoter transcribes while active. Recent studies have shown that the graded profiles of gene expression characteristic of the early patterning network in *Drosophila* carry significant spatial information (***Dubuis et al., 2013***; ***Petkova et al., 2016***). It is thought-provoking, then, to consider a scenario of development in which these exquisitely graded (analogue) patterns of expression are generated via simple binary control logic and, in turn, are interpreted by downstream promoters to drive binary decision-making regarding when and where transcription unfolds.

Materials and Methods

Cloning and transgenesis

This work employed the same *eve* stripe 2 reporter construct developed in a previous work (***Bothma et al., 2014***). This construct contains the *even-skipped* (eve) stripe 2 enhancer and promoter region (spanning -1.7 kbp to +50 bp) upstream of the *yellow* reporter gene. 24 repeats of the MS2 stem loop sequence were incorporated into the 5' end of the reporter gene.

Sample preparation and data collection

Sample preparation followed procedures described in ***Bothma et al.*** (***2014***) and ***Garcia and Gregor*** (***2018***). In short, female virgins of yw;His-RFP;MCP-GFP were crossed to males bearing the reporter gene. Embryos were collected and mounted in halocarbon oil between a semipermeable membrane and a coverslip. Data collection was performed using a confocal Leica SP8 Laser Scanning Confocal

Microscope. Average laser power on the specimen (measured at the output of a 10x objective) was 35 μW. Image resolution was 256 x 512 pixels, with a pixel size of 212nm and a pixel dwell time of 1.2μs. The signal from each frame as accumulated over three repetitions. At each time point, a stack of 21 images separated by 500 nm were collected. Image stacks were collected at a time resolution of 21 seconds. The MCP-GFP and Histone-RFP were excited with a laser wavelength of 488 and 556 nm, respectively. Fluorescence was detected with two separate Hybrid Detectors (HyD) using the 498-546nm and 566-669nm spectral windows. Specimens were imaged for a minimum of 40 minutes into nuclear cleavage cycle 14.

Image analysis

Image analysis of live imaging movies was performed based on the protocol found in (***Garcia et al., 2013***) with modifications to the identification of transcriptional spots, which were segmented using the Trainable Weka Segmentation plugin for FIJI using the FastRandomForest algorithm (***Schindelin et al., 2012***; ***Schneider et al., 2012***; ***Arganda-Carreras et al., 2017***; ***Witten et al., 2016***). In comparison with a previous algorithm based on Difference of Gaussians (***Little et al., 2013***; ***Garcia et al., 2013***; ***Bothma et al., 2014***, ***2015***), this alternative spot segmentation approach was found to be superior for the detection of dim transcription spots—a feature critical to establishing the precise timing of the cessation of activity at transcriptional loci.

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 AP 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 2–Figure Supplement 1***, used observed spatial patterns of fluorescence measured from 30 minutes into nc14—the approximate time at which mature stripe is first established (***Bothma et al., 2014***)—through to the time of last observation ($\geq$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 AP axis. Thus, a variety of orientations for the natural stripe axis were considered, ranging between ± 15 degrees of perpendicular with the manually specified AP axis. For each orientation, a sliding window 4% AP 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 the window defined a line through the field of view that was taken as the stripe center. All AP 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 step-over-step changes in fluorescence 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.

mHMM Inference

To account for finite RNAP elongation times, a compound state Markov formalism was developed in which the underlying 2 promoter system—assumed to have 3 states (see ***Figure 3***)—was transformed into a system with 3^w compound gene states, where w indicates the number of time steps

needed for a RNAP molecule to traverse the full transcript (see ***Appendix 4***). These compound gene states played the role of the "hidden" states within the traditional HMM formalism. See ***Appendix 3*** 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 for subsets of experimental traces (***Appendix 3***). This code is available at the GarciaLab/mHMM GitHub repository. Bootstrap sampling was used to estimate the standard error in our parameter estimates (***Appendix 3***). Subsets of 8,000 data points were used to generate time-averaged parameter estimates. Sample sizes for windowed inference varied due to data set limitations. When possible, samples of 4,000 points were used. Inference was not conducted for spatio-temporal regions for which fewer than 1,250 time points were available.

Absolute calibration of MS2 signal

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 (***Appendix 3***). 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} (***Garcia et al., 2013***) given by

$$\alpha = \frac{N_{FISH}}{F_{MS2}}. \tag{8}$$

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 ***Equation 8*** 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

$$F_1 = f_{max} \frac{\frac{1}{2}L_I + L_{II}}{v_{elong}}, \tag{9}$$

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

$$F_1 = \frac{1}{\alpha} = f_{max} \frac{\frac{1}{2}L_I + L_{II}}{v_{elong}} \tag{10}$$

such that

$$f_{max} = \frac{v_{elong}}{\alpha} \frac{1}{\frac{1}{2}L_I + L_{II}}. \tag{11}$$

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

$$F_{RNAP} = \frac{1}{t_{elong}} f_{max} \frac{\frac{1}{2}L_I + L_{II}}{v_{elong}} \tag{12}$$

$$= \frac{v_{elong}}{(L_I + L_{II})} f_{max} \frac{\frac{1}{2}L_I + L_{II}}{v_{elong}} \tag{13}$$

$$= f_{max} \frac{\frac{1}{2}L_I + L_{II}}{(L_I + L_{II})} \tag{14}$$

$$= \frac{v_{elong}}{\alpha} \frac{1}{(L_I + L_{II})} \tag{15}$$

resulting in

$$F_{RNAP} = \frac{v_{elong} F_{MS2}}{N_{FISH}} \frac{1}{(L_I + L_{II})}. \tag{16}$$

Measurements performed in ***Garcia et al.*** (***2013***) give N_{FISH} to be 220 ($\pm$ 30) mRNA per nucleus and v_{elong} to be 1.5 ($\pm$ 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 ($\pm$320) AU. 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 have:

$$F_{RNAP} = \frac{1.5 \times 9610}{220} \frac{1}{(1.275 + 4.021)} \tag{17}$$

$$= 13 \text{ AU/RNAP} \pm 1.7. \tag{18}$$

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.

Videos

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.

Video 2. Mean rate of transcription of *eve* stripe 2 reported by MS2. Mean transcriptional activity averaged over a 4 min time window as a function of time.

Video 3. 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.

Video 4. Real-time inferred promoter states. Inference of real-time promoter state in individual nuclei.

Video 5. Transcriptional time window. Duration of the transcriptional time window.

Acknowledgments

We thank Thomas Gregor and Lev Barinov for discussion about an initial implementation of the mHMM approach, Florian Jug for help with the spot segmentation using machine learning, and Elizabeth Eck, Maryam Kazemzadeh-Atoufi and Jonathan Liu for the P2 data used in the absolute MS2 calibration. We are also grateful to Jack Bateman, Jane Kondev, Rob Phillips, Allyson Sgro and Donald Rio for comments and discussion on the manuscript. HGG was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl & Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236). NL was supported by NIH Genomics and Computational Biology training grant 5T32HG000047-18. CW was supported by the NIH/NIC (U54 CA193313), CUNY (RFCUNY 40D14-A), and the NSF (IIS-1344668).

References

Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, Sebastian Seung H. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics. 2017 Aug; 33(15):2424–2426. https://academic.oup.com/bioinformatics/article/33/15/2424/3092362, doi: 10.1093/bioinformatics/btx180.

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. Mol Cell. 1998; 2(4):437–45. doi: S1097-2765(00)80143-4 [pii].

Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc Natl Acad Sci U S A. 2014; 111(29):10598–10603. doi: 10.1073/pnas.1410022111.

Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife. 2015; 4. doi: 10.7554/eLife.07956.

Bothma JP, Norstad MR, Alamos S, Garcia HG. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. Cell. 2018; doi: 10.1016/j.cell.2018.03.069.

Bronson JE, Fei J, Hofman JM, Jr RLG, Wiggins CH. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. arXiv:09073156v1. 2009; .

Christopher B. Pattern recognition and machine learning. New York: Springer; 2006.

Corrigan AM, Tunnacliffe E, Cannon D, Chubb JR. A continuum model of transcriptional bursting. Elife. 2016; 5. doi: 10.7554/eLife.13051.

Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. Elife. 2014; 3. doi: 10.7554/eLife.03939.

Coulon A, Larson DR. Fluctuation Analysis: Dissecting Transcriptional Kinetics with Signal Theory. Methods in Enzymology. 2016; .

Crocker J, Ilsley GR. Using synthetic biology to study gene regulatory evolution. Curr Opin Genet Dev. 2017; 47:91–101. doi: 10.1016/j.gde.2017.09.001.

Desponds J, Tran H, Ferraro T, Lucas T, Perez Romero C, Guillou A, Fradin C, Coppey M, Dostatni N, Walczak AM. Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos. PLoS Comput Biol. 2016; 12(12):e1005256. doi: 10.1371/journal.pcbi.1005256.

Dubuis JO, Samanta R, Gregor T. Accurate measurements of dynamics and reproducibility in small genetic networks. Mol Syst Biol. 2013; 9:639.

Edgar BA, Odell GM, Schubiger G. Cytoarchitecture and the patterning of fushi tarazu expression in the Drosophila blastoderm. Genes Dev. 1987; 1(10):1226–37.

Frasch M, Levine M. Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. Genes Dev. 1987; 1(9):981–95.

Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016; doi: 10.1016/j.cell.2016.05.025.

Fukaya T, Lim B, Levine M. Rapid Rates of Pol II Elongation in the Drosophila Embryo. Current biology : CB. 2017 5; 27(9):1387–1391. http://www.ncbi.nlm.nih.gov/pubmed/28457866http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5665007, doi: 10.1016/j.cub.2017.03.069.

Garcia HG, Tikhonov M, Lin A, Gregor T. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol. 2013; 23(21):2140–5. doi: 10.1016/j.cub.2013.08.054.

Garcia HG, Gregor T. In: Gaspar I, editor. Live Imaging of mRNA Synthesis in Drosophila New York, NY: Springer New York; 2018. p. 349–357. https://doi.org/10.1007/978-1-4939-7213-5_23, doi: 10.1007/978-1-4939-7213-5_23.

Gilbert SF. Developmental biology. 9th ed. Sunderland, Mass.: Sinauer Associates; 2010.

Gillespie DT. General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. Journal of Computational Physics. 1976; 22(4):403–434.

Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123(6):1025–36.

Halstead JM, Lionnet T, Wilbertz JH, Wippich F, Ephrussi A, Singer RH, Chao JA. Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science. 2015; 347(6228):1367–671. doi: 10.1126/science.aaa3380.

Hocine S, Raymond P, Zenklusen D, Chao JA, Singer RH. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nat Methods. 2012; doi: 10.1038/nmeth.2305.

Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014; 346(6216):1533–6. doi: 10.1126/science.1255301.

Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science. 2011; 332(6028):475–8. doi: 332/6028/475 [pii] 10.1126/science.1202142.

Lim B, Heist T, Levine M, Fukaya T. Visualization of Transvection in Living Drosophila Embryos. Mol Cell. 2018; 70(2):287–296 e6. doi: 10.1016/j.molcel.2018.02.029.

Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y, Wells AL, Chao JA, Park HY, de Turris V, Lopez-Jones M, Singer RH. A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods. 2011; doi: nmeth.1551 [pii] 10.1038/nmeth.1551.

Little SC, Tikhonov M, Gregor T. Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell. 2013; 154(4):789–800. doi: 10.1016/j.cell.2013.07.025.

