

1 Multimodal transcriptional control 2 of pattern formation in embryonic 3 development

4 **Nicholas C Lammers^{1†}, Vahe Galstyan^{2,3†}, Armando Reimer¹, Sean A Medin⁴, Chris
5 H Wiggins^{5,6,7,8*}, Hernan G Garcia^{1,4,9,10*}**

*For correspondence:

hggarcia@berkeley.edu (HGG);
chris.wiggins@columbia.edu (CHW)

¹These authors contributed equally to this work
6 Biophysics Graduate Group, University of California at Berkeley, Berkeley, California;
7 ²Biochemistry and Molecular Biophysics Option, California Institute of Technology,
8 Pasadena, California; ³Department of Physics, Columbia University, New York, New York;
9 ⁴Department of Physics, University of California at Berkeley, Berkeley, California;
10 ⁵Department of Applied Physics and Applied Mathematics, Columbia University, New
11 York, New York; ⁶Data Science Institute, Columbia University, New York, New York;
12 ⁷Department of Systems Biology, Columbia University, New York, New York; ⁸Department
13 of Statistics, Columbia University, New York, New York; ⁹Department of Molecular and
14 Cell Biology, University of California at Berkeley, Berkeley, California; ¹⁰Institute for
15 Quantitative Biosciences-QB3, University of California at Berkeley, Berkeley, California

16
17 **Abstract** Predicting developmental outcomes from regulatory DNA sequence and transcription
18 factor patterns remains an open challenge in physical biology. Using stripe 2 of the *even-skipped*
19 gene in *Drosophila* embryos as a case study, we dissect the regulatory forces underpinning a key
20 step along the developmental decision-making cascade: the generation of cytoplasmic mRNA
21 patterns via the control of transcription in individual cells. Using live imaging and computational
22 approaches, we found that the transcriptional burst frequency is modulated across the stripe to
23 control the mRNA production rate. However, we discovered that bursting alone cannot
24 quantitatively recapitulate the formation of the stripe, and that control of the *window of time* over
25 which each nucleus transcribes *even-skipped* plays a critical role in stripe formation. Theoretical
26 modeling revealed that these regulatory strategies—bursting and the time window—obey different
27 kinds of regulatory logic, suggesting that the stripe is shaped by the interplay of two distinct
28 underlying molecular processes.

29 30 **Introduction**

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

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

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

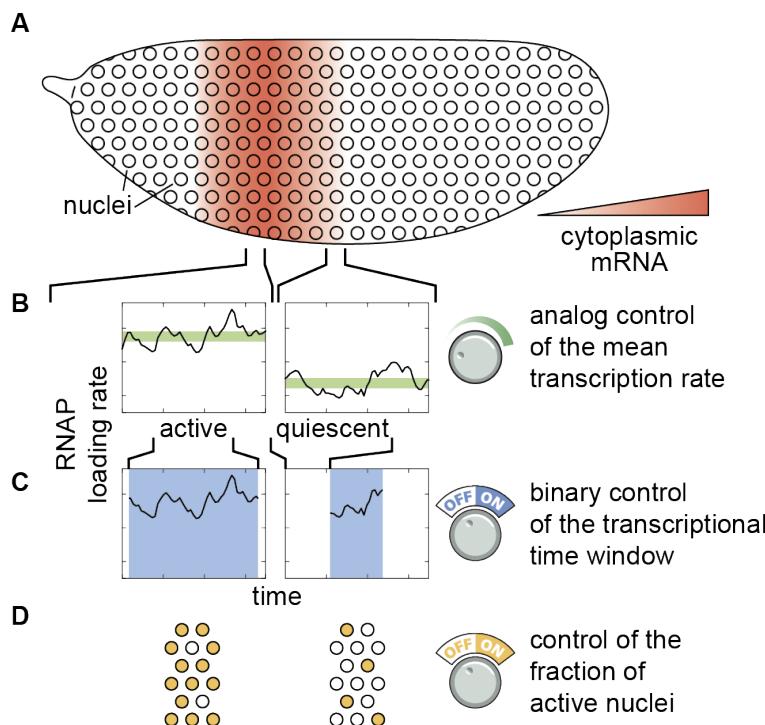


Figure 1. Models of pattern formation by single-cell transcriptional activity. (A) Cytoplasmic mRNA patterns could arise from transcription factors exerting control over (B) the mean transcription rate, (C) the transcriptional time window dictating when a nucleus is transcriptionally active or quiescent,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

62 In order to uncover the quantitative contribution of these three regulatory strategies to pattern
63 formation, and to determine whether other regulatory strategies are at play, it is necessary to
64 measure the rate of RNA polymerase loading in individual nuclei, in real time, in a living embryo.

65 However, to date, most studies have relied on fixed-tissue techniques such as mRNA FISH and
66 immunofluorescence in order to obtain snapshots of the cytoplasmic distributions of mRNA and
67 protein as development progresses (*Jaeger et al., 2004; Fakhouri et al., 2010; Parker et al., 2011;*
68 *Estrada et al., 2016; Crocker et al., 2016; Verd et al., 2017; Park et al., 2018*). Such techniques
69 are virtually silent regarding the regulation of single-cell gene expression over time, and are thus
70 ill-suited to the study of how spatiotemporal variations in transcriptional dynamics give rise to
71 patterns of cytoplasmic mRNA.

72 In this work, we investigated how single-cell transcriptional activity leads to the formation of
73 stripe 2 of the widely studied *even-skipped* (*eve*) gene in the developing fruit fly embryo (*Small*
74 *et al., 1992; Arnosti et al., 1996*). We combined single-cell live imaging with theoretical modeling
75 in order to study transcriptional activity at the single-cell level in real time, seeking a quantitative
76 connection between the spatiotemporal regulation of transcription and the formation of cytoplasmic
77 patterns of mRNA. Consistent with previous studies, we found that the rate of mRNA production
78 is elevated in the center of the stripe (*Bothma et al., 2014*). Strikingly, however, we discovered
79 that this analog control is alone insufficient to quantitatively recapitulate the formation of the
80 stripe; binary control of the transcriptional time window (*Figure 1C*) is also necessary. Furthermore,
81 we developed novel computational approaches to uncover the molecular underpinning of each
82 regulatory strategy. We employed a memory-adjusted hidden Markov model (mHMM) to uncover
83 variations in transcriptional dynamics in individual nuclei across space and time (*Suter et al.,*
84 *2011; Molina et al., 2013; Corrigan et al., 2016*). We showed that, consistent with previous results,
85 transcription factors control the rate of transcription by altering the frequency of transcriptional
86 bursts (*Fukaya et al., 2016; Zoller et al., 2018*). Finally, we utilized logistic regressions to correlate
87 *eve* stripe 2 transcriptional dynamics with changes in input transcription factor concentrations. This
88 analysis revealed that the transcriptional time window adheres to different regulatory logic than
89 transcriptional bursting: while repressor levels alone were sufficient to explain the early silencing of
90 nuclei in the anterior and posterior stripe flanks, the control of bursting among transcriptionally
91 engaged nuclei depends upon the input concentrations of both activators and repressors. Thus,
92 our findings point to the presence of two distinct regulatory mechanisms that control transcription
93 and gene expression patterns in early development, showcasing the potential for theoretical
94 modelling and biological numeracy to yield novel biological insights when coupled with precise and
95 quantitative experimental observation.

96 Results

97 Predicting cytoplasmic mRNA distributions from transcriptional activity

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

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

101 where $mRNA(x, t)$ indicates the mRNA concentration at position x along the embryo at time t , $R(x, t)$
102 corresponds to the mRNA synthesis rate averaged over multiple nuclei within the same position x ,
103 $p_{active}(x)$ is the fraction of active nuclei (corresponding to the regulatory strategy shown in *Figure 1D*)
104 and γ is the degradation rate (see *Appendix 1* for details of this derivation).

105 In order to examine the quantitative consequences of the three potential regulatory strategies
106 (*Figure 1B-D*), we adopted widespread assumptions in the modeling of transcriptional regulation
107 (*Phillips et al., 2013*). First, we assumed that the degradation rate γ is a constant and not under any
108 kind of spatiotemporal control. Comparisons between model predictions and empirically measured
109 levels of cytoplasmic mRNA suggest that this assumption is reasonable (see *Appendix 2*). Second,

110 we posited that at each position throughout the embryo the synthesis rate $R(x, t)$ does not vary
111 significantly in time such that it can be approximated by its time average $R(x) = \langle R(x, t) \rangle$. This
112 assumption will be revised later in the text in order to account for the time-dependent regulation of
113 the mean rate of transcription. Finally, we assumed that nuclei along the axis of the embryo start
114 transcribing at time $t_{\text{on}}(x)$, and stop transcribing and enter a state of transcriptional quiescence at
115 time $t_{\text{off}}(x)$. Under these assumptions, **Equation 1** can be solved analytically, resulting in

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

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

119 **Binary control of the transcriptional time window is the primary driver of stripe 120 formation**

121 In order to test the simple model of pattern formation put forward in **Equation 2**, we quantified
122 transcription of stripe 2 of *eve* in the fruit fly. We imaged the transcription of an *eve* stripe 2 reporter
123 using the MS2 system (*Garcia et al., 2013; Lucas et al., 2013; Bothma et al., 2014*). Transcripts of a
124 reporter gene driven by the *eve* stripe 2 enhancer and the *eve* promoter contain repeats of a DNA
125 sequence that, when transcribed, form stem loops (*Bertrand et al., 1998*). These stem loops are
126 recognized by maternally provided MS2 coat protein fused to GFP (**Figure 2A**). As a result, sites of
127 nascent transcript formation appear as fluorescent puncta within individual nuclei (**Figure 2B** and
128 Video 1). This fluorescence can be calibrated using single-molecule FISH in order to estimate the
129 number of RNA polymerase molecules actively transcribing the gene (see Materials and Methods
130 and *Garcia et al. (2013)*). By aligning multiple embryos (**Figure 2–Figure Supplement 1**), we obtained
131 the average number of actively transcribing RNA polymerase molecules as a function of time and
132 position throughout the embryo (**Figure 2C**).

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

140 Our data also revealed that the transcriptional time window is modulated along the stripe (**Fig-
141 ure 3–Figure Supplement 2A**). Whereas the time at which each nucleus becomes transcriptionally
142 active, $t_{\text{on}}(x)$, was constant across the stripe, with all nuclei becoming active 9 ± 4 min after the
143 previous anaphase (**Figure 3–Figure Supplement 2B**), the time at which nuclei stop transcribing and
144 become quiescent, $t_{\text{off}}(x)$, showed a strong modulation along the embryo's axis (**Figure 3–Figure
145 Supplement 2C**). As a result, the time window over which each punctum is engaged in transcription,
146 $\Delta t = t_{\text{off}} - t_{\text{on}}$, is sharply modulated along the stripe (**Figure 3C** and D and Video 3), with nuclei
147 in the stripe center transcribing for >30 min and nuclei on the boundaries only transcribing for
148 approximately 10 min.

