bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Multimodal transcriptional control of pattern formation in embryonic 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 Predicting developmental outcomes from regulatory DNA sequence and transcription factor patterns remains an open challenge in physical biology. Using stripe 2 of the *even-skipped* gene in *Drosophila* embryos as a case study, we dissect the regulatory forces underpinning a key step along the developmental decision-making cascade: the generation of cytoplasmic mRNA patterns via the control of transcription in individual cells. Using live imaging and computational approaches, we found that the transcriptional burst frequency is modulated across the stripe to control the mRNA production rate. However, we discovered that bursting alone cannot quantitatively recapitulate the formation of the stripe, and that control of the *window of time* over which each nucleus transcribes *even-skipped* plays a critical role in stripe formation. Theoretical modeling revealed that these two regulatory strategies—bursting and the time window—obey different kinds of regulatory logic, suggesting that the stripe is shaped by the interplay of two distinct underlying molecular processes.

Introduction

During embryonic development, tightly choreographed patterns of gene expression—shallow gradients, sharp steps, narrow stripes—specify cell fates (***Gilbert, 2010***). The correct positioning, sharpness, and amplitude of these patterns of cytoplasmic mRNA and protein ensure the reliable determination of animal body plans (***Peter and Davidson, 2015***). Yet, despite decades of work mapping the gene regulatory networks that drive development and extensive efforts to dissect the regulatory logic of the enhancer elements that dictate the behavior of these networks, the precise prediction of how gene expression patterns and developmental outcomes are driven by transcription factor concentrations remains a central challenge in the field (***Vincent et al., 2016***).

Predicting developmental outcomes demands a quantitative understanding of the flow of information along the central dogma: how input transcription factors dictate the output rate of

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

mRNA production, how this rate of mRNA production dictates cytoplasmic patterns of mRNA, and how these mRNA patterns lead to protein patterns that feed back into the gene regulatory network. While the connection between transcription factor concentration and output mRNA production rate has been the subject of active research over the last three decades (***Lawrence et al., 1987***; ***Driever and Nusslein-Volhard, 1988***; ***Small et al., 1991***; ***Struhl et al., 1992***; ***Jiang and Levine, 1993***; ***Gray et al., 1994***; ***Jaeger et al., 2004***; ***Segal et al., 2008***; ***Levine et al., 2014***; ***Garcia et al., 2016***; ***Vincent et al., 2016***; ***Sayal et al., 2016***), the connection between this output rate and the resulting cytoplasmic patterns of mRNA has remained largely unexplored. For example, a stripe of cytoplasmic mRNA within an embryo could arise as a result of radically different transcriptional dynamics at the single-nucleus level (***Figure 1***A). Specifically, if individual nuclei along this stripe modulate their average RNA polymerase loading rate, then graded control of the mean rate of transcription results: nuclei in the middle of the stripe transcribe at a higher average rate than nuclei on the stripe boundaries (***Figure 1***B). We identify this graded transcriptional control strategy with the analog control of gene expression. Alternatively, transcription factors could exert control over the length of time a nucleus is transcriptionally active (***Figure 1***C). In this binary control scheme—akin to an on/off switch that dictates whether a nucleus is transcriptionally active or quiescent—individual nuclei transcribe at the same average rate regardless of their position along the stripe, but for different lengths of time. Finally, some nuclei might not engage in transcription at all during the formation of the pattern (***Figure 1***D). Here, a larger fraction of nuclei engage in mRNA production in the stripe center than in the boundaries. Any of these scenarios, or some combination thereof, can explain the formation of a cytoplasmic mRNA pattern.

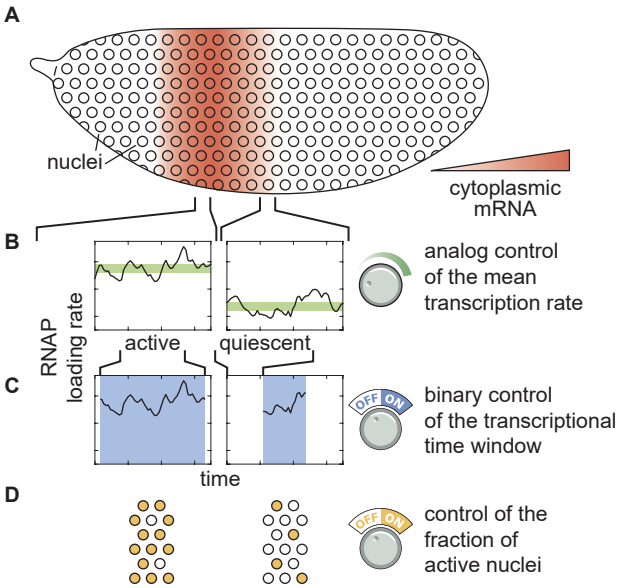


Figure 1. Models of pattern formation by single-cell transcriptional activity. (A) Cytoplasmic mRNA patterns could arise from transcription factors exerting control over **(B)** the mean transcription rate, **(C)** the transcriptional time window dictating when a nucleus is transcriptionally active or quiescent, **(D)** the fraction of active nuclei, or some combination thereof.

In order to uncover the quantitative contribution of these three regulatory strategies to pattern formation, and to determine whether other regulatory strategies are at play, it is necessary to measure the rate of RNA polymerase loading in individual nuclei, in real time, in a living embryo. However, to date, most studies have relied on fixed-tissue techniques such as mRNA FISH and immunofluorescence in order to obtain snapshots of the cytoplasmic distributions of mRNA and protein as development progresses (***Jaeger et al., 2004***; ***Fakhouri et al., 2010***; ***Parker et al., 2011***; ***Estrada et al., 2016***; ***Crocker et al., 2016***; ***Verd et al., 2017***; ***Park et al., 2018***). Such techniques

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

are virtually silent regarding the regulation of single-cell gene expression over time, and are thus ill-suited to the study of how spatiotemporal variations in transcriptional dynamics give rise to patterns of cytoplasmic mRNA.

In this work, we investigated how single-cell transcriptional activity leads to the formation of stripe 2 of the widely studied *even-skipped* (*eve*) gene in the developing fruit fly embryo (***Small et al., 1992***; ***Arnosti et al., 1996***). We combined single-cell live imaging with theoretical modeling in order to study transcriptional activity at the single-cell level in real time, seeking a quantitative connection between the spatiotemporal regulation of transcription and the formation of cytoplasmic patterns of mRNA. Consistent with previous studies, we found that the rate of mRNA production is elevated in the center of the stripe (***Bothma et al., 2014***). Strikingly, however, we discovered that this analog control is alone insufficient to quantitatively recapitulate the formation of the stripe; binary control of the transcriptional time window (***Figure 1***C) is also necessary. Furthermore, we developed novel computational approaches to uncover the molecular underpinning of each regulatory strategy. We employed a memory-adjusted hidden Markov model (mHMM) to uncover variations in transcriptional dynamics in individual nuclei across space and time (***Suter et al., 2011***; ***Molina et al., 2013***; ***Corrigan et al., 2016***). We showed that, consistent with previous results, transcription factors control the rate of transcription by altering the frequency of transcriptional bursts (***Fukaya et al., 2016***; ***Zoller et al., 2018***). Finally, we utilized logistic regressions to correlate *eve* stripe 2 transcriptional dynamics with changes in input transcription factor concentrations. This analysis revealed that the transcriptional time window adheres to different regulatory logic than transcriptional bursting: while repressor levels alone were sufficient to explain the early silencing of nucei in the anterior and posterior stripe flanks, the control of bursting among transcriptionally engaged nuclei depends upon the input concentrations of both activators and repressors. Thus, our findings point to the presence of two distinct regulatory mechanisms that control transcription and gene expression patterns in early development, showcasing the potential for theoretical modelling and biological numeracy to yield novel biological insights when coupled with precise and quantitative experimental observation.

Results

Predicting cytoplasmic mRNA distributions from transcriptional activity

To predict how the transcriptional activity of individual nuclei dictates the formation of cytoplasmic patterns of mRNA, we began with a simple model that considers the balance between the rate of mRNA synthesis and degradation

$$\frac{\mathrm{dmRNA}}{\mathrm{d}t}(x,t) = \underbrace{p_{\mathrm{active}}(x)}_{\substack{\text{fraction of}\\ \text{active nuclei}}} \underbrace{R(x,t)}_{\text{synthesis}} - \underbrace{\gamma\,\mathrm{mRNA}(x,t)}_{\text{degradation}}, \tag{1}$$

where $\mathrm{mRNA}(x,t)$ indicates the mRNA concentration at position x along the embryo at time t, $R(x,t)$ corresponds to the mRNA synthesis rate averaged over multiple nuclei within the same position x, $p_{\mathrm{active}}(x)$ is the fraction of active nuclei (corresponding to the regulatory strategy shown in ***Figure 1***D) and γ is the degradation rate (see ***Appendix 1*** for details of this derivation).

In order to examine the quantitative consequences of the three potential regulatory strategies (***Figure 1***B-D), we adopted widespread assumptions in the modeling of transcriptional regulation (***Phillips et al., 2013***). First, we assumed that the degradation rate γ is a constant and not under any kind of spatiotemporal control. Comparisons between model predictions and empirically measured levels of cytoplasmic mRNA suggest that this assumption is reasonable (see ***Appendix 2***). Second, we posited that at each position throughout the embryo the synthesis rate $R(x,t)$ does not vary significantly in time such that it can be approximated by its time average $R(x) = \langle R(x,t)\rangle$. This assumption will be revised later in the text in order to account for the time-dependent regulation of the mean rate of transcription. Finally, we assumed that nuclei along the axis of the embryo start

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

transcribing at time $t_{\text{on}}(x)$, and stop transcribing and enter a state of transcriptional quiescence at time $t_{\text{off}}(x)$. Under these assumptions, ***Equation 1*** can be solved analytically, resulting in

$$\text{mRNA}(x,t) = \underbrace{\frac{R(x)}{\gamma}}_{\substack{\text{analog control of}\\ \text{mean transcription rate}}} \times \underbrace{\left(e^{-\gamma(t-\min\{t_{\text{off}}(x),t\})} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\substack{\text{binary control of}\\ \text{transcriptional time window}}} \times \underbrace{p_{\text{active}}(x)}_{\substack{\text{control of the fraction}\\ \text{of active nuclei}}} . \quad (2)$$

This equation makes precise predictions about how each regulatory strategy contributes to the formation of the cytoplasmic mRNA pattern. Thus, measuring how each quantity is regulated across the stripe allows us to predict their relative contributions to pattern formation.

Binary control of the transcriptional time window is the primary driver of stripe formation

In order to test the simple model of pattern formation put forward in ***Equation 2***, we quantified transcription of stripe 2 of *eve* in the fruit fly. We imaged the transcription of an *eve* stripe 2 reporter using the MS2 system (***Garcia et al., 2013***; ***Lucas et al., 2013***; ***Bothma et al., 2014***). 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 maternally provided MS2 coat protein fused to 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 RNA polymerase molecules actively transcribing the gene (see Materials and Methods and ***Garcia et al.*** (***2013***)). By aligning multiple embryos (***Figure 2–Figure Supplement 1***), we obtained the average number of actively transcribing RNA polymerase molecules as a function of time and position throughout the embryo (***Figure 2***C).

Using the MS2 system, we quantified each potential regulatory strategy and determined its predicted contribution to pattern formation according to our model in ***Equation 2***. We first used our data to estimate the time-averaged rate of RNA polymerase loading, $R(x)$ (see ***Appendix 2*** for details). We found that this rate is modulated along the axis of the embryo (***Figure 3***A and B; see also Video 2, ***Figure 3–Figure Supplement 1*** and Materials and Methods): whereas in the center of the stripe RNA polymerase molecules are loaded at a rate of approximately 16 molecules/min, this loading rate decreases to about 8 molecules/min at the boundaries.

Our data also revealed that the transcriptional time window is modulated along the stripe (***Figure 3–Figure Supplement 2***A). Whereas the time at which each nucleus becomes transcriptionally active, $t_{\text{on}}(x)$, was constant across the stripe, with all nuclei becoming active 9 ± 4 min after the previous anaphase (***Figure 3–Figure Supplement 2***B), the time at which nuclei stop transcribing and become quiescent, $t_{\text{off}}(x)$, showed a strong modulation along the embryo's axis (***Figure 3–Figure Supplement 2***C). As a result, the time window over which each punctum is engaged in transcription, $\Delta t = t_{\text{off}} - t_{\text{on}}$, is sharply modulated along the stripe (***Figure 3***C and D and Video 3), with nuclei in the stripe center transcribing for >30 min and nuclei on the boundaries only transcribing for approximately 10 min.

Finally, our analysis also revealed the magnitude of the modulation of the fraction of active nuclei along the stripe. Most nuclei along the stripe were engaged in transcription. In the stripe center, around 90% of nuclei transcribed at some point during the nuclei cycle. This number reduced to about 70% at the boundaries (***Figure 3***E and F and Video 4).

The analysis in ***Figure 3***A-F reveals that each of the three regulatory strategies identified in ***Figure 1*** is at play in the embryo, and that they all have the potential to contribute to pattern formation. However, these measurements alone cannot inform us on *how much* each of these strategies contributes to the cytoplasmic mRNA pattern. To quantify the degree to which each regulatory strategy contributes to the formation of *eve* stripe 2, we employed the model described in ***Equation 2***.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

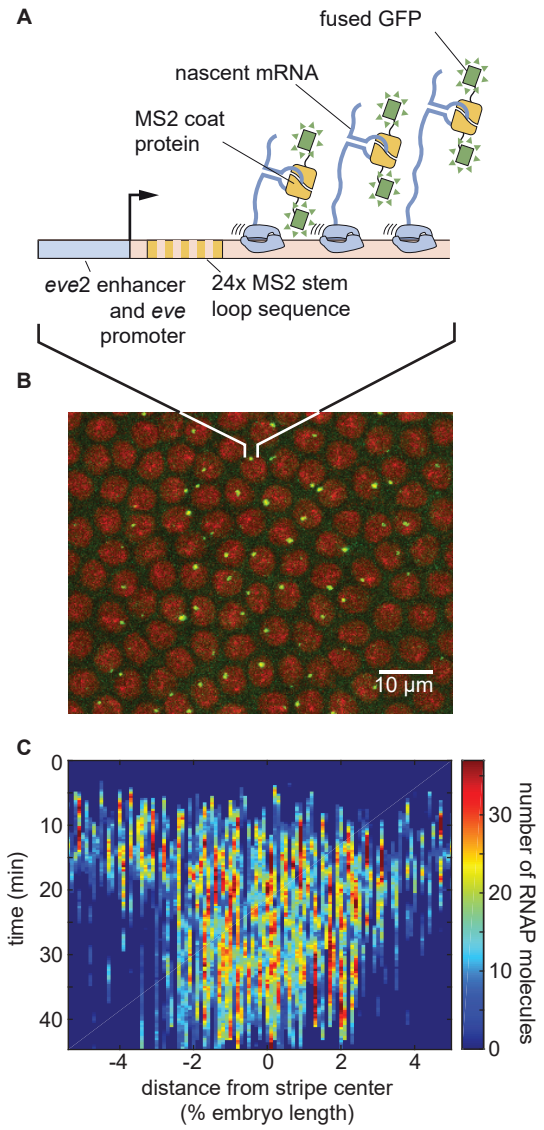


Figure 2. Measuring transcriptional dynamics of *eve* stripe 2 formation using the MS2 system. (A) MS2 stem loops introduced in an *eve* stripe 2 reporter gene are bound by MS2 coat protein fused to GFP. **(B)** Sites of nascent transcript formation appear as green fluorescent puncta whose intensity reports on the number of actively transcribing RNA polymerase molecules. Nuclei are visualized through a fusion of RFP to Histone. **(C)** Mean number of RNA polymerase molecules actively transcribing the gene as a function of space and time. (C, data averaged over 11 embryos).

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

Figure 2–Figure supplement 2. Integrating MS2 Spots.

Figure 3G indicates the quantitative contribution of each regulatory strategy (each term on the right-hand side of ***Equation 2***) to the formation of this cytoplasmic pattern. The cytoplasmic pattern of mRNA, corresponding to left-hand side of ***Equation 2***, was obtained from our live-imaging data (see ***Appendix 2*** for details). Regulation of the fraction of active nuclei along the embryo (***Figure 3***G, yellow) contributes negligibly to this mRNA pattern. In contrast, both the analog regulation of the mean rate (***Figure 3***G, green) and the binary control of the transcriptional time window (***Figure 3***G, blue) make significant contributions to the overall pattern, with binary control playing the dominant role. We thus concluded that the joint effect of these two strategies (***Figure 3***G, brown) is sufficient

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

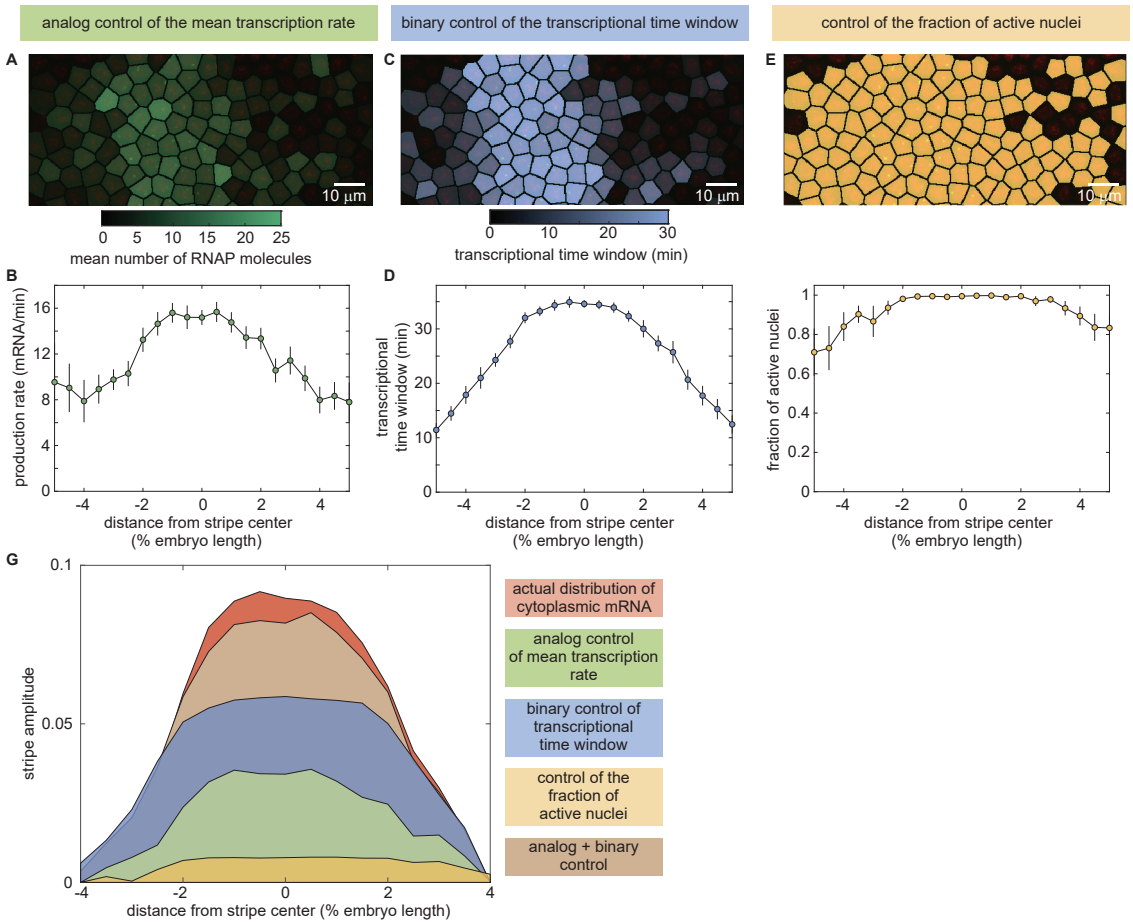


Figure 3. Regulatory strategies for pattern formation in *eve* stripe 2. (A,B) Time-averaged rate of mRNA production, **(C,D)** transcriptional time window, and **(E,F)** fraction of active nuclei as a function of position along the embryo. **(G)** Amplitude of the cytoplasmic mRNA distribution compared to the contributions to stripe formation of the analog control of the mean transcription rate, the binary control of the transcriptional time window, and the control of the fraction of active nuclei. The combined contribution from the analog and binary strategies is also shown. See ***Figure 3–Figure Supplement 3*** for details of how depicted profiles were derived from raw data. (A,C,E, representative snapshots of an individual embryo 40 min into nuclear cycle 14; B,D,F, average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean).

Figure 3–Figure supplement 1. Mean transcriptional activity over time.

Figure 3–Figure supplement 2. Regulation of the transcriptional time window.

Figure 3–Figure supplement 3. Definition of stripe amplitude.

Figure 3–Figure supplement 4. Joint effect of mean rate, binary control, and fraction of active nuclei.

to quantitatively recapitulate the stripe of cytoplasmic mRNA from single-cell transcriptional activity.

Mean transcription rate is dictated by bursting through modulation of the rate of promoter turn on

Are the binary and analog control strategies driven by distinct molecular mechanisms, or are they different manifestations of the same underlying process? To uncover the molecular mechanism behind the analog control of the mean rate of transcription, we analyzed the transcriptional activity of individual nuclei. Previous work demonstrated that the rate of gene expression at individual loci within the *eve* stripe 2 pattern is highly stochastic (***Bothma et al., 2014***). Indeed, as shown in ***Figure 4***A, our data revealed punctuated peaks and troughs in the number of active RNA polymerase molecules. These features have been related to the rate of RNA polymerase loading

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

at the *eve* promoter by assuming that promoter loading is "burst-like", with the promoter loading RNA polymerase molecules onto the gene at a constant rate over discrete periods of time (***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***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Zoller et al., 2018***), support a minimal two-state model of promoter switching (***Figure 4***B): promoters switch stochastically between ON and OFF states with rates k_{on} and k_{off}. In this model, promoters in the ON state engage in mRNA production at rate r.

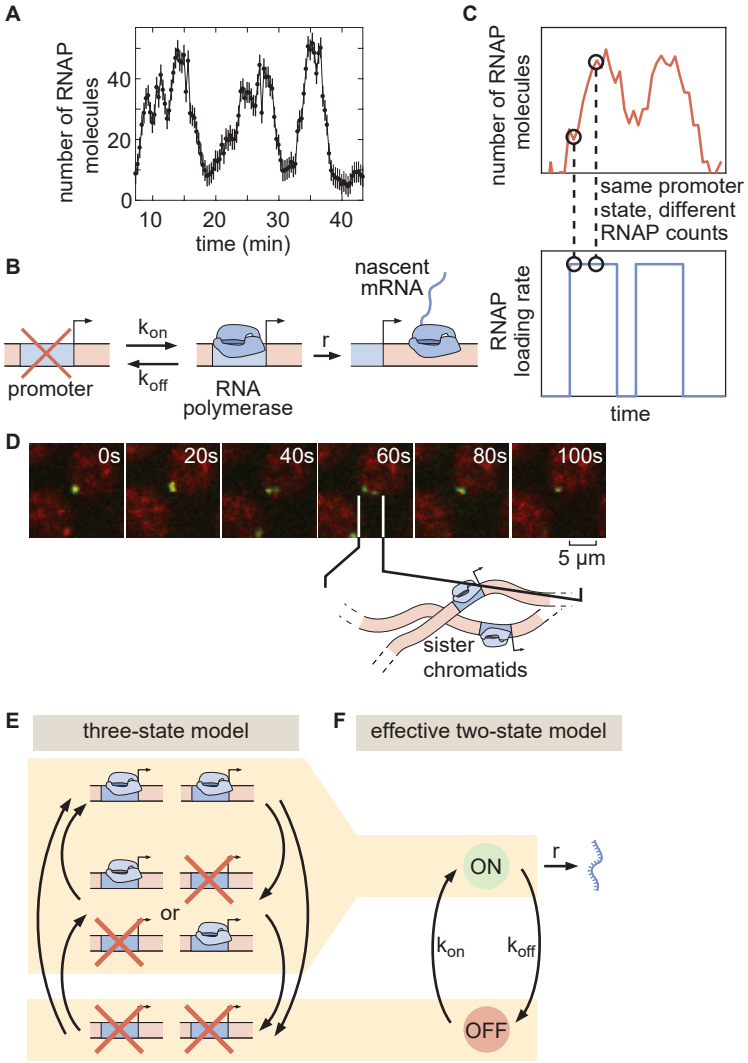


Figure 4. Transcriptional bursting in *eve* stripe 2. (A) Single-nucleus measurements reveal that nuclei transcribe in bursts. **(B)** Two-state model of bursting of a single promoter. **(C)** The same hidden rate of RNA polymerase loading (bottom) can correspond to different observable numbers of RNA polymerase molecules on the gene (top), such that standard Hidden Markov model approaches cannot be used to infer the hidden promoter state. **(D)** Fluorescent puncta are composed of two distinct transcriptional loci within a diffraction-limited spot, each corresponding to a sister chromatid. **(E)** Three-state model of promoter switching within a fluorescent punctum that accounts for the combined action of both sister chromatids. **(F)** Effective two-state model of transcriptional bursting. (A, error bars obtained from estimation background fluorescent fluctuations; Materials and Methods and ***Garcia et al.*** (***2013***).)