Little SC, Tkacik G, Kneeland TB, Wieschaus EF, Gregor T. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. PLoS Biol. 2011; 9(3):e1000596. doi: 10.1371/journal.pbio.1000596.

McKnight SL, Miller J O L. Electron microscopic analysis of chromatin replication in the cellular blastoderm Drosophila melanogaster embryo. Cell. 1977; 12(3):795–804.

Mir M, Reimer A, Haines JE, Li XY, Stadler M, Garcia H, Eisen MB, Darzacq X. Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. 2017; 31(17):1784–1794. doi: 10.1101/gad.305078.117.

Morisaki T, Lyon K, DeLuca KF, DeLuca JG, English BP, Zhang Z, Lavis LD, Grimm JB, Viswanathan S, Looger LL, Lionnet T, Stasevich TJ. Real-time quantification of single RNA translation dynamics in living cells. Science. 2016; 352(6292):1425–9. doi: 10.1126/science.aaf0899.

Pare A, Lemons D, Kosman D, Beaver W, Freund Y, McGinnis W. Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr Biol. 2009; 19(23):2037–42. doi: S0960-9822(09)01848-X [pii] 10.1016/j.cub.2009.10.028.

Peccoud J, Ycart B. Markovian modeling of gene product synthesis. Theor Popul Biol. 1995; 48:222–234.

Petkova MD, Tkačik G, Bialek W, Wieschaus EF, Gregor T. Optimal decoding of information from a genetic network. . 2016; https://arxiv.org/pdf/1612.08084.pdf.

Qin F, Auerbach A, Sachs F. Maximum likelihood estimation of aggregated Markov processes. Proc Biol Sci. 1997; 264(1380):375–83. doi: 10.1098/rspb.1997.0054.

Qin F, Auerbach A, Sachs F. A direct optimization approach to hidden Markov modeling for single channel kinetics. Biophys J. 2000; 79(4):1915–27. doi: 10.1016/S0006-3495(00)76441-1.

Rabinowitz M. Studies on the cytology and early embryology of the egg of Drosophila melanogaster. Journal of Morphology. 1941; 69(1):1–49. doi: DOI 10.1002/jmor.1050690102.

Rice GA, Chamberlin MJ, Kane CM. Contacts between mammalian RNA polymerase II and the template DNA in a ternary elongation complex. Nucleic Acids Research. 1993; 21(1):113–118.

Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. PLoS Comput Biol. 2011; 7(3):e1001100. doi: 10.1371/journal.pcbi.1001100.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. Nature Methods. 2012 Jul; 9(7):676–682. https://www.nature.com/articles/nmeth.2019, doi: 10.1038/nmeth.2019.

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012 Jun; 9:671. http://dx.doi.org/10.1038/nmeth.2089.

Senaratne TN, Joyce EF, Nguyen SC, Wu CT. Investigating the Interplay between Sister Chromatid Cohesion and Homolog Pairing in Drosophila Nuclei. PLoS Genet. 2016; 12(8):e1006169. doi: 10.1371/journal.pgen.1006169.

Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, Mueller F, Darzacq X. Transcription factors modulate c-Fos transcriptional bursts. Cell Rep. 2014; 8(1):75–83. doi: 10.1016/j.celrep.2014.05.053.

Shermoen AW, McCleland ML, O'Farrell PH. Developmental control of late replication and S phase length. Curr Biol. 2010; 20(23):2067–77. doi: 10.1016/j.cub.2010.10.021.

Shermoen AW, O'Farrell PH. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell. 1991; 67(2):303–10.

Small S, Blair A, Levine M. Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. 1992; 11(11):4047–57.

So LH, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. General properties of transcriptional time series in Escherichia coli. Nat Genet. 2011; 43(6):554–60. doi: 10.1038/ng.821.

Wang C, Han B, Zhou R, Zhuang X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell. 2016; 165(4):990–1001. doi: 10.1016/j.cell.2016.04.040.

Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann; 2016. Google-Books-ID: 1SylCgAAQBAJ.

Wu B, Eliscovich C, Yoon YJ, Singer RH. Translation dynamics of single mRNAs in live cells and neurons. Science. 2016; 352(6292):1430–5. doi: 10.1126/science.aaf1084.

Xu H, Sepulveda LA, Figard L, Sokac AM, Golding I. Combining protein and mRNA quantification to decipher transcriptional regulation. Nat Methods. 2015; 12(8):739–42. doi: 10.1038/nmeth.3446.

Yan X, Hoek TA, Vale RD, Tanenbaum ME. Dynamics of Translation of Single mRNA Molecules In Vivo. Cell. 2016; 165(4):976–89. doi: 10.1016/j.cell.2016.04.034.

Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008; 15(12):1263–71. doi: nsmb.1514 [pii] 10.1038/nsmb.1514.

Zoller B, Little SC, Gregor T. Diverse spatial expression patterns emerge from common transcription bursting kinetics. ArXiv e-prints. 2017; 1712.08215.

Appendix 1

Theoretical model of cytoplasmic mRNA levels in steady state

Here we provide a more detailed treatment of 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 active and quiescent periods. When the promoter is actively transcribing ($t_{\mathrm{on}} \leq t \leq t_{\mathrm{off}}$), the net rate of mRNA production is

$$\frac{\mathrm{dmRNA}}{\mathrm{d}t}(x,t) = \underbrace{r(x,t)\frac{k_{\mathrm{on}}(x,t)}{k_{\mathrm{on}}(x,t)+k_{\mathrm{off}}(x,t)}}_{\text{transcription rate}} - \underbrace{\gamma\,\mathrm{mRNA}(x,t)}_{\text{degradation rate}}, \tag{19}$$

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

$$\frac{\mathrm{dmRNA}}{\mathrm{d}t}(x,t) = -\gamma\,\mathrm{mRNA}(x,t) \tag{20}$$

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 $0.09\ \mu\mathrm{m}^2/\mathrm{s}$ (***Halstead et al., 2015***) and a typical mRNA degradation rate of 0.14 min^{-1} (***Edgar et al., 1987***). Given these numbers, we expect an *eve* mRNA molecule to diffuse approximately $6\ \mu$m, which corresponds to one nuclear diameter or 1% of the embryo length. Thus, given the overall width of the stripe mRNA profile of about 8% of the embryo length (***Figure 7***F), we expect diffusion to play a minimal role in stripe formation.

In both fixed-tissue and live-imaging studies of transcriptional bursting in development, it is common to assume that bursting is modulated along the axis of the embryo, but does not change in time (***Pare et al., 2009***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Fukaya et al., 2016***; ***Desponds et al., 2016***; ***Zoller et al., 2017***). We used this assumption of temporal independence as a starting point in the development of increasingly general descriptions of mRNA production. To begin, we considered a scenario in which transcriptional loci have been active for sufficiently long period of time to reach a steady state, such that the rate of transcript production is balanced by the rate of degradation, resulting in

$$\mathrm{mRNA}(x,t) = \underbrace{r(x)\frac{k_{\mathrm{on}}(x)}{k_{\mathrm{on}}(x)+k_{\mathrm{off}}(x)}}_{\text{transcription rate}}\frac{1}{\gamma}, \tag{21}$$

Note that that there is no explicit time dependence in ***Equation 21***. Thus, in this steady-state limit, the system is memory-less and any spatial variation in cytoplasmic mRNA levels is generated solely by the graded modulation of the mean rate of transcription. We identified this regulatory strategy as a realization of *analogue control*. We next considered that the period of transcriptional competence is preceded and succeeded by periods of inactivity, and that the timing of the onset and termination of transcription ($t_{\mathrm{on}}(x)$ and $t_{\mathrm{off}}(x)$) may also be subject to regulatory control. First, we considered the role of $t_{\mathrm{on}}(x)$ by envisioning a scenario where transcription begins at time $t_{\mathrm{on}}(x)$, but does not cease. In this scenario, the

accumulated mRNA is given by

$$\text{mRNA}_{\text{competent}}(x,t) = \underbrace{r(x)\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma}\left(1-e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{time window}}. \tag{22}$$

Note that if the system evolves for a long amount of time, the exponential term in the previous equation becomes large ($\gamma(t-t_{\text{on}}(x)) \gg 1$) such that it reaches steady state, resulting in ***Equation 21***. Finally, we considered the impact of regulating the timing with which nuclei cease transcriptional activity (t_{on}). Here, when $t > t_{\text{on}}(x)$, the amount of mRNA produced during the period of activity is subsumed within a decaying exponential envelope such that

$$\text{mRNA}_{\text{quiescent}}(x,t) = \underbrace{e^{-\gamma(t-t_{\text{off}}(x))}}_{\text{quiescent decay}} \left[\underbrace{r(x)\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma}\left(1-e^{-\gamma(t_{\text{on}}(x)-t_{\text{on}}(x))}\right)}_{\text{time window}}\right]. \tag{23}$$

Equation 23 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 (analogue control of transcriptional bursting). We refactor ***Equation 23*** to reflect this distinction and consider the case when $t < t_{\text{on}}$, giving

$$\text{mRNA}_{\text{full}}(x,t) = \underbrace{\frac{r(x)}{\gamma}\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\text{analogue control}} \times \underbrace{e^{-\gamma(t-\min(t_{\text{off}}(x),t))}\left(1-e^{-\gamma(\min(t_{\text{off}}(x),t)-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{24}$$

Which can be simplified slightly to yield

$$\text{mRNA}(x,t) = \underbrace{\frac{r(x)}{\gamma}\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\text{analogue control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{25}$$

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

The presence of two transcriptional loci within each observed fluorescent spot necessitates the extension of the 2-state model to describe 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 3***). We begin with a general expression for this scenario that takes the contribution from the analogue control term shown in ***Equation 25*** to be a sum over the output of each of the 3 activity states

$$\text{mRNA}(x,t) = \underbrace{\frac{1}{\gamma}\left(\sum_{i=0} r_i(x)\pi_i(x)\right)}_{\text{analogue control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}, \tag{26}$$

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 3-state case. The $\pi_i(x)$ terms denote the steady-state occupancies of each activity state and are a function of the rates with which the promoter switches between activity states as defined

in ***Figure 3***B. In general, the fractional occupancy of each activity state, p_i, may vary as a function of time

$$\frac{\partial \boldsymbol{p}(x,t)}{\partial t} = \boldsymbol{R}(x)\boldsymbol{p}(x,t). \tag{27}$$

Where **R** denotes the transition rate matrix that describes the system

$$\boldsymbol{R}(x) = \begin{bmatrix} -k_{01}(x) & k_{10}(x) & 0 \\ k_{01}(x) & -k_{10}(x) - k_{12}(x) & k_{21}(x) \\ 0 & k_{12}(x) & -k_{21}(x) \end{bmatrix} \tag{28}$$