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

153 The analysis in **Figure 3A–F** reveals that each of the three regulatory strategies identified in
154 **Figure 1** is at play in the embryo, and that they all have the potential to contribute to pattern

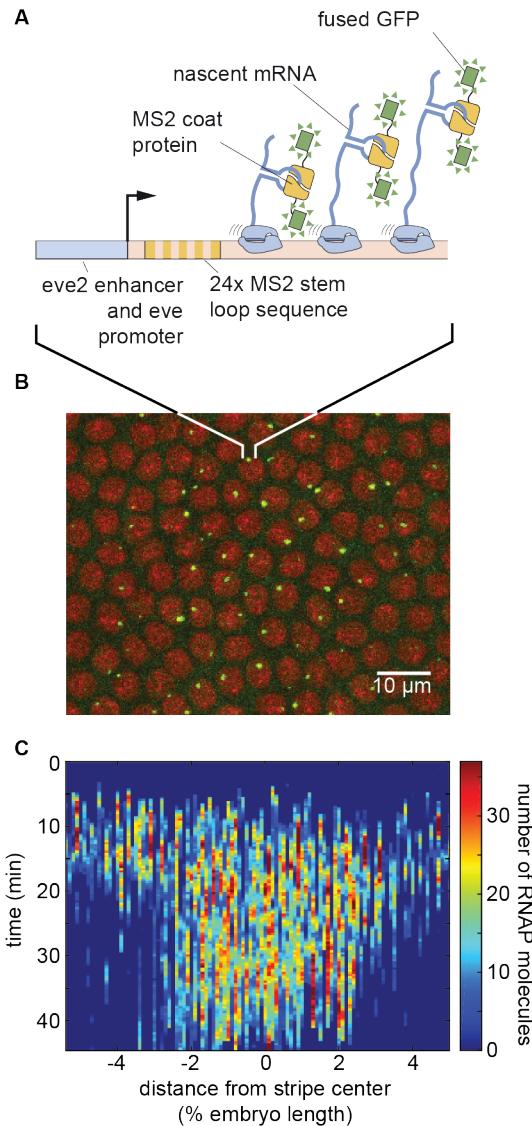


Figure 2. Measuring transcriptional dynamics of eve stripe 2 formation using the MS2 system. (A) MS2 stem loops introduced in an eve stripe 2 reporter gene are bound by MS2 coat protein fused to GFP. (B) Sites of nascent transcript formation appear as green fluorescent puncta whose intensity reports 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.

155 formation. However, these measurements alone cannot inform us on *how much* each of these
156 strategies contributes to the cytoplasmic mRNA pattern. To quantify the degree to which each
157 regulatory strategy contributes to the formation of eve stripe 2, we employed the model described
158 in **Equation 2**.

159 **Figure 3G** indicates the quantitative contribution of each regulatory strategy (each term on the
160 right-hand side of **Equation 2**) to the formation of this cytoplasmic pattern. The cytoplasmic pattern
161 of mRNA, corresponding to left-hand side of **Equation 2**, was obtained from our live-imaging data
162 (see **Appendix 2** for details). Regulation of the fraction of active nuclei along the embryo (**Figure 3G**,
163 yellow) contributes negligibly to this mRNA pattern. In contrast, both the analog regulation of the

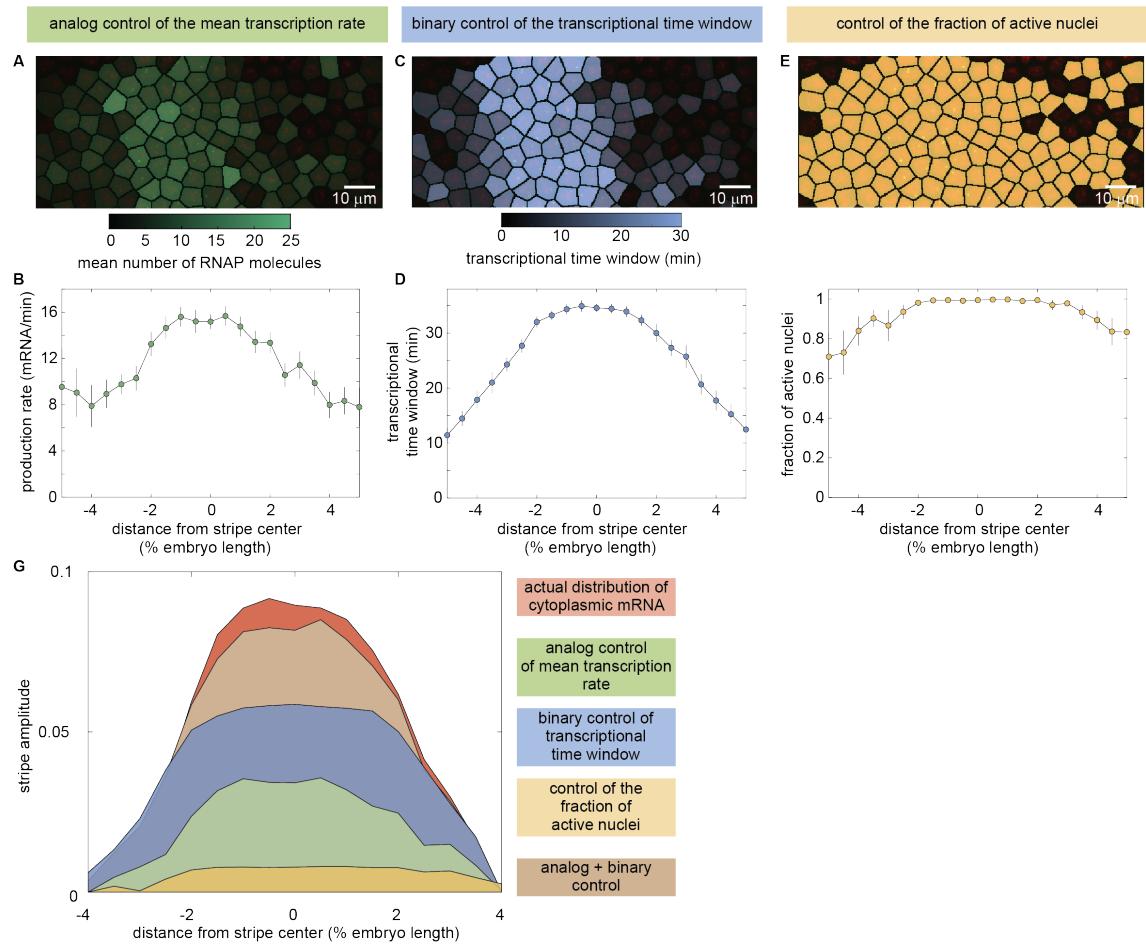


Figure 3. Regulatory strategies for pattern formation in *eve* stripe 2. (A,B) Time-averaged rate of mRNA production, (C,D) transcriptional time window, and (E,F) fraction of active nuclei as a function of position along the embryo. (G) Amplitude of the cytoplasmic mRNA distribution compared to the contributions 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.

mean rate (*Figure 3G*, green) and the binary control of the transcriptional time window (*Figure 3G*, 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 3G*, brown) is sufficient to quantitatively recapitulate the stripe of cytoplasmic mRNA from single-cell transcriptional activity.

168 Mean transcription rate is dictated by bursting through modulation of the rate of 169 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

174 individual loci within the *eve* stripe 2 pattern is highly stochastic (**Bothma et al., 2014**). Indeed, as
175 shown in **Figure 4A**, our data revealed punctuated peaks and troughs in the number of active RNA
176 polymerase molecules. These features have been related to the rate of RNA polymerase loading
177 at the *eve* promoter by assuming that promoter loading is “burst-like”, with the promoter loading
178 RNA polymerase molecules onto the gene at a constant rate over discrete periods of time (**Bothma**
179 **et al., 2014**). This and other evidence from live imaging (**Bothma et al., 2014; Fukaya et al., 2016;**
180 **Desponts et al., 2016**), as well as data from fixed-tissue approaches (**Pare et al., 2009; Little et al.,**
181 **2013; Xu et al., 2015; Zoller et al., 2018**), support a minimal two-state model of promoter switching
182 (**Figure 4B**): promoters switch stochastically between ON and OFF states with rates k_{on} and k_{off} . In
183 this model, promoters in the ON state engage in mRNA production at rate r .

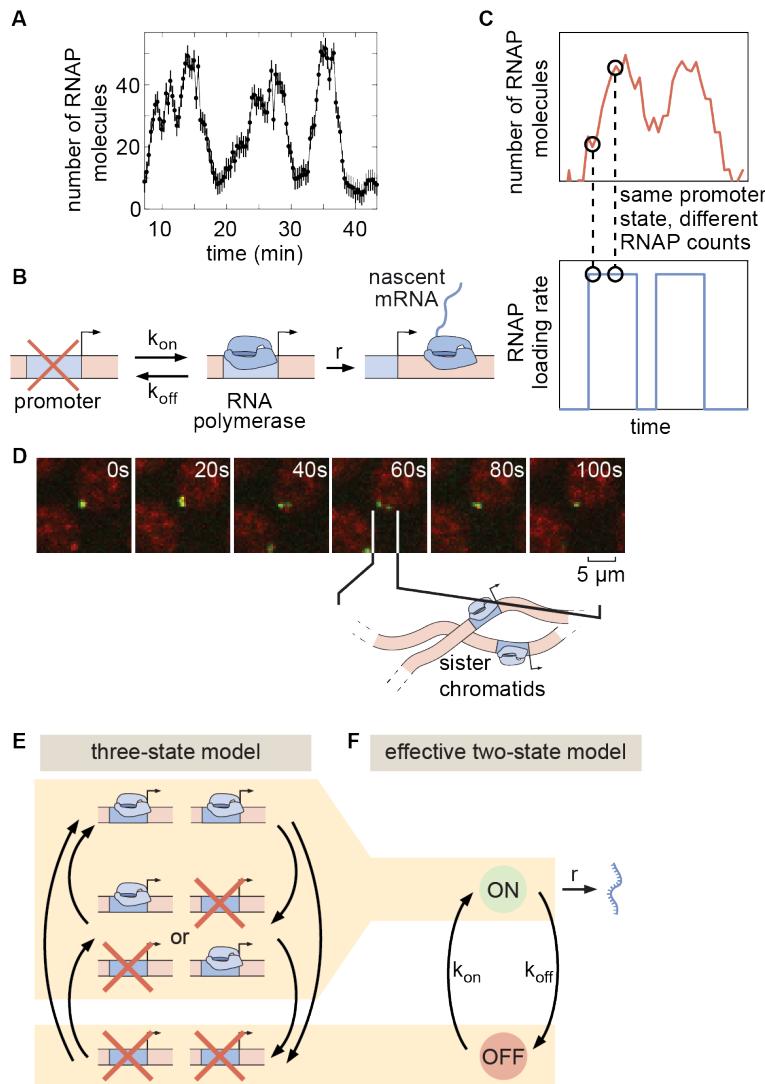


Figure 4. Transcriptional bursting in *eve* stripe 2. **(A)** Single-nucleus measurements reveal that nuclei transcribe in bursts. **(B)** Two-state model of bursting of a single promoter. **(C)** The same hidden rate of RNA polymerase loading (bottom) can correspond to different observable numbers of 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)**.)