In the bursting model, the mean rate of transcription is given by the product of the fraction of time spent in the ON state with the transcription rate in this active state (***Peccoud and Ycart, 1995***; ***Kepler and Elston, 2001***; ***Sasai and Wolynes, 2003***; ***Sanchez and Kondev, 2008***; ***Sanchez et al.,***

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

2011; Xu et al., 2016)

$$\underbrace{R(x)}_{\substack{\text{mean}\\ \text{transcription rate}}} = \underbrace{r(x)}_{\substack{\text{RNAP loading}\\ \text{rate}}} \times \underbrace{\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\substack{\text{fraction of time}\\ \text{in ON state}}}, \tag{3}$$

where all parameters are allowed to vary as a function of position along the embryo, x (see Appendix 1 for details of this derivation). Thus, within this framework, the observed modulation of the mean rate of transcription across the stripe (Figure 3G, green) implies that one or more of these bursting parameters is subject to spatially controlled regulation. However, the mean rate trend alone is not sufficient to identify *which* of the three bursting parameters (k_{on}, k_{off}, and r) is being regulated by the input transcription factors in order to control the average transcription rate.

Typically, the *in vivo* molecular mechanism of transcription factor action 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; Padovan-Merhar et al., 2015; Skinner et al., 2016; Bartman et al., 2016; Zoller et al., 2018; Hendy et al., 2017). In contrast, MS2-based live imaging can directly inform on the dynamics of transcriptional bursting in real time. The MS2 approach, however, reports on the total number of actively transcribing RNA polymerase molecules and not on the instantaneous rate of RNA polymerase loading at the promoter, which is the relevant quantity for estimating k_{on}, k_{off}, and r. To date, approaches for extracting bursting parameters from such data in multicellular organisms have mainly relied on the manual analysis of single-nucleus transcriptional dynamics (Bothma et al., 2014; Fukaya et al., 2016) or autocorrelation-based methods that infer mean bursting parameters across ensembles of traces (Larson et al., 2011; Coulon et al., 2014; Desponds et al., 2016). A computational method for inferring the rates of RNA polymerase loading (Figure 4C, bottom) from the total number of actively transcribing RNA polymerase molecules in single cells (Figure 4C, top) is thus needed to obtain the bursting parameters.

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). However, our observable (the MS2 signal) does not correspond to the hidden variable of interest (the promoter state) in a one-to-one fashion (compare Figure 4C top and bottom). Instead, the observable MS2 signal reflects the net effect of promoter switching over a period equal to the time that an RNA polymerase molecule takes to transcribe the whole gene. Thus, instantaneous fluorescence 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 (Choubey et al., 2015; Xu et al., 2016; Corrigan et al., 2016; Choubey, 2018; Choubey et al., 2018). 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 (details about implementation of the method are given in Appendix 3). Similar approaches were recently introduced to study transcriptional dynamics in cell culture and tissue samples (Suter et al., 2011; Molina et al., 2013; Zechner et al., 2014; Zoller et al., 2015; Hey et al., 2015; Bronstein et al., 2015; Corrigan et al., 2016; Featherstone et al., 2016). We used simulated data to establish that mHMM reliably extracts the kinetic parameters of transcriptional bursting from live-imaging data (Appendix 4), providing an ideal tool for dissecting the contributions from individual bursting parameters to observed patterns of transcriptional activity across space and time.

Before applying our model to real-time transcriptional data, we had to account for the rapid replication of the *D. melanogaster* genome at the beginning of each nuclear cycle (Rabinowitz, 1941; Shermoen et al., 2010), which leads to the presence of two distinct *eve* loci within each fluorescent spot (Figure 4D and Video 5). The first evidence of resolved chromatids appears as early as 8

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

minutes into nuclear cycle 14 (***Appendix 5–Figure 1***)—coincident with the average onset time of transcription (***Figure 3–Figure Supplement 2***B). Moreover, our analysis indicates that replication of the relevant portion of the genome likely occurs in all *eve*-expressing nuclei by no later than 10 minutes following mitosis (***Appendix 5–Figure 2***). Thus, we conclude that the vast majority of our data feature two distinct *eve* loci within each diffraction-limited transcription spot. Moreover, while the distance between sister loci varies over time (see, *e.g.* ***Figure 4***D), they nonetheless stay in relatively close proximity to ensure their proper segregation from each other at the next mitosis (***Senaratne et al., 2016***) such that the fluorescent intensity signals extracted from our data reflect the integral over both loci (***Figure 2–Figure Supplement 2***). As a result, if we assume that each locus can be well-represented by a two-state model of transcriptional bursting, then an effective three-state model is needed to capture *eve* dynamics (***Figure 4***E). For ease of exposition, we present our main results in the context of an effective two-state model, in which, as detailed in ***Appendix 1***, the system is considered to be in the ON state so long as *either* chromatid is bursting (***Figure 4***F). Note that none of our conclusions below are affected by this choice of an effective model as shown in ***Appendix 6*** where we present full results for the three-state model.

A typical experimental trace for a nucleus in the core of the stripe is shown in ***Figure 5***A, along with its best fit, which corresponds to the mHMM-inferred promoter trajectory in ***Figure 5***B. Our ability to infer the instantaneous promoter state in individual nuclei throughout development is further illustrated in ***Figure 5***C and Video 6. These data revealed that, as development progresses and the stripe sharpens, the *eve* promoter continuously fluctuates between the ON and OFF states on a time scale of approximately 1-2 minutes.

In order to infer time-averaged bursting parameter values, we grouped traces by position along the anterior-posterior axis. The rate of RNA polymerase loading, r, remained constant throughout the stripe (***Figure 5***D), suggesting that none of the transcription factors regulating *eve* stripe 2 act on this kinetic parameter. Similarly, we noted no significant spatial modulation of the rate of switching out of the ON state, k_{off} (***Figure 5***E). In contrast, the rate of switching into the ON state (also known as burst frequency), k_{on}, was strongly up-regulated in the stripe center (***Figure 5***E). These observations suggested that, in order to enact analog control of the mean transcription rate, transcription factors act primarily on the rate of promoter turning on, consistent with previous results both in embryos (***Xu et al., 2015***; ***Desponds et al., 2016***; ***Fukaya et al., 2016***) and in single cells (***So et al., 2011***; ***Senecal et al., 2014***; ***Padovan-Merhar et al., 2015***; ***Bartman et al., 2016***). This regulatory modality increases the fraction of time that loci near the stripe center spend in the ON state (***Figure 5–Figure Supplement 1*** and ***Zoller et al.*** (***2018***)).

Binary control of the transcriptional time window is independent of transcriptional bursting

Having determined that analog control of the mean transcriptional rate is realized by the modulation of the burst frequency, k_{on}, we next sought to uncover the molecular mechanism by which the binary regulation of the transcriptional time window is implemented. In one possible scenario, the onset of transcriptional quiescence at the end of the transcriptional time window would reflect a fundamental change to the molecular character of the transcriptional locus such that the bursting framework no longer applies. For instance, repressing transcription factors could induce an irreversible change in the local chromatin landscape that precludes further activator-mediated bursting, effectively silencing transcription (***Figure 6***A, top; ***Allis et al.*** (***2015***)). Alternatively, if the rates of promoter switching vary in time, then the time window could be explained without invoking an extra silenced state that is mechanistically distinct from the processes driving transcriptional bursting. 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, abolishing all activity at the locus and leading to the observed quiescence. Such modulation could be achieved by downregulating k_{on}, downregulating r, and/or upregulating k_{off} (***Figure 6***A, bottom).

In order to discriminate between these two possible scenarios, we split the stripe into the

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

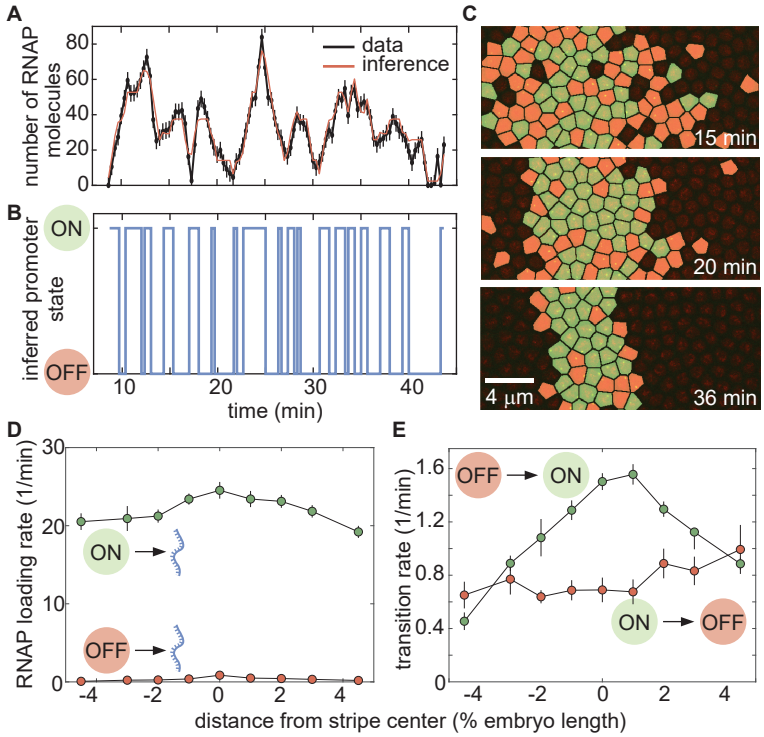


Figure 5. Inferring bursting dynamics using a memory-adjusted Hidden Markov model. (A) Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (colors as in B). **(D)** The rate of initiation for each transcriptional state is not significantly modulated along the embryo. **(E)** Our mHMM reveals that the transition rate between the OFF and ON state (equivalent to the burst frequency) is up-regulated in the stripe center. (A, error bars obtained from estimation of background fluorescent fluctuations, as described in Materials and Methods and ***Garcia et al.*** (***2013***); D,E, error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

Figure 5–Figure supplement 1. Fraction of time spent in each transcriptional state.

five regions shown in ***Figure 6***B. For each region, we sought to determine whether the bursting dynamics varied over time in a manner that could explain the dynamics of entry into quiescence of individual nuclei (***Figure 6***C). To probe for this time-dependence in transcriptional bursting, we extended our mHMM method to obtain promoter-bursting parameters over discrete periods of time by performing inference on our live-imaging data using a sliding window (see ***Appendix 3*** for details). Our inference revealed that the rate of promoter turn on, k_{on}, varied significantly in time (***Figure 6***D). Specifically, k_{on} decreased in both the anterior and posterior stripe boundaries (black and red curves) as development progressed and the fraction of active nuclei decreased (grey shaded region), while loci in the stripe center (green and yellow curves) exhibited a significant *increase* in k_{on}. Further, while relatively constant at most positions along the stripe, both the rate of RNA polymerase loading when in the ON state, r, and the rate of promoter turn off, k_{off}, decreased slightly in (***Figure 6***E and F).

These findings confirmed our time-averaged inference results (***Figure 5***D and E) indicating that k_{on} was the primary kinetic pathway through which transcription factors influence *eve* stripe 2 transcription dynamics. Moreover, the coincidence of the decrease in k_{on} in flank nuclei with the onset of transcriptional quiescence (grey shaded region in ***Figure 6***D) seemed to suggest that, at least in part, quiescence in the stripe flanks could be driven by the temporal modulation of bursting parameters (***Figure 6***A, bottom). However, other trends in our data results were not consistent with the view that a decrease in k_{on} drives transcriptional quiescence. Although 70% and 50% of nuclei in the regions directly anterior and posterior of the stripe center were quiescent by 40 min into

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-4.0 International license.

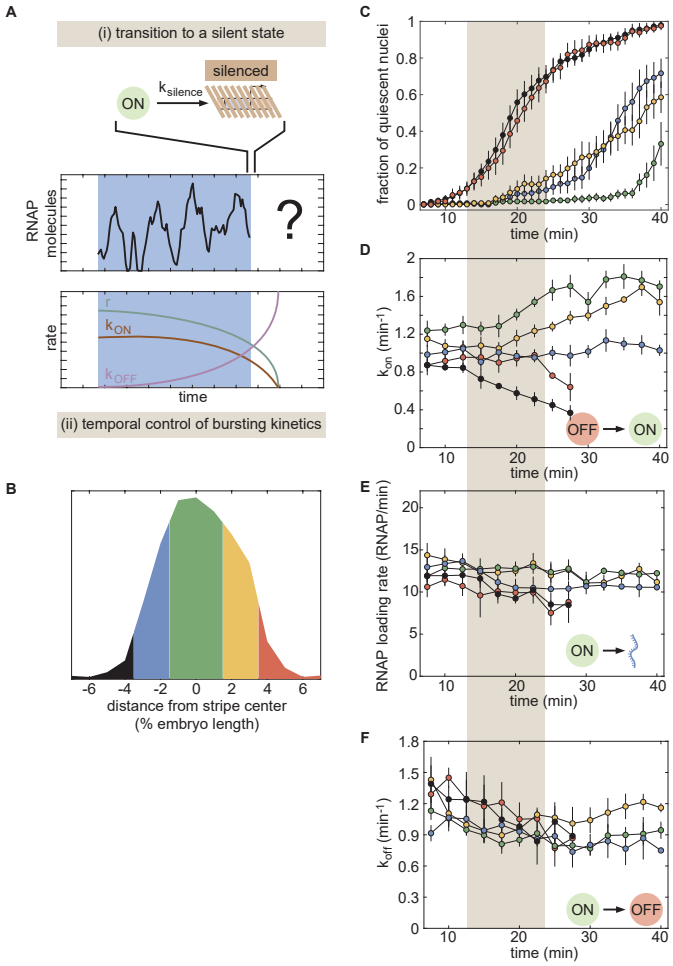


Figure 6. Investigating the molecular character of transcriptional quiescence. (A) Two hypotheses explaining promoter quiescence onset: (i) an irreversible transition into an alternative, transcriptionally silent state and (ii) the modulation of one or more bursting parameters over time. **(B)** Division of the stripe into five regions for our analysis of: **(C)** the fraction of quiescent nuclei, **(D)** the transition rate from OFF to ON, **(E)** the rate of RNA polymerase loading when the promoter is in the ON state, and (**F**) the transition rate from ON to OFF as a function of time and position along the stripe. Grey shaded region indicates the onset of transcriptional quiescence. (C, error bars indicate bootstrap estimate of the standard error of the mean. D-F, error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrapped samples of experimental data. See Materials and Methods for details)

the nuclear cycle (blue and yellow curves in ***Figure 6***C), we detected no corresponding decrease in k_{on}. In fact, k_{on} actually *increased* in some inner regions of the stripe (***Figure 6***D) —a trend that would increase overall transcriptional activity and would therefore go against the establishment of transcriptional quiescence.

The divergent outcomes observed in the central stripe regions, with the rate of transcriptional bursting remaining constant or increasing at *eve* loci within the engaged population of nuclei even as loci in neighboring nuclei turn off for good, runs counter to the hypothesis that quiescence is driven by the temporal modulation of the promoter switching parameters. It is conceivable that temporal changes in bursting parameters associated with the onset of quiescence occur too rapidly to be captured by our model. However, as discussed in ***Appendix 7***, these changes would need to occur on the same time scale as bursting itself (1 to 3 min). Given that both the other temporal trends detected by our inference (***Figure 6***) and the shifts in the input transcription factors themselves (***Appendix 8***) unfold on significantly slower timescales (5-15 min), we concluded that while possible, a scenario were bursting dynamics are changing too quickly to detect is unlikely.

The contradictory trends observed in the stripe center suggested that entry into transcriptional

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

quiescence might be akin to an irreversible transition into a silent state (***Figure 6***A, top), thus suggesting that binary control of the transcriptional time window and the transcriptional bursting driving the analog control of the mean transcription rate may arise from distinct molecular processes.

Input-output analysis reveals distinct regulatory logic for bursting and the transcriptional time window

eve stripe 2 is mainly established by the combined action of two activators, Bicoid and Hunchback, and two repressors, Giant and Krüppel (***Frasch and Levine, 1987***; ***Small et al., 1992***; ***Arnosti et al., 1996***). If transcriptional bursting and the transcriptional time window are controlled by distinct molecular mechanisms, then distinct forms of regulatory logic may be at play. For example, the Bicoid and Hunchback activators could control transcriptional bursting, while the Giant and Krüppel repressors could dictate the entry into the quiescent state. In order to reveal the molecular logic controlling each regulatory strategy, we sought to correlate the fraction of nuclei that have entered the quiescent state (***Figure 7***A) and the fraction of nuclei in the bursting ON state (***Figure 7***B) with the corresponding spatiotemporal patterns in the input concentrations of these four transcription factors.

We measured Bicoid concentration profiles using a a well-established Bicoid-GFP fusion (***Gregor et al., 2007***) and obtained spatiotemporal concentration profiles for Krüppel, Giant, and Hunchback from published immunofluorescence data (***Dubuis et al., 2013***). We combined these data with our live-imaging data of *eve* stripe 2 transcriptional activity to generate an "average embryo" in which the concentration of all relevant inputs and the output transcriptional activity at each point in time and space were known (***Figure 7***C and Video 7). Building upon previous work (***Ilsley et al., 2013***), we utilized logistic regressions to probe the regulatory role played by each of these four factors in the spatiotemporal control of transcriptional bursting and the transcriptional time window. Logistic regression is a widely used method of inferring predictive models in processes with binary outcomes. For example, in order to query the regulatory logic behind the control of the transcriptional time window, the model probes the impact of each transcription factor on the relative likelihood of a locus entering the quiescent state versus the likelihood of remaining transcriptionally engaged such that

$$\log\left(\frac{\text{likelihood (transcriptionally quiescent)}}{\text{likelihood (transcriptionally engaged)}}\right) = \beta_0 + \beta_1\,[\text{Bicoid}] + \beta_2\,[\text{Hunchback}] + \beta_3\,[\text{Giant}] + \beta_4\,[\text{Krüppel}]\,, \tag{4}$$

where the coefficients β_n indicate the magnitude and nature (activating or repressing) of the transcription factor's regulatory function. In estimating these coefficients, we used prior knowledge about the function of each transcription factor, requiring Bicoid and Hunchback to play activating roles, and Krüppel and Giant to play repressing roles (***Small et al., 1991***, ***1992***). We used an analogous model to investigate the regulatory logic controlling transcriptional bursting by inferring the factors that determine the relative likelihood that nuclei are in the bursting ON versus the OFF state, likelihood (ON state)/likelihood (OFF state).

Our analysis of the fraction of nuclei in the quiescent state revealed that no single transcription factor can explain quiescence dynamics (***Figure 7***D and E). However, a simple model in which increasing levels of the repressors Giant and Krüppel drive the onset of transcriptional quiescence in the anterior and posterior stripe flanks, respectively, recapitulated experimentally observed trends. The further addition of Hunchback and/or Bicoid had no impact on the model's predictive power, suggesting that activator concentrations have no influence over the molecular processes responsible for silencing. Relaxing constraints on the functional role of each transcription factor–for instance, allowing the presumed activators to function as repressors–also provided no significant improvement over models presented here as shown in ***Appendix 8***.

We next turned our attention to the relationship between transcription factor levels and the fraction of nuclei in the ON state (***Figure 7***B). Unlike the transcriptional time window, repressor levels

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-4.0 International license.

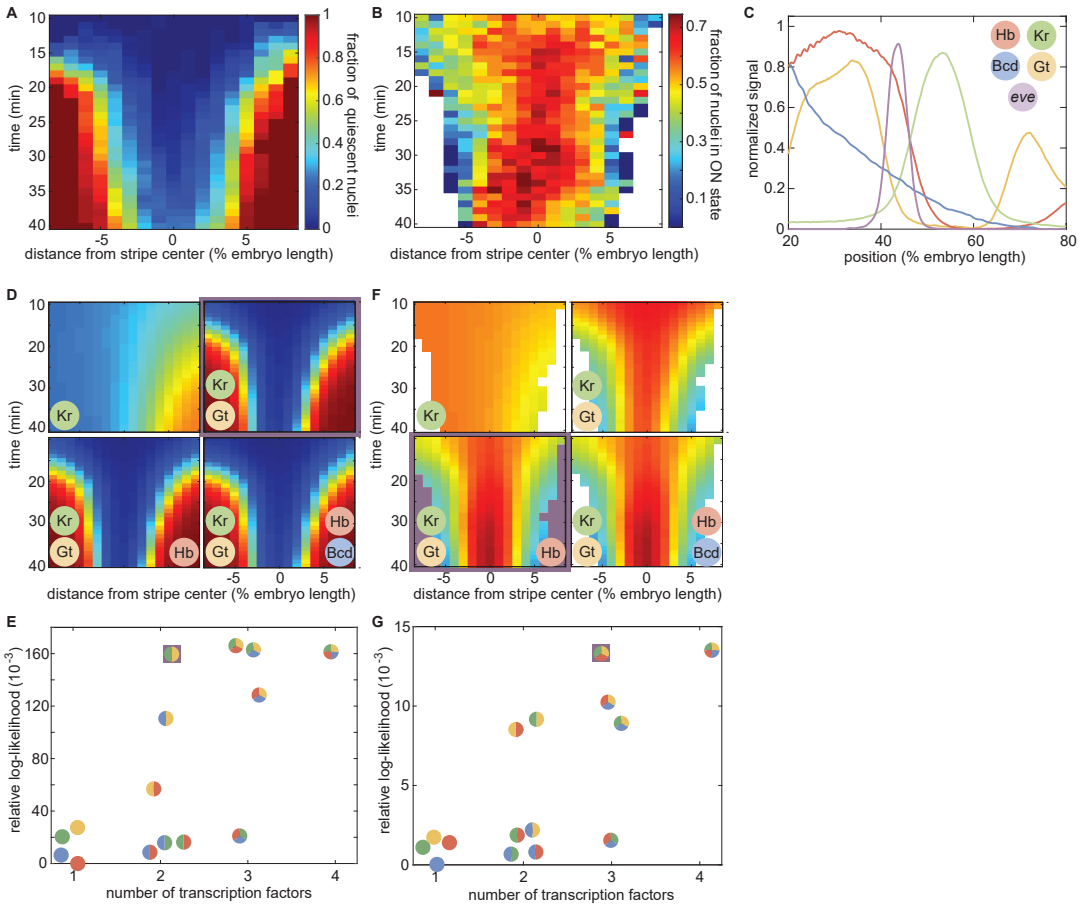


Figure 7. Probing the regulatory logic of bursting and the transcriptional time window. (A) Fraction of nuclei in the transcriptionally quiescent state and **(B)** fraction of nuclei in the bursting ON state as a function of time and position along the embryo. **(C)** Snapshot of input transcription factor levels and predicted *eve* mRNA levels of our "average" embryo. **(D)** Predicted fraction of quiescent nuclei for progressively more complex regression models. The simplest model with the highest likelihood is outlined in purple. **(E)** Model likelihood indicating that Krüppel and Giant levels are sufficient to recapitulate the fraction of quiescent nuclei in (D). **(F)** Predicted fraction of nuclei in the ON state. The simplest and most likely model is highlighted in purple. **(G)** Model scores reveal that Giant, Krüppel, and Hunchback recapitulate the bursting behavior in (F).

alone could not recapitulate the observed bursting profile; Hunchback levels were also necessary in order to fully capture the spatiotemporal bursting dynamics (***Figure 7***E and G). Specifically, we linked a rise in Hunchback concentration to an observed rise in the fraction of nuclei in the ON state in the stripe center between 30 and 35 min into the nuclear cycle (***Figure 7***B and F).

Our input-output analysis thus revealed that bursting and the transcriptional time window exhibit significantly different forms of regulator logic: whereas repressor levels alone are sufficient to explain the transcriptional time window, the joint action of activators and repressors appears necessary to explain the observed patterns of transcriptional bursting. These results are consistent with the hypothesis that regulation of bursting and of the transcriptional time window occur via distinct molecular processes, therefore supporting a model in which the long-lived trancriptionally silent state observed in flank nuclei constitutes a distinct molecular state outside of the bursting model.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Discussion

In *Drosophila* development, information encoded in a handful of maternally deposited protein gradients propagates through increasingly complex layers of interacting genes, culminating in the specification of the adult body plan. The prediction of this cascade of developmental outcomes requires a quantitative understanding of the mechanisms that facilitate the flow of information along the central dogma. Here, we utilized live imaging in conjunction with theoretical modelling to shed light on a critical link in this cascade: how the regulation of transcriptional activity at the single-nucleus level gives rise to a spatiotemporal pattern of cytoplasmic mRNA.