Where, consistent with our inference results, we take the corner terms to be equal to 0. Thus, the $\pi_i(x)$ terms comprise the occupancy vector, $\boldsymbol{\pi}(\boldsymbol{x})$, that adheres to the following condition

$$0 = \boldsymbol{R}(x)\boldsymbol{\pi}(x). \tag{29}$$

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 Markov chain such that, upon reaching $\boldsymbol{\pi}$, no further shifts in the mean fraction of time spent in each activity state. ***Equation 29*** leads to a system of three equations

$$0 = -k_{01}\pi_0 + k_{10}\pi_1 \tag{30}$$

$$0 = \pi_0 k_{01} - \pi_1\left(k_{10} + k_{12}\right) + \pi_2 k_{21} \tag{31}$$

$$0 = \pi_1 k_{12} - \pi_2 k_{21} \tag{32}$$

Before proceeding, we note that, since $\boldsymbol{\pi}$ is a probability distribution, we can eliminate one of our unknowns by enforcing normalization

$$1 = \pi_0 + \pi_1 + \pi_2. \tag{33}$$

With this in mind, we can solve ***Equation 30*** for π_1

$$\pi_1 k_{10} = \pi_0 k_{01} \tag{34}$$

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}}. \tag{35}$$

Next, we use the normalization condition to eliminate π_2 from ***Equation 32***

$$\pi_1 k_{12} = \pi_2 k_{21} \tag{36}$$

$$\pi_1 k_{12} = (1 - \pi_0 - \pi_1)k_{21}. \tag{37}$$

By combining this result with ***Equation 35*** we obtain

$$\pi_0 \frac{k_{01}}{k_{10}} k_{12} = (1 - \pi_0 - \pi_0 \frac{k_{01}}{k_{10}})k_{21} \tag{38}$$

$$\pi_0 \frac{k_{01}k_{12}}{k_{10}k_{21}} = 1 - \pi_0 \frac{k_{10} + k_{01}}{k_{10}} \tag{39}$$

$$\pi_0 = \frac{k_{10}k_{21}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}}. \tag{40}$$

With ***Equation 40*** in hand, it is then straight forward to solve for the remaining π_i terms. First we obtain π_1 by plugging ***Equation 40*** into ***Equation 35***

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}} \tag{41}$$

$$\pi_1 = \frac{k_{01}k_{21}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}}. \tag{42}$$

And finally π_2

$$\pi_2 = 1 - \pi_0 - \pi_1 \tag{43}$$

$$\pi_2 = \frac{k_{01}k_{12}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}}. \tag{44}$$

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

$$\text{mRNA}(x,t) = \underbrace{\frac{1}{\gamma}\left(r_1(x)\frac{k_{01}(x)k_{21}(x)}{\kappa(x)} + r_2(x)\frac{k_{01}(x)k_{12}(x)}{\kappa(x)}\right)}_{\text{analogue control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{45}$$

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 40***, ***Equation 42***, ***Equation 44***

$$\kappa = k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}. \tag{46}$$

Thus, from ***Equation 45*** we see that, while there are more terms comprising the analogue control expression, the expression nonetheless takes on the same essential form as in ***Equation 25***.

Deriving expressions for cytoplasmic mRNA levels away from steady state

For the majority of this work, we operated under the simplifying assumption that bursting parameters do not vary in time; however, the results of our windowed mHMM inference make it clear that bursting parameters do, in fact, exhibit significant temporal variation. While beyond the scope of the present work, we note here that the general solution to ***Equation 19*** and ***Equation 20*** takes the form

$$\text{mRNA}(x,t) = \int_{t_{\text{on}}(x)}^{min(t_{\text{on}}(x),t)} \underbrace{r(x,\tau)\frac{k_{\text{on}}(x,\tau)}{k_{\text{on}}(x,\tau) + k_{\text{off}}(x,\tau)}}_{\text{analogue control}} \underbrace{e^{-\gamma(t-\tau)}}_{\text{binary control}}. \tag{47}$$

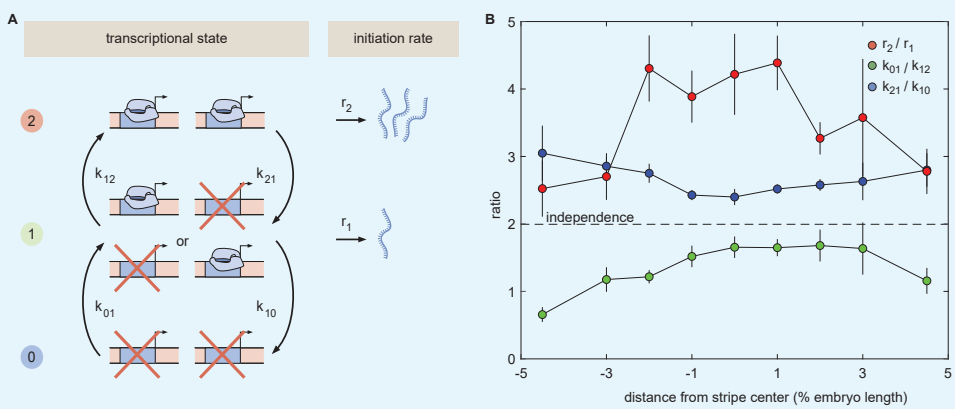
This expression makes it possible to calculate the mRNA accumulation as a function of space and time for arbitrary dependence of all the model parameters.

Appendix 2

Effect of sister chromatid correlation on model transition rates

As illustrated in ***Appendix 2–Figure 1***A, our 3-state kinetic model assumed that each observed fluorescence spot in comprised of two distinct promoters. The model imposed no assumptions regarding the nature or strength of the coupling between transcriptional activity at these sister loci. In addition to permitting greater flexibility, this agnostic approach also meant that the structure of the kinetic model returned by our mHMM 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). If the two loci are completely independent, all three ratios should be equal to 2. In the case of the initiation rates, this expectation arises because state (2) should correspond to two identical loci actively initiating transcripts, whereas state (1) has only one active locus. For the transition rates, we expect k_{01} and k_{21} to be twice as large as k_{12} and k_{10}, respectively, because, for any switch out of states (0) or (2), *both* loci are eligible to transition, whereas one and only one is eligible for any switch out of state (1) (see ***Figure 1***A). Thus, the effective rates of switching out of states (0) and (2) should be equal to twice the single promoter on and off rates ($2k_{\mathrm{on}}$ and $2k_{\mathrm{off}}$, respectively, as defined in ***Figure 1***A), while the rates of switching out of state (1) would be expected to take on the single promoter values (k_{on} and k_{off}). Any deviation from these expectations indicates that transcriptional activity at the two neighboring loci is coupled in some way. ***Appendix 2–Figure 1***B summarizes our findings.



Appendix 2 Figure 1. Probing the coupling between sister loci. (A) Schematic of general 3-state kinetic model inferred for transcriptional loci. (B) Summary of bursting parameter ratios. All three bursting parameter ratios deviate from their expected value under the independence assumption given by the horizontal dashed line.

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, $r_1(x)$ (***Appendix 2–Figure 1***B, red). 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}, (***Appendix 2–Figure 1***B, blue). This indicates that each promoter is more likely to switch off when its sister 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 (***Appendix 2–Figure 1***B, green).

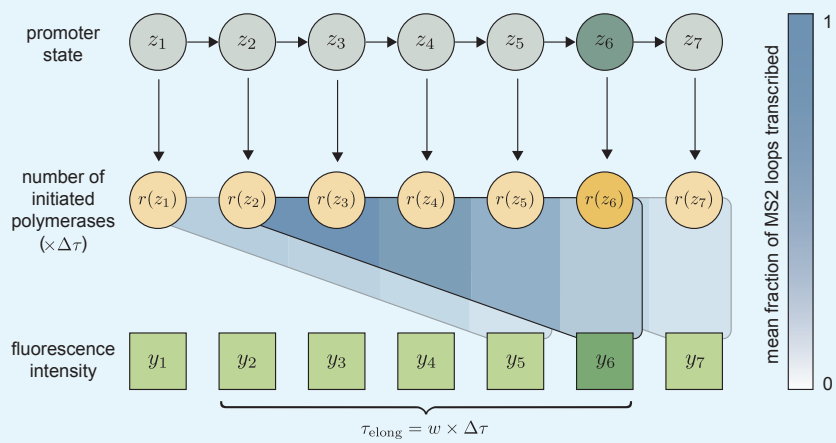
Further experiments in which the sister chromatids are labeled in an orthogonal manner are needed to confirm and elaborate upon these results. One important consideration to address is the fact that the spatial proximity of the two loci appears to fluctuate significantly over time (*see*, *e.g.*, ***Figure 3***A). 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.

Appendix 3

The memory-adjusted hidden Markov model

Model introduction

To model the dynamics of an observed fluorescence series, $\boldsymbol{y} = \{y_1, y_2, ..., y_T\}$, 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 $\boldsymbol{A} = p(z_t|z_{t-1})$, where A_{kl} is the probability of transitioning from the lth state into the kth state in the time interval $\Delta\tau$, where $\Delta\tau$ is the data sampling resolution. We assign a characteristic polymerase 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 polymerases will initially grow linearly ($\tau \leq \tau_{\text{MS2}}$) and will stay constant for the remaining 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).



Appendix 3 Figure 1. Schematic overview of the mHMM 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 transcribed by polymerases at varying positions along the gene at the moment of observation.

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. Thus, 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,

$\{z_t, z_{t-1}, ..., z_{t-w+1}\}$. 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, $s_t = \{z_t, z_{t-1}, ..., z_{t-w+1}\}$, which, together with the set of model parameters, $\boldsymbol{\theta}$, is sufficient to define the probability distribution of the observation y_t, thereby satisfying the Markov condition. Since $z_t \in \{1, ..., K\}$, each compound state can take one of K^w different values, $s_t \in \{1, ..., K^w\}$. 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 $s_{t+1} = \{z_{t+1}, z_t, ..., z_{t-w+2}\}$ are present in s_t as well. The schematic overview of the mHMM architecture is shown in ***Appendix 3–Figure 1***.

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 $p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta})$ of the series of hidden compound states, $\boldsymbol{s} = \{s_1, s_2, ..., s_T\}$, and fluorescence values, $\boldsymbol{y} = \{y_1, y_2, ..., y_T\}$, is given by

$$p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = p(s_1|\boldsymbol{\pi}) \prod_{t=1}^{T} p(y_t|s_t, \boldsymbol{\upsilon}, \sigma) \prod_{t=2}^{T} p(s_t|s_{t-1}, \boldsymbol{A}). \tag{48}$$

Here $\boldsymbol{\pi}$ is a K-element vector, with π_k being the probability that the trace starts at the kth effective promoter state, and $\boldsymbol{\upsilon}$ is a K-element vector of fluorescence emission values per time step.

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

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{49}$$

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

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{\boldsymbol{s}=\{s_1, s_2, ..., s_T\}} p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}). \tag{50}$$

Note that the summation is performed over all possible choices of $\boldsymbol{s}$ - a vector of T elements, each of which can take K^w possible values. The total number of terms in the sum in 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 (***Corrigan et al., 2016***) 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, making our model more versatile and with fewer parameters. In the "Continuous vs. Poisson promoter loading" section of ***Appendix 3*** we demonstrate that relaxing the continuous RNAP loading assumption when generating synthetic data does not significantly affect the accuracy of the mHMM inference.