184 In the bursting model, the mean rate of transcription is given by the product of the fraction
185 of time spent in the ON state with the transcription rate in this active state (**Peccoud and Ycart,**
186 **1995; Kepler and Elston, 2001; Sasai and Wolynes, 2003; Sanchez and Kondev, 2008; Sanchez et al.,**
187 **2011; Xu et al., 2016**)

$$\underbrace{R(x)}_{\text{mean transcription rate}} = \underbrace{r(x)}_{\text{RNAP loading rate}} \times \underbrace{\frac{k_{\text{on}}(x)}{k_{\text{on}}(x) + k_{\text{off}}(x)}}, \quad (3)$$

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

194 Typically, the *in vivo* molecular mechanism of transcription factor action is inferred from mea-
195 surements of transcriptional noise obtained through snapshots of dead and fixed embryos or
196 cells using theoretical models (**Zenklusen et al., 2008; So et al., 2011; Little et al., 2013; Jones et al.,**
197 **2014; Senecal et al., 2014; Xu et al., 2015; Padovan-Merhar et al., 2015; Skinner et al., 2016; Bart-**
198 **man et al., 2016; Zoller et al., 2018; Hendy et al., 2017**). In contrast, MS2-based live imaging can
199 directly inform on the dynamics of transcriptional bursting in real time. The MS2 approach, however,
200 reports on the total number of actively transcribing RNA polymerase molecules and not on the
201 instantaneous rate of RNA polymerase loading at the promoter, which is the relevant quantity for
202 estimating k_{on} , k_{off} , and r . To date, approaches for extracting bursting parameters from such data in
203 multicellular organisms have mainly relied on the manual analysis of single-nucleus transcriptional
204 dynamics (**Bothma et al., 2014; Fukaya et al., 2016**) or autocorrelation-based methods that infer
205 mean bursting parameters across ensembles of traces (**Larson et al., 2011; Coulon et al., 2014;**
206 **Desponts et al., 2016**). A computational method for inferring the rates of RNA polymerase loading
207 (**Figure 4C, bottom**) from the total number of actively transcribing RNA polymerase molecules in
208 single cells (**Figure 4C, top**) is thus needed to obtain the bursting parameters.

209 Hidden Markov models (HMMs) are widely used to uncover the dynamics of a system as it
210 transitions through states that are not directly accessible to the observer (**Bronson et al., 2009**).
211 However, our observable (the MS2 signal) does not correspond to the hidden variable of interest
212 (the promoter state) in a one-to-one fashion (compare **Figure 4C top and bottom**). Instead, the
213 observable MS2 signal reflects the net effect of promoter switching over a period equal to the
214 time that an RNA polymerase molecule takes to transcribe the whole gene. Thus, instantaneous
215 fluorescence does not just depend on the current promoter state; it exhibits a dependence on
216 how active the promoter has been over a preceding window of time, which effectively constitutes
217 a memory for recent promoter states (**Choubey et al., 2015; Xu et al., 2016; Corrigan et al., 2016;**
218 **Choubey, 2018; Choubey et al., 2018**). Classic HMM approaches cannot account for this kind of
219 system memory.

220 In order to model the process of transcription and extract the kinetic parameters of promoter
221 switching, we augmented classic HMMs to account for memory (details about implementation
222 of the method are given in **Appendix 3**). Similar approaches were recently introduced to study
223 transcriptional dynamics in cell culture and tissue samples (**Suter et al., 2011; Molina et al., 2013;**
224 **Zechner et al., 2014; Zoller et al., 2015; Hey et al., 2015; Bronstein et al., 2015; Corrigan et al.,**
225 **2016; Featherstone et al., 2016**). We used simulated data to establish that mHMM reliably extracts
226 the kinetic parameters of transcriptional bursting from live-imaging data (**Appendix 4**), providing an
227 ideal tool for dissecting the contributions from individual bursting parameters to observed patterns
228 of transcriptional activity across space and time.

229 Before applying our model to real-time transcriptional data, we had to account for the rapid
230 replication of the *D. melanogaster* genome at the beginning of each nuclear cycle (**Rabinowitz, 1941**;

231 *Shermoen et al., 2010*), which leads to the presence of two distinct *eve* loci within each fluorescent
232 spot (*Figure 4D* and Video 5). The first evidence of resolved chromatids appears as early as 8
233 minutes into nuclear cycle 14 (*Appendix 5–Figure 1*)—coincident with the average onset time of
234 transcription (*Figure 3–Figure Supplement 2B*). Moreover, our analysis indicates that replication
235 of the relevant portion of the genome likely occurs in all *eve*-expressing nuclei by no later than
236 10 minutes following mitosis (*Appendix 5–Figure 2*). Thus, we conclude that the vast majority of
237 our data feature two distinct *eve* loci within each diffraction-limited transcription spot. Moreover,
238 while the distance between sister loci varies over time (see, e.g. *Figure 4D*), they nonetheless stay in
239 relatively close proximity to ensure their proper segregation from each other at the next mitosis
240 (*Senaratne et al., 2016*) such that the fluorescent intensity signals extracted from our data reflect
241 the integral over both loci (*Figure 2–Figure Supplement 2*). As a result, if we assume that each
242 locus can be well-represented by a two-state model of transcriptional bursting, then an effective
243 three-state model is needed to capture *eve* dynamics (*Figure 4E*). For ease of exposition, we present
244 our main results in the context of an effective two-state model, in which, as detailed in *Appendix 1*,
245 the system is considered to be in the ON state so long as *either* chromatid is bursting (*Figure 4F*).
246 Note that none of our conclusions below are affected by this choice of an effective model as shown
247 in *Appendix 6* where we present full results for the three-state model.

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

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

266 **Binary control of the transcriptional time window is independent of transcriptional 267 bursting**

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

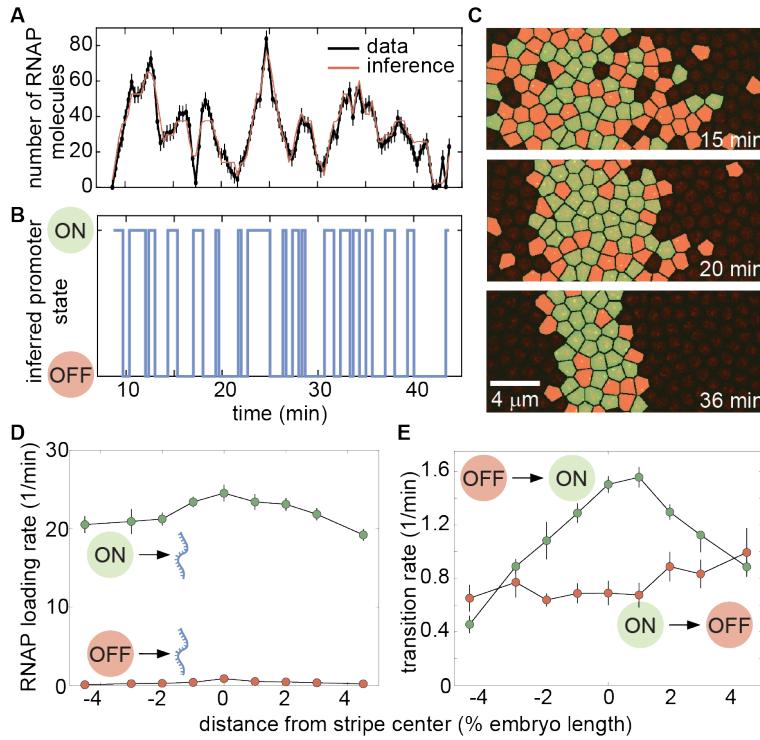


Figure 5. Inferring bursting dynamics using a memory-adjusted Hidden Markov model. (A)

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.

281 be achieved by downregulating k_{on} , downregulating r , and/or upregulating k_{off} ([Figure 6A](#), bottom).

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

295 These findings confirmed our time-averaged inference results ([Figure 5D](#) and E) indicating that
 296 k_{on} was the primary kinetic pathway through which transcription factors influence *eve* stripe 2
 297 transcription dynamics. Moreover, the coincidence of the decrease in k_{on} in flank nuclei with the
 298 onset of transcriptional quiescence (grey shaded region in [Figure 6D](#)) seemed to suggest that, at