A priori, there are several distinct regulatory strategies at the single-cell level capable of generating spatially differentiated patterns of cytoplasmic mRNA (***Figure 1***), each with distinct implications for the nature of the underlying molecular processes at play. Several recent studies have revealed that the average rate of transcription is mainly modulated across the embryo by tuning the frequency of transcriptional bursting (***Lionnet et al., 2011***; ***Bothma et al., 2014***; ***Xu et al., 2015***; ***Desponds et al., 2016***; ***Fukaya et al., 2016***; ***Zoller et al., 2018***). Yet it has remained unclear whether this modulation of the rate of transcription (and thereby mRNA production) is the dominant modality by which input concentrations of transcription factors drive the formation of patterns of gene expression, or if, instead, it is simply the most readily apparent mechanism among multiple distinct control strategies.

In this work, we derived a simple theoretical model that predicts how the interplay between regulatory strategies at the single-cell level dictates the accumulation of cytoplasmic mRNA and the subsequent formation of a gene expression pattern (***Equation 2***), and tested its predictions experimentally by performing single-cell live-imaging measurements in developing embryos using the MS2 system. Our results revealed that the modulation of the mean rate of transcription alone is insufficient to recapitulate the formation of a sharp gene expression stripe (***Figure 3***F, green). We discovered that the window of time over which promoters engage in transcription is sharply controlled along the axis of the embryo (***Figure 3***C and D) and that the joint action of the analog control of the rate of transcription and the binary control of the duration of transcription is necessary and sufficient to quantitatively recapitulate most of the the full stripe profile (***Figure 3***F, brown).

Here, we used biological numeracy as a driver for biological discovery: our discovery of the key role of the binary control of the transcriptional time window in pattern formation was only made possible by going beyond the *qualitative* description of pattern formation and demanding a *quantitative* agreement between our theoretical predictions and the experimental data (***Phillips, 2015***). Further, our work emphasizes how the regulation of gene expression *timing* in development is as important as the regulation of the spatial extent of these patterns along the embryo. Thus, in order to make progress toward a quantitative and predictive picture of pattern formation in development, 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 the study of snapshots of fixed embryos (***Pare et al., 2009***; ***Little et al., 2013***; ***Xu et al., 2015***; ***Zoller et al., 2018***) and embrace a *dynamical* description that acknowledges that development is a process that occurs outside of steady state (***Berrocal et al., 2018***).

To determine whether the same molecular mechanisms dictate the analog control of the mean transcription rate and the binary control of the transcriptional time window, we utilized a variety of theoretical and computational tools in conjunction with our live-imaging data. Specifically, to uncover how the mean rate of transcription is regulated across the stripe, we developed a mHMM that is capable of inferring the instantaneous activity state of individual gene loci from MS2 traces. We used this mHMM to infer average promoter-switching parameters across the stripe (***Figure 5***). In agreement with previous measurements of various gene expression patterns (***Xu et al., 2015***; ***Desponds et al., 2016***; ***Fukaya et al., 2016***; ***Zoller et al., 2018***), our results revealed that the the burst frequency (k_{on}) is the main bursting parameter regulated by the input transcription factors

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

across *eve* stripe 2. This increase in k_{on} in the stripe center increases the fraction of time that nuclei spend in the active transcriptional state.

Importantly, our mHMM algorithm is not limited to the *eve* stripe 2 system and should prove useful to infer the underlying regulatory kinetics 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, the method 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 method will be useful for the broader biophysical analysis of *in vivo* cellular processes at the single-molecule level.

Having identified k_{on} as the primary kinetic mode by which transcription factors modulate the mean rate of expression across *eve* stripe 2, we next sought to probe the relationship between bursting and the transcriptional time window (***Figure 6***A). We adapted our mHMM to go beyond time-independent models of promoter switching to infer the regulation of these rates across both space and time. We observed striking temporal trends indicating that the burst frequency responds dynamically to time-varying transcription factor inputs. However, we noted a significant disconnect between temporal trends in the burst frequency and the onset of transcriptional quiescence. In particular, k_{on} either increased or remained constant near the stripe center even as a significant fraction of *eve* nuclei transitioned into quiescence (***Figure 6***C and D). We reasoned that the onset of transcriptional quiescence is likely not the result of a progressive reduction in burst frequency, amplitude, or duration, and that quiescence is instead driven by molecular processes that are distinct from those that regulate transcriptional bursting such as transcriptional silencing.

To test this hypothesis, we utilized a logistic regression framework and time-resolved data for the primary regulators of *eve* stripe 2 to query the regulatory logic exhibited by the time window and bursting, respectively (***Appendix 8***). Consistent with our time-resolved mHMM results, the two regulatory strategies responded to transcription factor concentrations in different ways. On the one hand, increasing levels of Giant and Krüppel were sufficient to explain the onset of transcriptional quiescence in the stripe flanks (***Figure 7***A and D). This observation points to a model in which repressor levels act unilaterally—without respect to coincident levels of activator proteins—to shut off transcription at loci in an (at least effectively) irreversible fashion. Conversely, the joint action of Giant, Krüppel, and Hunchback was necessary to recapitulate the observed pattern of transcriptional bursting (***Figure 7***B and F).

This difference in the regulatory logic observed for the two strategies dissected in this work suggests that control of the transcriptional time window and the modulation of the average transcription rate arise from two distinct, orthogonal molecular mechanisms. Further, the striking absence of a direct functional role for Bicoid in the regulation of either phenomenon suggests that, while Bicoid is almost certainly necessary for the expression of *eve* stripe 2 (***Small et al., 1992***), it does not play a direct role in dictating the magnitude or duration of *eve* stripe 2 transcription. In this interpretation, Bicoid functions like a general transcription factor, facilitating the transcription of *eve* 2 without directly conferring spatiotemporal information.

While the results of our input-output analysis provide valuable insights into the mechanisms driving the regulation of transcription of the *eve* stripe 2 enhancer, key questions remain about the molecular character of the underlying processes. For instance, while loci engaged in transcriptional bursting appear to continuously sense changes in transcription factor concentrations, it remains an open question whether loci continue to actively read out transcription factor concentrations following the onset of transcriptional quiescence. While the transition *appears* irreversible in our data, it is possible that quiescence is, in fact, reversible in principle, but simply not observed in practice because repressor levels increase over time in our region of interest. In such a scenario, the direct action of relatively short-lived repressor binding could function to silence the locus, and a reduction in repressor concentration would lead to a rapid reactivation of transcription. Alternatively, if repressor levels function more akin to a trigger to, say, induce a change in the chromatin state at the *eve* locus, this would imply that loci, once quiescent, cease to sense transcription

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

factor concentrations and would fail to reactivate even if repressor levels decreased. Of course, intermediate cases could also be imagined.

In order to further test these and other molecular hypotheses , it will be critical to move beyond spatiotemporal averages for transcription factor inputs (***Figure 7***C) and, instead, use live single-nucleus measurements to directly correlate input transcription factor concentration dynamics with the corresponding transcriptional activity at the single-cell level (***Holloway and Spirov, 2017***). Experimentally, we recently demonstrated the simultaneous measurement of inputs and outputs in single nuclei of a living fly embryo using novel genetically encoded LlamaTags (***Bothma et al., 2018***). We believe that utilizing this novel technique, in conjunction with the theoretical methods presented here, to query the effects of targeted disruptions to transcription factor binding domains on regulatory enhancers will constitute a powerful assay for querying transcription factor function at the molecular level. Thus, there are clear experimental and theoretical paths to uncovering the detailed quantitative mechanisms behind the molecular control of transcriptional bursting and quiescence in development. Such a quantitative description is a necessary step toward a predictive understanding of developmental decision-making that makes it possible to calculate developmental outcomes from knowledge of the nature of the transcription factor interactions within gene regulatory networks.

Materials and methods

Cloning and transgenesis

This work employed the same *eve* stripe 2 reporter construct developed by ***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 (MCP, MS2 coat protein) were crossed to males bearing the reporter gene. Embryos were collected and mounted in halocarbon oil 27 between a semipermeable membrane (Lumox film, Starstedt) and a coverslip. Data collection was performed using a 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 256×512 pixels, with a pixel size of 212 nm and a pixel dwell time of 1.2 μs. The signal from each frame was 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 using a White Light Laser, respectively. Fluorescence was detected with two separate Hybrid Detectors (HyD) using the 498-546 nm and 566-669 nm spectral windows. Specimens were imaged for a minimum of 40 minutes into nuclear cleavage cycle 14.

Image analysis

Image analysis of live embryo 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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Data processing

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

This routine, described graphically in ***Figure 2–Figure Supplement 1***, used observed spatial patterns of fluorescence measured from 30 minutes into nuclear cycle 14—the approximate time at which the mature stripe is first established (***Bothma et al., 2014***)— 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 anterior posterior axis. Thus, a variety of orientations for the natural stripe axis were considered, ranging between $\pm$ 15 degrees from the line perpendicular to the stripe with the manually specified anterior posterior axis. For each orientation, a sliding window of 4% embryo length in width was used to find the position along the proposed orientation that captured the largest fraction of the total fluorescence emitted by the mature stripe. The orientation and position that maximized the amount of fluorescence captured within this window defined a line through the field of view that was taken as the stripe center. All anterior-posterior positions used for subsequent analyses were defined relative to this center line.

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

mHMM inference

To account for finite RNA polymerase elongation times, a compound state Markov formalism was developed in which the underlying tw-promoter system—assumed to have three states (see ***Figure 4***E,F)—was transformed into a system with 3^w compound gene states, where w indicates the number of time steps needed for an RNA polymerase molecule to traverse the full transcript (see ***Appendix 9***). 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 from subsets of experimental traces (***Appendix 3***). The scripts are available at the GarciaLab/mHMM GitHub repository. Bootstrap sampling was used to estimate the standard error in our parameter estimates. Subsets of 8,000 data points were used to generate time-averaged parameter estimates. In order to accurately capture the time-averaged dynamics across the entirety of nuclear cycle 14, the full length of each experimental trace (that is, the full set of time points for which activity was observed) was used for time averaged inference. Sample sizes for windowed inference varied due to data set limitations. When possible, samples of 4,000 points were used. Only data points falling within a 15 minute window centered about the time point of interest were included in windowed inference runs. Inference was not conducted for spatiotemporal regions for which fewer than 1,250 time points were available. A minimum of 10 bootstrap samples were used to estimate each parameter value reported in this work. Reported values represent the median taken across bootstrap samples.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Input-output logistic regressions

The input-output analysis presented in ***Figure 7*** utilized input transcription factor data from immunostaining experiments presented in ***Dubuis et al.*** (***2013***), as well as live measurements of a Bicoid-GFP fusion courtesy of Jonathan Liu and Elizabeth Eck. Logistic regression parameters were estimated in Matlab using the *fmincon* function. See ***Appendix 8*** for further details.

Bootstrap error calculation

Bootstrap resampling was used frequently throughout this work to estimate the standard error in a variety of reported quantities, from trends estimated directly from raw experimental data in ***Figure 1*** to mHMM inference results presented in ***Figure 5*** and ***Figure 6***. In this procedure, multiple bootstrap replicates, y^i_{boot} are generated by sampling with replacement from the pool of available experimental data, Y (see, *e.g.* ***Efron and Hastie*** (***2016***)). The parameter of interest (say, $t_{on}(x)$) is then calculated for each replicate and the mean of these estimates is taken as the bootstrap estimate of the parameter value, $\hat{t_{on}}(x)$, while the standard deviation across the pool of bootstrap parameter estimates is used to approximate the standard error in our estimate of $t_{on}(x)$. In our case, simply performing this procedure across the available pool of nuclei failed to account for biological variability that exists from embryo to embryo. To account for this, we introduced a hierarchical bootstrapping procedure. The first step in this procedure was to draw bootstrap samples from across the 11 embryos used in this study. Because these samples were taken with replacement, most bootstrap samples excluded some embryos out of the original set of 11 and included duplicates (or triplicates, etc.) of others. Each embryo-level bootstrap defined a subset of nuclei. The final set of nuclei used for parameter estimation was generated by performing another round of bootstrap sampling on this pool. Bootstrap averages and standard errors were then calculated as described above. This two-step procedure thus accounts for both embryo-to-embryo and nucleus-to-nucleus variability.

We note that the limited number of data points available for many spatiotemporal regions prevented us from performing this two-tiered bootstrap procedure in the case of our time-dependent mHMM inference (***Figure 6***D-F and ***Appendix 6–Figure 3***D-E). In these cases, we used all available sets (essentially skipping the first bootstrap resampling step) and took bootstrap samples from amongst available nuclei as in step two of the procedure described above.

Absolute calibration of MS2 signal

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

$$\alpha = \frac{N_{\mathrm{FISH}}}{F_{\mathrm{MS2}}}. \tag{5}$$

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 5*** 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,

$$F_1 = f_{\max} \frac{\frac{1}{2} L_I + L_{II}}{v_{\mathrm{elong}}}, \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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 RNA polymerase molecules along the gene. We can solve for $f_{\max}$ using α, namely,

$$F_1 = \frac{1}{\alpha} = f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}}, \tag{7}$$

such that

$$f_{\max} = \frac{v_{\text{elong}}}{\alpha} \frac{1}{\frac{1}{2}L_I + L_{II}}. \tag{8}$$

Here, we recognize that the cumulative fluorescence per RNA polymerase 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 RNA polymerase molecule, that is,

$$\begin{aligned} F_{\text{RNAP}} &= \frac{1}{\tau_{\text{elong}}} f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}} \\ &= \frac{v_{\text{elong}}}{(L_I + L_{II})} f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}} \\ &= f_{\max} \frac{\frac{1}{2}L_I + L_{II}}{(L_I + L_{II})} \\ &= \frac{v_{\text{elong}}}{\alpha} \frac{1}{(L_I + L_{II})}, \end{aligned} \tag{9}$$

resulting in

$$F_{\text{RNAP}} = \frac{v_{\text{elong}} F_{\text{MS2}}}{N_{\text{FISH}}} \frac{1}{(L_I + L_{II})}. \tag{10}$$

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 minutes. For the reporter gene used to take these measurements, L_I and L_{II} are 1.275 kb and 4.021 kb, respectively. Thus, we obtain

$$\begin{aligned} F_{\text{RNAP}} &= \frac{1.5 \times 9610}{220} \frac{1}{(1.275 + 4.021)} \\ &= 13 \pm 1.7 \text{ AU/RNAP}. \end{aligned} \tag{11}$$

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 RNA polymerase 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 RNA polymerase molecules.

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

Video 3. Transcriptional time window. Nuclei along the stripe false colored after the duration of their transcriptional time window.

Video 4. Fraction of active nuclei. Nuclei along the stripe false colored according to whether they engaged in transcription at any time point during the nuclear cycle.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Video 5. Fluorescent puncta contain sister chromatids. Fluorescent puncta transiently separate to reveal the presence of sister chromatids as shown by the white circles throughout the movie.

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

Video 7. Average embryo containing all inputs and the output. Average concentrations of Bicoid (blue), Hunchback (red), Kŕuppel (green) and Giant (yellow) combined with the average transcriptional activity of the *eve* reporter (purple). (Hunchback, Kŕuppel and Giant data obtained from ***Dubuis et al.*** (***2013***)).

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 and for the Bicoid-GFP data used in the logistic regression. We are also grateful to Jack Bateman, Jane Kondev, Rob Phillips, Allyson Sgro, Donald Rio and Richard Harland 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

Allis CD, Caparros ML, Jenuwein T, Reinberg D. Epigenetics. Second edition. ed. Cold Spring Harbor, New York: CSH Press, Cold Spring Harbor Laboratory Press; 2015.

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.

Arnosti DN, Barolo S, Levine M, Small S. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development. 1996; 122(1):205–14.

Bartman CR, Hsu SC, Hsiung CC, Raj A, Blobel GA. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol Cell. 2016; 62(2):237–247. doi: 10.1016/j.molcel.2016.03.007.

Berrocal A, Lammers NC, Garcia HG, Eisen MB. Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene. bioRxiv. 2018; doi: 10.1101/335901.

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].

Bishop C. Pattern recognition and machine learning. New York: Springer; 2006.

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, Gonzalez RL, Wiggins CH. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. Biophysical Journal. 2009; 97(12):3196–3205. doi: 10.1016/j.bpj.2009.09.031.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Bronstein L, Zechner C, Koeppl H. Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population. Methods. 2015; 85:22–35. doi: 10.1016/j.ymeth.2015.05.012.

Choubey S. Nascent RNA kinetics: Transient and steady state behavior of models of transcription. Phys Rev E. 2018; 97(2-1):022402. doi: 10.1103/PhysRevE.97.022402.

Choubey S, Kondev J, Sanchez A. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. PLoS Comput Biol. 2015; 11(11):e1004345. doi: 10.1371/journal.pcbi.1004345.

Choubey S, Kondev J, Sanchez A. Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells. Biophys J. 2018; 114(9):2072–2082. doi: 10.1016/j.bpj.2018.03.031.

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, Stern DL. Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nat Genet. 2016; 48(3):292–8. doi: 10.1038/ng.3509.

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.

Driever W, Nusslein-Volhard C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell. 1988; 54(1):95–104.

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.

Efron B, Hastie T. Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. 1st ed. New York, NY, USA: Cambridge University Press; 2016.

Estrada J, Ruiz-Herrero T, Scholes C, Wunderlich Z, DePace AH. SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences. PLoS One. 2016; 11(3):e0151740. doi: 10.1371/journal.pone.0151740.

Fakhouri WD, Ay A, Sayal R, Dresch J, Dayringer E, Arnosti DN. Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. Mol Syst Biol. 2010; 6:341. doi: msb200997 [pii] 10.1038/msb.2009.97.

Featherstone K, Hey K, Momiji H, McNamara AV, Patist AL, Woodburn J, Spiller DG, Christian HC, McNeilly AS, Mullins JJ, Finkenstadt BF, Rand DA, White MR, Davis JR. Spatially coordinated dynamic gene transcription in living pituitary tissue. Elife. 2016; 5:e08494. doi: 10.7554/eLife.08494.

Fowlkes CC, Hendriks CL, Keranen SV, Weber GH, Rubel O, Huang MY, Chatoor S, DePace AH, Simirenko L, Henriquez C, Beaton A, Weiszmann R, Celniker S, Hamann B, Knowles DW, Biggin MD, Eisen MB, Malik J. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell. 2008; 133(2):364–74. doi: 10.1016/j.cell.2008.01.053.

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; 166(2):358–368. doi: 10.1016/j.cell.2016.05.025.

Fukaya T, Lim B, Levine M. Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr Biol. 2017; 27(9):1387–1391. doi: 10.1016/j.cub.2017.03.069.

Garcia HG, Brewster RC, Phillips R. Using synthetic biology to make cells tomorrow's test tubes. Integr Biol (Camb). 2016; 8(4):431–50. doi: 10.1039/c6ib00006a.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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.

Gray S, Szymanski P, Levine M. Short-Range Repression Permits Multiple Enhancers to Function Autonomously within a Complex Promoter. Genes Development. 1994; 8(15):1829–1838.

Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and nuclear dynamics of the bicoid morphogen gradient. Cell. 2007; 130(1):141–52.

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.

Hendy O, Campbell J J, Weissman JD, Larson DR, Singer DS. Differential context-specific impact of individual core promoter elements on transcriptional dynamics. Mol Biol Cell. 2017; 28(23):3360–3370. doi: 10.1091/mbc.E17-06-0408.

Hey KL, Momiji H, Featherstone K, Davis JR, White MR, Rand DA, Finkenstadt B. A stochastic transcriptional switch model for single cell imaging data. Biostatistics. 2015; 16(4):655–69. doi: 10.1093/biostatistics/kxv010.

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.

Holloway DM, Spirov AV. Transcriptional bursting in Drosophila development: Stochastic dynamics of eve stripe 2 expression. PLoS One. 2017; 12(4):e0176228. doi: 10.1371/journal.pone.0176228.

Ilsley GR, Fisher J, Apweiler R, De Pace AH, Luscombe NM. Cellular resolution models for even skipped regulation in the entire Drosophila embryo. eLife. 2013 aug; 2:e00522. http://www.ncbi.nlm.nih.gov/pubmed/23930223http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3736529, doi: 10.7554/eLife.00522.

Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov KN, Manu, Myasnikova E, Vanario-Alonso CE, Samsonova M, Sharp DH, Reinitz J. Dynamic control of positional information in the early Drosophila embryo. Nature. 2004; 430(6997):368–71. doi: 10.1038/nature02678.

Jiang J, Hoey T, Levine M. Autoregulation of a segmentation gene in Drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. Genes Dev. 1991; 5(2):265–77.

Jiang J, Levine M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell. 1993; 72(5):741–52.

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.

Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. 2001; 81:3116–36.

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.

Lawrence PA, Johnston P, Macdonald P, Struhl G. Borders of parasegments in Drosophila embryos are delimited by the fushi tarazu and even-skipped genes. Nature. 1987; 328(6129):440–2. doi: 10.1038/328440a0.

Levine M, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. Cell. 2014; 157(1):13–25. doi: 10.1016/j.cell.2014.02.009.

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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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.

Lucas T, Ferraro T, Roelens B, De Las Heras Chanes J, Walczak AM, Coppey M, Dostatni N. Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol. 2013; 23(21):2135–9. doi: 10.1016/j.cub.2013.08.053.

Luengo Hendriks CL, Keranen SV, Fowlkes CC, Simirenko L, Weber GH, DePace AH, Henriquez C, Kaszuba DW, Hamann B, Eisen MB, Malik J, Sudar D, Biggin MD, Knowles DW. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome Biology. 2006; 7(12):R123. doi: 10.1186/gb-2006-7-12-r123.

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.

McKnight SL, Miller J O L. Post-replicative nonribosomal transcription units in D. melanogaster embryos. Cell. 1979; 17(3):551–63.

Medin SA, Lammers NC, Garcia HG. In preparation. 2019; .

Molina N, Suter DM, Cannavo R, Zoller B, Gotic I, Naef F. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proc Natl Acad Sci U S A. 2013; 110(51):20563–8. doi: 10.1073/pnas.1312310110.

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.

Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol Cell. 2015; 58(2):339–52. doi: 10.1016/j.molcel.2015.03.005.

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.

Park J, Estrada J, Johnson G, Ricci-Tam C, Bragdon M, Shulgina Y, Cha A, Gunawardena J, DePace AH. Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. bioRxiv. 2018; doi: 10.1101/408708.

Parker DS, White MA, Ramos AI, Cohen BA, Barolo S. The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. Sci Signal. 2011; 4(176):ra38. doi: 10.1126/scisignal.2002077.

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

Peter IS, Davidson EH. Genomic control process : development and evolution. London, UK ; San Diego, CA, USA: Academic Press is an imprint of Elsevier; 2015.

Petkova MD, Tkacik G, Bialek W, Wieschaus EF, Gregor T. Optimal Decoding of Cellular Identities in a Genetic Network. Cell. 2019; 176(4):844–855 e15. doi: 10.1016/j.cell.2019.01.007.

Phillips R. Theory in Biology: Figure 1 or Figure 7? Trends Cell Biol. 2015; doi: 10.1016/j.tcb.2015.10.007.

Phillips R, Kondev J, Theriot J, Garcia HG. Physical Biology of the Cell, 2nd Edition. New York: Garland Science; 2013.

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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Sanchez A, Kondev J. Transcriptional control of noise in gene expression. Proc Natl Acad Sci U S A. 2008; 105(13):5081–6.

Sasai M, Wolynes PG. Stochastic gene expression as a many-body problem. Proc Natl Acad Sci U S A. 2003; 100(5):2374–9. doi: 10.1073/pnas.2627987100.

Sayal R, Dresch JM, Pushel I, Taylor BR, Arnosti DN. Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early Drosophila embryo. Elife. 2016; 5. doi: 10.7554/eLife.08445.

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.

Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature. 2008; 451(7178):535–40. doi: nature06496 [pii] 10.1038/nature06496.

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.

Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. Single-cell analysis of transcription kinetics across the cell cycle. Elife. 2016; 5:e12175. doi: 10.7554/eLife.12175.

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

Small S, Kraut R, Hoey T, Warrior R, Levine M. Transcriptional regulation of a pair-rule stripe in Drosophila. Genes Dev. 1991; 5(5):827–39.

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.

Struhl G, Johnston P, Lawrence PA. Control of Drosophila Body Pattern by the Hunchback Morphogen Gradient. Cell. 1992; 69(2):237–249.

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011; 332(6028):472–4. doi: science.1198817 [pii] 10.1126/science.1198817.

Verd B, Crombach A, Jaeger J. Dynamic Maternal Gradients Control Timing and Shift-Rates for Drosophila Gap Gene Expression. PLoS Comput Biol. 2017; 13(2):e1005285. doi: 10.1371/journal.pcbi.1005285.

Vincent BJ, Estrada J, DePace AH. The appeasement of Doug: a synthetic approach to enhancer biology. Integr Biol (Camb). 2016; 8(4):475–84. doi: 10.1039/c5ib00321k.