Expectation-maximization (EM) algorithm

Consistent with standard EM approaches (cf. Bishop (***Christopher, 2006***), 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,

$$\hat{\boldsymbol{\theta}}_{k+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{51}$$

$$\mathcal{L}(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{\boldsymbol{s}=\{s_1, s_2, \ldots, s_T\}} p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k)\, \log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) \leq \log\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{52}$$

Here $\mathcal{L}(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k)$ is the objective function, $\hat{\boldsymbol{\theta}}_k$ is the estimate of the model parameters in the kth 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, $\log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta})$, can be written as

$$\log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = \log p(s_1|\boldsymbol{\pi}) + \sum_{t=1}^{T} \log p(y_t|s_t, \boldsymbol{v}, \sigma) + \sum_{t=2}^{T} \log p(s_t|s_{t-1}, \boldsymbol{A}). \tag{53}$$

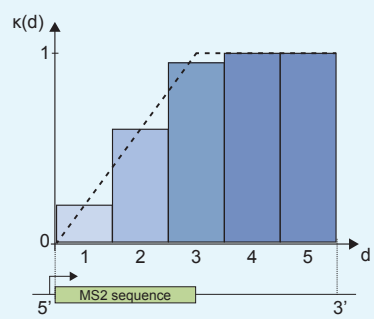
Now, we introduce several notations - $s_t^i := 1$ if and only if $s_t = i$; $\Delta(s_t, d) :=$ the dth digit of the promoter state sequence $s_t = \{z_t, z_{t-1}, \ldots, z_{t-(w-1)}\}$, starting from the left end; $C_{zs} = 1$ if and only if $\Delta(s, 1) = z$; $B_{s',s} = 1$ if and only if the transition $s \to s'$ 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 53*** can be rewritten as

$$\log p(s_1|\boldsymbol{\pi}) = \sum_{i=1}^{K^w} \sum_{k=1}^{K} s_1^i C_{ki} \log \pi_k, \tag{54}$$

$$\log p(y_t|s_t, \boldsymbol{v}, \sigma) = \frac{1}{2} \sum_{i=1}^{K^w} s_t^i \left(\log \lambda - \log(2\pi) - \lambda(y_t - V_i(\boldsymbol{v}))^2\right), \tag{55}$$

$$\log p(s_t|s_{t-1}, \boldsymbol{A}) = \sum_{i,j=1}^{K^w} \sum_{k,l=1}^{K} B_{ij} s_t^i s_{t-1}^j C_{ki} C_{lj} \log A_{kl}. \tag{56}$$

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



Appendix 3 Figure 2. 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: $\tau_{\text{elong}} = 100$ sec, $\tau_{\text{MS2}} = 50$ sec, $\Delta\tau = 20$ sec, $w = \tau_{\text{elong}}/\Delta\tau = 5$, $n_{\text{MS2}} = \tau_{\text{MS2}}/\Delta\tau = 2.5$.

Because of the finite time τ_{MS2} it takes 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 at the dth 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 ***Appendix 3–Figure 2***.

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

$$V_i(\boldsymbol{v}) = F_{i,:}\boldsymbol{v}, \tag{57}$$

where the ith row of the $K^w \times K$ matrix $\boldsymbol{F}$ is the number of times each promoter state is present in the ith 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 $\boldsymbol{F}$ 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, $\log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta})$, we obtain a final expression for the objective function, namely,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = & \sum_{i=1}^{K^w} \sum_{k=1}^{K} \langle s_1^i \rangle C_{ki} \log \pi_k \\ & + \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle \left(\log \lambda - \log(2\pi) - \lambda (y_t - F_{i,:}\boldsymbol{v})^2 \right) \\ & + \sum_{t=1}^{T} \sum_{i,j=1}^{K^w} \sum_{k,l=1}^{K} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{lj} \log A_{kl}. \end{aligned} \tag{58}$$

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

$$\langle s_t^i \rangle = \sum_{\boldsymbol{s}=\{s_1, s_2, \ldots, s_T\}} s_t^i \, p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{59}$$

$$\langle s_t^i s_{t-1}^j \rangle = \sum_{\boldsymbol{s}=\{s_1, s_2, \ldots, s_T\}} s_t^i s_{t-1}^j \, p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k). \tag{60}$$

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 59*** and ***Equation 60***, we obtain

$$\langle s_t^i \rangle = \sum_{s_1, s_2, \ldots, s_T} s_t^i \, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t} s_t^i \, p(s_t|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{61}$$

$$\langle s_t^i s_{t-1}^j \rangle = \sum_{s_1, s_2, \ldots, s_T} s_t^i s_{t-1}^j \, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t, s_{t-1}} s_t^i s_{t-1}^j \, p(s_t, s_{t-1}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k). \tag{62}$$

Following the conventional implementation of the forward-backward algorithm (cf. Bishop (***Christopher, 2006***), Chapter 13), we use the Markov property of the promoter

state dynamics, together with the sum and products rules of probability, to write

$$p(s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k) = \frac{\alpha_t(s_t)\beta_t(s_t)}{p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_k)}, \tag{63}$$

$$p(s_{t-1},s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k) = \frac{\alpha_{t-1}(s_{t-1})\,p(y_t|s_t,\hat{\boldsymbol{\theta}}_k)\,p(s_t|s_{t-1},\hat{\boldsymbol{\theta}}_k)\beta_t(s_t)}{p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_k)}, \tag{64}$$

$$\alpha_t(i) = p(y_1,\ldots,y_t,s_t=i|\hat{\boldsymbol{\theta}}_k), \tag{65}$$

$$\beta_t(i) = p(y_{t+1},\ldots,y_T|s_t=i,\hat{\boldsymbol{\theta}}_k). \tag{66}$$

Here $\alpha_t(i)$ is the joint probability of observing the fluorescence emission values in the first t steps and being at the ith 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 65*** and ***Equation 66***. We evaluate the elements of α and β matrices recursively as

$$\alpha_t(i) = p(y_t|s_t=i,\hat{\boldsymbol{\theta}}_k)\sum_{j=1}^{K^w}\alpha_{t-1}(j)\,p(s_t=i|s_{t-1}=j,\hat{\boldsymbol{\theta}}_k), \tag{67}$$

$$\beta_t(i) = \sum_{j=1}^{K^w}\beta_{t+1}(j)\,p(y_{t+1}|s_{t+1}=j,\hat{\boldsymbol{\theta}}_k)\,p(s_{t+1}=j|s_t=i,\hat{\boldsymbol{\theta}}_k). \tag{68}$$

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

$$\alpha_1(i) = p(y_1|s_1=i,\hat{\boldsymbol{\theta}}_k)\,p(s_1=i|\hat{\boldsymbol{\theta}}_k), \tag{69}$$

$$\beta_T(i) = 1, \tag{70}$$

where the first follows the definition of $\alpha_t(i)$, and the second is obtained from ***Equation 63*** by setting $t=T$. Having evaluated the α and β matrices, the likelihood $p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_k)$ that appears in the denominator of ***Equation 63*** and ***Equation 64*** can be found by setting $t=T$ in ***Equation 63*** and summing over s_T, namely,

$$\left(\sum_{s_T=1}^{K^w} p(s_T|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k)\right)p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_k) \equiv p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_k) = \sum_{s_T=1}^{K^w}\alpha_T(s_T). \tag{71}$$

With the probabilities $p(s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k)$ and $p(s_{t-1},s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k)$ known, the expectation coefficients follow directly from ***Equation 61*** and ***Equation 62***.

The optimal model parameters in the $(k+1)^{\text{th}}$ step of the EM algorithm are obtained by maximizing the objective function $\mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k)$ in ***Equation 58*** with respect to $\{\boldsymbol{\pi},\boldsymbol{\nu},\lambda,\boldsymbol{A}\}$, subject to the probability constraints $\sum_{k=1}^{K}\pi_k = 1$ and $\sum_{k=1}^{K}A_{kl} = 1$, $1 \le l \le K$. The update

equations for the model parameters are found as

$$\text{initial state pmf: } \hat{\pi}_m = \frac{\sum_{i=1}^{K^w} \langle s_1^i \rangle C_{mi}}{\sum_{k=1}^{K} \sum_{i=1}^{K^w} \langle s_1^i \rangle C_{ki}}, \tag{72}$$

$$\text{fluorescence emission rates: } \hat{\boldsymbol{\upsilon}} = \boldsymbol{M}^{-1}\boldsymbol{b}, \text{ where} \tag{73}$$

$$M_{mn} = \sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle F_{in} F_{im}, \tag{74}$$

$$b_m = \sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle y_t F_{im}, \tag{75}$$

$$\text{noise: } \frac{1}{\hat{\lambda}} = \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle (y_t - F_{i,:}\hat{\boldsymbol{\upsilon}})^2}{\sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle}, \tag{76}$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{t=1}^{T} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{mi} C_{nj}}{\sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{nj}}. \tag{77}$$

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 $\boldsymbol{y}_{1:N}$ are N different fluorescence traces with corresponding trace lengths $T_{1:N}$, and $\boldsymbol{s}_{1:N}$ are the hidden compound state sequences corresponding to each trace, from the independence criterion we obtain

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{s}_{1:N} | \boldsymbol{\theta}) = \prod_{n=1}^{N} p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}), \tag{78}$$

$$p(\boldsymbol{s}_n | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) = p(\boldsymbol{s}_n | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k), \quad 1 \le n \le N. \tag{79}$$

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

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) &= \sum_{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N} p(\boldsymbol{s}_{1:N} | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_{1:N}, \boldsymbol{s}_{1:N} | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \sum_{\boldsymbol{s}_n} p(\boldsymbol{s}_n | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \sum_{\boldsymbol{s}_n} p(\boldsymbol{s}_n | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \mathcal{L}_n(\boldsymbol{\theta} | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k).
\end{aligned} \tag{80}$$

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 58***), we find

$$\begin{aligned}\mathcal{L}(\theta\,|\,\boldsymbol{y}_{1:N},\hat{\theta}_k) = &\sum_{n=1}^{N}\sum_{i=1}^{K^w}\sum_{k=1}^{K}\langle s_1^i(n)\rangle C_{ki}\log\pi_k \\ &+\frac{1}{2}\sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{i=1}^{K^w}\langle s_t^i(n)\rangle\left(\log\lambda-\log(2\pi)-\lambda(y_t(n)-F_{i,:}\boldsymbol{v})^2\right) \\ &+\sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{i,j=1}^{K^w}\sum_{k,l=1}^{K}B_{ij}\langle s_t^i(n)s_{t-1}^j(n)\rangle C_{ki}C_{lj}\log A_{kl},\end{aligned} \tag{81}$$

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 nth fluorescence trace via the forward-backward algorithm, and $y_t(n)$ is the fluorescence at time t in the nth trace. The update equations are then derived analogous to the single-trace case, with an additional summation performed over all traces, namely,

$$\text{initial state pmf: } \hat{\pi}_m = \frac{\sum_{h=1}^{N}\sum_{i=1}^{K^w}\langle s_1^i(h)\rangle C_{mi}}{\sum_{k=1}^{K}\sum_{h=1}^{N}\sum_{i=1}^{K^w}\langle s_1^i(h)\rangle C_{ki}}, \tag{82}$$

$$\text{fluorescence emission rates: } \hat{\boldsymbol{v}} = \boldsymbol{M}^{-1}\boldsymbol{b}, \quad \text{where} \tag{83}$$

$$M_{mn} = \sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle F_{in}F_{im}, \tag{84}$$

$$b_m = \sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle y_t(h)F_{im}, \tag{85}$$

$$\text{noise: } \frac{1}{\hat{\lambda}} = \hat{\sigma}^2 = \frac{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle(y_t(h)-F_{i,:}\hat{\boldsymbol{v}})^2}{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle}, \tag{86}$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i,j=1}^{K^w}B_{ij}\langle s_t^i(h)s_{t-1}^j(h)\rangle C_{mi}C_{nj}}{\sum_{k=1}^{K}\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i,j=1}^{K^w}B_{ij}\langle s_t^i(h)s_{t-1}^j(h)\rangle C_{ki}C_{nj}}. \tag{87}$$

Execution of the mHMM method

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

$$f(\mathbf{x}) \sim \frac{\Gamma\left(\sum_{k=1}^{K}u_k\right)}{\prod_{k=1}^{K}\Gamma(u_k)}\prod_{k=1}^{K}x_k^{u_k-1}. \tag{88}$$

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, $\boldsymbol{r}$, and the Gaussian precision parameter, $\lambda = 1/\sigma^2$, we first treat the fluorescence data $\boldsymbol{y}_{1:N}$ as i.i.d. and use a simplified time-independent EM algorithm to find their optimal values (cf. Bishop (***Christopher, 2006***), 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 (25 in our implementation), 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. $s_t = \{z_t, z_{t-1}, ..., z_1\}$.