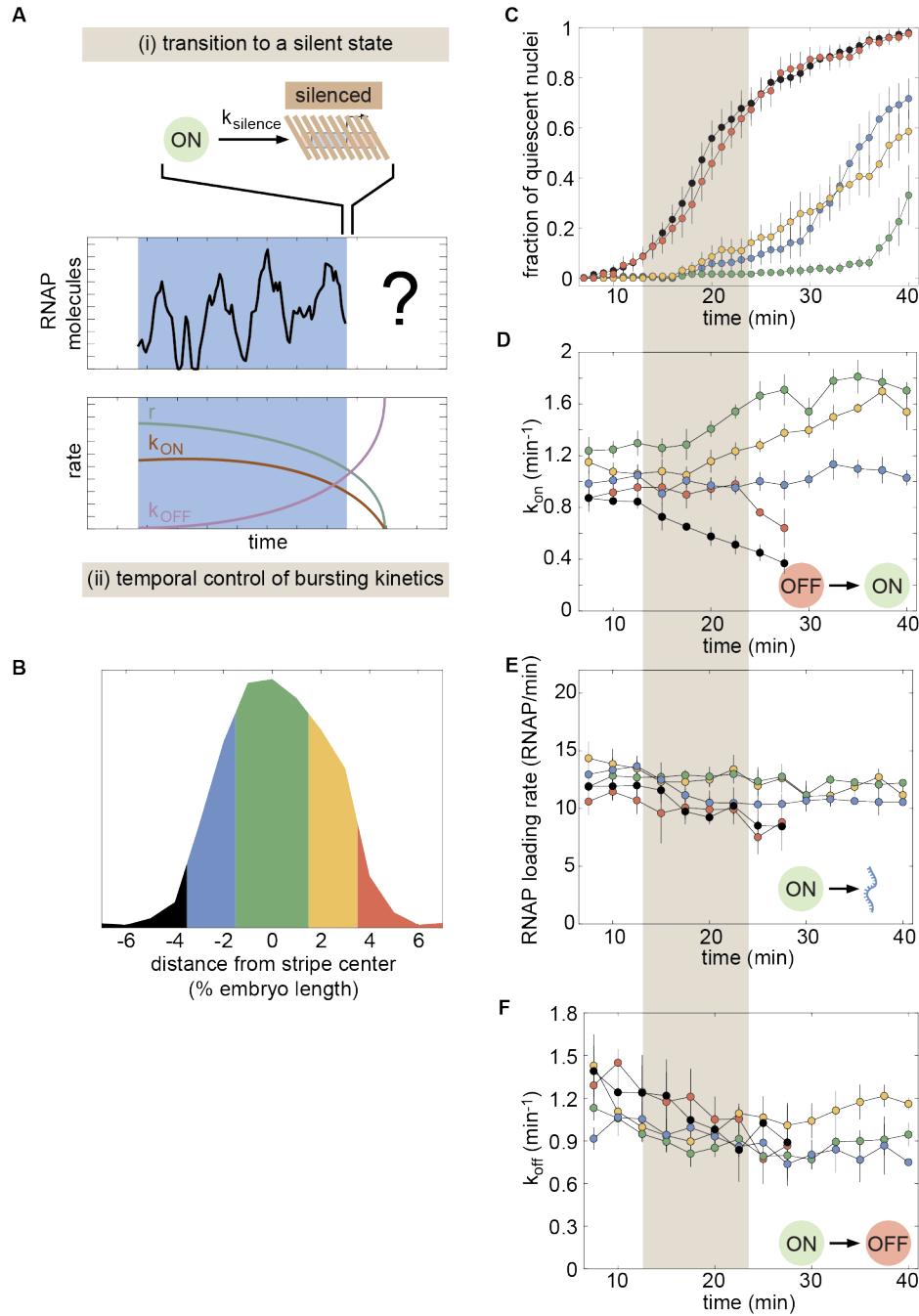


Figure 6. Investigating the molecular character of transcriptional quiescence. (A) Two hypotheses explaining promoter quiescence onset: (i) an irreversible transition into an alternative, transcriptionally silent state and (ii) the modulation of one or more bursting parameters over time. (B) Division 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)

299 least in part, quiescence in the stripe flanks could be driven by the temporal modulation of bursting
 300 parameters (**Figure 6A**, bottom). However, other trends in our data results were not consistent with
 301 the view that a decrease in k_{on} drives transcriptional quiescence. Although 70% and 50% of nuclei

302 in the regions directly anterior and posterior of the stripe center were quiescent by 40 min into
303 the nuclear cycle (blue and yellow curves in **Figure 6C**), we detected no corresponding decrease
304 in k_{on} . In fact, k_{on} actually *increased* in some inner regions of the stripe (**Figure 6D**) —a trend that
305 would increase overall transcriptional activity and would therefore go against the establishment of
306 transcriptional quiescence.

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

317 The contradictory trends observed in the stripe center suggested that entry into transcriptional
318 quiescence might be akin to an irreversible transition into a silent state (**Figure 6A**, top), thus suggesting
319 that binary control of the transcriptional time window and the transcriptional bursting driving
320 the analog control of the mean transcription rate may arise from distinct molecular processes.

321 **Input-output analysis reveals distinct regulatory logic for bursting and the trans-
322 scriptional time window**

323 *eve* stripe 2 is mainly established by the combined action of two activators, Bicoid and Hunchback,
324 and two repressors, Giant and Krüppel (*Frasch and Levine, 1987; Small et al., 1992; Arnosti et al.,
325 1996*). If transcriptional bursting and the transcriptional time window are controlled by distinct
326 molecular mechanisms, then distinct forms of regulatory logic may be at play. For example, the
327 Bicoid and Hunchback activators could control transcriptional bursting, while the Giant and Krüppel
328 repressors could dictate the entry into the quiescent state. In order to reveal the molecular logic
329 controlling each regulatory strategy, we sought to correlate the fraction of nuclei that have entered
330 the quiescent state (**Figure 7A**) and the fraction of nuclei in the bursting ON state (**Figure 7B**) with
331 the corresponding spatiotemporal patterns in the input concentrations of these four transcription
332 factors.

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

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

346 where the coefficients β_n indicate the magnitude and nature (activating or repressing) of the
347 transcription factor’s regulatory function. In estimating these coefficients, we used prior knowledge

348 about the function of each transcription factor, requiring Bicoid and Hunchback to play activating
349 roles, and Krüppel and Giant to play repressing roles (*Small et al., 1991, 1992*). We used an
350 analogous model to investigate the regulatory logic controlling transcriptional bursting by inferring
351 the factors that determine the relative likelihood that nuclei are in the bursting ON versus the OFF
352 state, likelihood (ON state)/likelihood (OFF state).

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

362 We next turned our attention to the relationship between transcription factor levels and the
363 fraction of nuclei in the ON state (*Figure 7B*). Unlike the transcriptional time window, repressor levels
364 alone could not recapitulate the observed bursting profile; Hunchback levels were also necessary
365 in order to fully capture the spatiotemporal bursting dynamics (*Figure 7E and G*). Specifically, we
366 linked a rise in Hunchback concentration to an observed rise in the fraction of nuclei in the ON
367 state in the stripe center between 30 and 35 min into the nuclear cycle (*Figure 7B and F*).

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

376 Discussion

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

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

394 In this work, we derived a simple theoretical model that predicts how the interplay between
395 regulatory strategies at the single-cell level dictates the accumulation of cytoplasmic mRNA and
396 the subsequent formation of a gene expression pattern (*Equation 2*), and tested its predictions

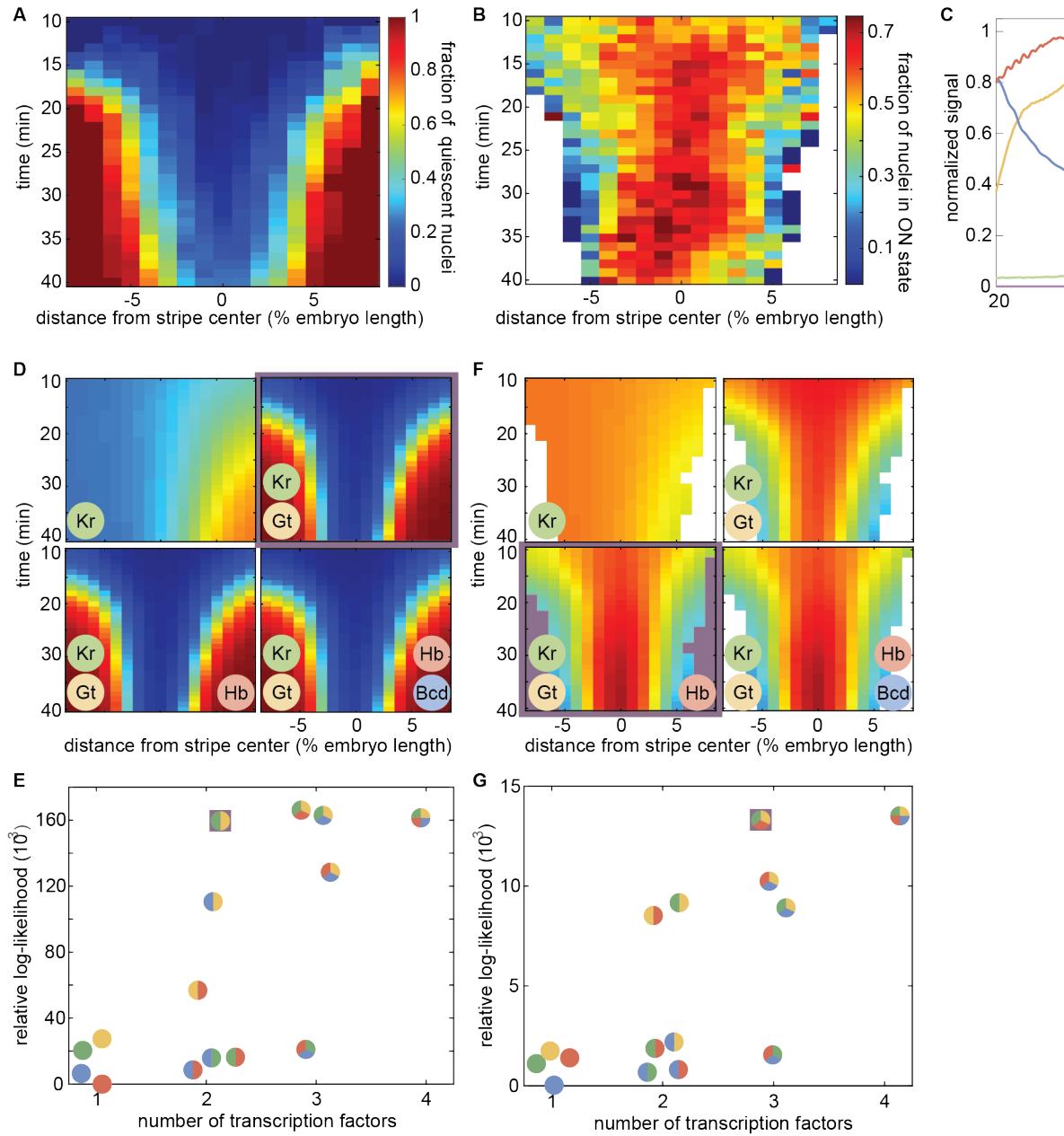


Figure 7. Probing the regulatory logic of bursting and the transcriptional time window. **(A)** Fraction of nuclei in the transcriptionally quiescent state and **(B)** fraction of nuclei in the bursting ON state as a function of time and position along the embryo. **(C)** Snapshot of input transcription 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).

397 experimentally by performing single-cell live-imaging measurements in developing embryos using
398 the MS2 system. Our results revealed that the modulation of the mean rate of transcription
399 alone is insufficient to recapitulate the formation of a sharp gene expression stripe (**Figure 3F**,
400 green). We discovered that the window of time over which promoters engage in transcription is
401 sharply controlled along the axis of the embryo (**Figure 3C** and D) and that the joint action of the
402 analog control of the rate of transcription and the binary control of the duration of transcription is
403 necessary and sufficient to quantitatively recapitulate most of the the full stripe profile (**Figure 3F**,
404 brown).