Viterbi AJ. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans Inform Theory. 1967; 13(2):260–269. doi: 10.1109/TIT.1967.1054010.

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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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.

Xu H, Skinner SO, Sokac AM, Golding I. Stochastic Kinetics of Nascent RNA. Phys Rev Lett. 2016; 117(12). doi: 10.1103/PhysRevLett.117.128101.

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.

Zechner C, Unger M, Pelet S, Peter M, Koeppl H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. Nat Methods. 2014; 11(2):197–202. doi: 10.1038/nmeth.2794.

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 Unified Kinetics of Transcriptional Bursting. Cell. 2018; 175(3):835–847 e25. doi: 10.1016/j.cell.2018.09.056.

Zoller B, Nicolas D, Molina N, Naef F. Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol Syst Biol. 2015; 11(7):823. doi: 10.15252/msb.20156257.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 1

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

Derivation details

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

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = \underbrace{R(x,t)}_{\text{transcription rate}} - \underbrace{\gamma\,\text{mRNA}(x,t)}_{\text{degradation rate}}, \tag{12}$$

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

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = -\gamma\,\text{mRNA}(x,t), \tag{13}$$

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

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

$$R(x) \approx \langle R(x,t) \rangle_t. \tag{14}$$

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

$$\text{mRNA}_{\text{active}}(x,t) = \underbrace{R(x)}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma}\left(1 - e^{-\gamma(t - t_{\text{on}}(x))}\right)}_{\text{time window}}. \tag{15}$$

Note that if the system evolves for a long amount of time, the second term in the parenthetical in ***Equation 15*** becomes vanishingly small ($\gamma(t - t_{\text{on}}(x)) \gg 1$) such that all time dependence drops out of the expression and we recover the familiar expression for mRNA levels in steady state

$$\text{mRNA}_{\text{active}}(x,t) = \frac{R(x)}{\gamma}, \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

where mRNA production and degradation are balanced.

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

$$\text{mRNA}_{\text{quiescent}}(x,t) = \underbrace{e^{-\gamma(t-t_{\text{off}}(x))}}_{\text{quiescent decay}} \left[\underbrace{R(x)}_{\text{transcription rate}} \times \underbrace{\frac{1}{\gamma}\left(1 - e^{-\gamma(t_{\text{off}}(x)-t_{\text{on}}(x))}\right)}_{\text{time window}} \right]. \tag{17}$$

Equation 17 represents a scenario in which the accumulation of cytoplasmic mRNA results from the interplay between two distinct regulatory strategies: the modulation of when the transcription starts and stops (binary control of the transcription time window) and the average rate with which transcription occurs within this time window (analog control of transcriptional bursting). We refactor ***Equation 17*** to reflect this distinction and consider the case when $t > t_{\text{on}}$, giving

$$\text{mRNA}_{\text{full}}(x,t) = \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{e^{-\gamma(t-\min(t_{\text{off}}(x),t))}\left(1 - e^{-\gamma(\min(t_{\text{off}}(x),t)-t_{\text{on}}(x))}\right)}_{\text{binary control}}, \tag{18}$$

which can be simplified slightly to yield

$$\text{mRNA}(x,t) = \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{19}$$

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

$$\text{mRNA}(x,t) = p_{\text{active}}(x) \times \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{20}$$

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

Accounting for multiple transcriptional states

In the main text, ***Equation 3*** expresses the mean rate of mRNA production, $R(x)$, as a function of the bursting parameters k_{on}, k_{off}, and r. We can combine this equation with ***Equation 20*** to obtain an expression for the predicted amount of cytoplasmic mRNA that includes the burst parameters inferred by our mHMM

$$\text{mRNA}(x,t) = p_{\text{active}}(x) \times \underbrace{\frac{r(x)}{\gamma}\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}. \tag{21}$$

While we present our results in terms of an effective two-state model in the main text, the presence of two transcriptional loci within each observed fluorescent spot suggests that the system is more naturally described using a three-state kinetic model. Here, we extend the framework presented in ***Equation 21*** to a scenario in which there are three distinct system states: 0 promoters on (0), 1 promoter on (1), and both promoters on (2) (see ***Figure 4***). We begin with a general expression for this scenario that takes the contribution from the analog

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

control term shown in ***Equation 20*** to be a sum over the output of each of the 3 activity states, namely,

$$\text{mRNA}(x,t) = p_{\text{active}}(x) \times \underbrace{\frac{1}{\gamma}\left(\sum_{i=0}^{2} r_i(x)\pi_i(x)\right)}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\right)}_{\text{binary control}}, \tag{22}$$

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

The fractional occupancies of the activity states ($\pi_i(x)$ terms in ***Equation 22***) are a function of the rates with which the promoter switches between activity states. In general, the fractional occupancy of each activity state, π_i, may vary as a function of time; however we focus on their steady state values here, such that:

$$0 = \boldsymbol{R}(x)\boldsymbol{\pi}(x), \tag{23}$$

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

$$\boldsymbol{R}(x) = \begin{bmatrix} -k_{01}(x) & k_{10}(x) & 0 \\ k_{01}(x) & -k_{10}(x) - k_{12}(x) & k_{21}(x) \\ 0 & k_{12}(x) & -k_{21}(x) \end{bmatrix}. \tag{24}$$

Together, ***Equation 23*** and ***Equation 24*** allow us to solve for the fractional occupancy of each activity state as a function of the transition rates that describe the system.

For the remainder of this derivation, we will drop the explicit x and t dependencies for ease of notation. Intuitively, the steady state (or stationary) distribution represents a limiting behavior of the system such that, upon reaching $\boldsymbol{\pi}$, no further shifts occur in the mean fraction of time spent in each activity state. ***Equation 23*** leads to a system of three equations:

$$0 = -\pi_0 k_{01} + \pi_1 k_{10} \tag{25}$$

$$0 = \pi_0 k_{01} - \pi_1\left(k_{10} + k_{12}\right) + \pi_2 k_{21} \tag{26}$$

$$0 = \pi_1 k_{12} - \pi_2 k_{21} \tag{27}$$

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

$$1 = \pi_0 + \pi_1 + \pi_2. \tag{28}$$

With this in mind, we can solve ***Equation 25*** for π_1 to find

$$\pi_1 k_{10} = \pi_0 k_{01} \tag{29}$$

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}}. \tag{30}$$

Next, we use the normalization condition to eliminate π_2 from ***Equation 27***:

$$\begin{aligned} \pi_1 k_{12} &= \pi_2 k_{21} \\ &= (1 - \pi_0 - \pi_1)k_{21}. \end{aligned} \tag{31}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

By combining this result with ***Equation 30***, we obtain

$$\pi_0 \frac{k_{01}}{k_{10}} k_{12} = (1 - \pi_0 - \pi_0 \frac{k_{01}}{k_{10}}) k_{21} \tag{32}$$

$$\pi_0 \frac{k_{01} k_{12}}{k_{10} k_{21}} = 1 - \pi_0 \frac{k_{10} + k_{01}}{k_{10}} \tag{33}$$

$$\pi_0 = \frac{k_{10} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \tag{34}$$

With ***Equation 34*** in hand, it is then straightforward to solve for the remaining π_i terms. First we obtain π_1 by substituting ***Equation 34*** into ***Equation 30***:

$$\begin{aligned} \pi_1 &= \pi_0 \frac{k_{01}}{k_{10}} \\ &= \frac{k_{01} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \end{aligned} \tag{35}$$

And finally π_2:

$$\begin{aligned} \pi_2 &= 1 - \pi_0 - \pi_1 \\ &= \frac{k_{01} k_{12}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \end{aligned} \tag{36}$$

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

$$\text{mRNA}(x,t) = p_{\text{active}}(x) \underbrace{\frac{1}{\gamma}\left(r_1(x) \frac{k_{01}(x) k_{21}(x)}{\kappa(x)} + r_2(x) \frac{k_{01}(x) k_{12}(x)}{\kappa(x)} \right)}_{\text{analog control}} \times \underbrace{\left(e^{-\gamma(t - \min(t_{\text{off}}(x), t))} - e^{-\gamma(t - t_{\text{on}}(x))} \right)}_{\text{binary control}}, \tag{37}$$

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 34***, ***Equation 35*** and ***Equation 36***, namely,

$$\kappa = k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}. \tag{38}$$

Thus, from ***Equation 37*** we see that, while there are more terms comprising the analog control expression, the expression nonetheless takes on the same essential form as in ***Equation 20***.

Mapping the three-state model into an effective two-state model

Here we provide expressions relating the effective two-state parameters presented in the main text to parameters from the full three-state model. As we have done throughout this work, we take the transition rates between states (0) and (2) of the 3-state model to be negligible (consistent with inference results, see ***Appendix 6***). First, the on rate, $k_{\text{on}}^{\text{eff}}$ is directly equivalent to the transition rate between states (0) and (1), that is,

$$k_{on}^{\text{eff}} = k_{01}. \tag{39}$$

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

$$k_{\text{off}}^{\text{eff}} = \frac{\pi_1 k_{10}}{\pi_1 + \pi_2} \tag{40}$$

$$= \frac{k_{01} k_{21} k_{10}}{k_{01} k_{21} + k_{01} k_{12}} \tag{41}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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

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

$$r^{\text{eff}} = \frac{\pi_1 r_1 + \pi_2 r_2}{\pi_1 + \pi_2} \tag{43}$$

$$= \frac{r_1 k_{01} k_{21} + r_2 k_{01} k_{12}}{k_{01} k_{21} + k_{01} k_{12}} \tag{44}$$

$$= \frac{r_1 k_{21} + r_2 k_{12}}{k_{21} + k_{12}}. \tag{45}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 2

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 parameter estimates returned by mHMM, was used to calculate the total cytoplasmic mRNA levels per nucleus shown in ***Figure 3***G (red), as well as the "binary control" of the duration of the transcriptional time window contribution ***Figure 3***G (blue).

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 RNA polymerase molecules. Thus, after a period equal to the amount of time needed for an RNA polymerase 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***), that is,

$$F(t) \propto M(t + \tau_{\text{elong}}) - M(t), \tag{46}$$

where $M(t)$ indicates the total number of mRNA molecules that have been produced up to time t. We relate this fluorescence signal to absolute numbers of RNA polymerase molecules using the calibration procedure described in the Materials and 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 2***. 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 at time t to be approximately equal to the observed fluorescence at time $t - \frac{\tau_{\text{elong}}}{2}$

$$F\left(t - \frac{\tau_{\text{elong}}}{2}\right) \approx \beta \frac{\mathrm{d}M(t)}{\mathrm{d}t}. \tag{47}$$

Here, the $\frac{\tau_{\text{elong}}}{2}$ term accounts for 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 47*** as one of equality. For ***Figure 3***G, 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 (active, quiescent, and those that never engaged in transcription) within the region of interest was used as a proxy for the instantaneous rate of mRNA production per nucleus, given by

$$\begin{aligned} \frac{\mathrm{d}M(x,t)}{\mathrm{d}t} &= \frac{\beta}{N} \sum_{i=1}^{N} F_i(x,t) \\ &= \beta \langle F(t) \rangle_x. \end{aligned} \tag{48}$$

Here, $F_i(x,t)$ is the fluorescence of nucleus i at time t. The x subscript in ***Equation 48*** indicates that the average is taken over all nuclei falling within the same anterior-posterior region within the *eve* stripe 2 pattern. Having obtained an expression for the rate of mRNA production as a function of space and time, we next sought to account for the degradation of mRNA over time. As indicated in the main text, we assumed a constant rate of mRNA decay, γ, over space and time. The next section in this appendix provides evidence for the validity of

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

this assumption. For a constant mRNA decay rate, calculating the average concentration of mRNA amounts to taking a weighted sum over all preceding production rates for a position of interest, where the weight terms account for the effects of mRNA decay and are of the form $e^{(-\gamma t)}$. Thus, we summed over all time points for each region of interest to estimate the total amount of cytoplasmic mRNA present on average, yielding the quantity on the left-hand side of ***Equation 2***, namely,

$$\text{mRNA}(x,t) = \beta \sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \left\langle F(t-n\Delta\tau) \right\rangle. \tag{49}$$

Here $\Delta\tau$ is the experimental time resolution, and $T = \frac{t}{\Delta\tau}$ denotes the number of measurements taken through time t. The exponential term within the summand on the right-hand side captures the effects of mRNA decay (*see* ***Appendix 1***). Finally, to calculate the normalized mRNA profile shown in ***Figure 3***G (red), the estimates for the total accumulated mRNA per nucleus found using ***Equation 49*** must be divided by the sum across all spatial regions considered

$$\text{mRNA}_{\text{norm}}(x_j,t) = \frac{\sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \left\langle F(t-n\Delta\tau) \right\rangle_{x_j}}{\sum_{i\in X} \sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \left\langle F(t-n\Delta\tau) \right\rangle_{x_i}}, \tag{50}$$

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 $\text{mRNA}_{\text{norm}}$. As a final step, we subtract the minimum across the AP region considered to remove any basal offset such that

$$\text{mRNA}_{\text{full}}(x_j,t) = \text{mRNA}_{\text{norm}}(x_i,t) - \min_{x} \left(\text{mRNA}_{\text{norm}}(x_j,t) \right). \tag{51}$$

Calculating mRNA profiles due to the binary control of the transcriptional time window

The predicted profile due to binary control of the transcriptional time window alone (***Figure 3***G, blue) 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 $t < t^i_{\text{off}}$ and 0 otherwise. Additionally, only nuclei that were active at *some point* during nuclear cycle 14 were included, to distinguish the effects of the transcriptional time window (***Figure 1***C) from the control of the fraction of active nuclei (***Figure 1***D). Thus, in this scenario, the "average rate" of mRNA production is equivalent to the fraction of nuclei engaged in transcriptional activity at a given point in time such that the rate of mRNA production is given by

$$\begin{aligned} \frac{\mathrm{d}M_{\text{binary}}(x,t)}{\mathrm{d}t} &= \frac{1}{N(x)} \sum_{i=1}^{N} (x,t) f_i(t) \\ &= \langle f(x,t) \rangle \\ &= \frac{N_c(x,t)}{N(x)}, \end{aligned} \tag{52}$$

where $N_c(t)$ indicates the number of transcriptionally competent nuclei at time t. The binary equivalent to ***Equation 49*** takes the form of a time-weighted sum of the fraction of active

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

nuclei within a region

$$\mathrm{mRNA}_{\mathrm{binary}}(x,t) = \sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \frac{N_c(x, n\Delta t)}{N(x)}. \tag{53}$$

The steps for calculating the the normalized binary mRNA levels comprising the blue profile in ***Figure 3***G from ***Equation 53*** are identical to those shown for the full mRNA profile in ***Equation 50*** and ***Equation 51*** and are therefore not repeated here.

Comparison between predicted and measured cytoplasmic mRNA profiles

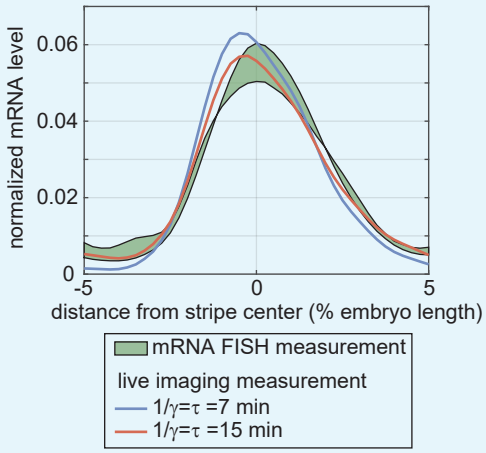
As a check for the validity of our approach to predicting levels of cytoplasmic mRNA from live imaging data (***Equation 50*** and ***Equation 51***), we sought to compare our model's predictions to existing mRNA FISH data for the endogenous *eve* stripe 2 (***Fowlkes et al., 2008***). For this comparison, we elected to use live imaging data for *eve* stripe 2 activity that was driven by a BAC containing the full *eve* locus (see ***Berrocal et al.*** (***2018***) for details). This was done to minimize potential differences with the activity of the endogenous gene. Most notably, the reporter construct used for the majority of this work does not contain an enhancer sequence that is responsible for driving *eve* expression late in nuclear cycle 14 (***Jiang et al., 1991***).

Appendix 2 Figure 1 summarizes the results of this comparison. To account for uncertainty regarding the precise dorsal-ventral (DV) orientation of embryos within our live-imaging set, we compared our model's predictions to mRNA measurements for a range of DV positions, encompassed by the green-shaded profile. We found a high degree of agreement between model predictions and reported levels of cytoplasmic mRNA. This conclusion is relatively insensitive to our assumptions regarding the average lifetime of *eve* mRNA as shown by the lines in the figure. While substantial uncertainties regarding the precise timing of the mRNA measurements prevented us from leveraging this comparison to, for instance, infer the rate of *eve* mRNA decay, we nonetheless concluded that it is sufficient to establish our modelling assumptions. In particular, the relative insensitivity of the distribution of cytoplasmic mRNA to the decay rate suggests that, while it is possible that the precise rate of mRNA decay is regulated across space or time, such phenomena—if they exist—would not impact the core conclusions presented in this work. Moreover, as discussed below, this paper's findings are also relatively insensitive to our choice of decay rate γ, with the basic dynamics of stripe formation remaining consistent even in the limits near instantaneous and infinitely slow mRNA decay (***Figure 2***).

As alluded to above, several variables limited our ability to carry out a precise comparison between our model predictions and empirical measurements. The most significant of these was the lack of sufficiently precise temporal information for the empirical mRNA measurements. The authors used the percent invagination of cellular membranes through cellularization as a means to break individual embryos into rough temporal cohorts (***Luengo Hendriks et al., 2006***). We cross-referenced the invagination ranges for each temporal group with data provided by ***Dubuis et al.*** (***2013***) to obtain estimates for the range of times encapsulated by each of these cohorts. This calibration revealed that most time-averaged cohorts spanned too broad a range of times to allow for reasonable comparison. We elected to use the cohort comprised of embryos with ages ranging between 38 to 48 minutes into nuclear cycle 14 both because this range was much narrower than that spanned by the preceding cohort and because we had established that the stripe appeared to be relatively stable during this time period. An additional complication with establishing the precise timing of each cohort was the fact that the authors of ***Luengo Hendriks et al.*** (***2006***) measured invagination on the ventral surface of the embryo, while the authors in ***Dubuis et al.*** (***2013***) used the dorsal surface. However, while invagination is known to proceed more rapidly on

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

the ventral side of the embryo,the authors in ***Luengo Hendriks et al.*** (***2006***) reported that this discrepancy is minimal up to the point where cell membrane extension has progressed to approximately 40% of its eventual full extent. The lower and upper bounds on the percent membrane invagination for the chosen cohort are 26% and 50% respectively. Thus, we expect the time estimate derived for the beginning of the period to be reasonably accurate, since Dorsal and Ventral membrane progression was reported to be comparable during this period. On the other hand, to the degree that Ventral invagination outpaces Dorsal invagination at the end of our period of interest, this would result in an over-estimation of ending time. Thus, it is possible that the true temporal window encompassed by the selected cohort is actually tighter than 10 minutes, since the ending time might in fact be earlier than 48 minutes into nuclear cycle 14. Given the relative stability of the stripe profile during this period of development, we do not expect this potential discrepancy to have a material impact on our conclusions.



Appendix 2 Figure 1. Comparison of predicted cytoplasmic mRNA by live-imaging measurements to direct measurements by FISH. In an effort to check the validity of our modelling assumptions, we compared the predictions of our mRNA model stemming from live-imaging measurements of stripe 2 of an *eve* reporter from a BAC containing the full *eve* locus to direct measurements of *eve* cytoplasmic mRNA levels using FISH (***Luengo Hendriks et al., 2006***). Here, the blue lines indicate our model's predictions under two different assumptions for the rate of mRNA degradation, and the shaded green profile indicates the range of directly measured mRNA levels. Comparisons indicate a high degree of agreement between prediction and measurement, indicating that our modelling assumptions are justified.

Sensitivity of results to mRNA lifetime assumption

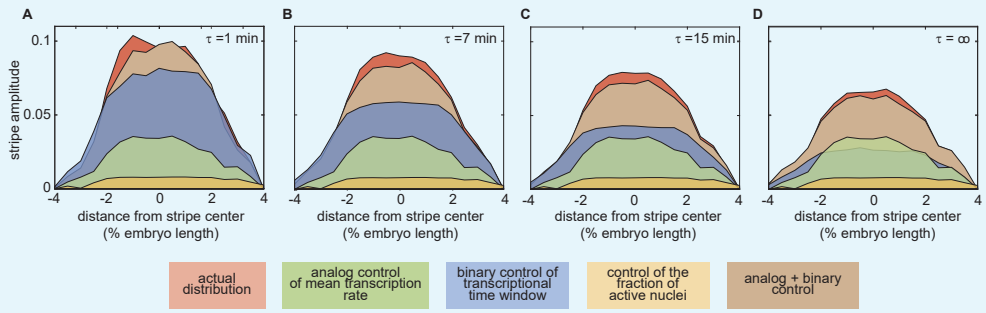
In the main text we assume a degradation rate for *eve* of 0.14 min^{-1} (corresponding to a lifetime of roughly $\tau = 7$ min). Since, to our knowledge, the decay rate of *eve* mRNA has not been measured directly, we follow ***Bothma et al.*** (***2014***) and base this estimate off of measurements for another of the pair rule genes, *fushi tazu* (*ftz*, ***Edgar et al.*** (***1987***)). In this section, we examine the degree to which the apparent contributions of each regulatory strategy (***Figure 1***) change under different assumptions for *eve* mRNA lifetime. Rather than conducting an exhaustive survey, we instead focus primarily on two limiting cases: rapid mRNA decay ($\tau = 1$ min) and *no* mRNA decay ($\tau = \infty$).

Appendix 2-Figure 2 summarizes the results of our analysis. We find that, regardless of the assumed mRNA lifetime, our model predicts that *eve* stripe 2 is formed almost entirely via the interplay between the binary control of the transcriptional time window and the analog modulation of the mean rate of transcription (compare brown and red profiles in ***Appendix 2-Figure 2***). However, we find that the relative importance of each factor depends,

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

somewhat, on the assumed decay rate. In the case of rapid mRNA decay, as well as for the decay rate assumed in the main text, the time window (blue profile) is clearly the dominant factor in driving pattern formation (***Appendix 2-Figure 2***A and B). If we assume the true mRNA lifetime is 15 minutes, slightly more than double our best guess of 7 minutes, we find that the time window is still predicted to contribute slightly more to stripe formation, but that the two contributions are now of order with one another (***Appendix 2-Figure 2***C). Finally, in the limit where there is effectively no mRNA decay, the effects of the mean rate and time window are roughly equivalent (***Appendix 2-Figure 2***D). This result can be explained by the fact that the mean rate strategy is insensitive to the decay rate, whereas the effect of the time window is enhanced by the action of mRNA decay.

Thus, overall, we found that our model's prediction that the control of the transcriptional time window plays a primary role in stripe formation holds for mRNA lifetimes less than or equal to 15 minutes, which is more than double the measured life time of *ftz* mRNA (***Edgar et al., 1987***). Perhaps more importantly, both factors are found to play a significant role, *irrespective* of mRNA decay rate, indicating that our central finding is robust to our assumption regarding mRNA decay dynamics.



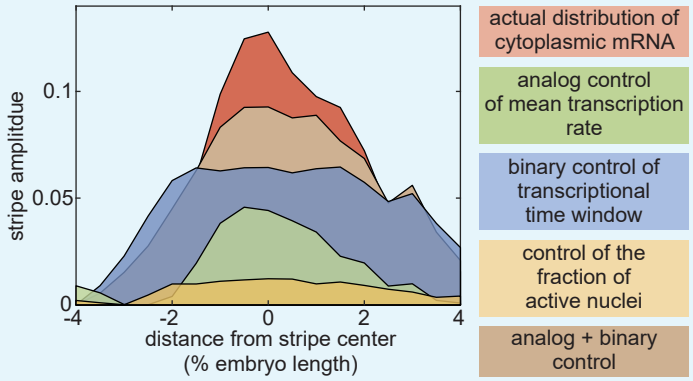
Appendix 2 Figure 2. Sensitivity of regulatory strategy contribution to assumed mRNA lifetime. The average lifetime of *eve* mRNA is a significant assumed parameter in our model. This figure compares the predicted contributions of each regulatory strategy for the mRNA lifetime assumed in the main text ($\tau = 7$ min) to limiting cases in which mRNA is assumed to decay almost instantaneously ($\tau = 1$ min) on the one hand, and infinitely slowly on the other ($\tau = \infty$). Even at these extremes, the central conclusion that the stripe is formed via the join action of mean rate modulation (green profile) and the time window (blue profile) remains intact. As expected, the relative contribution of the time window is sensitive to the assumed τ, yet even in the limit of no significant mRNA decay, its impact is still of order with the effect of mean rate modulation.