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 mHMM inferences with different initialization conditions on a machine with 24 CPU cores. Users of the mHMM method are advised to have this metric as a reference when estimating the computational cost of their inference.

Statistical validation of mHMM

To validate mHMM and produce ***Figure 4*** and ***Figure 4–Figure Supplement 1***, we generated synthetic trajectories of effective promoter states ($K = 3$) using the Gillespie algorithm (***Gillespie, 1976***) and added Gaussian noise to obtain synthetic activities traces. Parameters in ***Appendix 3–Table 1*** 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. We used the relation between the transition rate matrix, $\mathbf{R}$, and the inferred transition probability matrix, $\mathbf{A}$, to obtain estimates of the transition rates, namely,

$$\mathbf{A} = e^{\mathbf{R}\Delta\tau}, \tag{89}$$

$$R_{ij} = \left(\frac{1}{\Delta\tau}\log\mathbf{A}\right)_{ij}. \tag{90}$$

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.

Appendix 3 Table 1. 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 mHMM inference on real 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) min^{-1}
RNAP initiation rates (r_0, r_1, r_2)	(0, 18.5, 46) 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

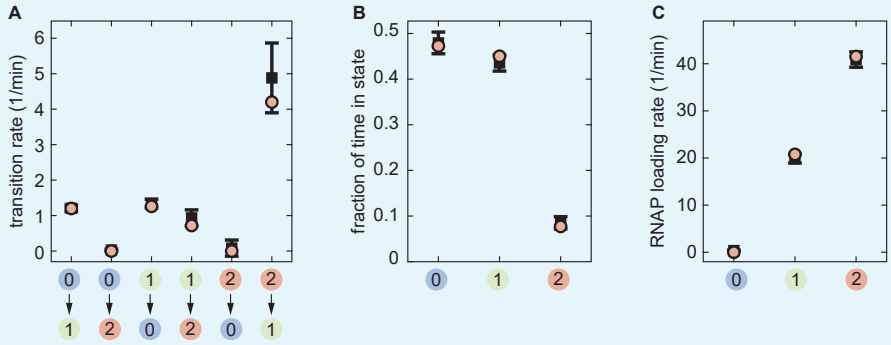
Continuous vs. Poisson promoter loading

To demonstrate the validity of our choice to use continuous RNAP initiation rates in the transcription model, we explicitly accounted for individual RNAP loading events when generating the traces. We assumed that individual polymerase molecules traverse at a constant elongation rate (v_{elong} = 46 bp/sec, ***Appendix 4***) 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 that has a finite footprint size of $l_{\text{RNAP}} = 50$ bp (***Rice et al., 1993***). With this information in hand, we calculated the mean arrival time of polymerases as

$$\langle \tau_{\text{arrival}} \rangle = \frac{1}{r_1} - \frac{l_{\text{RNAP}}}{v_{\text{elong}}}, \tag{91}$$

where r_1 is the mean RNAP loading rate at a single promoter. $\langle \tau_{\text{arrival}} \rangle$ was then used in simulating the arrival events of individual polymerases.

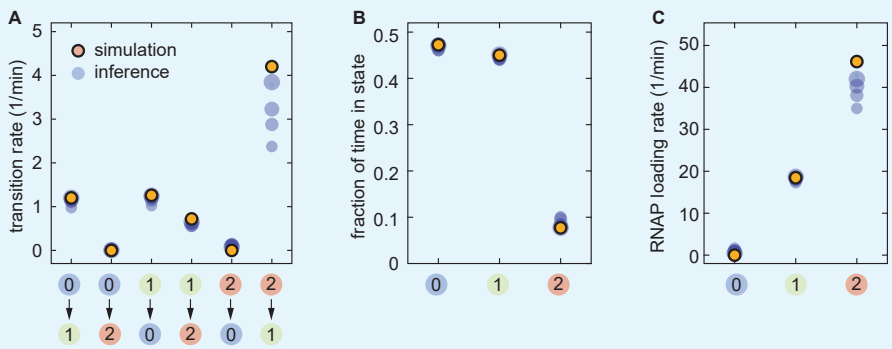
We perform an inference on these simulated traces using mHMM with the objective of determining whether a Poisson loading rate has an effect on the obtained parameters. As shown in ***Appendix 3–Figure 3***, when the data is generated using Poisson loading, mHMM slightly overestimates the high transition rate, but otherwise manages to accurately recover the model parameters.



Appendix 3 Figure 3. Validation of mHMM on Poisson promoter loading data. (A) Transition rates, **(B)** state occupancies and **(C)** RNAP loading rates inferred from 20 independently generated datasets assuming Poisson loading of RNAP. Error bars represent one standard devision from the mean inference values.

Sensitivity of mHMM to data sampling resolution

In our mHMM 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 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 mHMM, we conducted inference on synthetic activity traces generated with varying sampling rates. Since the memory of the system ($w = \tau_{\text{elong}}/\Delta\tau$) needs to be an integer, we varied w in the $[3, 7]$ range, correspondingly changing the sampling resolution from low ($\tau_{\text{elong}}/3 \approx 46$s) to high ($\tau_{\text{elong}}/7 = 20$s). We used the values in ***Appendix 3–Table 1*** for the remaining model parameters.



Appendix 3 Figure 4. Sensitivity of mHMM 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 transparent circles corresponds to higher data sampling resolutions (largest: 20s, smallest: 46s).

Appendix 3–Figure 4 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 (***Appendix 3–Figure 4***A). The accuracy of inferred state occupancies is also remarkably high, making it robust to variations in the data sampling rate (***Appendix 3–Figure 4***B). 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 RNA polymerases per time step (***Appendix 3–Figure 4***C). 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 become comparable.

Windowed mHMM

To investigate temporal trends in bursting parameters, we extended the mHMM 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 53*** 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,

$$\log p(\boldsymbol{y}_{T_1:T_2}, \boldsymbol{s}_{T_1:T_2}|\boldsymbol{\theta}) = \log p(s_{T_1}|\boldsymbol{\pi}^{(T_1-w+1)}, \boldsymbol{A}) + \sum_{t=T_1}^{T_2} \log p(y_t|s_t, \boldsymbol{r}, \sigma) + \sum_{t=T_1}^{T_2} \log p(s_t|s_{t-1}, \boldsymbol{A}), \tag{92}$$

$$\begin{aligned}\log p(s_{T_1}|\boldsymbol{\pi}^{(T_1-w+1)}, \boldsymbol{A}) &= \log \left(p(z_{T_1-w+1}|\boldsymbol{\pi}^{(T_1-w+1)}) \prod_{t=T_1-w+2}^{T_1} p(z_t|z_{t-1}, \boldsymbol{A}) \right) \\ &= \sum_{i=1}^{K^w} \sum_{n=1}^{K} s^i_{T_1} D^w_{ni} \log \pi_n^{(T_1-w+1)} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \sum_{k,l=1}^{K} s^i_{T_1} D^{d-1}_{ki} D^d_{li} \log A_{kl}.\end{aligned} \tag{93}$$

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^d_{ni} is an indicator variable which takes the value 1 only if the promoter state in the dth position of the ith compound state is n.

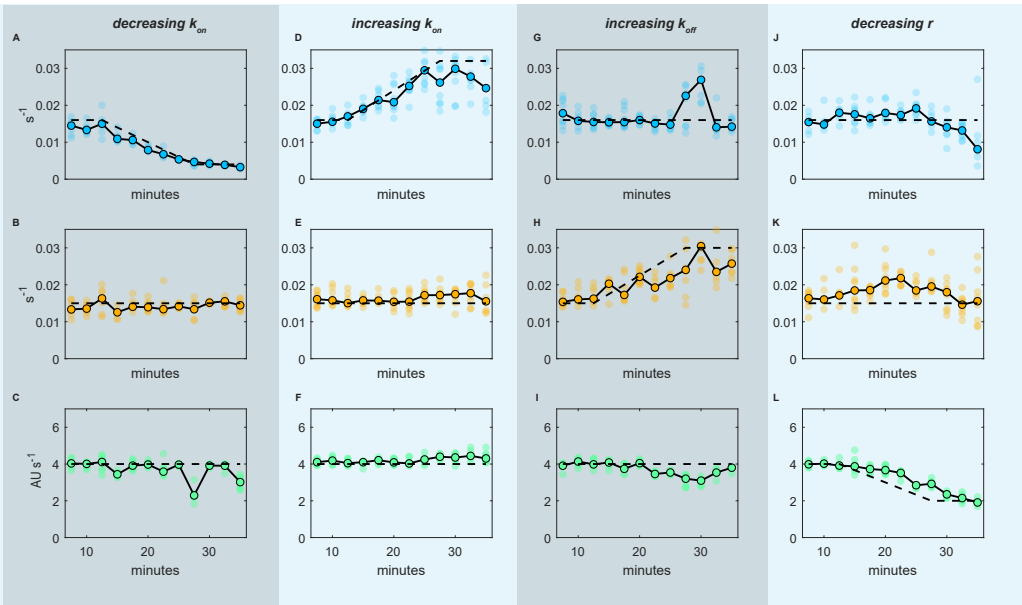
The modified expression for the joint probability distribution does not change the functional form of the equations used for calculating the expectation coefficients. Maximization equations for the emission rates and the noise also remain intact. Only the maximization equation for the transition probabilities is revised from ***Equation 77*** into

$$\hat{A}_{mn} = \frac{\sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{mi} C_{nj} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{mi} D^d_{ni} \log A_{mn}}{\sum_{k=1}^{K} \sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{ki} C_{nj} + \sum_{k=1}^{K} \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{ki} D^d_{ni} \log A_{kn}}. \tag{94}$$

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. $\boldsymbol{A}\,\boldsymbol{\pi}^{(T_1-w+1)} = \boldsymbol{\pi}^{(T_1-w+1)}$. We therefore use the current estimate of $\boldsymbol{A}$ to evaluate $\boldsymbol{\pi}^{(T_1-w+1)}$ at each EM iteration, instead of performing a maximization step.

To check that our extended mHMM 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. ***Appendix 3–Figure 5*** summarizes the results for these validation tests.

For each test, 100 simulated traces, 40 minutes in length, were generated ($\Delta\tau$ =20 sec) 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 lead to inference groups consisting of 4500 data points, with the exception of the first and last time points, which each had 3700 data points (first and last $w+1$ points are excluded from inference). Transition and initiation rates shown in ***Appendix 3–Figure 5*** are associated with state (1) of the 3-state model (k_{01}, k_{10} and r in ***Appendix 2–Figure 1***A), as these were found to provide the most faithful indication of underlying system trends.



Appendix 3 Figure 5. Validating windowed mHMM 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 parameter value at each time point across 10 distinct replicates. Translucent points indicate 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 mHMM 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 (***Appendix 3–Figure 5***A-C,D-E), windowed mHMM 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} (***Appendix 3–Figure 5***G-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 (***Appendix 3–Figure 5***J-L) we observe a $\sim$ 5 minute delay in the model response. We attribute this delay to the finite dwell time of RNAP molecules on the gene (in this case τ_{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 ***Appendix 3–Figure 5***J, K).