405 Here, we used biological numeracy as a driver for biological discovery: our discovery of the
406 key role of the binary control of the transcriptional time window in pattern formation was only
407 made possible by going beyond the *qualitative* description of pattern formation and demanding a
408 *quantitative* agreement between our theoretical predictions and the experimental data (**Phillips**,
409 **2015**). Further, our work emphasizes how the regulation of gene expression *timing* in development
410 is as important as the regulation of the spatial extent of these patterns along the embryo. Thus,
411 in order to make progress toward a quantitative and predictive picture of pattern formation in
412 development, it is necessary to go beyond the widespread steady-state, *static* picture of pattern
413 formation in development put forward by previous single-cell transcriptional activity studies that
414 focused on the study of snapshots of fixed embryos (**Pare et al., 2009; Little et al., 2013; Xu et al.,**
415 **2015; Zoller et al., 2018**) and embrace a *dynamical* description that acknowledges that development
416 is a process that occurs outside of steady state (**Berrocal et al., 2018**).

417 To determine whether the same molecular mechanisms dictate the analog control of the mean
418 transcription rate and the binary control of the transcriptional time window, we utilized a variety
419 of theoretical and computational tools in conjunction with our live-imaging data. Specifically, to
420 uncover how the mean rate of transcription is regulated across the stripe, we developed a mHMM
421 that is capable of inferring the instantaneous activity state of individual gene loci from MS2 traces.
422 We used this mHMM to infer average promoter-switching parameters across the stripe (**Figure 5**).
423 In agreement with previous measurements of various gene expression patterns (**Xu et al., 2015;**
424 **Desponts et al., 2016; Fukaya et al., 2016; Zoller et al., 2018**), our results revealed that the the
425 burst frequency (k_{on}) is the main bursting parameter regulated by the input transcription factors
426 across *eve* stripe 2. This increase in k_{on} in the stripe center increases the fraction of time that nuclei
427 spend in the active transcriptional state.

428 Importantly, our mHMM algorithm is not limited to the *eve* stripe 2 system and should prove
429 useful to infer the underlying regulatory kinetics of any gene that is tagged using approaches such
430 as the MS2 or PP7 systems in any organism (**Larson et al., 2011; Hocine et al., 2012; Fukaya et al.,**
431 **2016**). Further, the method could be used to infer the state of the ribosome as mRNA is being
432 translated into protein in novel single-molecule *in vivo* translation assays (**Morisaki et al., 2016;**
433 **Wang et al., 2016; Yan et al., 2016; Wu et al., 2016**). Thus, we envision that our method will be
434 useful for the broader biophysical analysis of *in vivo* cellular processes at the single-molecule level.

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

447 To test this hypothesis, we utilized a logistic regression framework and time-resolved data for

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

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

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

479 In order to further test these and other molecular hypotheses, it will be critical to move beyond
480 spatiotemporal averages for transcription factor inputs (**Figure 7C**) and, instead, use live single-
481 nucleus measurements to directly correlate input transcription factor concentration dynamics
482 with the corresponding transcriptional activity at the single-cell level (**Holloway and Spirov, 2017**).
483 Experimentally, we recently demonstrated the simultaneous measurement of inputs and outputs
484 in single nuclei of a living fly embryo using novel genetically encoded LlamaTags (**Bothma et al.,**
485 **2018**). We believe that utilizing this novel technique, in conjunction with the theoretical methods
486 presented here, to query the effects of targeted disruptions to transcription factor binding domains
487 on regulatory enhancers will constitute a powerful assay for querying transcription factor function
488 at the molecular level. Thus, there are clear experimental and theoretical paths to uncovering
489 the detailed quantitative mechanisms behind the molecular control of transcriptional bursting
490 and quiescence in development. Such a quantitative description is a necessary step toward a
491 predictive understanding of developmental decision-making that makes it possible to calculate
492 developmental outcomes from knowledge of the nature of the transcription factor interactions
493 within gene regulatory networks.

494 **Materials and methods**

495 **Cloning and transgenesis**

496 This work employed the same *eve* stripe 2 reporter construct developed by *Bothma et al. (2014)*.
497 This construct contains the *even-skipped* (*eve*) stripe 2 enhancer and promoter region (spanning
498 -1.7 kbp to +50 bp) upstream of the *yellow* reporter gene. 24 repeats of the MS2 stem loop sequence
499 were incorporated into the 5' end of the reporter gene.

500 **Sample preparation and data collection**

501 Sample preparation followed procedures described in *Bothma et al. (2014)* and *Garcia and Gregor*
502 (*2018*). In short, female virgins of *yw*;His-RFP;MCP-GFP (MCP, MS2 coat protein) were crossed to
503 males bearing the reporter gene. Embryos were collected and mounted in halocarbon oil 27
504 between a semipermeable membrane (Lumox film, Starsteds) and a coverslip. Data collection
505 was performed using a Leica SP8 Laser Scanning Confocal Microscope. Average laser power
506 on the specimen (measured at the output of a 10x objective) was 35 μ W. Image resolution was
507 256 \times 512 pixels, with a pixel size of 212 nm and a pixel dwell time of 1.2 μ s. The signal from each
508 frame was accumulated over three repetitions. At each time point, a stack of 21 images separated
509 by 500 nm were collected. Image stacks were collected at a time resolution of 21 seconds. The
510 MCP-GFP and Histone-RFP were excited with a laser wavelength of 488 and 556 nm using a White
511 Light Laser, respectively. Fluorescence was detected with two separate Hybrid Detectors (HyD)
512 using the 498-546 nm and 566-669 nm spectral windows. Specimens were imaged for a minimum
513 of 40 minutes into nuclear cleavage cycle 14.

514 **Image analysis**

515 Image analysis of live embryo movies was performed based on the protocol found in *Garcia et al.*
516 (*2013*) with modifications to the identification of transcriptional spots, which were segmented
517 using the Trainable Weka Segmentation plugin for FIJI using the FastRandomForest algorithm
518 (*Schindelin et al., 2012; Schneider et al., 2012; Arganda-Carreras et al., 2017; Witten et al., 2016*).
519 In comparison with a previous algorithm based on Difference of Gaussians (*Little et al., 2013*;
520 *Garcia et al., 2013; Bothma et al., 2014, 2015*), this alternative spot segmentation approach was
521 found to be superior for the detection of dim transcription spots—a feature critical to establishing
522 the precise timing of the cessation of activity at transcriptional loci.

523 **Data processing**

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

531 This routine, described graphically in *Figure 2–Figure Supplement 1*, used observed spatial
532 patterns of fluorescence measured from 30 minutes into nuclear cycle 14—the approximate time at
533 which the mature stripe is first established (*Bothma et al., 2014*)—to the time of last observation
534 (\geq 40 min) to find the natural position and orientation of the mature stripe. Generally, the *eve* stripes
535 run roughly perpendicular to the anterior-posterior (AP) axis of the embryo; however, the approach
536 allowed for the possibility that the true orientation of the *eve* 2 stripe deviated from the orientation
537 implied by manual estimates of the anterior posterior axis. Thus, a variety of orientations for the
538 natural stripe axis were considered, ranging between \pm 15 degrees from the line perpendicular to
539 the stripe with the manually specified anterior posterior axis. For each orientation, a sliding window
540 of 4% embryo length in width was used to find the position along the proposed orientation that

541 captured the largest fraction of the total fluorescence emitted by the mature stripe. The orientation
542 and position that maximized the amount of fluorescence captured within this window defined a
543 line through the field of view that was taken as the stripe center. All anterior-posterior positions
544 used for subsequent analyses were defined relative to this center line.

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

551 **mHMM inference**

552 To account for finite RNA polymerase elongation times, a compound state Markov formalism
553 was developed in which the underlying tw-promoter system—assumed to have three states (see
554 *Figure 4E,F*)—was transformed into a system with 3^w compound gene states, where w indicates
555 the number of time steps needed for an RNA polymerase molecule to traverse the full transcript
556 (see *Appendix 9*). These compound gene states played the role of the “hidden” states within
557 the traditional HMM formalism. See *Appendix 3* for details regarding the model’s architecture.
558 Following this transformation from promoter states to compound gene states, it was possible to
559 employ a standard version of the expectation-maximization (EM) algorithm, implemented using
560 custom-written scripts in Matlab, to estimate bursting parameters from subsets of experimental
561 traces (*Appendix 3*). The scripts are available at the GarciaLab/mHMM GitHub repository. Bootstrap
562 sampling was used to estimate the standard error in our parameter estimates. Subsets of 8,000
563 data points were used to generate time-averaged parameter estimates. In order to accurately
564 capture the time-averaged dynamics across the entirety of nuclear cycle 14, the full length of each
565 experimental trace was used for time averaged inference. Sample sizes for windowed inference
566 varied due to data set limitations. When possible, samples of 4,000 points were used. Only data
567 points falling within a 15 minute window centered about the time point of interest were included in
568 windowed inference runs. Inference was not conducted for spatiotemporal regions for which fewer
569 than 1,250 time points were available. A minimum of 10 bootstrap samples were used to estimate
570 each parameter value reported in this work. Reported values represent the median taken across
571 bootstrap samples.

572 **Input-output logistic regressions**

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

577 **Bootstrap error calculation**

578 Bootstrap resampling was used frequently throughout this work to estimate the standard error
579 in a variety of reported quantities, from trends estimated directly from raw experimental data
580 in *Figure 1* to mHMM inference results presented in *Figure 5* and *Figure 6*. In this procedure,
581 multiple bootstrap replicates, y_{boot}^i are generated by sampling with replacement from the pool
582 of available experimental data, Y (see, e.g. *Efron and Hastie (2016)*). The parameter of interest
583 (say, $t_{on}(x)$) is then calculated for each replicate and the mean of these estimates is taken as the
584 bootstrap estimate of the parameter value, $\hat{t}_{on}(x)$, while the standard deviation across the pool of
585 bootstrap parameter estimates is used to approximate the standard error in our estimate of $t_{on}(x)$.
586 In our case, simply performing this procedure across the available pool of nuclei failed to account
587 for biological variability that exists from embryo to embryo. To account for this, we introduced
588 a hierarchical bootstrapping procedure. The first step in this procedure was to draw bootstrap

589 samples from across the 11 embryos used in this study. Because these samples were taken with
590 replacement, most bootstrap samples excluded some embryos out of the original set of 11 and
591 included duplicates (or triplicates, etc.) of others. Each embryo-level bootstrap defined a subset of
592 nuclei. The final set of nuclei used for parameter estimation was generated by performing another
593 round of bootstrap sampling on this pool. Bootstrap averages and standard errors were then
594 calculated as described above. This two-step procedure thus accounts for both embryo-to-embryo
595 and nucleus-to-nucleus variability.