Control strategy contributions for *eve* BAC

A key question regarding the results in the main text is whether and to what degree the relative contributions of the regulatory control strategies we identified in ***Figure 1*** and ***Figure 3*** for the reporter containing only the *eve* stripe 2 enhancer hold true for the formation of the stripe in the endogenous context. While we cannot directly query activity at the endogenous *eve* locus, we were able to examine the dynamics of stripe formation for an *eve* BAC used in the companion paper to this manuscript (***Berrocal et al., 2018***). Since this BAC contains the full *eve* regulatory locus, it likely provides a better proxy for stripe formation in the endogenous context than the isolated *eve* 2 reporter. ***Appendix 2—Figure 3*** shows the results of this analysis. As with the reporter construct used in the main text, we find that the stripe is formed primarily through the interplay between two regulatory strategies: the modulation of the average rate of production (green) and of the duration of transcriptional activity (blue). As with the reporter, the binary control of the transcriptional time window is the dominant driver of stripe formation (compare with ***Figure 1***G). Interestingly, unlike the reporter construct, the full predicted profile (red profile) that accounts for the interplay

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

between mRNA decay temporal fluctuations in the mean rate of mRNA production differs substantially from the simpler model (brown profile) that approximates mRNA production as constant over time. We speculate that this difference is attributable to the influence of the "late enhancer"—which is present in the *eve* BAC but *not* in the reporter—that takes over control of *eve* activity late in nc14. Further work will be necessary to fully elucidate the regulatory impact of this late element on the formation of the mature *eve* stripe pattern.



Appendix 2 Figure 3. Regulatory strategy contributions to *eve* stripe 2 formation in endogenous context. As with the reporter construct, the formation of *eve* stripe 2 in the context of the full *eve* regulatory locus is dominated by the interplay between mean rate modulation (green) and control of the time window of transcriptional activity (blue).

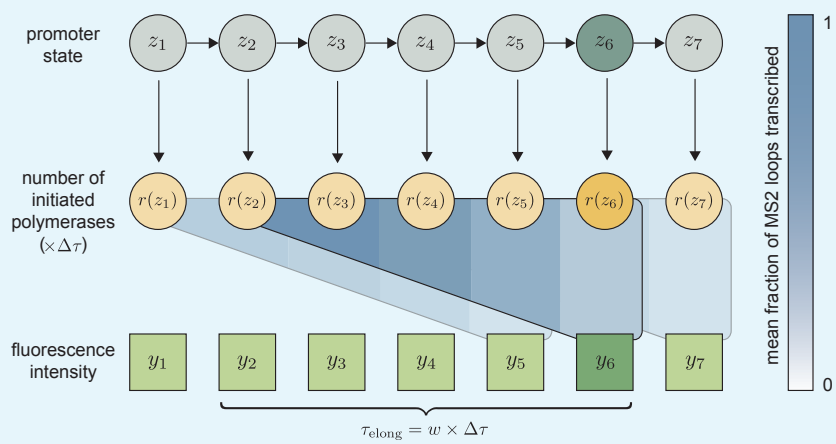
bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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 RNA polymerase initiation rate, $r(k)$, with units of RNA polymerase per minute, to each effective promoter state, $z(k)$, $1 \leq k \leq K$. Thus, the number of polymerases initiated between time steps $t-1$ and t will be $r(z_t)\Delta\tau$. Because the fluorescence intensity contributed by each polymerase depends on the number of transcribed MS2 stem loops, the contribution will vary with the position of the polymerase on the gene. In our transcription model we assume that polymerase elongation takes place at a constant rate. Therefore, if τ_{MS2} is the time it takes to transcribe the MS2 loops, the fluorescence contribution of an RNA polymerase molecule will initially grow linearly ($\tau \leq \tau_{\text{MS2}}$) and will then stay constant for the remainder of transcription ($\tau_{\text{MS2}} \leq \tau \leq \tau_{\text{elong}}$). Given this time dependence, we define a maximum fluorescence emission per time step for each state as $\upsilon(k) = F_{\text{RNAP}} r(k)\Delta\tau$, $1 \leq k \leq K$, 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 that have been transcribed and contribute fluorescence at the moment of observation. The color corresponding to the more recently initiated polymerases is therefore lighter (fewer loops transcribed) than that corresponding to polymerases initiated at earlier times (more loops transcribed).

The instantaneous fluorescence intensity is the cumulative contribution from polymerases initiated in the previous w time steps, where $w = \tau_{\text{elong}}/\Delta\tau$ is the system-dependent integer

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

memory. Here $\Delta\tau$ indicates the observational time resolution, a quantity set by experimental conditions. The time required for an RNA polymerase molecule to transcribe our reporter gene (τ_{elong}) is *a priori* unknown. We developed an autocorrelation-based method to estimate τ_{elong} directly from our experimental data (see ***Appendix 9*** and ***Coulon and Larson*** (***2016***)). 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

$$p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = p(s_1|\boldsymbol{\pi}) \prod_{t=1}^{T} p(y_t|s_t, \boldsymbol{\nu}, \sigma) \prod_{t=2}^{T} p(s_t|s_{t-1}, \boldsymbol{A}). \tag{54}$$

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 $\boldsymbol{\nu}$ is a K-element vector of fluorescence emission values per time step.

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

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{55}$$

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,

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{\boldsymbol{s}=\{s_1, s_2, ..., s_T\}} p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}). \tag{56}$$

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 is thus equal to K^{wT}, which grows exponentially with the number of time points. To make the estimation of the model parameters tractable, we use an approximate inference method, the expectation-maximization (EM) algorithm.

We note that the notion of a compound state was also introduced in an earlier work (***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 RNA polymerase molecules, but instead, assign a continuous RNA polymerase initiation rate to each promoter state. Additionally, our model estimates the magnitude of the background noise present in the experimentally measured fluorescence signal, whereas the model presented in (***Corrigan et al., 2016***) takes this quantity as an input, requiring that it be estimated separately. We believe that these differences serve to

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

make our model more flexible. Moreover, by eliminating the need for absolute calibration and noise estimation, we hoped to facilitate the use of our model in a wide variety of experimental contexts, for which one or the other quantity may not be readily obtainable. In the "Continuous vs. Poisson promoter loading" section of ***Appendix 4*** we demonstrate that relaxing the continuous RNA polymerase 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 (***Bishop*** (***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,

$$\hat{\boldsymbol{\theta}}_{k+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{57}$$

$$\mathcal{L}(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{\boldsymbol{s}=\{s_1,s_2,\ldots,s_T\}} p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k)\, \log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) \leq \log\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{58}$$

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

$$\log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = \log p(s_1|\boldsymbol{\pi}) + \sum_{t=1}^{T} \log p(y_t|s_t, \boldsymbol{v}, \sigma) + \sum_{t=2}^{T} \log p(s_t|s_{t-1}, \boldsymbol{A}). \tag{59}$$

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 59*** can be rewritten as

$$\log p(s_1|\boldsymbol{\pi}) = \sum_{i=1}^{K^w} \sum_{k=1}^{K} s_1^i C_{ki} \log \pi_k, \tag{60}$$

$$\log p(y_t|s_t, \boldsymbol{v}, \sigma) = \frac{1}{2} \sum_{i=1}^{K^w} s_t^i \left(\log \lambda - \log(2\pi) - \lambda(y_t - V_i(\boldsymbol{v}))^2\right), \tag{61}$$

$$\log p(s_t|s_{t-1}, \boldsymbol{A}) = \sum_{i,j=1}^{K^w} \sum_{k,l=1}^{K} B_{ij} s_t^i s_{t-1}^j C_{ki} C_{lj} \log A_{kl}. \tag{62}$$

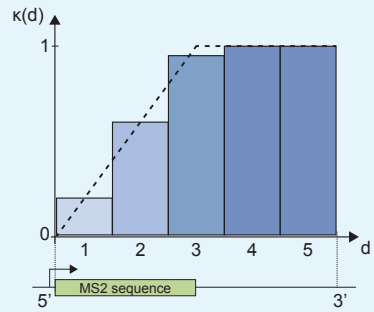
Here $\lambda = 1/\sigma^2$ is the Gaussian precision parameter, and $V_i(\boldsymbol{v})$ is the aggregate fluorescence produced in the w consecutive promoter states of the ith compound state.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

where $\lceil n_{\mathrm{MS2}} \rceil$ and $\lfloor n_{\mathrm{MS2}} \rfloor$ 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***.



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: $\tau_{\mathrm{elong}} = 100$ sec, $\tau_{\mathrm{MS2}} = 50$ sec, $\Delta\tau = 20$ sec, $w = \tau_{\mathrm{elong}}/\Delta\tau = 5$, $n_{\mathrm{MS2}} = \tau_{\mathrm{MS2}}/\Delta\tau = 2.5$.

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

$$V_i(\boldsymbol{v}) = F_{i,:}\boldsymbol{v}, \tag{63}$$

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,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = &\sum_{i=1}^{K^w}\sum_{k=1}^{K}\langle s_1^i\rangle C_{ki}\log\pi_k \\ &+ \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{K^w}\langle s_t^i\rangle\left(\log\lambda - \log(2\pi) - \lambda(y_t - F_{i,:}\boldsymbol{v})^2\right) \\ &+ \sum_{t=1}^{T}\sum_{i,j=1}^{K^w}\sum_{k,l=1}^{K} B_{ij}\langle s_t^i s_{t-1}^j\rangle C_{ki}C_{lj}\log A_{kl}.\end{aligned} \tag{64}$$

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

$$\langle s_t^i\rangle = \sum_{\boldsymbol{s}=\{s_1,s_2,\ldots,s_T\}} s_t^i\, p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{65}$$

$$\langle s_t^i s_{t-1}^j\rangle = \sum_{\boldsymbol{s}=\{s_1,s_2,\ldots,s_T\}} s_t^i s_{t-1}^j\, p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k). \tag{66}$$

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 65*** and ***Equation 66***, we obtain

$$\langle s_t^i\rangle = \sum_{s_1,s_2,\ldots,s_T} s_t^i\, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t} s_t^i\, p(s_t|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{67}$$

$$\langle s_t^i s_{t-1}^j\rangle = \sum_{s_1,s_2,\ldots,s_T} s_t^i s_{t-1}^j\, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t,s_{t-1}} s_t^i s_{t-1}^j\, p(s_t, s_{t-1}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k). \tag{68}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Following the conventional implementation of the forward-backward algorithm (cf. ***Bishop*** (***2006***), Chapter 13), we use the Markov property of the promoter state dynamics, together with the sum and products rules of probability, to write

$$p(s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k) = \frac{\alpha_t(s_t)\beta_t(s_t)}{p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k)}, \tag{69}$$

$$p(s_{t-1},s_t|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k) = \frac{\alpha_{t-1}(s_{t-1})\,p(y_t|s_t,\hat{\boldsymbol{\theta}}_k)\,p(s_t|s_{t-1},\hat{\boldsymbol{\theta}}_k)\beta_t(s_t)}{p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k)}, \tag{70}$$

$$\alpha_t(i) = p(y_1,\ldots,y_t,s_t=i|\,\hat{\boldsymbol{\theta}}_k), \tag{71}$$

$$\beta_t(i) = p(y_{t+1},\ldots,y_T|s_t=i,\hat{\boldsymbol{\theta}}_k). \tag{72}$$

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 71*** and ***Equation 72***. We evaluate the elements of α and β matrices recursively as

$$\alpha_t(i) = p(y_t|s_t=i,\hat{\boldsymbol{\theta}}_k)\sum_{j=1}^{K^w}\alpha_{t-1}(j)\,p(s_t=i|s_{t-1}=j,\hat{\boldsymbol{\theta}}_k), \tag{73}$$

$$\beta_t(i) = \sum_{j=1}^{K^w}\beta_{t+1}(j)\,p(y_{t+1}|s_{t+1}=j,\hat{\boldsymbol{\theta}}_k)\,p(s_{t+1}=j|s_t=i,\hat{\boldsymbol{\theta}}_k). \tag{74}$$

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

$$\alpha_1(i) = p(y_1|s_1=i,\hat{\boldsymbol{\theta}}_k)\,p(s_1=i|\,\hat{\boldsymbol{\theta}}_k), \tag{75}$$

$$\beta_T(i) = 1, \tag{76}$$

where the first follows from the definition of $\alpha_t(i)$, and the second is obtained from ***Equation 69*** 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 69*** and ***Equation 70*** can be found by setting $t=T$ in ***Equation 69*** and summing over s_T, namely,

$$\left(\sum_{s_T=1}^{K^w} p(s_T|\boldsymbol{y},\hat{\boldsymbol{\theta}}_k)\right)p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k) \equiv p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k) = \sum_{s_T=1}^{K^w}\alpha_T(s_T). \tag{77}$$

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 67*** and ***Equation 68***.

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 64*** with respect to $\{\boldsymbol{\pi},\boldsymbol{\upsilon},\lambda,\boldsymbol{A}\}$, 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

$$\text{initial state probabilities: } \hat{\pi}_m = \frac{\sum_{i=1}^{K^w}\langle s_1^i\rangle C_{mi}}{\sum_{k=1}^{K}\sum_{i=1}^{K^w}\langle s_1^i\rangle C_{ki}}, \tag{78}$$

$$\text{fluorescence emission rates: } \hat{\boldsymbol{\upsilon}} = \boldsymbol{M}^{-1}\boldsymbol{b}, \text{ where} \tag{79}$$

$$M_{mn} = \sum_{t=1}^{T}\sum_{i=1}^{K^w}\langle s_t^i\rangle F_{in}F_{im}, \tag{80}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

$$b_m = \sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle y_t F_{im}, \tag{81}$$

$$\text{noise: } \frac{1}{\hat{\lambda}} = \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle (y_t - F_{i,:}\hat{\boldsymbol{v}})^2}{\sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle}, \tag{82}$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{t=1}^{T} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{mi} C_{nj}}{\sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i,j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{nj}}. \tag{83}$$

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, we obtain

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{s}_{1:N} | \boldsymbol{\theta}) = \prod_{n=1}^{N} p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}), \tag{84}$$

$$p(\boldsymbol{s}_n | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) = p(\boldsymbol{s}_n | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k), \quad 1 \le n \le N. \tag{85}$$

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

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) &= \sum_{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N} p(\boldsymbol{s}_{1:N} | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_{1:N}, \boldsymbol{s}_{1:N} | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \sum_{\boldsymbol{s}_n} p(\boldsymbol{s}_n | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \sum_{\boldsymbol{s}_n} p(\boldsymbol{s}_n | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k) \log p(\boldsymbol{y}_n, \boldsymbol{s}_n | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \mathcal{L}_n(\boldsymbol{\theta} | \boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k).
\end{aligned} \tag{86}$$

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 64***), we find

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} | \boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) = &\sum_{n=1}^{N} \sum_{i=1}^{K^w} \sum_{k=1}^{K} \langle s_1^i(n) \rangle C_{ki} \log \pi_k \\
&+ \frac{1}{2} \sum_{n=1}^{N} \sum_{t=1}^{T_n} \sum_{i=1}^{K^w} \langle s_t^i(n) \rangle \left(\log \lambda - \log(2\pi) - \lambda (y_t(n) - F_{i,:}\boldsymbol{v})^2 \right) \\
&+ \sum_{n=1}^{N} \sum_{t=1}^{T_n} \sum_{i,j=1}^{K^w} \sum_{k,l=1}^{K} B_{ij} \langle s_t^i(n) s_{t-1}^j(n) \rangle C_{ki} C_{lj} \log A_{kl},
\end{aligned} \tag{87}$$

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 step 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,

$$\text{initial state probabilities: } \hat{\pi}_m = \frac{\sum_{h=1}^{N} \sum_{i=1}^{K^w} \langle s_1^i(h) \rangle C_{mi}}{\sum_{k=1}^{K} \sum_{h=1}^{N} \sum_{i=1}^{K^w} \langle s_1^i(h) \rangle C_{ki}}, \tag{88}$$

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

fluorescence emission rates: $\hat{\boldsymbol{v}} = \boldsymbol{M}^{-1}\boldsymbol{b}$, where (89)

$$M_{mn} = \sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle F_{in}F_{im}, \tag{90}$$

$$b_m = \sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle y_t(h)F_{im}, \tag{91}$$

$$\text{noise: } \frac{1}{\hat{\lambda}} = \hat{\sigma}^2 = \frac{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle(y_t(h) - F_{i,:}\hat{\boldsymbol{v}})^2}{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i=1}^{K^w}\langle s_t^i(h)\rangle}, \tag{92}$$

$$\text{transition probabilities: } \hat{A}_{mn} = \frac{\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i,j=1}^{K^w} B_{ij}\langle s_t^i(h)s_{t-1}^j(h)\rangle C_{mi}C_{nj}}{\sum_{k=1}^{K}\sum_{h=1}^{N}\sum_{t=1}^{T_h}\sum_{i,j=1}^{K^w} B_{ij}\langle s_t^i(h)s_{t-1}^j(h)\rangle C_{ki}C_{nj}}. \tag{93}$$

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

$$f(\mathbf{x}) \sim \frac{\Gamma\left(\sum_{k=1}^{K} u_k\right)}{\prod_{k=1}^{K}\Gamma(u_k)}\prod_{k=1}^{K} x_k^{u_k-1}. \tag{94}$$

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 identical and independently distributed (i.i.d.) and use a simplified time-independent EM algorithm to find their optimal values (cf. ***Bishop*** (***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 (10-20 in our implementations), and the globally optimal set of values is chosen in the end. In the Matlab implementation of the EM algorithm, the variables are all stored in logarithmic forms to avoid overflow and underflow issues, which could occur when recursively evaluating the elements of the α and β matrices. Also, special care is taken when accounting for time points less than the elongation time, i.e. $t < w$, in which case the compound state is a collection of not w, but t promoter states, i.e. $s_t = \{z_t, z_{t-1}, ..., z_1\}$.

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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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 59*** 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,

$$\log p(\boldsymbol{y}_{T_1:T_2}, \boldsymbol{s}_{T_1:T_2}|\boldsymbol{\theta}) = \log p(s_{T_1}|\boldsymbol{\pi}^{(T_1-w+1)}, \boldsymbol{A}) + \sum_{t=T_1}^{T_2} \log p(y_t|s_t, \boldsymbol{r}, \sigma) + \sum_{t=T_1}^{T_2} \log p(s_t|s_{t-1}, \boldsymbol{A}), \tag{95}$$

$$\begin{aligned}\log p(s_{T_1}|\boldsymbol{\pi}^{(T_1-w+1)}, \boldsymbol{A}) &= \log \left(p(z_{T_1-w+1}|\boldsymbol{\pi}^{(T_1-w+1)}) \prod_{t=T_1-w+2}^{T_1} p(z_t|z_{t-1}, \boldsymbol{A}) \right) \\ &= \sum_{i=1}^{K^w} \sum_{n=1}^{K} s^i_{T_1} D^w_{ni} \log \pi_n^{(T_1-w+1)} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \sum_{k,l=1}^{K} s^i_{T_1} D^{d-1}_{ki} D^d_{li} \log A_{kl}.\end{aligned} \tag{96}$$

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 83*** into

$$\hat{A}_{mn} = \frac{\sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{mi} C_{nj} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{mi} D^d_{ni} \log A_{mn}}{\sum_{k=1}^{K} \sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{ki} C_{nj} + \sum_{k=1}^{K} \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{ki} D^d_{ni} \log A_{kn}}. \tag{97}$$

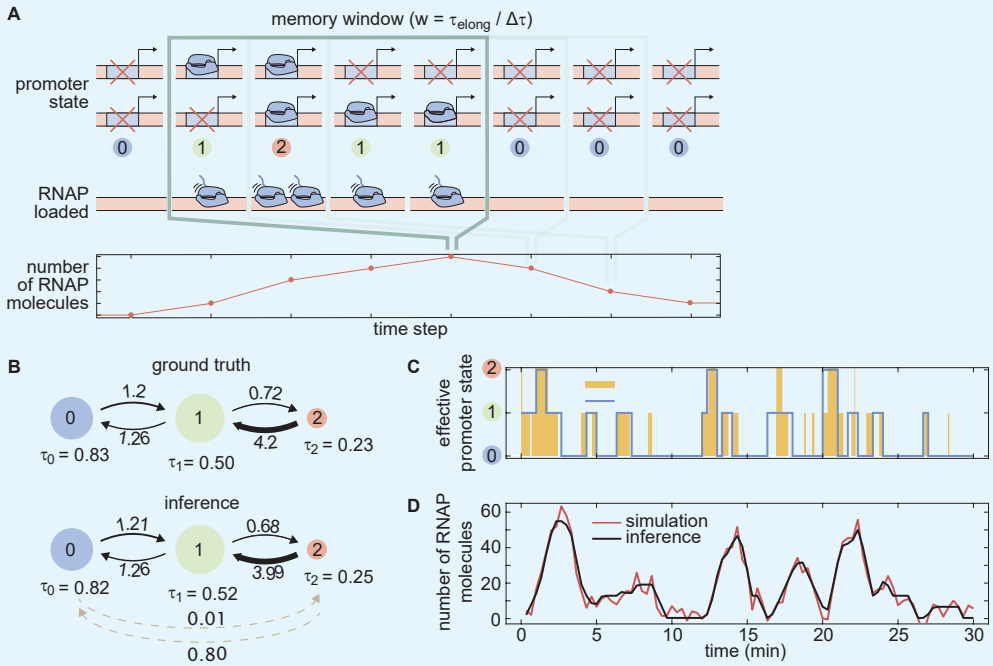
We make a steady-state assumption within the sliding window and choose $\boldsymbol{\pi}^{(T_1-w+1)}$ to be the stationary distribution of the current transition probability matrix, i.e. $\boldsymbol{A}\, \boldsymbol{\pi}^{(T_1-w+1)} = \boldsymbol{\pi}^{(T_1-w+1)}$. We therefore use the current estimate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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 4

Statistical validation of mHMM

We validated mHMM for the three-state ($K = 3$) architecture schematically illustrated in ***Appendix 4–Figure 1***A by generating synthetic trajectories of effective promoter states using the Gillespie algorithm (***Gillespie, 1976***) and adding Gaussian noise to the resulting activity traces. Parameters in ***Appendix 4–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. The top panel of ***Appendix 4–Figure 1***B shows the kinetic architecture used to simulate the promoter trajectory in ***Appendix 4–Figure 1***C (yellow) as it switches through the multiple possible states. This promoter trajectory leads to the simulated trace of the number of RNA polymerase molecules actively transcribing the gene in ***Appendix 4–Figure 1***D (red). Using mHMM, we found the best fitted path for our observable (***Appendix 4–Figure 1***D, black) and the corresponding most likely promoter state trajectory (***Appendix 4–Figure 1***C, blue).



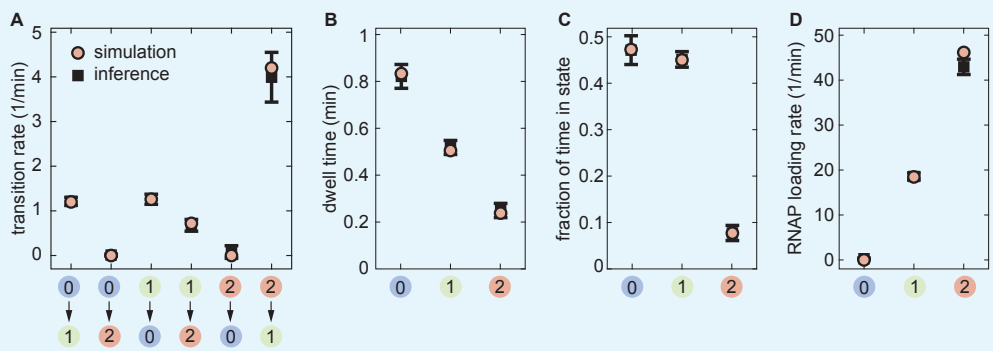
Appendix 4 Figure 1. Statistical validation of mHMM. (A) Three-state mHMM architecture where ON and OFF promoter states on each sister chromatid result in an effective three-state model. The trajectory of effective promoter states over the memory time window given by the elongation time dictates the number of RNA polymerase molecules loaded onto the gene. **(B)** Flow diagrams of promoter states and transition rates for the true parameters used to simulate trajectories (top) and corresponding average inference results obtained from 20 independent datasets (bottom). The area of each state circle is proportional to the relative state occupancy, and the thickness of the arrows is proportional to the transition rates. Dashed lines correspond to inferred transitions with very slow rates that were absent in the simulation. Rates are in min^{-1} and dwell times are in min. Error bars for the mean inferred parameters are shown in ***Appendix 4–Figure 2***. **(C)** Sample simulated promoter activity trace (yellow) generated using the parameters in (B), overlaid with the best fitted trace (blue) obtained using the Viterbi algorithm (***Viterbi, 1967***). **(D)** Simulated and best fitted observable number of RNA polymerase molecules corresponding to the promoter trajectory shown in (C).