Overall, we conclude that the windowed mHMM 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.

Appendix 4

Determining the RNAP dwell time using autocorrelation

In order to conduct mHMM inference, it is necessary to specify the number of time steps w required for an RNAP molecule to traverse the reporter gene.

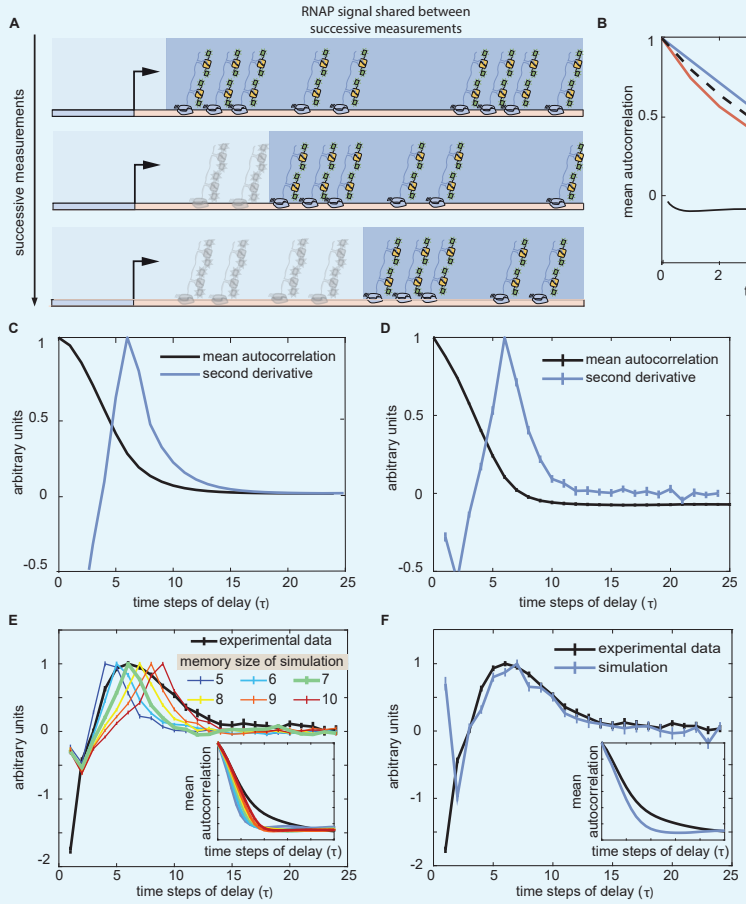
$$w = \frac{\tau_{elong}}{\Delta t}, \tag{95}$$

While Δt is set by the temporal resolution of our data acquisition, the elongation time (τ_{elong}) is *a priori* unknown. 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 $1.4 - 1.7$ kb min^{-1} (***Garcia et al.*** (***2013***)). However, a recent study of the selfsame regulatory element reported elongation rates of $2.4 - 3.0$ kb min^{-1}—nearly twice as fast (***Fukaya et al.*** (***2017***)). 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 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

$$R_F(\tau) = \frac{E[(F(t) - \mu_f)(F(t-\tau) - \mu_f)]}{\sigma_f^2}, \tag{96}$$

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 ***Appendix 4–Figure 1***A, 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 $\tau < \tau_{elong}$ 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 $\tau = \tau_{elong}$ (***Appendix 4–Figure 1***B, blue curve).



Appendix 4 Figure 1. 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). Theoretical analysis of the autocorrelation function (C) and stochastic simulations (D) 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 times steps (140s). (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)

This structural break in the autocorrelation function that occurs at $\tau = \tau_{elong}$ can be used to estimate the elongation time of the system; however, it is not the only feature present in ***Equation 96***. 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 "dynamics" 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 (***Appendix 4–Figure 1***B, red curve)

$$R_P(\tau) \sim e^{-\lambda_2 \tau}. \tag{97}$$

Thus, the observed autocorrelation function contains, at a minimum, signatures of both the finite RNAP dwell time (τ_{elong}) and to promoter switching dynamics. As a result, inferring elongation times from the structural break 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 use to read out the value of τ_{elong}. ***Appendix 4–Figure 1***C 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 (***Appendix 4–Figure 1***D). 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 ***Appendix 4–Figure 1***E indicates the form of the autocorrelation second derivative for the set of traces used for mHMM 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 $w = 7 = 140$ s best characterized our data (***Figure 1***E, green curve). This implies that

$$v_{elong} = \frac{6444\text{bp}}{140\text{s}} \tag{98}$$

$$v_{elong} = 46\text{bp s}^{-1} \tag{99}$$

$$v_{elong} = 2.8\text{kb min}^{-1}. \tag{100}$$

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 $2.4 - 3.0$ kb min^{-1} range reported in ***Fukaya et al.*** (***2017***).

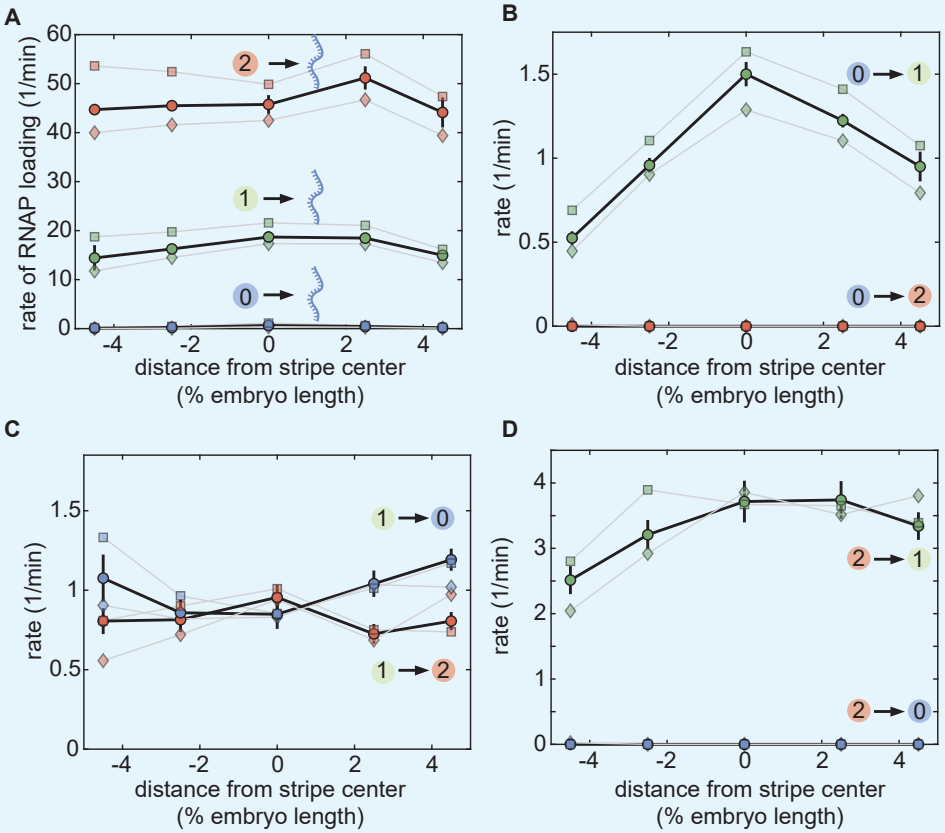
Appendix 4–Figure 1F shows how a simple adjustment to our simulation approach, wherein the elongation times for individual RNAP molecules were drawn from a Gaussian distribution with mean $\mu_w = 7$ and standard deviation $\sigma_w = 2.5$ time steps can qualitatively reproduce the wider profile observed 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 are inference was robust to our choice of τ_{elong}, testing cases where $\tau_{elong} = 120$ and $\tau_{elong} = 160$ (see below).

mHMM 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 ***Appendix 4–Figure 2***, we conducted mHMM 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 ***Appendix 4–Figure 1***E). While small quantitative difference are apparent across these three cases (median $CV = 11\%$), the offsets between sensitivities were generally found to be consistent, such that qualitative trends were largely robust to the assumed w value.



Appendix 4 Figure 2. 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)-(C) Transition rates also exhibit high degree of robustness to the w used for inference, although we observed some variability in the transition rates out of the middle state (1) for the stripe flank regions as shown in (C).

Appendix 5

Measuring the amount of produced mRNA

Here, we outline the approach that was used to estimate the total amount of mRNA produced by *eve* stripe 2 nuclei from MS2 traces. This approach, which is independent of the bursting parameters estimates returned by mHMM was used to estimate the total cytoplasmic mRNA levels per nucleus shown in ***Figure 7***F (green), as well as the "binary control" of the duration of the transcriptional time window contribution (yellow).

Calculating full mRNA profiles

The observed fluorescent 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 a RNAP molecule to transcribe the gene, τ_{elong}, the number of new mRNAs added to the cytoplasm will be proportional to $F(t)$ (***Bothma et al.*** (***2014***))

$$F(t) \propto M(t + t_{elong}) - M(t). \tag{101}$$

Where $M(t)$ indicates the total number of mRNA molecules taht have bee produced up to time t. We relate this fluorescence signal to absolute numbers of RNAP molecules using the calibration procedure described in the Materials & Methods. However, only the *relative* amounts of mRNA present across the *eve* stripe 2 pattern are needed in order to calculate the relative contributions from the different regulatory strategies identified in ***Equation 7***. 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. Drawing from the derivation provided in the SI Methods of ***Bothma et al.*** (***2014***), we take the rate of mRNA production time at time t, to be approximately equal to the observed fluorescence at time $t - \frac{t_{elong}}{2}$

$$F\left(t - \frac{t_{elong}}{2}\right) \approx \beta \frac{\mathrm{d}M(t)}{\mathrm{d}t}. \tag{102}$$

Where the $\frac{t_{elong}}{2}$ term accounts for the fact the time lag between the number of transcribing nascent mRNA and the rate of mRNA release into the cytoplasm. For ease of notation, we will ignore this offset factor for the remainder of this section. We will also treat the relationship in ***Equation 102*** as one of equality. For ***Figure 7***F, the metric of interest is the amount of mRNA produced *per nucleus*. Thus for a given region along the axis of the embryo, the average observed fluorescence across *all* N nuclei (both active and quiescent) within the region of interest was used as a proxy for the instantaneous rate of mRNA production per nucleus, given by

$$\frac{\mathrm{d}M(x,t)}{\mathrm{d}t} = \frac{\beta}{N} \sum_{i=1}^{N} F_i(x,t) \tag{103}$$

$$= \beta \langle F(t) \rangle_x. \tag{104}$$

Here, $F_i(x,t)$ is the fluorescence of nucleus i at time t. The x subscript in ***Equation 104*** indicates that the average is taken over all nuclei falling within the same spatial 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 summed over all time points for each region of interest

to estimate the total amount of cytoplasmic mRNA present on average in individual nuclei, yielding the quantity on the left-hand side of ***Equation 7***.

$$\text{mRNA}(x,t) = \beta \sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \left\langle F(n\Delta t) \right\rangle \tag{105}$$

Where Δt is the experimental time resolution and $T = \frac{t}{\Delta t}$ denotes the number of measurements taken through time t. The exponential term within the summand on the RHS captures the effects of mRNA decay (*see* ***Appendix 1***). Finally, to calculate the normalized mRNA profile shown in ***Figure 7*** (green), the estimates for the total accumulated mRNA per nucleus found using ***Equation 7*** must be divided by the sum across all spatial regions considered

$$\text{mRNA}^{j}_{norm}(x,t) = \frac{\sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \left\langle F(n\Delta t) \right\rangle_j}{\sum_{i\in X} \sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \left\langle F(n\Delta t) \right\rangle_i}, \tag{106}$$

where 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 mRNA_{norm}.