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

601 **Absolute calibration of MS2 signal**

602 In order to frame our results with respect to units with a clear physical interpretation, we calibrated
603 our fluorescence measurements in terms of absolute numbers of mRNA molecules. This calibration
604 was also used to inform our Poisson loading sensitivities (*Appendix 3*). To calculate this calibration
605 for our *eve* stripe 2 data, we relied on measurements reported by a previous study that utilized MS2
606 in conjunction with single molecule FISH to establish a calibration factor, α , between the integrated
607 MS2 signal, F_{MS2} , and the number of mRNA molecules produced at a single transcriptional locus,
608 N_{FISH} , (*Garcia et al., 2013*) given by

$$\alpha = \frac{N_{\text{FISH}}}{F_{\text{MS2}}}. \quad (5)$$

609 This calibration factor can be used to estimate the average contribution of a single mRNA molecule
610 to the observed (instantaneous) fluorescent signal. While the values for the parameters in *Equa-*
611 *tion 5* reported here pertain to the transcriptional output driven by the Bicoid activated P2 enhancer
612 and promoter during nuclear cycle 13, the calibration should generalize to all measurements taken
613 using the same microscope.

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

$$F_1 = f_{\text{max}} \frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}}, \quad (6)$$

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

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

621 such that

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

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,

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

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

622 resulting in

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

Measurements performed in *Garcia et al. (2013)* give N_{FISH} to be 220 (± 30) mRNA per nucleus and v_{elong} to be 1.5 (± 0.14) kb/min. Experimental measurements on the P2 enhancer (courtesy of Elizabeth Eck, Maryam Kazemzadeh-Atoufi and Jonathan Liu) indicate that the total fluorescence per nucleus, F_{MS2} , is 9,600 (± 320) AU minutes. For the reporter gene used to take these measurements, L_I and L_{II} are 1.275 kb and 4.021 kb, respectively. Thus, we obtain

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

623 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
624 results, and none of our core results depend upon having access to a precise calibration of the
625 observed signal in terms of absolute numbers of RNA polymerase molecules.

627 Videos

628 **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.

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

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

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

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

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

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

644 Acknowledgments

645 We thank Thomas Gregor and Lev Barinov for discussion about an initial implementation of the
646 mHMM approach, Florian Jug for help with the spot segmentation using machine learning, and
647 Elizabeth Eck, Maryam Kazemzadeh-Atoufi and Jonathan Liu for the P2 data used in the absolute
648 MS2 calibration and for the Bicoid-GFP data used in the logistic regression. We are also grateful
649 to Jack Bateman, Jane Kondev, Rob Phillips, Allyson Sgro, Donald Rio and Richard Harland for
650 comments and discussion on the manuscript. HGG was supported by the Burroughs Wellcome
651 Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers

652 Science Program, the Searle Scholars Program, the Shurl & Kay Curci Foundation, the Hellman
653 Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER
654 Award (1652236). NL was supported by NIH Genomics and Computational Biology training grant
655 5T32HG000047-18. CW was supported by the NIH/NIC (U54 CA193313), CUNY (RFCUNY 40D14-A),
656 and the NSF (IIS-1344668).

657 **References**

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

660 **Arganda-Carreras I**, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, Sebastian Seung H. Trainable
661 Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics*. 2017 Aug;
662 33(15):2424-2426. <https://academic.oup.com/bioinformatics/article/33/15/2424/3092362>, doi: 10.1093/bioinformatics/btx180.

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

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

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

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

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

673 **Bothma JP**, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. Dynamic regulation of eve stripe 2 ex-
674 pression reveals transcriptional bursts in living Drosophila embryos. *Proc Natl Acad Sci U S A*. 2014;
675 111(29):10598-10603. doi: 10.1073/pnas.1410022111.

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

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

680 **Bronson JE**, Fei J, Hofman JM, Gonzalez RL, Wiggins CH. Learning rates and states from biophysical time
681 series: A Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*. 2009;
682 97(12):3196-3205. doi: 10.1016/j.bpj.2009.09.031.

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

685 **Choubey S**. Nascent RNA kinetics: Transient and steady state behavior of models of transcription. *Phys Rev E*.
686 2018; 97(2-1):022402. doi: 10.1103/PhysRevE.97.022402.

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

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

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

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

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

697 **Crocker J**, Ilsley GR, Stern DL. Quantitatively predictable control of *Drosophila* transcriptional enhancers in vivo
698 with engineered transcription factors. *Nat Genet*. 2016; 48(3):292-8. doi: [10.1038/ng.3509](https://doi.org/10.1038/ng.3509).

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

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

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

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

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

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

712 **Fakhouri WD**, Ay A, Sayal R, Dresch J, Dayringer E, Arnosti DN. Deciphering a transcriptional regulatory code:
713 modeling short-range repression in the *Drosophila* embryo. *Mol Syst Biol*. 2010; 6:341. doi: [msb200997](https://doi.org/10.1371/journal.msb.200997) [pii]
714 [10.1371/journal.msb.200997](https://doi.org/10.1371/journal.msb.200997).

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

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

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

724 **Fukaya T**, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. *Cell*. 2016; 166(2):358-368. doi:
725 [10.1016/j.cell.2016.05.025](https://doi.org/10.1016/j.cell.2016.05.025).

726 **Fukaya T**, Lim B, Levine M. Rapid Rates of Pol II Elongation in the *Drosophila* Embryo. *Curr Biol*. 2017;
727 27(9):1387-1391. doi: [10.1016/j.cub.2017.03.069](https://doi.org/10.1016/j.cub.2017.03.069).

728 **Garcia HG**, Brewster RC, Phillips R. Using synthetic biology to make cells tomorrow's test tubes. *Integr Biol*
729 (Camb). 2016; 8(4):431-50. doi: [10.1039/c6ib00006a](https://doi.org/10.1039/c6ib00006a).

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

732 **Garcia HG**, Gregor T. In: Gaspar I, editor. *Live Imaging of mRNA Synthesis in Drosophila* New York, NY: Springer
733 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](https://doi.org/10.1007/978-1-4939-7213-5_23).

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

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

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

740 **Gregor T**, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and nuclear dynamics of the bicoid mor-
741 phogen gradient. *Cell*. 2007; 130(1):141-52.

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

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

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

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

752 **Holloway DM**, Spirov AV. Transcriptional bursting in *Drosophila* development: Stochastic dynamics of eve stripe
753 2 expression. *PLoS One*. 2017; 12(4):e0176228. doi: [10.1371/journal.pone.0176228](https://doi.org/10.1371/journal.pone.0176228).

754 **Ilsley GR**, Fisher J, Apweiler R, De Pace AH, Luscombe NM. Cellular resolution models for even skipped
755 regulation in the entire *Drosophila* embryo. *eLife*. 2013 aug; 2:e00522. <http://www.ncbi.nlm.nih.gov/pubmed/23930223> http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=23930223&dopt=Abstract doi: [10.7554/eLife.00522](https://doi.org/10.7554/eLife.00522).

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

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

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

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

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

768 **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](https://doi.org/10.1126/science.1202142).

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

773 **Levine M**, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. *Cell*. 2014;
774 157(1):13–25. doi: [10.1016/j.cell.2014.02.009](https://doi.org/10.1016/j.cell.2014.02.009).

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

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

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

783 **Lucas T**, Ferraro T, Roelens B, De Las Heras Chanes J, Walczak AM, Coppey M, Dostatni N. Live imaging
784 of bicoid-dependent transcription in *Drosophila* embryos. *Curr Biol*. 2013; 23(21):2135–9. doi:
785 [10.1016/j.cub.2013.08.053](https://doi.org/10.1016/j.cub.2013.08.053).

786 **Luengo Hendriks CL**, Keranen SV, Fowlkes CC, Simirenko L, Weber GH, DePace AH, Henriquez C, Kaszuba DW,
787 Hamann B, Eisen MB, Malik J, Sudar D, Biggin MD, Knowles DW. Three-dimensional morphology and gene
788 expression in the *Drosophila* blastoderm at cellular resolution I: data acquisition pipeline. *Genome Biology*.
789 2006; 7(12):R123. doi: [10.1186/gb-2006-7-12-r123](https://doi.org/10.1186/gb-2006-7-12-r123).

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

792 **McKnight SL**, Miller J O L. Post-replicative nonribosomal transcription units in *D. melanogaster* embryos. *Cell*.
793 1979; 17(3):551–63.

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

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

797 **Morisaki T**, Lyon K, DeLuca KF, DeLuca JG, English BP, Zhang Z, Lavis LD, Grimm JB, Viswanathan S, Looger LL,
798 Lionnet T, Stasevich TJ. Real-time quantification of single RNA translation dynamics in living cells. *Science*.
799 2016; 352(6292):1425–9. doi: [10.1126/science.aaf0899](https://doi.org/10.1126/science.aaf0899).

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

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

806 **Park J**, Estrada J, Johnson G, Ricci-Tam C, Bragdon M, Shulgina Y, Cha A, Gunawardena J, DePace AH. Dissecting
807 the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. *bioRxiv*.
808 2018; doi: [10.1101/408708](https://doi.org/10.1101/408708).

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

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

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

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

817 **Phillips R**. Theory in Biology: Figure 1 or Figure 7? *Trends Cell Biol*. 2015; doi: [10.1016/j.tcb.2015.10.007](https://doi.org/10.1016/j.tcb.2015.10.007).

818 **Phillips R**, Kondev J, Theriot J, Garcia HG. Physical Biology of the Cell, 2nd Edition. New York: Garland Science;
819 2013.

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

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

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

826 **Sanchez A**, Kondev J. Transcriptional control of noise in gene expression. *Proc Natl Acad Sci U S A*. 2008;
827 105(13):5081–6.

828 **Sasai M**, Wolynes PG. Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci U S A*. 2003;
829 100(5):2374–9. doi: [10.1073/pnas.2627987100](https://doi.org/10.1073/pnas.2627987100).

830 **Sayal R**, Dresch JM, Pushel I, Taylor BR, Arnosti DN. Quantitative perturbation-based analysis of gene expression
831 predicts enhancer activity in early *Drosophila* embryo. *Elife*. 2016; 5. doi: [10.7554/elife.08445](https://doi.org/10.7554/elife.08445).

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

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

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

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

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

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

847 **Skinner SO**, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. Single-cell analysis of transcription kinetics
848 across the cell cycle. *Elife*. 2016; 5:e12175. doi: [10.7554/elife.12175](https://doi.org/10.7554/elife.12175).

849 **Small S**, Blair A, Levine M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J*. 1992;
850 11(11):4047–57.

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

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

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

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

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

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

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

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

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

870 **Wu B**, Eliscovich C, Yoon YJ, Singer RH. Translation dynamics of single mRNAs in live cells and neurons. *Science*.
871 2016; 352(6292):1430–5. doi: [10.1126/science.aaf1084](https://doi.org/10.1126/science.aaf1084).