Appendix 4 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 experimental data shown in ***Figure 5***.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

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

As shown in ***Appendix 4–Figure 1***B and ***Appendix 4–Figure 2***, comparison of the simulated and inferred parameters indicates that we reliably recovered the parameters used to generate our simulated data with high precision. We accurately inferred transition rates, dwell times, fraction of time spent in each state, and the rates of RNA polymerase loading over 20 independent datasets of simulated traces.



Appendix 4 Figure 2. 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)** RNA polymerase loading rates are compared. Statistics on the inferred values are obtained from 20 independently generated datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates).

Validation details

We used the relation between the transition rate matrix, $\mathbf{R}$, and the inferred transition probability matrix, $\mathbf{A}$, defined in ***Appendix 3*** to obtain estimates of the transition rates, namely,

$$\mathbf{A} = e^{\mathbf{R}\Delta\tau}, \tag{98}$$

$$R_{ij} = \left(\frac{1}{\Delta\tau}\log\mathbf{A}\right)_{ij}. \tag{99}$$

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.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

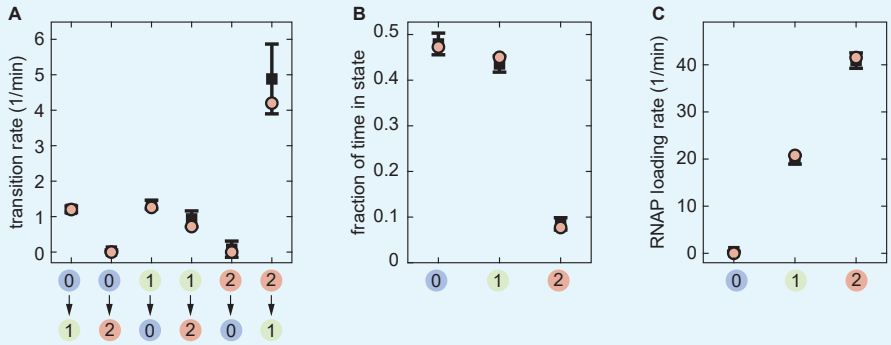
Continuous vs. Poisson promoter loading

To demonstrate the validity of our choice to use continuous RNA polymerase initiation rates in the transcription model (***Appendix 3***), we repurposed our simulation to, instead of considering a constant rate of RNA polymerase loading, explicitly account for individual RNA polymerase loading events when generating the traces. We assumed that individual polymerase molecules traverse at a constant elongation rate (v_{elong} = 46 bp/sec, ***Appendix 9***) and that their arrival to the promoter region has a Poisson waiting time distribution, provided that the promoter is cleared from the previous polymerase molecule which has a finite footprint size of $l_{\text{RNAP}} = 50$ bp (***Rice et al., 1993***). This led to a two-step model for the process of RNA polymerase initiation, with Poisson-distributed wait times for the recruitment of RNA polymerase to the promoter followed by a finite wait period as the RNA polymerase cleared the promoter—a process taken to be approximately deterministic. With this information in hand, we expressed the mean loading time of RNA polymerase at a single promoter (r_1^{-1}) as the sum of the mean time of polymerase arrival at an empty promoter, $\langle \tau_{\text{arrival}} \rangle$, and the time required to clear it after arrival, $\frac{l_{\text{RNAP}}}{v_{\text{elong}}}$, that is,

$$\frac{1}{r_1} = \langle \tau_{\text{arrival}} \rangle + \frac{l_{\text{RNAP}}}{v_{\text{elong}}}. \tag{100}$$

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

We performed inference on these simulated traces using mHMM with the objective of determining whether a Poisson loading rate had an effect on the obtained parameters. As shown in ***Appendix 4–Figure 3***, when the data is generated using Poisson RNA polymerase loading, mHMM slightly overestimates the high transition rate, but otherwise manages to accurately recover the model parameters. This therefore justifies our modeling approach of assigning continuous RNA polymerase initiation rates to each promoter state, instead of explicitly modeling the recruitment of individual polymerases.



Appendix 4 Figure 3. Validation of mHMM on Poisson RNA polymerase loading data. (A) Transition rates, **(B)** state occupancies and **(C)** RNA polymerase loading rates inferred from 15 independently generated datasets assuming Poisson loading of RNA polymerase. (Error bars represent one standard deviation calculated across these 15 independent replicates.)

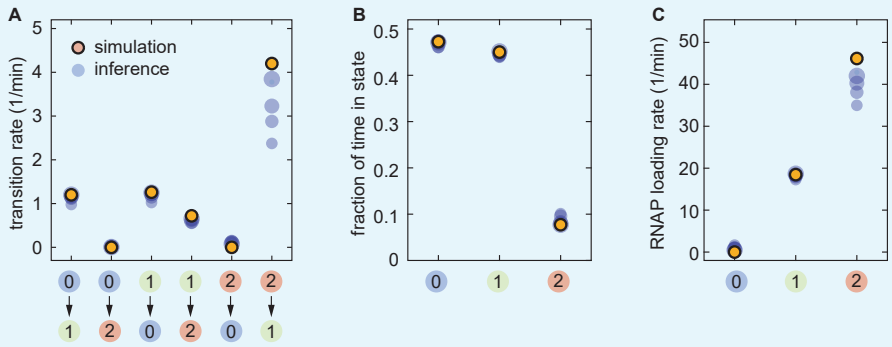
Sensitivity of mHMM to data sampling resolution

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 experimental time step ($\Delta\tau$), and that transitions to the next promoter state can occur only at the end of each step. This means that, if the fastest promoter switching rate is greater than the data sampling rate ($1/\Delta\tau$), our model might be unable to capture all those transitions. To study this possible limitation of

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

mHMM, we conducted inference on synthetic activity traces generated with varying sampling rates. Since the system memory ($w = \tau_{\text{elong}}/\Delta\tau$) needs to be an integer, we varied w in the $[3, 7]$ range, correspondingly changing the sampling resolution from low ($\tau_{\text{elong}}/3 \approx 46$s) to high ($\tau_{\text{elong}}/7 = 20$s). We used the values in ***Appendix 4–Table 1*** for the remaining model parameters.

Appendix 4–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 4–Figure 4***A). The accuracy of inferred state occupancies is also remarkably high, making it robust to variations in the data sampling rate (***Appendix 4–Figure 4***B). The high RNA polymerase loading rate tends to be underestimated for slower sampling resolutions, which is reasonable since the chances of promoter leaving state (2) during a single time step become greater, effectively reducing the net rate of loaded RNAP molecules per time step (***Appendix 4–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 are comparable.



Appendix 4 Figure 4. Sensitivity of mHMM to data sampling resolution. (A) Transition rates, **(B)** state occupancies and **(C)** RNA polymerase loading rates inferred from datasets generated with varying time resolutions. Transparent circles represent averages over 20 independently generated samples. The increasing size of the blue circles corresponds to higher data sampling resolutions (largest: 20s, smallest: 46s).

Performance of mHMM in different kinetic regimes

Thus far, the validation of mHMM was performed on datasets that were generated using parameters similar to those inferred for the *eve* promoter. These parameters have characteristic low ON rates (k_{01}, k_{12}) and a high OFF rate (k_{21}), where “low" and “high” are relative to the data sampling frequency, which for our experimental setup is 3/min. To assess the utility of our inference method for a generic choice of model parameters, we performed additional inference studies in three different parameter regimes: low ON rates and low OFF rates (***Appendix 4–Figure 5***A-C), high ON rates and low OFF rates (***Appendix 4–Figure 5***D-F), and high ON rates and high OFF rates (***Appendix 4–Figure 5***G-I).

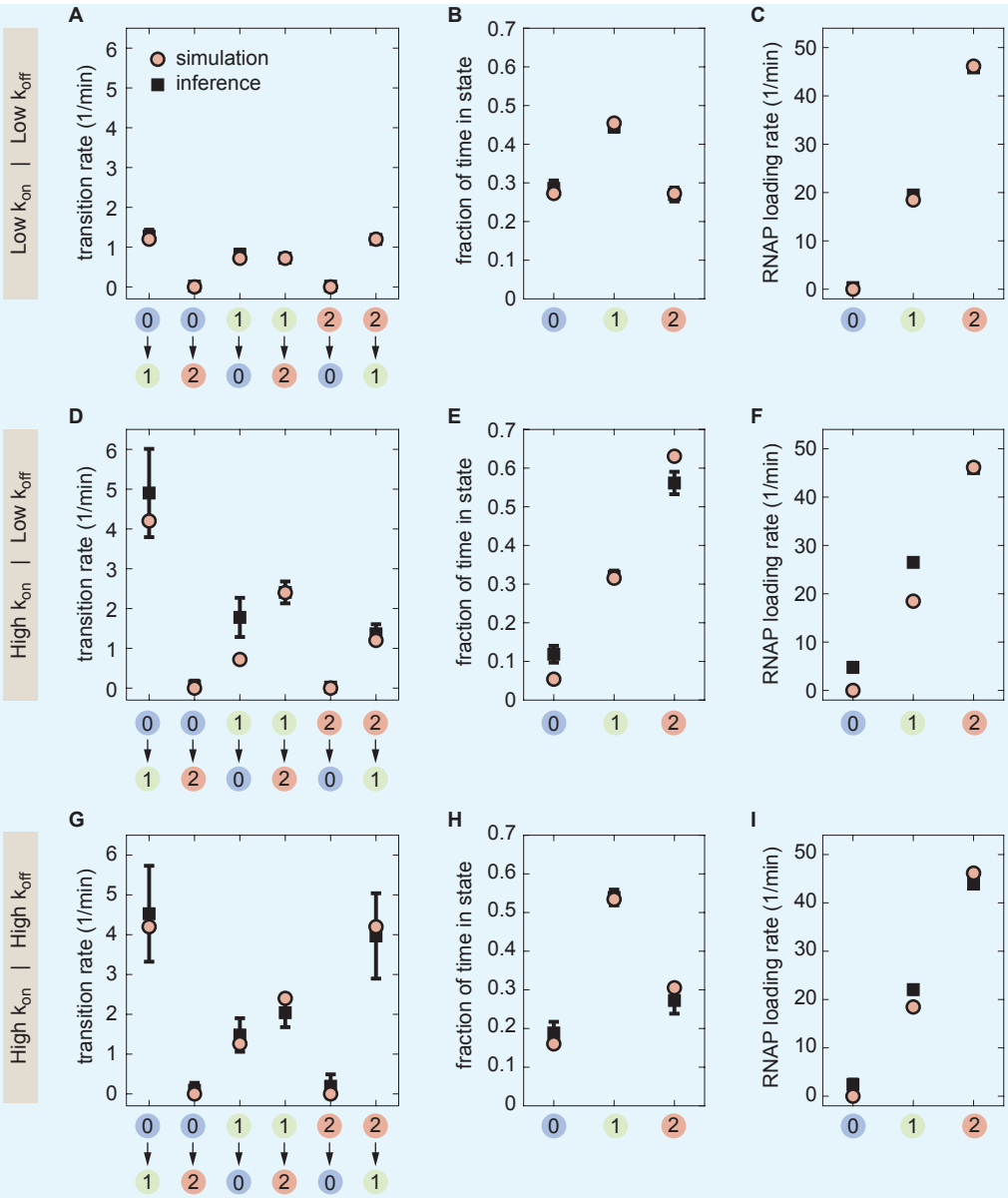
As expected, the inference is the most accurate when the data sampling frequency is greater than the transition rates (***Appendix 4–Figure 5***A-C), in which case multiple transitions within a single time frame occur only rarely, making our discrete Markovian representation of the state dynamics a valid approximation. The largest deviations of the inferred model parameters from their ground truth values occur when the ON rates are high and the OFF

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

rates are low (***Appendix 4–Figure 5***D-F). Since the promoter rarely remains in the lower initiation states (0 or 1) for the entire duration of a frame and tends to rapidly transfer into a higher initiation state (1 or 2, respectively), the rates of RNA polymerase loading for states 0 and 1 are significantly overestimated (***Appendix 4–Figure 5***F). Despite the inaccuracies in estimating the RNA polymerase loading rates, all transition rates, with the exception of k_{10}, are inferred with a high accuracy (***Appendix 4–Figure 5***D). Remarkably, the deviations caused by the high ON rates get substantially suppressed when the OFF rates are also made comparably high (***Appendix 3–Figure 5***G-I). This can be thought of as a consequence of an effective counterbalancing between unwanted ON and OFF transitions within a single time frame.

Overall, these additional studies, together with the statistical validation studies discussed earlier (***Appendix 4–Figure 2***), elucidate the domain of applicability of mHMM: the method performs accurate inference when the ON/OFF transition rates are respectively slow/slow, slow/high, or high/high; and is not successful in accurately inferring some of the model parameters when the ON rates are high, but the OFF rates are low. We hope that these characteristics of the method will be useful in informing the design of promoter architectures and new experiments.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.



Appendix 4 Figure 5. Study of mHMM performance for different choices of the ON/OFF transition rates Comparison of inference performance for different ON/OFF rates using a data sampling frequency of 3/min. **(A-C)** low/low, (**D-F**) high/low, (**G-I**) high/high. The statistics of inferred model parameter values is obtained from 20 independent datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates.)

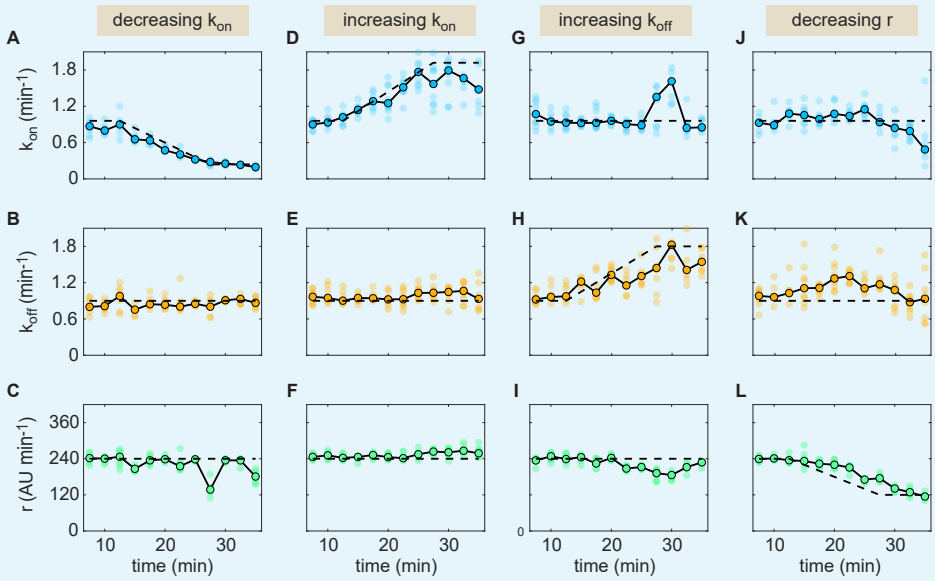
Windowed mHMM

To check that our windowed 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 4–Figure 6*** summarizes the results for these validation tests.

For each test, 100 simulated traces, 40 minutes in length, were generated ($\Delta\tau$ =20 s) that

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

exhibited the desired parameter trends. Consistent with our approach to the experimental data, a sliding window of 15 minutes was used for inference, meaning that for each inference time, τ_{inf}, all data points within 7.5 minutes of τ_{inf} were included in the inference. This led to inference groups consisting of 4500 data points, with the exception of the first and last time points, each of which had 3700 data points (first and last $w+1$ points were excluded from inference). Transition and initiation rates shown in ***Appendix 4–Figure 6*** are associated with state (1) of the three-state model ($k_{\mathrm{on}} = k_{01}/2$, $k_{\mathrm{off}} = k_{10}$ and $r = r_1$ in ***Appendix 5–Figure 2***A), as these were found to provide the most faithful indication of underlying system trends.



Appendix 4 Figure 6. Validation of 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 inferred parameter value at each time point across 10 distinct replicates. Translucent points indicate inference values from individual replicates. Thus, the dispersion of these replicates at a given time point indicates the precision of the inference.

For each scenario, we assessed whether and to what degree the windowed 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 4–Figure 6***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 4–Figure 6***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 4–Figure 6***J-L) we observe a ~ 5 min delay in the model response. We attribute this delay to the finite dwell time of RNA polymerase 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 4–Figure 6***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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

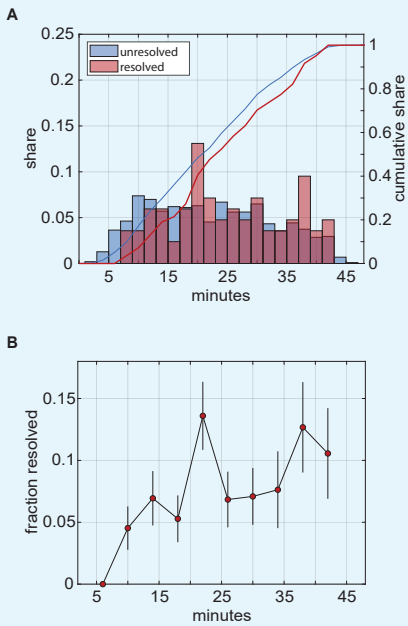
parameter trends the model is capable of inferring. Changes that occur on shorter time scales will be registered, but the temporal averaging introduced by the sliding window will lead to underestimates of the rate of the parameter changes in the underlying system.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 5

Sister chromatids

Detection of sister chromatid appearance



Appendix 5 Figure 1. Live imaging data indicate timing of sister chromatid appearance. (A) Distribution of observation times for frames in which chromatids were resolveable (red) and diffraction-limited (blue). Bars indicate emprical probability distribution function. Lines indicate cumulative density function. Data indicate the presence of chromatids by no later than 7-8 minutes into nuclear cycle 14. **(B)** Fraction of frames featuring resolved chromatids as a function of time. Trend suggests replication of relevant portion of genome across all observed nuclei is completed by approximately 10 minutes into nuclear cycle 14. Initial lag is likely attributable—at least in part—to stochastic turn-on times between sister *eve* loci and lower fluorescence levels early on in the nuclear cycle.

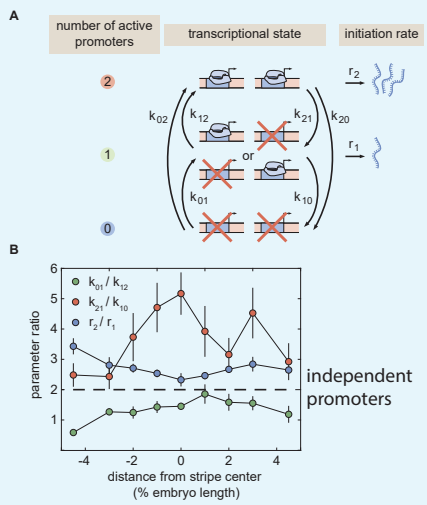
Previous studies have indicated that the *D. melanogaster* genome is quickly replicated at the beginning of each nuclear cycle in early development (***Rabinowitz, 1941***; ***Shermoen et al., 2010***) , suggesting that each diffraction-limited spot in our imaging data likely contains two distinct *eve* loci. We sought use our live imaging data to verify whether genome replication occurred early enough in the nuclear cycle such that the presence of the replicated promoters would have to be taken into account. While the two *eve* loci are located within a diffraction-limited spot for the majority of frames in our data, there are a subset of frames in which two distinct puncta can be clearly observed due to fluctuations in the separation between chromatids (see ***Figure 4***D). We reasoned that, by tracking the frequency of frames with resolved puncta over time, we could ascertain how the timing of genome replication compares to the onset of transcription. If replication precedes the onset of transcription, then the fraction of resolved frames should be relatively stable over for the duration of *eve* expression in nuclear cycle 14. If, on the other hand, replication happens after the onset of transcription, we should see a significant increase in the frequency of resolved sister chromatids over time as development progresses.

To pursue this question, we randomly selected snapshots of transcriptional loci in 100 different nuclei for each of the 11 embryos used in this study. We then determined the fraction of these sampled snapshots in which two distinct puncta were clearly visible by eye and observed how these instances of resolved chromatids were distributed in time. As indicated in ***Appendix 5–Figure 1***A, we see evidence for resolved puncta by around 7 minutes into nuclear cycle 14. This is well within the average range for turn-on times observed throughout the stripe (see ***Figure 3–Figure Supplement 2***B). Our results indicate that, at the very least, the genomic region containing our *eve* stripe 2 reporter is replicated within *some* nuclei by 6-8 minutes into nuclear cycle 14. ***Appendix 5–Figure 1***B tracks the share of total observations for which we detected resolved puncta as a function of time. A systematic delay in DNA replication would be expected to result in a progressive increase

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

in this metric over time. However, such a trend is not evident. While we see *no* resolved sister loci between 4 and 8 minutes (first point in the plot in ***Appendix 5–Figure 1***B), this absence could be attributed to other factors at play early on in nuclear cycle 14. For example, part of this apparent lag could be attributable to the fact that loci are, on average, dimmer early on in the nuclear cycle, which could mask the presence of two *eve* loci by reducing the probability of both producing observable amounts of fluorescence at the same time. It is also possible that the precise timing of locus turn-on varies for each sister locus, as it does for loci in different nuclei. Regardless, even if the initial rise between 6 and 10 minutes in ***Appendix 5–Figure 1***B *is* reflective of the replication of the locus during this period of time, the relative stability of the frequency of resolved loci from 10 minutes onward indicates that this process is restricted to the first few minutes of transcription. Additional experiments are needed to further elucidate the interplay between DNA replication and the onset of transcription. Regardless, the examination of our live imaging data supports the conclusion that the majority of our data consist of diffraction limited spots containing two distinct *eve* loci.

Probing for interactions between sister chromatids



Appendix 5 Figure 2. Probing combined transcription of sister chromatids. (A) Revised three-state model of promoter switching within a fluorescent punctum that accounts for the combined action of both sister chromatids. **(B)** Summary of bursting parameter ratios. All three bursting parameter ratios deviate from their expected values under the independence assumption given by the horizontal dashed line. (Error bars indicate magnitude of difference between first and third quartiles of mHMM inference results for bootstrap samples of experimental data over multiple embryos. See Materials and Methods for details)

If each fluorescent punctum contains two promoters (***Figure 4***D), then it is necessary to revisit the widely used two-state model of transcriptional bursting. In this revised scenario, each promoter on one of the sister chromatids undergoes fast ON/OFF switching. Therefore, each spot (encompassing two identical loci) can be in one of three states: (0) both promoters OFF, (1) one promoter ON and the other OFF, and (2) both promoters ON (***Appendix 5–Figure 2***B). States (1) and (2) are expected to exhibit different rates of RNA polymerase loading, r_1 and r_2, respectively. See ***Appendix 1*** and ***Appendix 3*** and for details regarding the implementation of this three-state model.

The presence of two transcriptional loci within each fluorescent punctum suggests three constraints on the relationship between bursting parameters in the model shown in ***Appendix 5–Figure 2***A. First, if these two promoters transcribe independently, then state (2) will have double the loading rate of state (1) such that $r_2 = 2r_1$. Second, the probability of both promoters transitioning simultaneously should be negligible; we expect no transitions between states (0) and (2) such that $k_{02} = k_{20} = 0$. Finally, if the promoters switch between their states in an independent manner, then there will be an extra constraint on their 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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

on. This condition sets the constraint $k_{01} = 2k_{12}$. Similarly, $k_{10} = k_{21}/2$.

While the independence of sister chromatids is supported by recent single-molecule FISH experiments (***Little et al., 2011***; ***Zoller et al., 2018***), classic electron microscopy work suggests a scenario in which sister chromatids are tightly correlated in their transcriptional activity (***McKnight and Miller, 1977***, ***1979***). Given this uncertainty regarding chromatid independence, we elected to employ a general three-state model that makes no assumptions about the nature and strength of sister chromatid interactions. In addition to permitting greater flexibility, this agnostic approach also meant that the structure of the kinetic model returned by 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). A deviation from these expectations would indicate either that the two sister loci do not initiate RNA polymerase independently (first constraint), or that they do not transition between activity states independently (second and third constraint).

Overall, our results suggest that the two loci are coupled to a nontrivial degree. We observe that the rate of initiation for the high state, $r_2(x)$, (corresponding to two active promoters) is consistently greater than twice the middle state, $r_1(x)$ (***Appendix 5–Figure 2***B, bliue). This trend suggests some sort of synergy in the RNA polymerase 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 5–Figure 2***C, red). This indicates that each promoter is more likely to switch off when its sister locus is also active. This anti-correlation is consistent with some form of competition between the loci, a scenario that could arise, for instance, if local concentrations of activating TFs are limiting. In addition, we observe substantial variation in the relationship between the high and low on rates (k_{01} and k_{12}, respectively), ranging from one of near equality in the anterior flank to nearly the 2-to-1 ratio that would be expected of independent loci in the stripe center and posterior (***Appendix 5–Figure 2***C, green). Finally, as shown in ***Appendix 6–Figure 1***, we observe no transitions between the (0) and (2) states, lending support to the hypothesis that, despite their correlation, our spots do contain two promoters.