Calculating mRNA profiles due to binary control

The predicted profile due to binary switching alone (***Figure 7***F, yellow) was calculated following the same procedure as for the full mRNA profile described above, save for the fact that, in this case, instantaneous fluorescent values for individual nuclei were converted to binary indicator variables ($f_i(t)$) that were set equal to 1 if $t > t^{i}_{off}$ and 0 otherwise. In this scenario the "average rate" of mRNA production is equivalent to the fraction of active nuclei at a given point in time such that the rate of mRNA production is given by

$$\frac{\mathrm{d}M_{bin}(x,t)}{\mathrm{d}t} = \frac{1}{N(x)} \sum_{i=1}^{N} (x,t) f_i(t) \tag{107}$$

$$= \langle f(x,t) \rangle \tag{108}$$

$$= \frac{N_c(x,t)}{N(x,t)}, \tag{109}$$

where $N_c(t)$ indicates the number of transcriptionally competent nuclei at time t. The binary equivalent to ***Equation 105*** takes the form of a time-weighted sum of the fraction of active nuclei within a region

$$\text{mRNA}_{bin}(x,t) = \sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \frac{N_c(x,n\Delta t)}{N(x,t)}. \tag{110}$$

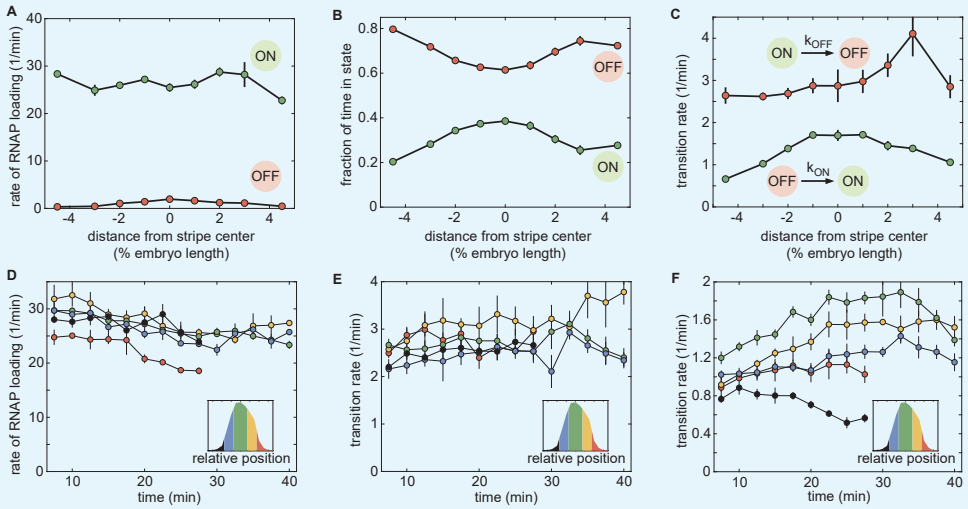
The expression for the normalized binary mRNA levels comprising the yellow profile in ***Figure 7***F takes the same for as ***Equation 106***.

Appendix 6

2-state Inference Results

Although the presence to sister chromatids indicated that the a 3-state model was most appropriate for the *eve* stripe 2 system, we wanted to check that our inference results were robust to this assumption. To do this, we conducted time-averaged and windowed inference assuming a simpler, 2-state model (*see, e.g.* ***Figure 1***A).

Most of our findings remained unchanged in the context of the 2-state model. Consistent with 3-state case, 2-state time-averaged mHMM 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 (***Appendix 6–Figure 1***A-C). Moreover, as with the 3-state case, 2-state results indicated the bulk of this variation resulted from modulation in k_{on} (***Appendix 6–Figure 1***C, green). Interestingly, whereas we did see a degree of spatial dependence in k_{off} for 3-states, observed no such trend for 2-states (***Appendix 6–Figure 1***C, blue). 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 3 state system, then we would expect the 2 state k_{off} estimate to reflect the joint action of the k_{10}, k_{21}, and k_{12} rates from the 3 state model.



Appendix 6 Figure 1. 2-state mHMM inference. Error bars indicate bootstrapped estimates of the standard error in mHMM inference. (A)-(C) Time-averaged 2-state inference results. (A) Consistent with 3-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 (C)) was modulated along the AP axis to vary the amount of time the promoter spent in the ON state (green curve (B)). In a departure from the 3-state case, we observed no significant AP trend in k_{off}, though we noted a spike in k_{off} at AP -3. (D)-(F) Time-resolved (windowed) 2-state mHMM results. (D) Consistent with the 3-state inference, we saw little to no modulation in r over time, although we noted a mild downward trend across all AP bins that was most pronounced in posterior flank (red curve). (E) 2-state inference indicated no significant temporal trends in k_{off}. (F) k_{on} time trends largely agreed with the 3-state case, although we noted that the decrease in k_{on} in the posterior flank that was apparent in the 3-state results was not observable in the 2-state case (***Figure 8***E, red). (Error bars indicate bootstrap estimates of the standard error in mHMM inference as described in ***Appendix 3***)

As with the time-averaged case, we found that results for 2-state windowed mHMM were generally consistent with 3-state trends. A notable exception to this rule was the absence of any significant decrease in k_{on} in the posterior stripe flank (***Appendix 6–Figure 1***F, red). This is not entirely surprising, as the trend returned by 3-state inference as relatively mild (***Figure 8***E,

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 3-state model allowed it to register a subtle shift in the activation rate that was convolved with countervailing features in the 2-state case. Future work will seek to elucidate the source of this discrepancy and further test the validity of the trend uncovered in the 3-state case.

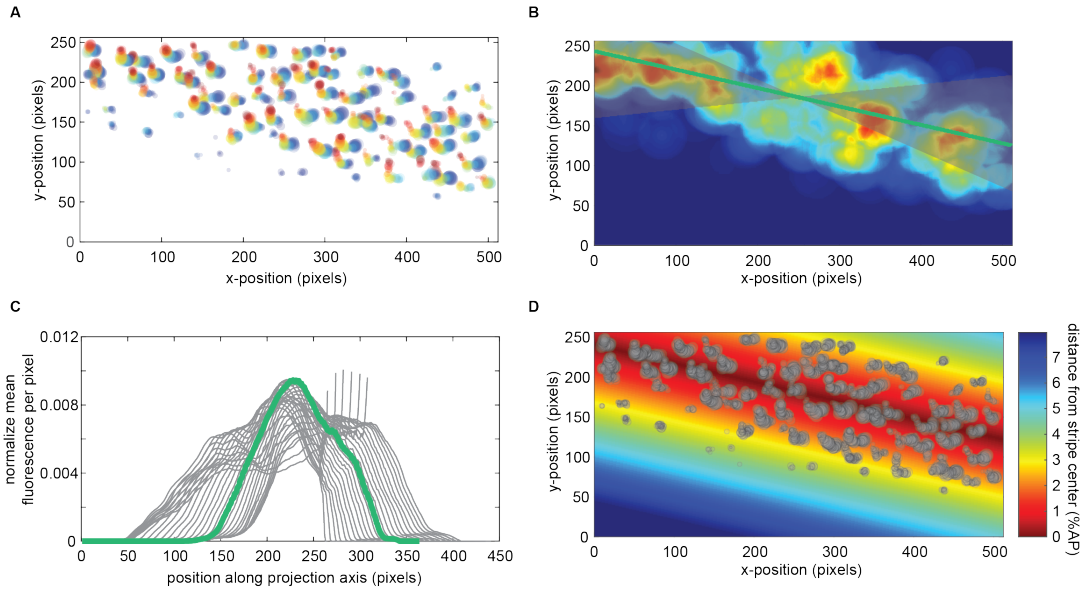


Figure 2–Figure supplement 1. Aligning stripes from multiple embryos. In order to minimize alignment errors when combining data from across multiple *Drosophila* embryos, an automated routine was employed to define a new experimental axis for each data set based upon the spatial distribution of transcriptional activity in the mature *eve* stripe 2 pattern. (A) Example of spatial distribution of observed fluorescence for an experimental data set. Each circle corresponds to the fluorescence from a single locus at a single point in time. Only observations after 30 minutes into nc14 were used. Circle size indicates fluorescent intensity. Color indicates temporal ordering: 30 min (blue) to 47 min (red). (B) A Gaussian filter was convolved with the raw data points shown in (A). This filtering was found to ameliorate stripe fitting artifacts that arose due to the relative sparsity of the raw data. Shaded gray region indicates potential stripe orientations that were tested during the stripe fitting procedure. The green line indicates the optimal stripe orientation returned by the algorithm. (C) For each proposed orientation, a 1D stripe profile was generated by calculating the average fluorescence per pixel for each position along the projection axis—defined as the direction perpendicular to the proposed stripe orientation. The integral of this projected profile was used as a baseline for the comparison of potential stripe center positions. For each proposed orientation, the position along the projection axis that maximized the fraction of the integrated profile captured within a 4% AP window was taken as the optimal center. The orientation with the highest such fraction metric across all those tested was taken as the stripe axis (green profile). Together, the optimal stripe center and orientation constitute a new, empirically determined, stripe position. (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.

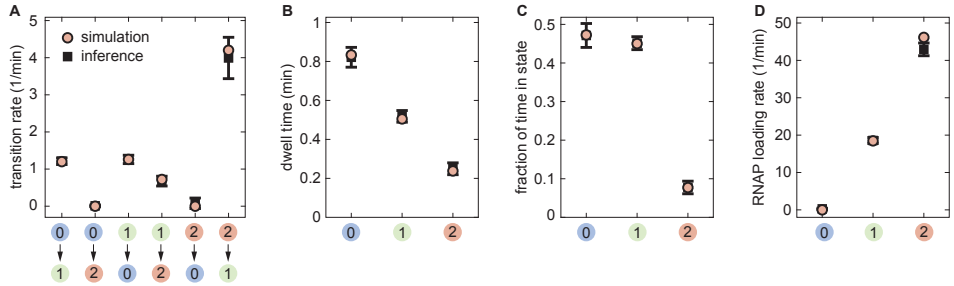


Figure 4–Figure supplement 1. Inference statistics for the mHMM validation. The true and inferred values of **(A)** transition rates, **(B)** dwell times in states, **(C)** state occupancies, and **(D)** RNAP loading rates are compared. Statistics on the inferred values are obtained from 20 independently generated datasets. (Error bars indicate one standard deviation from the mean inference values).

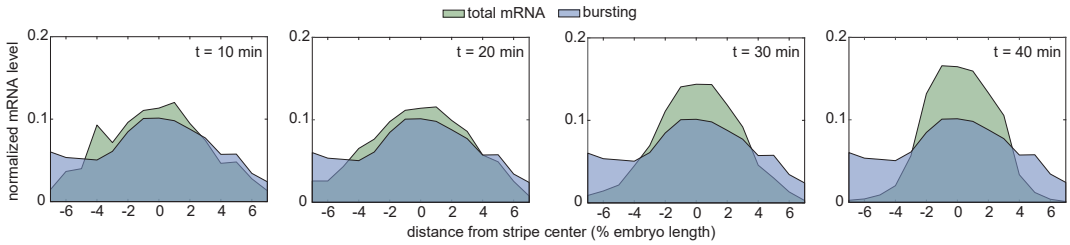


Figure 6–Figure supplement 1. Contributions of transcriptional bursting to *eve* stripe 2 formation over time. Bursting alone is largely sufficient to recapitulate the observed mRNA profile through 20 minutes; however from 30 minutes into nc14 onward the observed stripe becomes markedly sharper than what can be explained by the spatial modulation in cytoplasmic mRNA levels due to bursting alone.