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

874 **Xu H**, Skinner SO, Sokac AM, Golding I. Stochastic Kinetics of Nascent RNA. *Phys Rev Lett*. 2016; 117(12). doi:
875 [10.1103/PhysRevLett.117.128101](https://doi.org/10.1103/PhysRevLett.117.128101).

876 **Yan X**, Hoek TA, Vale RD, Tanenbaum ME. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell*. 2016;
877 165(4):976–89. doi: [10.1016/j.cell.2016.04.034](https://doi.org/10.1016/j.cell.2016.04.034).

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

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

882 **Zoller B**, Little SC, Gregor T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional
883 Bursting. *Cell*. 2018; 175(3):835–847 e25. doi: [10.1016/j.cell.2018.09.056](https://doi.org/10.1016/j.cell.2018.09.056).

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

886 Appendix 1

887 888 Theoretical model to predict cytoplasmic mRNA levels given from *in* 889 *vivo* measurements of transcriptional activity

890 891 Derivation details

892 893 894 895 896 897 898 899 900 890 891 892 893 894 895 896 897 898 899 900 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

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

901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 where γ is the mRNA degradation rate constant. For a promoter that has entered a transcriptionally quiescent state ($t > t_{\text{off}}$), we have

$$\frac{dmRNA}{dt}(x, t) = -\gamma mRNA(x, t), \quad (13)$$

901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 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 $6 \mu\text{m}$, which corresponds to one nuclear diameter or 1% of the embryo length, before being degraded. Thus, given the overall width of the stripe mRNA profile of about 8% of the embryo length (*Figure 3G*), we expect diffusion to play a minimal role in stripe formation. Finally, note that we are also ignoring the delay between transcriptional initiation and the delivery of an mRNA molecule to the cytoplasm as a result of nuclear export. This delay would affect the *timing* of pattern formation, but would leave our conclusions about the relative role of transcriptional bursting and the regulation of the duration of the transcriptional time window unaffected.

901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 To make progress, as in the main text, we make the simplifying assumption that the instantaneous rate of transcription can be well approximated by the time average at each position given by

$$R(x) \approx \langle R(x, t) \rangle_t. \quad (14)$$

901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 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

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

901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 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

$$mRNA_{\text{active}}(x, t) = \frac{R(x)}{\gamma}, \quad (16)$$

932

933

934

935

936

937

938

939

940

941

where mRNA production and degradation are balanced.

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

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

942

943

944

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

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

945

946

947

948

949

which can be simplified slightly to yield

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

950

951

952

953

954

955

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

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

956

957

958

959

960

961

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

962

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

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

963

964

965

966

967

968

969

970

971

972

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

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

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

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

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:

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

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

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

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:

$$0 = -\pi_0 k_{01} + \pi_1 k_{10} \quad (25)$$

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

$$0 = \pi_1 k_{12} - \pi_2 k_{21} \quad (27)$$

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

$$1 = \pi_0 + \pi_1 + \pi_2. \quad (28)$$

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

$$\pi_1 k_{10} = \pi_0 k_{01} \quad (29)$$

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}}. \quad (30)$$

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

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

1011

1012

1013

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

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

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

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

$$\pi_0 = \frac{k_{10} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \quad (34)$$

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**:

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

And finally π_2 :

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

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

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

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,

$$\kappa = k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}. \quad (38)$$

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,

$$k_{\text{on}}^{\text{eff}} = k_{01}. \quad (39)$$

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:

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

$$= \frac{k_{01} k_{21} k_{10}}{k_{01} k_{21} + k_{01} k_{12}} \quad (41)$$

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

$$= \frac{k_{21}k_{10}}{k_{21} + k_{12}}. \quad (42)$$

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

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

$$= \frac{r_1 k_{01} k_{21} + r_2 k_{01} k_{12}}{k_{01} k_{21} + k_{01} k_{12}} \quad (44)$$

$$= \frac{r_1 k_{21} + r_2 k_{12}}{k_{21} + k_{12}}. \quad (45)$$

1065 **Appendix 2**

1066

Measuring the amount of produced mRNA

1067

1068

1069

1070

1071

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 3G** (red), as well as the “binary control” of the duration of the transcriptional time window contribution **Figure 3G** (blue).

1072

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,

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

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 $t - \frac{\tau_{\text{elong}}}{2}$

$$F\left(t - \frac{\tau_{\text{elong}}}{2}\right) \approx \beta \frac{dM(t)}{dt}. \quad (47)$$

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 3G**, 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

$$\begin{aligned} \frac{dM(x, t)}{dt} &= \frac{\beta}{N} \sum_{i=1}^N F_i(x, t) \\ &= \beta \langle F(t) \rangle_x. \end{aligned} \quad (48)$$

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

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129

1130

1131

1132

1133

1134

1135

1136

1137

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

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

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 3G** (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

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

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

$$\text{mRNA}_{\text{full}}(x_j, t) = \text{mRNA}_{\text{norm}}(x_j, t) - \min_x (\text{mRNA}_{\text{norm}}(x_j, t)). \quad (51)$$

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

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

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

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

nuclei within a region

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

The steps for calculating the the normalized binary mRNA levels comprising the blue profile in **Figure 3G** 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

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

1209

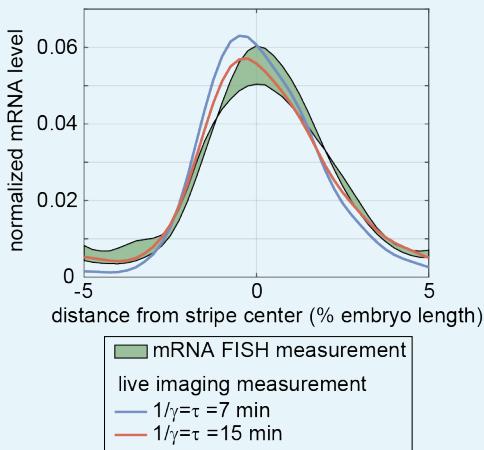
1210

1211

1212

1213

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.



1214

1215

1216

1217

1218

1219

1220

1221

1222

1223

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.

1225

1226

1227

1228

1229

1230

1231

1232

1233

Sensitivity of results to mRNA lifetime assumption

In the main text we assume a degradation rate for *eve* of 0.14 min^{-1} (corresponding to a lifetime of roughly $\tau = 7 \text{ min}$). Since, to our knowledge, the decay rate of *eve* mRNA has not been measured directly, we follow *Bothma et al. (2014)* and base this estimate off of measurements for another of the pair rule genes, *fushi tarazu* (*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 ($\tau = 1 \text{ min}$) 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, somewhat, on the assumed decay rate. In the case of rapid mRNA decay, as well as for the

1237

Manuscript submitted to bioRxiv

1238

1239

1240

1241

1242

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260

1261

1262

1263

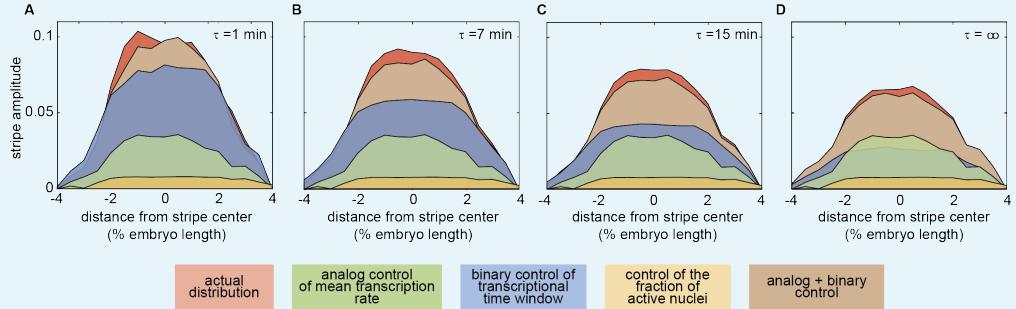
1264

1265

1266

decay rate assumed in the main text, the time window (blue profile) is clearly the dominant factor in driving pattern formation (**Appendix 2—Figure 2A** 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 2C**). 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 2D**). 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 joint action of mean rate modulation (green profile) and the time window (blue profile) remains intact. As expected, the relative contribution of the time window is sensitive to the assumed τ , yet even in the limit of no significant mRNA decay, its impact is still of order with the effect of mean rate modulation.

Control strategy contributions for eve BAC

A key question regarding the results in the main text is whether and to what degree the relative contributions of the regulatory control strategies we identified in **Figure 1** and **Figure 3** for the reporter containing only the *eve* stripe 2 enhancer hold true for the formation of the stripe in the endogenous context. While we cannot directly query activity at the endogenous *eve* locus, we were able to examine the dynamics of stripe formation for an *eve* BAC used in the companion paper to this manuscript (**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 1G**). Interestingly, unlike the reporter construct, the full predicted profile (red profile) that accounts for the interplay between mRNA decay temporal fluctuations in the mean rate of mRNA production differs

1279

1280

1281

1282

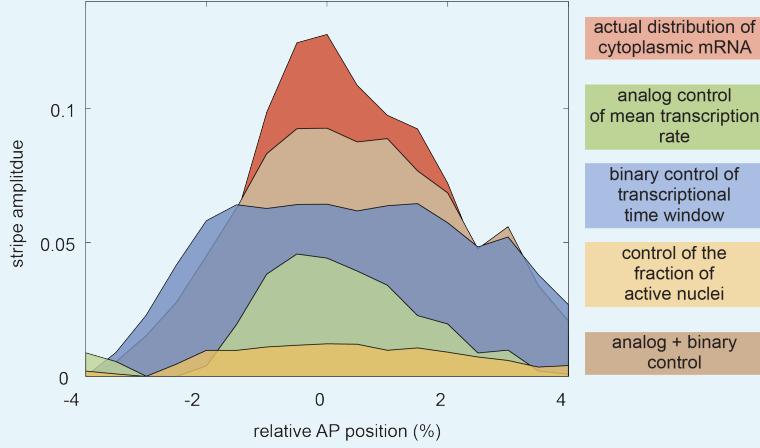
1283

1284

1285

1286

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.