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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

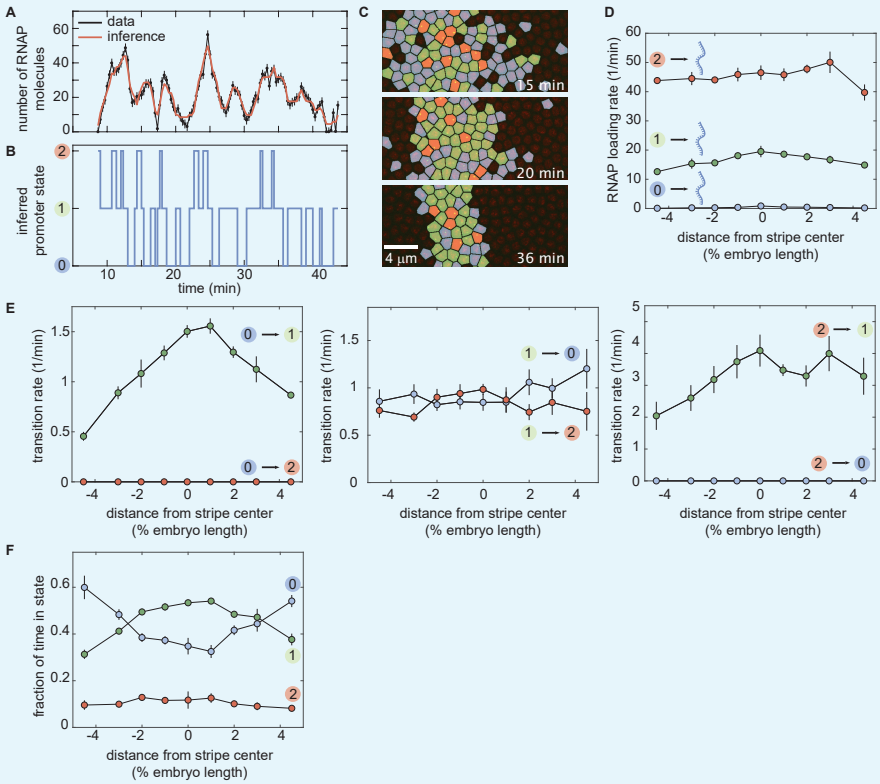
Appendix 6

mHMM inference sensitivities

Full three-state inference results

For the sake of simplicity, we presented our inference results in the main text using an effective two-state model in which two distinct active transcriptional states were combined into a single effective ON state (see ***Figure 4***E and F). Here, for completeness, we include time-averaged and time-resolved inference results for the full three-state model where, as shown in ***Appendix 5–Figure 2***, (0) corresponds to the state where both promoters are in the OFF state, (1) indicates the state where either promoter is in the ON state, and (2) represents the states where both promoters are in the ON state.

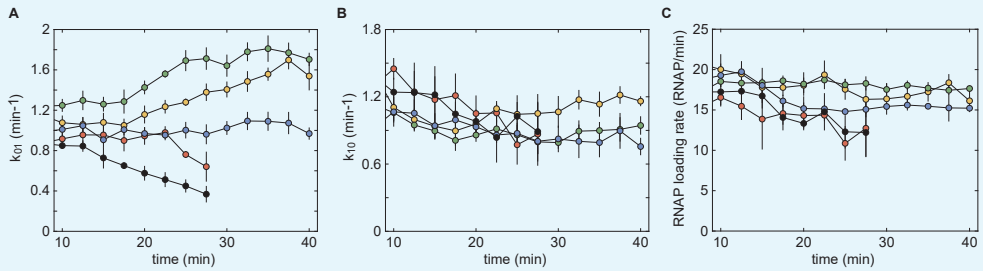
As indicated in the main text, the full three-state results (***Appendix 6–Figure 1***) exhibited the same trends as were evident in the effective two-state plots (***Figure 5***). In agreement with the effective two-state model, the rate of transcript initiation is not modulated to a significant degree across the stripe (***Appendix 6–Figure 1***D). Moreover, we once again see that activation rates, and specifically the rate of switching from OFF to the middle ON rate (states 0 and 1 in ***Appendix 6–Figure 1***E) are strongly elevated in the stripe center.



bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 6 Figure 1. Full three-state results for time-averaged mHMM inference. (A) Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (colors as in B). **(D)** The rate of initiation for each transcriptional state is not significantly modulated along the embryo. **(E)** Our mHMM revealed that the transition rate between the OFF (0) and middle ON state (1) is up-regulated in the stripe center. In contrast, the rates of switching out of the middle and high ON states show little to no significant AP-dependent modulation. (**F**) The modulation of the rate of switching from 0 to 1 acts to increase the fraction of time the promoter spends in the active states in the stripe center. (A, error bars obtained from estimation of background fluorescent fluctuations, as described in Materials and Methods and ***Garcia et al.*** (***2013***); D, E, and F, error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

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



Appendix 6 Figure 2. Full three-state results for time-dependent mHMM inference. (A) Transition rate from transcriptionally inactive state (0) to the first active state (1). Same trends evident as for effective 2 state model. **(B)** Transition rate from first on state (1) to OFF state (0). **(C)** Rate of transcript initiation in first on state (1) as a function of time. (Error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

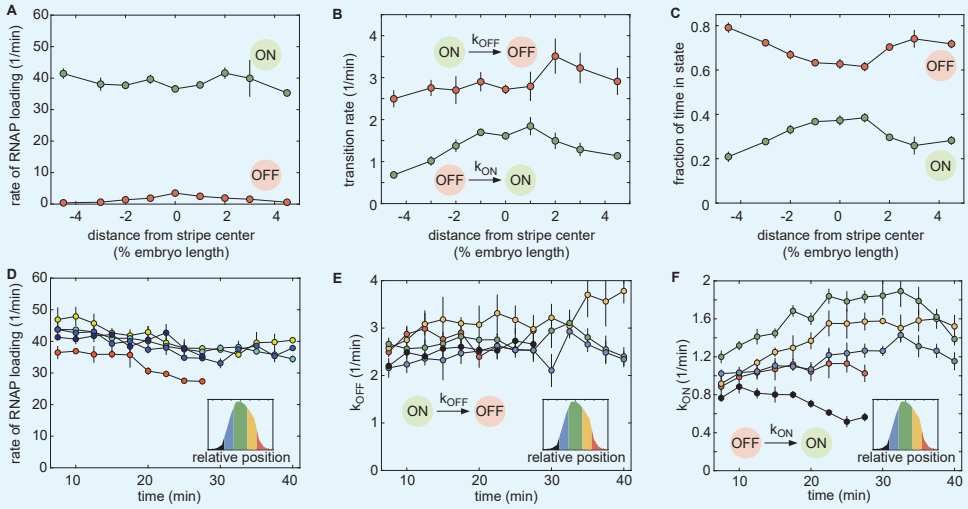
Two-state inference results

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

Most of our findings remained unchanged in the context of the two-state model. Consistent with the three-state case, the two-state time-averaged mHMM inference indicated that the fraction of time spent in an active state, rather than the rate of RNA polymerase initiation, drives the difference in mRNA production rates across the stripe (***Appendix 6–Figure 3***A-C). Moreover, as with the three-state case, two-state results indicated that the bulk of this variation stem from modulation in k_{on} (***Appendix 6–Figure 3***C, green). Interestingly, whereas we did see a degree of spatial dependence in k_{off} for 3-states, we observed no such trend for 2-states (***Appendix 6–Figure 3***C, red). In general, this is not surprising, as our use of a simpler model likely means that multiple switching rates are being projected onto the k_{off}

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

parameter. Specifically, if the *eve* stripe 2 system is indeed a true three-state system, then we would expect the two-state k_{off} estimate to reflect the joint action of the k_{10}, k_{21}, and k_{12} rates from the three-state model. As a result, the spatial dependence of each one of these rates would get averaged out when combined onto k_{off}.



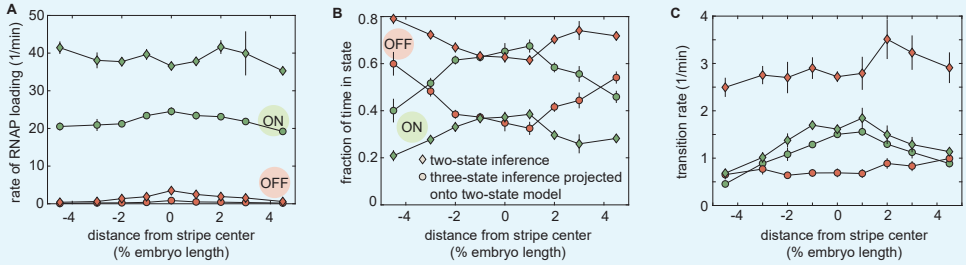
Appendix 6 Figure 3. Two-state mHMM inference. **(A-C)** Time-averaged 2-state inference results. (A) Consistent with three-state inference results, we observed no significant modulation in the rate of initiation along the axis of the embryo. Moreover, we found that k_{on} (green plot in (B)) was modulated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

As with the time-averaged case, we found that results for two-state windowed mHMM were generally consistent with three-state trends. A notable exception to this rule was the absence of any significant decrease in k_{on} in the posterior stripe flank (***Appendix 6 Figure 3***F, red). This is not entirely surprising, as the trend returned by the three-state inference was relatively mild (***Figure 6***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 three-state model allowed it to register a subtle shift in the activation rate that was convolved with countervailing features in the two-state case. Future work will seek to elucidate the source of this discrepancy and further test the validity of the trend uncovered in the three-state case.

Comparing true and effective two-state inference results

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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.



Appendix 6 Figure 4. Comparing two- and three-state mHMM inference results. Three-state inference results can be presented in terms of a two-state model in which states (1) and (2) are aggregated into a single ON state (see ***Figure 4***E and F). Here, color schemes are consistent with those employed in ***Appendix 6–Figure 3***A-C. Squares indicate true two-state results (presented in the previous section) and circles indicate effective two-state trends derived from the three-state results presented in ***Figure 5***. **(A)** Anterior-posterior-dependent trends in the rate of RNA polymerase initiation are nearly identical between the true and effective initiation rates, however the initiation rate returned by two-state mHMM inference (green squares) is roughly twice as large as that implied by the three state results (green circles). **(B)** As with the initiation rates, we observe similar trends between the true and effective cases, but substantial differences in magnitude. The effective two-state model recovers an ON state occupancy that is roughly double that returned by two state mHMM inference. **(C)** While the ON rate trends and magnitudes are nearly identical, the OFF rate returned by two-state mHMM inference is roughly triple that implied by three-state inference. Thus it is clear that this difference in OFF rate underlies the observed departures in both state occupancies (B) and state initiation rates (A). (Error bars indicate magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

As ***Appendix 6–Figure 4*** makes clear, while anterior-posterior-dependent parameter trends are by and large consistent between the true and effective two state models, we do observe substantial differences in the absolute magnitudes of parameter values. These differences originate (directly or indirectly) from the three-fold difference in the value of k_{off} between the true and effective models (***Appendix 6–Figure 4***C, red squares and circles, respectively). The k_{off} value for the effective two-state model is defined as

$$k_{\text{off}} = \frac{k_{10}k_{21}}{k_{21} + k_{12}}. \tag{101}$$

See ***Appendix 1*** for expressions for all three effective two-state bursting parameters (k_{on}, k_{off}, and r) in terms of these three-state transition rates. This value represents the inverse of the mean amount of time the system, upon switching out of state (0), spends in one of the active states before returning to (0), and we can see that it is necessarily less than or equal to k_{10}.

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

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

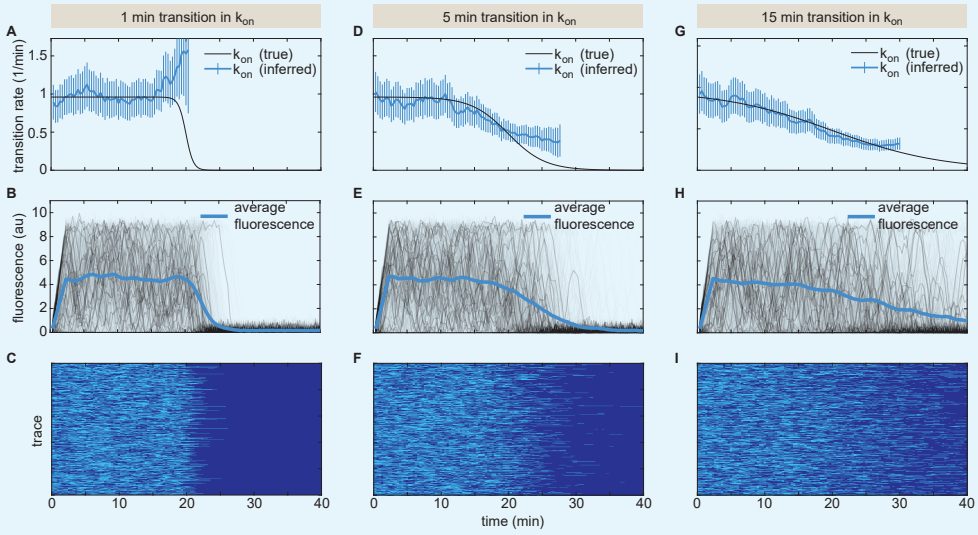
bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 7

Inherent limits of bursting parameter inference

By definition, the onset of transcriptional quiescence coincides with the cessation of observable bursting activity. In the main text, we argue that this cessation appears to be driven by processes that are mechanistically distinct from those driving transcriptional bursting. It remains possible, however, that quiescence is instead driven by changes in the bursting machinery itself as illustrated in scenario (ii) in ***Figure 6***A. If this is the case, it is important to note that fundamental limits exist to the time-scale of shifts in the bursting parameters that could be detected in any sort of time-dependent burst parameter inference (see, *e.g.*, ***Figure 6***): changes of order-with or faster-than the time scale on which transcriptional bursts occur (1-3 min for *eve* stripe 2) cannot be detected. Notably, this is not a limit of the mHMM method, but rather reflects an inherent limitation set by system itself—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 a resolution limit for any kind of bursting parameter inference.

To illustrate this limitation, we simulated three scenarios for a two-state transcriptional system in which k_{on} decreases to 0 s^{-1} over periods 15, 5, and 1 min in length. We then sought to recover the trend in k_{on}. To emphasize that the limitations are not specific to mHMM, we used the *true* promoter trajectories generated by our simulation algorithm 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. The results indicate that, as expected, a transition in k_{on} that happens in the span of 1 minute is not detectable from a burst inference perspective (***Appendix 7–Figure 1***A-C). This indicates that, at this timescale, a shift in burst parameters (scenario ii in ***Figure 6***A) would be indistinguishable from an abrupt, change in which the promoter entered a silent state outside of those considered by the bursting model (scenario (i) in ***Figure 6***A). Interestingly, results for 5 and 15 minute k_{on} transitions (***Appendix 7–Figure 1***D-I) also indicate that even transitions that occur over longer periods of time cannot be fully recovered due to the fact that bursting behavior is observed over a limited window of time (40 minutes in our case). Thus, once the burst frequency decreases to a sufficiently low level, there simply are not enough bursts observed within the window of observation to estimate the burst frequency from the data.



bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 7 Figure 1. Inherent limits of bursting parameter inference. (A-C) Simulating a 1 min transition in k_{on}. **(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 faster than the bursting timescale, it is not 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 fast 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 showing how the k_{on} transition occurs on the timescale of a single burst. Light blue indicates ON periods and dark blue indicates OFF periods. Since there are almost no active traces after the transition of k_{on} to perform an inference. it would be impossible to determine if a modulation in the bursting parameters—as opposed to a transition into some alternative, silent state—drives the onset of quiescence. **(D-F)** Simulation of a 5 min transition in k_{on}. **(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)** Because the transition happens slower than in the 1 min case shown in (A-C), there are some bursts that unfold during the transition and, hence, we have some reference points with which to infer the underlying trend. **(G-I)** Simulating a 15 min transition in k_{on}. **(G)** The mHMM can reliably infer the temporal variation in k_{on}. **(H,I)** The observation that bursts of activity are interspersed throughout the k_{on} transition makes it possible to recover the temporal trend. (A,D,G, error bars indicate 95% confidence interval of exponential fits used to estimate k_{on}).

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 8

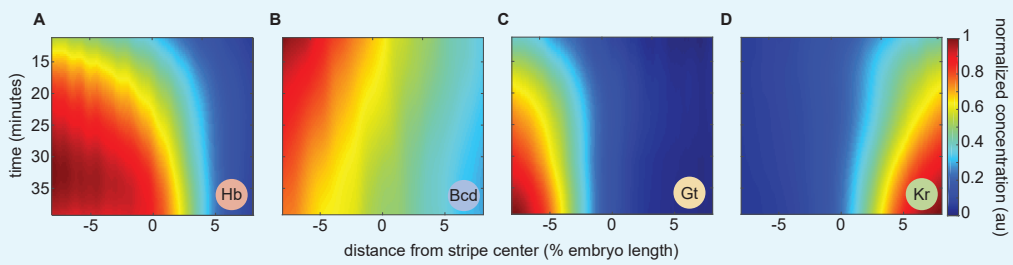
Input-Output analysis details

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

Input transcription factor data

Data sources

The input-output analysis presented in the main text made use of previously published data sets for the spatiotemporal concentration profiles of the gap genes Hunchback, Krüppel, and Giant (***Appendix 8–Figure 1***A, C and D). These data derive from elegant experiments in which individual embryos were co-immunostained for transcription factors of interest and precisely staged by measuring progressive cellularization over the course of nuclear cycle 14 to generate a time series of protein concentration profiles spanning the course of this period of development (***Dubuis et al., 2013***). The Bicoid concentration data used for this analysis derives from live imaging experiments using a Bicoid-GFP fusion established by ***Gregor et al.*** (***2007***). These data come courtesy of Jonathan Liu and Elizabeth Eck (***Appendix 8–Figure 1***B).



Appendix 8 Figure 1. Spatiotemporal transcription factor concentration maps. Heatmaps indicate normalized concentration profiles for the *eve* stripe 2 regulators **(A)** Hunchback, **(B)** Bicoid, **(C)** Giant, and **(D)** Krüppel as a function of space and time. In each case, levels were normalized relative to the maximum concentration observed within the spatiotemporal window of interest.

Data processing

To prepare the Krüppel, Giant, and Hunchback profiles for use in our logistic regression analysis, we adopted an approach similar to that described in ***Dubuis et al.*** (***2013***). Dorso-ventral orientation of embryos was found to have negligible effect on calculated intensity profiles and was ignored (i.e. all embryos were included, regardless of orientation). For each time point in nuclear cycle 14, a weighted temporal average was calculated using a sliding Gaussian kernel with $\sigma_t = 5$ min. For each time point, the minimum observed value across all anterior-posterior positions was then calculated and subtracted in order to remove background fluorescence. Normalized profiles were then calculated using the formula

$$I_{\text{norm}} = \frac{I_{\text{raw}}}{\max(I_{\text{raw}}) - \min(I_{\text{raw}})} \tag{102}$$

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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Finally, we discovered that the anterior-posterior axes in our live imaging data (both for *eve* stripe 2 and Bicoid-GFP) were inconsistent with the axes employed by the fixed data reported by the authors in ***Dubuis et al.*** (***2013***). We addressed this issue by using *eve* stripe 2 as a fiduciary mark to register the positions of the fixed and live data sets. Specifically, we aligned the mRNA peak predicted by our model at 40 minutes into nuclear cycle 14 with the peak in second stripe of the *eve* protein profile at 40 minutes, as reported in ***Petkova et al.*** (***2019***).

Logistic regression framework

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

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

Inference details: transcriptional time window

For the time window input-output analysis, we considered only loci that were transcriptionally active for one or more time steps in nuclear cycle 14. Loci were classified as transcriptionally active for all time points between the first and last time points for which they exhibited detectable levels of transcriptional activity and silent for all time points following their final shut-off for which their nuclei were still present in the experimental field of view. Time points preceding the onset of activity were discarded. ***Appendix 8–Figure 2***A illustrates how this quantity varies over space and time in our experimental data. We considered a class of logistic regression models in which each transcription factor was permitted to appear at most once, thus requiring that each factor act on *eve2* in a uniform manner through space and time; i.e., the same protein could not activate expression on one stripe flank and repress on the other.

Inference details: transcriptional bursting

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

Results of unconstrained inference

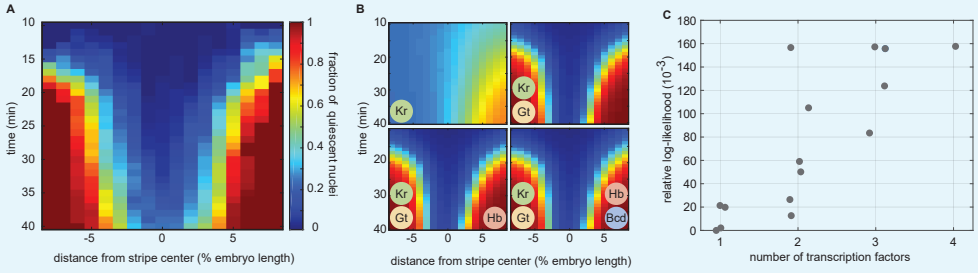
For the input-output inference results presented in the main text (***Figure 7***), we used prior knowledge about the regulatory function of each input transcription factor to constrain its range of permissible values in our inference. Specifically, we constrained the activators Bicoid and Hunchback to play activating roles in our model and, likewise, required that the

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

repressors Krüppel and Giant played repressing roles. In several cases, this constrained inference led to models in which one or more transcription factors played no significant regulatory role (Bicoid and Hunchback for the time window and Bicoid for transcriptional bursting). In this section, we tested the sensitivity of the conclusions presented in the main text to our use of functional constraints by conducting unconstrained input-output inference runs.

Transcriptional time window

The results of our unconstrained input-output inference for the transcriptional time window are identical to those presented in the main text. Despite the fact that no limitations were imposed on the regulatory function of each factor, we nonetheless recovered a model in which the two repressors, Giant and Krüppel, are necessary and sufficient to explain the onset of transcriptional quiescence in the stripe flanks. In agreement with the constrained case, we found that the addition of Hunchback and Bicoid to this two-repressor model had no qualitative effect on the output profile predicted by the model (***Appendix 8–Figure 2***B). A quantitative comparison of model fit scores confirmed that the addition of Hunchback and Bicoid did nothing to improve model fit (***Appendix 8–Figure 2***C). Thus, we conclude that our finding that the transcriptional time window can be explained entirely by the joint repressive action of Krüppel and Giant is insensitive to our choice to impose functional constraints.



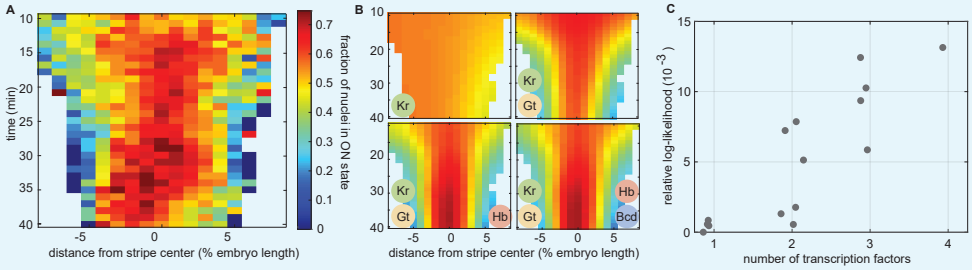
Appendix 8 Figure 2. Unconstrained inference results for the transcriptional time window. (A) Observed fraction of quiescent nuclei as a function of space and time. Identical data to that presented in ***Figure 7***A. **(B)** Relaxing constraints on the functional nature of each transcription factor had no appreciable effect on the inference results. Profiles shown here are indistinguishable from those shown in ***Figure 7***D. Once again, we find that the joint action of the repressors Giant and Krüppel is sufficient to explain the progressive onset of transcriptional quiescence in the stripe flanks. **(C)** A quantitative comparison of model fits reinforces the qualitative conclusions drawn from (B). Models including 3 and 4 transcription factors cannot improve on the fit achieved by the simpler double repressor model. Here blue dots indicate models for which only Giant and Krüppel make significant contributions to the model fit. This indicates that, while the 3 and 4 transcription factor models include additional parameters, these do not contribute appreciably to overall model fit, emphasizing the fact that these models behave, effectively, as double repressor models.

Transcriptional bursting

In the context of the transcriptional bursting input-output analysis, the removal of functional constraints led to a significantly more complex landscape of inferred regulatory models. While the functional roles of Krüppel, Giant, and Hunchback were consistent with the constrained case (repressing, repressing, and activating, respectively), Bicoid was consistently inferred to play a repressing role. Despite this complication, the three-factor Krüppel-Giant-Hunchback model favored by the constrained inference remained the best-fitting three-factor model (***Appendix 8–Figure 3***C, red circle). While the addition of Bicoid as a repressor to create a model dependent on all four input transcription factors led to a small improvement in model fit (***Appendix 8–Figure 3***C), comparison of this four-factor model's predicted activity profile with that of the Krüppel-Giant-Hunchback model revealed no material improvement

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

in the model's agreement with the experimental data (***Appendix 8–Figure 3***B, bottom left vs. bottom right). Moreover, there is (to our knowledge) no experimental evidence for Bicoid playing a repressive role in the regulation of *eve* stripe 2. Indeed, there is strong evidence that Bicoid is necessary for *eve* stripe 2 activity (***Small et al., 1992***). We thus conclude the Krüppel-Giant-Hunchback model remains the most plausible option in the unconstrained case.