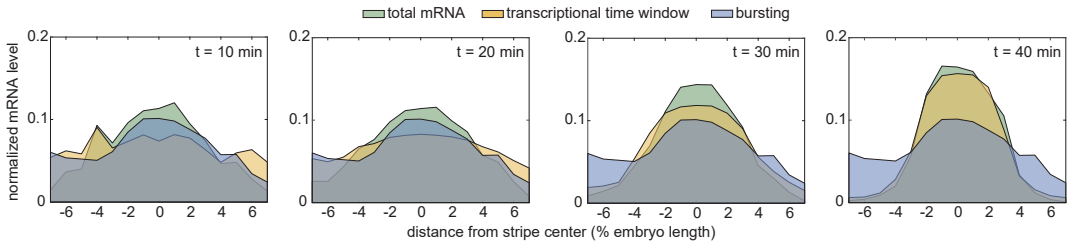


Figure 7–Figure supplement 1. Relative contributions to stripe formation as a function of time. For the first 30 minutes of nuclear cycle 14, the transcriptional time window (yellow) plays a minimal role in driving cytoplasmic mRNA levels (green). Instead, the difference in the mean rate of mRNA production along the AP axis (blue) explains the bulk of the (modest) differential mRNA along the AP axis. By 30 minutes, however, a significant fraction of flank nuclei have transitioned into a transcriptionally quiescent state and the time window strategy begins to play a larger role in dictating cytoplasmic mRNA levels. By 40 minutes into nc14, the time window is the dominant driver of *eve* stripe 2 pattern.

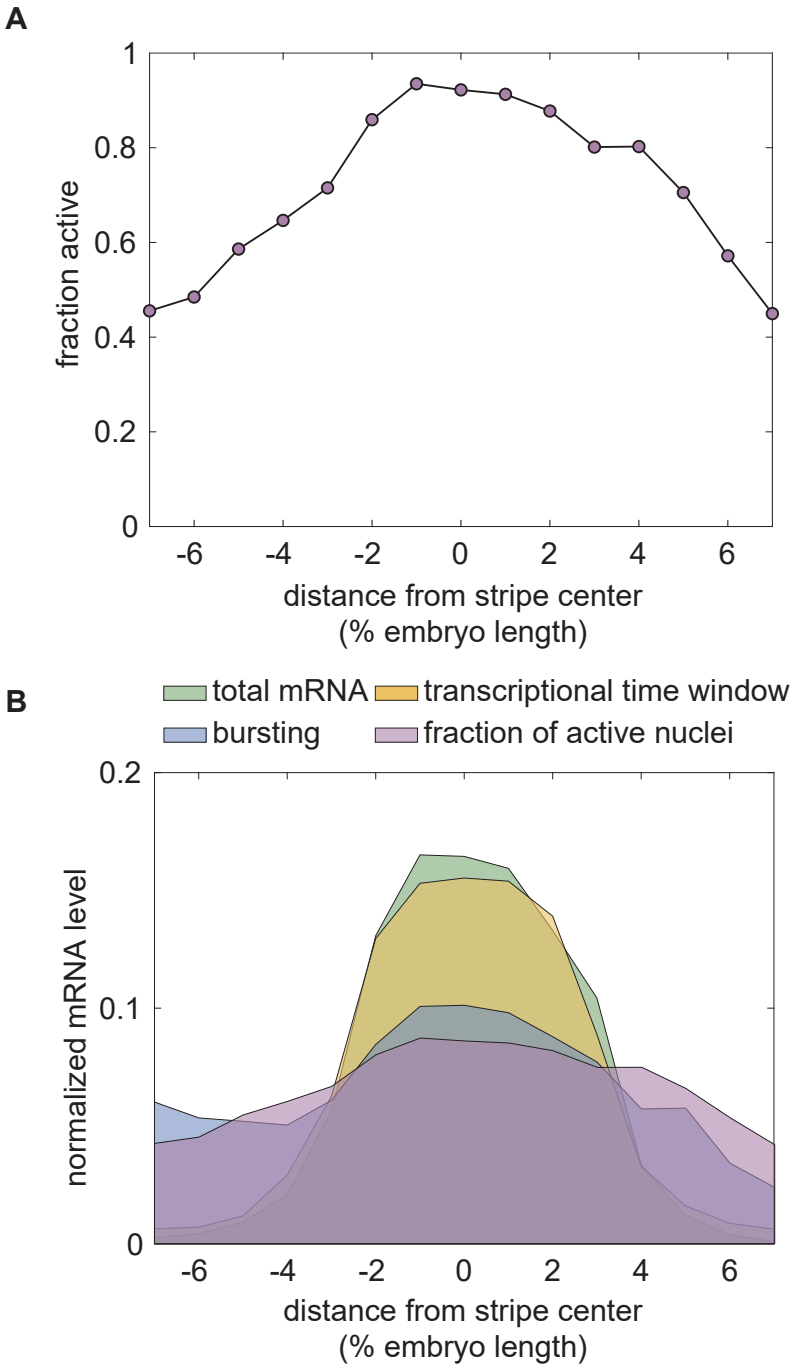


Figure 7–Figure supplement 2. Contributions of inactive nuclei to stripe formation. (A) Our data indicate that a significant fraction of nuclei never turn on in stripe flank regions, as has been the case for other genes (***Garcia et al., 2013***). We detect a difference of roughly a factor of two between the fraction of nuclei ever on in the stripe center and the far stripe flanks. (B) Comparison between the contributions of transcriptional bursting, the transcriptional time window, and the fraction of nuclei that never engage in transcription to stripe formation. (Bootstrap error bars in (A) are smaller than corresponding data markers)

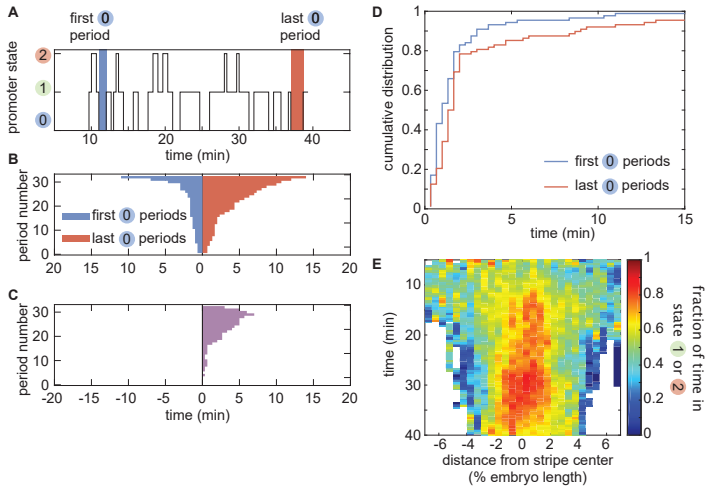


Figure 8–Figure supplement 1. Temporal regulation of bursting dynamics. (A) We examined the promoter trajectories inferred by mHMM for nuclei in the anterior boundary of the stripe (between -7 and -4 % of the embryo length relative to the stripe center) and measured the duration of the first and last periods over which fluorescent puncta were in the (0) state. (B) Rank-ordered plot of durations of the first and last (0) periods. (C) The difference between the duration of the first and last (0) periods shown in (B) and (D) the cumulative distribution of the duration of first and last (0) periods show that (0) periods become longer as development progresses. (E) Heatmap of the fractional occupancy of the ON state (states (1) and (2) in **Figure 3**B) across space and time. We observed a clear rise in the ON state occupancy in the stripe center. We also noted a subtler decrease in the ON state occupancy on the stripe flanks. Regions with fewer than 10 observations were not included (shown as white).

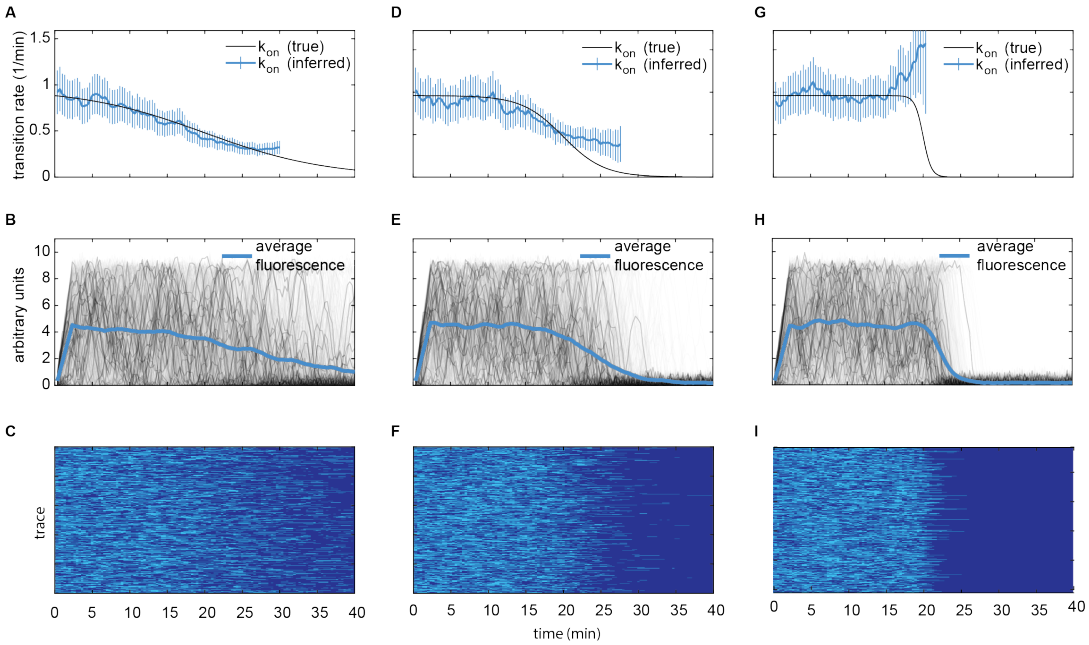


Figure 8–Figure supplement 2. By definition, the onset of transcriptional quiescence coincides with the cessation of observable bursting activity. If this cessation is driven by changes in the bursting parameters as in scenario (ii) in **Figure 8**A, temporal dynamics of bursting parameters that of the same order or faster than the characteristic timescale of bursting cannot be detected. Notably, this is not a limit of the mHMM 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 ineffable resolution limit for any kind of bursting parameter inference. To illustrate this limitation, we simulated 3 scenarios in which k_{on} decreases to 0 over 15, 5, and 1 minute periods. We then sought to recover the trend in k_{on}. To emphasize that the limitations are not specific to mHMM, but, rather, are an inevitable consequence of the structure of the system, we used the *true* promoter trajectories 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. (A-C) 15 minute 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 fact that bursts of activity are interspersed throughout the k_{on} transition makes it possible to recover the temporal trend. (D-F) 5 minute 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 minute 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 minute 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. 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. (A,D,G, error bars indicate 95 % confidence interval of exponential fits used to estimate k_{on}).