Appendix 8 Figure 3. Unconstrained inference results for transcriptional bursting. (A) Observed fraction of transcriptionally active nuclei in the ON (bursting) state. Identical data to that presented in ***Figure 7***B. **(B)** As with time window, relaxing the constraints on the functional nature of each transcription factor did little to alter the inference results presented in the main text (compare to ***Figure 7***E). As with the constrained results, the joint action of Giant, Krüppel, and Hunchback appears sufficient to explain the spatiotemporal activity pattern revealed by mHMM inference. **(C)** A quantitative comparison of model fits.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

Appendix 9

Determining the RNA polymerase dwell time using autocorrelation

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

$$w = \frac{\tau_{\text{elong}}}{\Delta\tau}, \tag{103}$$

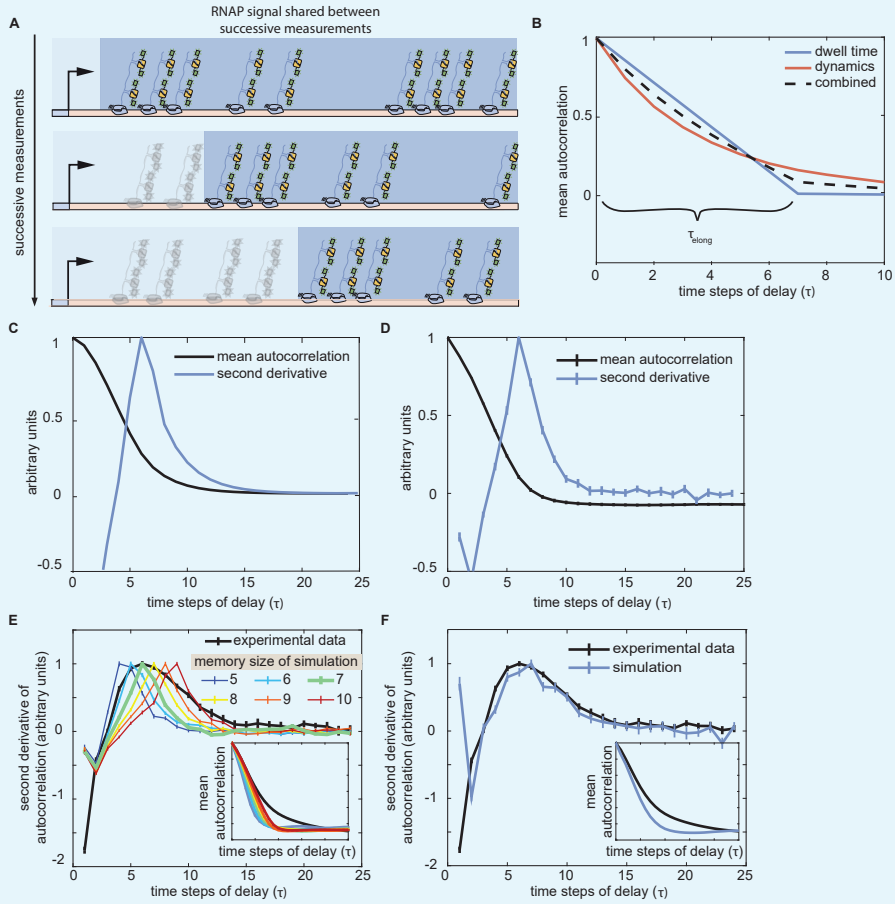
where $\Delta\tau$ is set by the temporal resolution of our data acquisition and τ_{elong} is the elongation time which is unknown *a priori*. Past studies have estimated elongation rates for other systems involved in early patterning in the *Drosophila* embryo, but there is substantial disparity between the reported values. A live imaging study of transcriptional activity driven by the *hunchback* P2 enhancer reported an elongation rate of $1.4 - 1.7$ kb min^{-1} (***Garcia et al., 2013***). However, a recent study of the same 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 RNA polymerase elongation rates measured for other systems might not apply to our *eve* stripe 2 reporter. Thus, in order to ensure the validity of our inference, we developed an approach that uses the mean autocorrelation function of experimental fluorescence traces to estimate the elongation time directly from our data.

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

$$R_F(\tau) = \frac{E[(F(t)-\mu_{\text{f}})(F(t-\tau)-\mu_{\text{f}})]}{\sigma_{\text{f}}^2}, \tag{104}$$

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 9–Figure 1***A, the fact that it takes RNA polymerase molecules some finite amount of time to traverse the gene implies that the observed fluorescence at a transcriptional locus at some time t, $F(t)$, will be correlated with preceding fluorescence values $F(t-\tau)$ so long as $\tau < \tau_{\text{elong}}$ because the two time points will share a subset of the same elongating RNA polymerase molecules. As τ increases, the correlation between $F(t)$ and $F(t-\tau)$ due to these shared RNA polymerase molecules will decay in a linear fashion (since the average number of shared RNA polymerase molecules decreases linearly with τ) until it reaches zero when $\tau = \tau_{\text{elong}}$ (***Appendix 9–Figure 1***B, blue curve).

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.



Appendix 9 Figure 1. Using the autocorrelation of the fluorescence signal to estimate RNA polymerase dwell time. (A) It takes a finite amount of time for RNA polymerase molecules to transcribe the full length of the reporter gene. As a result, successive fluorescence measurements will contain some of the same GFP-tagged RNA polymerase molecules. Dark blue-shaded regions indicate the subset of RNA polymerase 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 RNA polymerase elongation dynamics. For instance, successive time points also exhibit correlation due to the promoter switching dynamics (red curve). **(C)** Theoretical analysis of the autocorrelation function and **(D)** stochastic simulations indicate that the second derivative of the mean autocorrelation function (dark blue curves) can be used to find the structural break in the function (black curves) that corresponds to τ_{elong}. Here, a peak at 6 time steps of delay indicates an elongation time of 7 times steps (140 s). **(E)** Simulated traces with elongation time of 7 time steps (green curve) exhibit a peak in the second derivative that coincides with the maximum of the experimental curve. Inset plots show corresponding mean autocorrelation curves for experimental data and simulations. **(F)** Stochastic simulations in which we allow for variation in elongation times distributed around a mean of 7 time steps qualitatively recapitulates the observed curve. (C-F, second derivative profiles depicted here are normalized relative to their maximum value for ease of depiction.)

The dramatic change in the slope of the autocorrelation function that occurs at $\tau = \tau_{\text{elong}}$ can be used to estimate the elongation time of the system; however, it is not the only feature present in ***Equation 104***. Because the time series of promoter states constitutes a Markov chain, the instantaneous promoter state and, therefore, the instantaneous rate of RNA polymerase 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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

correlated with itself. Thus, we see that the rates of promoter switching dictate the speed with which this "dynamic" autocorrelation decreases with increasing τ. More precisely, the dynamics autocorrelation will take the form of a decaying exponential in τ, with the time scale set, approximately, by the second largest eigenvalue of the Markov chain's transition rate matrix (***Appendix 9–Figure 1***B, red curve)

$$R_P(\tau) \sim e^{-\lambda_2 \tau}. \tag{105}$$

Where λ_2 denotes the second larrgest eigenvalue of the transtion rate matrix. Thus, the observed autocorrelation function contains, at a minimum, signatures of both the finite RNA polymerase dwell time (τ_{elong}) and of promoter switching dynamics. As a result, inferring elongation times from the change in slope in the mean autocorrelation is often relatively subtle in practice.

A theoretical analysis of $R_F(\tau)$ indicated that the second derivative of the mean autocorrelation function reliably exhibits a peak that can be use to read out the value of τ_{elong} (***Medin et al., 2019***). ***Appendix 9–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 9–Figure 1***D). These simulated traces included the expected contributions from both the Markov dynamics (red profile in ***Appendix 9–Figure 1***B) and the finite RNA polymerase dwell time (blue profile in ***Appendix 9–Figure 1***C). 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 9–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 $w = 7$ ($\tau_{\text{elong}} = 140\,\text{s}$) best characterized our data (***Appendix 9–Figure 1***E, green curve). This implies that

$$\begin{aligned} v_{\text{elong}} &= \frac{6444\,\text{bp}}{140\,\text{s}} \\ &= 46\,\text{bp}\,\text{s}^{-1} \\ &= 2.8\,\text{kb}\,\text{min}^{-1}, \end{aligned} \tag{106}$$

where the length used represents the distance from the start of the MS2 step loop sequence to the end of the 3′ end of the construct. Interestingly, this elongation rate falls within the $2.4 - 3.0\,\text{kb}\,\text{min}^{-1}$ range reported in ***Fukaya et al.*** (***2017***).

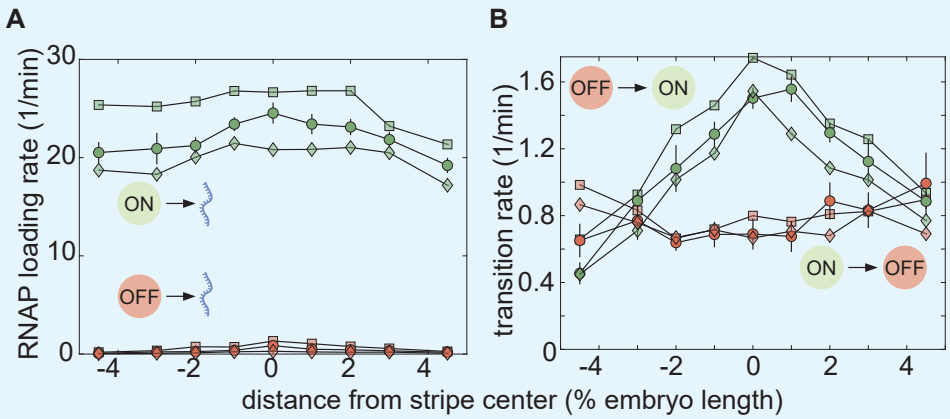
Appendix 9–Figure 1F shows how a simple adjustment to our simulation approach, wherein the elongation time steps w for individual RNA polymerase molecules were drawn from a Gaussian distribution with mean $\mu_w = 7$ and standard deviation $\sigma_w = 2.5$ time steps can qualitatively reproduce the wider profile observed in experimental data, indicating that our observations are indeed consistent with the presence of variability in RNA polymerase elongation times. Additional experimental and theoretical work will be necessary to uncover the biological source of this variability.

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

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

mHMM inference is insensitive to small changes in RNA polymerase dwell time

Due to the uncertainty in our estimate of τ_{elong}, we conducted sensitivity studies to ensure that our inference results were robust to our input assumption for w. As shown in ***Appendix 9–Figure 2***, we conducted time-averaged 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 9–Figure 1***E). While small quantitative difference are apparent across these three cases, the results for different values of w generally showed a constant offset throughout the embryo , such that qualitative trends were largely robust to the assumed w value.



Appendix 9 Figure 2. Elongation time sensitivities. Square, circle, and diamond symbols denote inference results for memory time window values w of 6, 7, and 8 (τ_{elong} of 120s, 140s, and 160s), respectively. Parameter plots for $w = 7$ case are bolded. Bootstrap errors are shown for $w = 7$ case (error magnitudes are comparable across conditions). **(A)** Although the absolute magnitude of the inferred effective initiation rate varies by approximately 10 to 25% across the three conditions, we found that the AP trends (or lack thereof) are robust to our choice of memory. **(B)** Transition rates also exhibit a high degree of robustness to the w used for inference. While we observed moderate variation in the inferred magnitude of k_{on} (green markers), AP trends are insenstive to w assumed for inference within the range tested. Very little variation was observed in k_{off} (red markers) across conditions. (Error bars indicate magnitude of difference between first and third quartiles of mHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

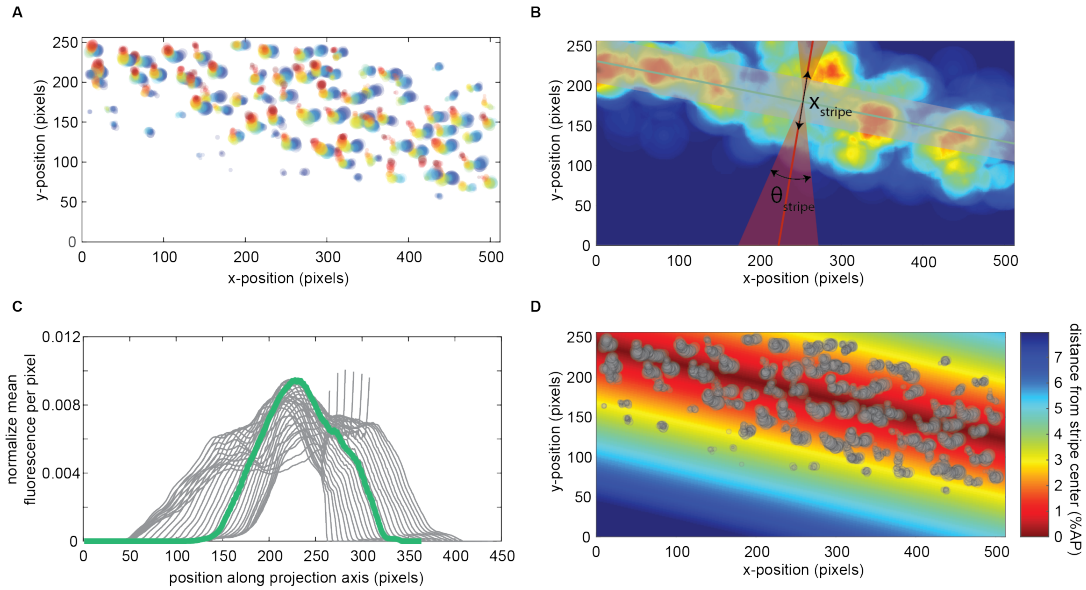


Figure 2–Figure supplement 1. Aligning stripes from multiple embryos. In order to minimize alignment errors when combining data from across multiple *Drosophila* embryos, an automated routine was employed to define a new experimental axis for each data set based upon the spatial distribution of transcriptional activity in the mature *eve* stripe 2 pattern. **(A)** Example of the spatiotemporal distribution of observed fluorescence for an individual embryo. Each circle corresponds to the fluorescence from a single locus at a single point in time. Only observations after 30 min into nuclear cycle 14 were used. Circle size indicates fluorescence intensity. Color indicates temporal ordering: 30 min (blue) to 47 min (red). **(B)** A Gaussian filter was convolved with the raw data points in (A). This filtering ameliorated stripe fitting artifacts that arose due to the relative sparsity of the raw data. The fitting procedure considered both a range of possible stripe orientations (θ_{stripe}) and, within each orientation, a range of possible positions of the stripe along the anterior-posterior axis (x_{stripe}) that, together constituted a set of possibilities for the new stripe center position and orientation. Here, the shaded red region indicates the range of values for θ_{stripe} that were considered. The red line indicates the best stripe axis inferred by the algorithm and the green line indicates the corresponding optimal stripe center. No constraints were placed on x_{stripe}, save for the limits of the experimental field of view. **(C)** For each proposed stripe orientation (θ_{stripe}), a projected stripe profile was generated by taking the average pixel intensity for each position, x_i, along the proposed stripe axis. To determine the optimal center location for each orientation, a sliding window with a width equal to 4% of the embryo length was used to determine the fraction of the total profile fluorescence that fell within 2% embryo length of the stripe center. For example, the gray shaded region in (B) illustrates what this range would be for the green stripe center line (B). This fraction of the total profile was used as a baseline for the comparison of potential stripe center positions. The θ_{stripe} and x_{stripe} that maximized this metric (green profile in (C)) were taken to define a new, empirically determined stripe center. **(D)** This inferred stripe position defined an experimental axis for each embryo that was used to aggregate observations from across embryos. Gray circles indicate experimental observations (size corresponds to intensity as in (A)) and shading indicates distance from inferred stripe center.

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

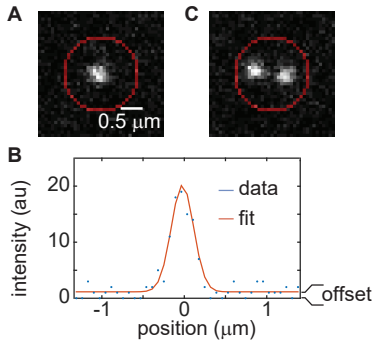


Figure 2–Figure supplement 2. Integrating MS2 Spots. (A) Fluorescence of a site of nascent transcript formation is measured by integrating raw pixel intensities in a circular region around the fluroescent MS2 spot of a predefined area (indicated by the red circle) and then subtracting off the background intensity obtained as outlined in (B). **(B)** X-Z projection of 2D Gaussian function fitted to MS2 spot shown in (A). Background intensity is estimated using the offset value fo this Gaussian fit. The per-pixel offset is then multiplied by the area of the integration region. This background value is then subtracted from the fluorescence integrated across the area shown in (A). **(C)** The radius was chosen to be large enough to integrate the intensities from both sister chromatids, even when they are spatially separated and distinguishable .

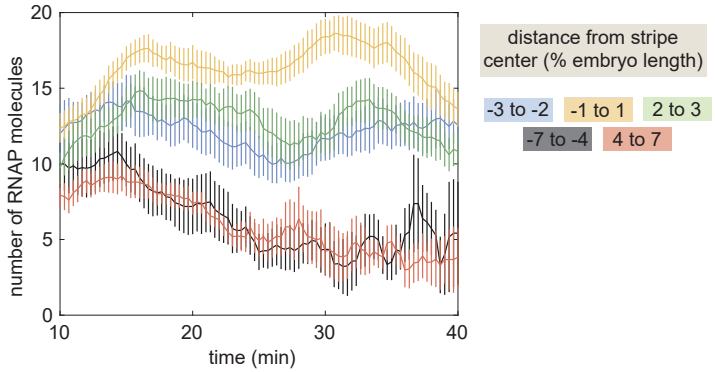


Figure 3–Figure supplement 1. Mean transcriptional activity. Mean transcriptional activity as a function of time for different positions along the stripe. (Average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-4.0 International license.

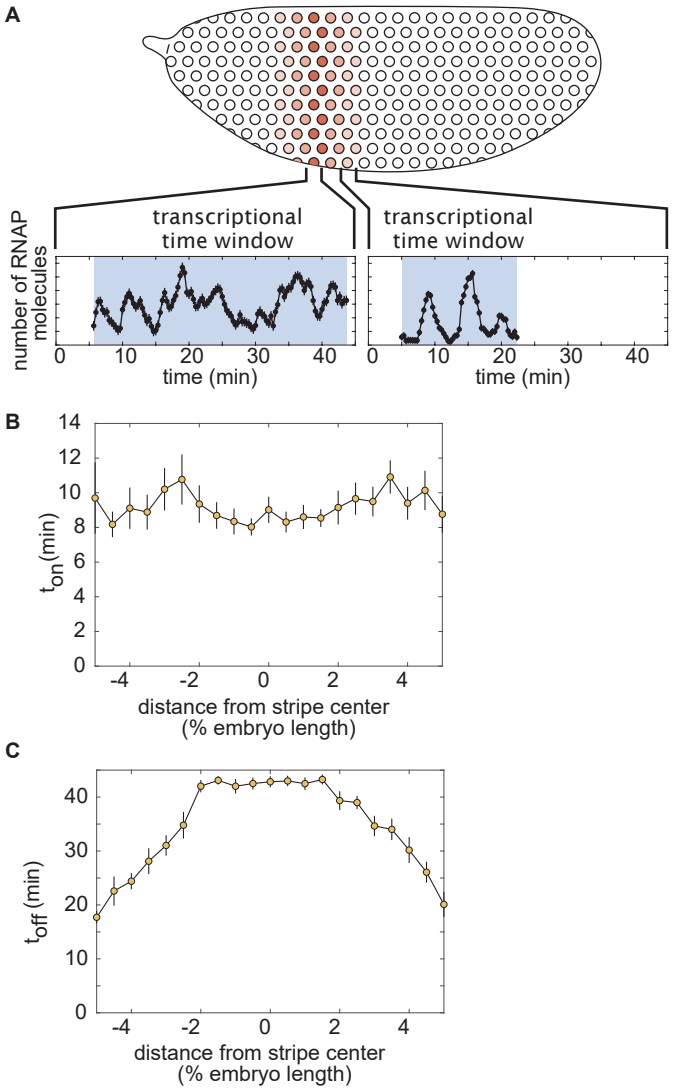


Figure 3–Figure supplement 2. Regulation of the transcriptional time window. (A) Single-nucleus measurements reveal that the duration of transcription is modulated along the stripe and that nuclei transcribe in a burst-like fashion. **(B)** Time for nuclei to activate transcription after mitosis, t_{on}, as a function of position along the stripe. **(C)** Time for nuclei to enter the quiescent transcriptional state, t_{off}. (B,C, average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

bioRxiv preprint doi: https://doi.org/10.1101/335919; this version posted April 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-4.0 International license.

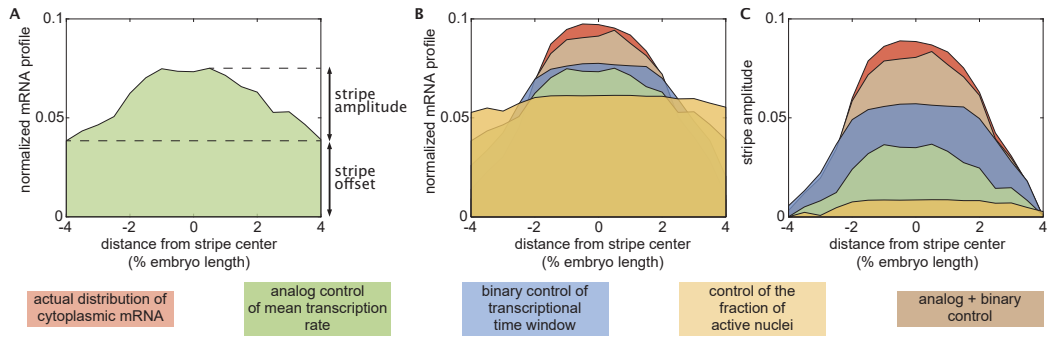


Figure 3–Figure supplement 3. Definition of stripe amplitude. (A) The normalized mRNA profile for the stripe can be separated into an offset and an amplitude. **(B)** Normalized mRNA profiles and **(C)** stripe amplitude for the cytoplasmic pattern of mRNA as well as for the contributions from the various regulatory strategies.

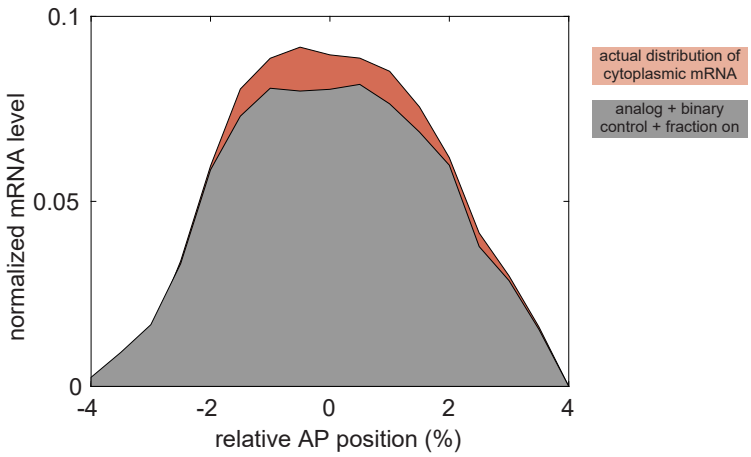


Figure 3–Figure supplement 4. Joint effect of mean rate, binary control, and fraction of active nuclei. Including of the predicted effect of anterior-posterior-dependent modulation of the fraction of active nuclei has little effect on the predicted cytoplasmic mRNA profile (compare brown profile in **Figure 1**G, gray profile above). The remaining difference between the full profile (red) and the gray profile can be attributed the effects of temporal variations in the mean rate of transcription.

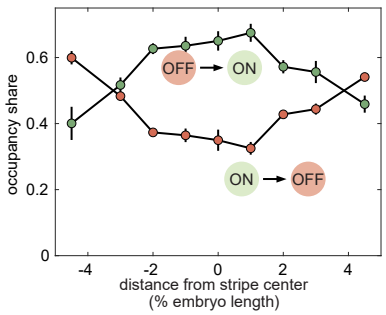


Figure 5–Figure supplement 1. Fraction of time spent in each transcriptional state. Fraction of time spent in the ON and OFF states as a function of the position along the stripe. (Error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)