1287

1288

1289

1290

1291

1292

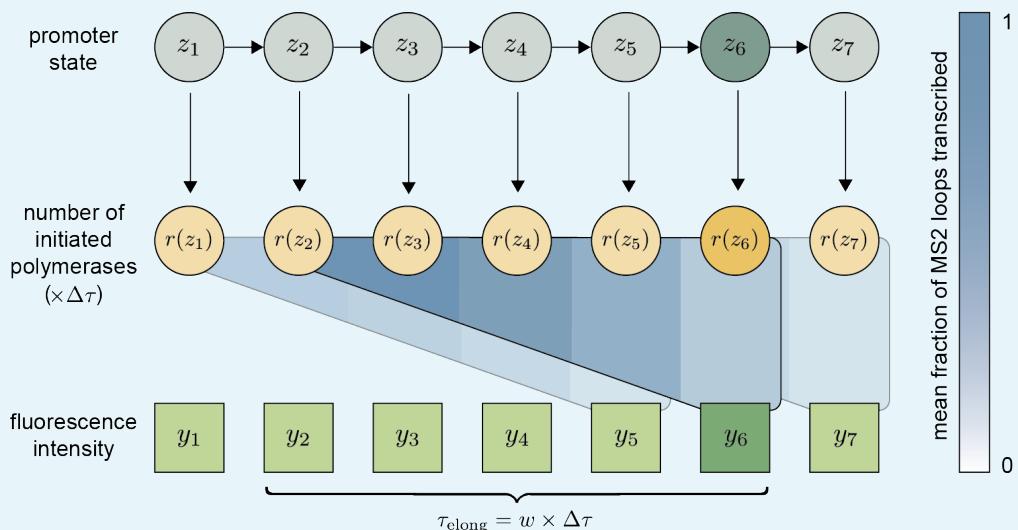
Appendix 2 Figure 3. Regulatory strategy contributions to eve stripe 2 formation in endogenous context. As with the reporter construct, the formation of eve stripe 2 in the context of the full eve regulatory locus is dominated by the interplay between mean rate modulation (green) and control of the time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

Appendix 3

The memory-adjusted hidden Markov model

Model introduction

To model the dynamics of an observed fluorescence series, $\mathbf{y} = \{y_1, y_2, \dots, y_T\}$, where T is the number of data points in a trace, we assume that, at each time step, the sister promoters can be in one of K effective states. In the analysis of *eve* stripe 2 data, we use a simple model with the number of effective states equal to three ($K = 3$). The method, however, allows for more complex transcription architectures with higher numbers of states. Transitions between the effective promoter states are assumed to be Markovian, meaning that the hidden promoter state z_t at time step t is conditionally dependent only on the state in the previous time step. This dependency is modeled through a $K \times K$ transition probability matrix $\mathbf{A} = p(z_t | z_{t-1})$, where A_{kl} is the probability of transitioning from the l^{th} state into the k^{th} state in the time interval $\Delta\tau$, where $\Delta\tau$ is the data sampling resolution. We assign a characteristic 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 $v(k) = F_{\text{RNAP}}r(k)\Delta\tau$, $1 \leq k \leq K$, where F_{RNAP} is the fluorescence calibration factor determined using smFISH experiments (see Materials and Methods).



Appendix 3 Figure 1. Schematic overview of the mHMM architecture. The sister promoters are modeled as undergoing a series of Markovian transitions between effective transcriptional states (z_i). Each promoter state uniquely determines the number of polymerases initiated in a single time step ($r(z_i)\Delta t$). 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).

1328 The instantaneous fluorescence intensity is the cumulative contribution from polymerases
1329 initiated in the previous w time steps, where $w = \tau_{\text{elong}}/\Delta\tau$ is the system-dependent integer
1330 memory. Here $\Delta\tau$ indicates the observational time resolution, a quantity set by experimental
1331 conditions. The time required for an RNA polymerase molecule to transcribe our reporter
1332 gene (τ_{elong}) is *a priori* unknown. We developed an autocorrelation-based method to estimate
1333 τ_{elong} directly from our experimental data (see **Appendix 9** and **Coulon and Larson (2016)**).
1334 The observation y_t at time step t conditionally depends not only on the hidden promoter state
1335 z_t , but also on the hidden states in the previous w time steps, $\{z_t, z_{t-1}, \dots, z_{t-w+1}\}$. To be able to
1336 describe the observed system dynamics through a hidden Markov model, the observation at
1337 time step t needs to be conditionally independent from the states at earlier time steps. We
1338 therefore introduce the concept of a compound state, $s_t = \{z_t, z_{t-1}, \dots, z_{t-w+1}\}$, which, together
1339 with the set of model parameters, θ , is sufficient to define the probability distribution of the
1340 observation y_t , thereby satisfying the Markov condition. Since $z_t \in \{1, \dots, K\}$, each compound
1341 state can take one of K^w different values, $s_t \in \{1, \dots, K^w\}$. While the number of possible
1342 compound states is K^w , only K different transitions are allowed between them, since the
1343 most recent $w - 1$ promoter states are deterministically passed from one compound state to
1344 the next, i.e. the last $w - 1$ elements in $s_{t+1} = \{z_{t+1}, z_t, \dots, z_{t-w+2}\}$ are present in s_t as well. The
1345 schematic overview of the mHMM architecture is shown in **Appendix 3-Figure 1**.

1346 We model the fluorescence emission probabilities corresponding to each hidden com-
1347 pound state as Gaussian distributions with a standard deviation σ , which we learn during
1348 inference. The joint probability distribution $p(\mathbf{y}, \mathbf{s}|\theta)$ of the series of hidden compound states,
1349 $\mathbf{s} = \{s_1, s_2, \dots, s_T\}$, and fluorescence values, $\mathbf{y} = \{y_1, y_2, \dots, y_T\}$, is given by

$$1350 \quad p(\mathbf{y}, \mathbf{s}|\theta) = p(s_1|\boldsymbol{\pi}) \prod_{t=1}^T p(y_t|s_t, \boldsymbol{\nu}, \sigma) \prod_{t=2}^T p(s_t|s_{t-1}, \mathbf{A}). \quad (54)$$

1353 Here $\boldsymbol{\pi}$ is a K -element vector, with π_k being the probability that the trace starts at the k^{th}
1354 effective promoter state, and $\boldsymbol{\nu}$ is a K -element vector of fluorescence emission values per
1355 time step.

1357 Our goal is to find an estimate of the model parameters, $\hat{\theta} = \{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\nu}}, \hat{\mathbf{A}}, \hat{\sigma}\}$, which maximizes
1358 the likelihood $p(\mathbf{y}|\theta)$ of observing the fluorescence data, namely,

$$1359 \quad \hat{\theta} = \underset{\theta}{\operatorname{argmax}} p(\mathbf{y}|\theta). \quad (55)$$

1361 The likelihood can be obtained by marginalizing the joint probability distribution, $p(\mathbf{y}, \mathbf{s}|\theta)$,
1362 over the hidden compound states, that is,

$$1364 \quad p(\mathbf{y}|\theta) = \sum_{\mathbf{s}=\{s_1, s_2, \dots, s_T\}} p(\mathbf{y}, \mathbf{s}|\theta). \quad (56)$$

1366 Note that the summation is performed over all possible choices of \mathbf{s} — a vector of T elements,
1367 each of which can take K^w possible values. The total number of terms in the sum is thus
1368 equal to K^{wT} , which grows exponentially with the number of time points. To make the
1369 estimation of the model parameters tractable, we use an approximate inference method,
1370 the expectation-maximization (EM) algorithm.

1371 We note that the notion of a compound state was also introduced in an earlier work
1372 (**Corrigan et al., 2016**) to account for the memory effect in hidden Markov modeling of *actin*
1373 transcription and then an EM methodology was applied to learn the kinetic parameters from
1374 MS2-based transcription data. Unlike their approach, however, we do not explicitly model
1375 the recruitment of individual RNA polymerase molecules, but instead, assign a continuous
1376 RNA polymerase initiation rate to each promoter state. Additionally, our model estimates the
1377 magnitude of the background noise present in the experimentally measured fluorescence

1375

1376

1377

1378

1379

1380

1381

1382

1383

1384

1385

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 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.

1386

Expectation-maximization (EM) algorithm

1387

1388

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,

1389

1390

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

1391

1392

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

1393

1394

1395

1396

1397

1398

1399

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

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

1400

Now, we introduce several notations: $s_t^i := 1$ if and only if $s_t = i$; $\Delta(s_t, d) :=$ the d^{th} digit of the promoter state sequence $s_t = \{z_t, z_{t-1}, \dots, z_{t-(w-1)}\}$, starting from the left end; $C_{zs} = 1$ if and only if $\Delta(s, 1) = z$; $B_{s',s} = 1$ if and only if the transition $s \rightarrow s'$ between the compound states s and s' is allowed, which happens when the latest $(w-1)$ promoter states in the compound state s match the earliest $(w-1)$ promoter states of the compound state s' . With these notations in hand, the terms in *Equation 59* can be rewritten as

1401

1402

1403

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

1404

1405

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

1406

1407

1408

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

1409

1410

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

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

1411

1412

1413

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

1426

1427

1428

1429

1430

1431

1432

1433

1434

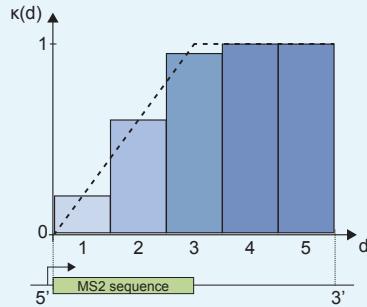
1435

1436

window will be given by

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

where $\lceil n_{MS2} \rceil$ and $\lfloor n_{MS2} \rfloor$ 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: $\tau_{\text{elong}} = 100$ sec, $\tau_{MS2} = 50$ sec, $\Delta\tau = 20$ sec, $w = \tau_{\text{elong}}/\Delta\tau = 5$, $n_{MS2} = \tau_{MS2}/\Delta\tau = 2.5$.

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

$$V_i(\mathbf{v}) = F_{i,:}\mathbf{v}, \quad (63)$$

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

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

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

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

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

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

Using the current estimate of the model parameters, $\hat{\theta}_k$, 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

1445

1446

1447

1448

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

1449

1450

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

1451

1452

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

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

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

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

1453

1454

1455

$$\alpha_t(i) = p(y_1, \dots, y_t, s_t = i | \hat{\theta}_k), \quad (71)$$

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

1456

1457

1458

1459

1460

1461

Here $\alpha_t(i)$ is the joint probability of observing the fluorescence emission values in the first t steps and being at the i^{th} compound state at step t ; while $\beta_t(i)$ is the conditional probability of observing fluorescence values from the time point $(t + 1)$ till the end of the series, given that the compound state at time t is i . Note that α and β can be treated as $K^w \times T$ matrices, where each column is a vector of length K^w , accounting for the K^w possible values of i in **Equation 71** and **Equation 72**. We evaluate the elements of α and β matrices recursively as

1462

1463

1464

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

1465

1466

1467

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

1468

1469

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

1470

1471

1472

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

$$\beta_T(i) = 1, \quad (76)$$

1473

1474

1475

1476

1477

1478

1479

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 $p(\mathbf{y} | \hat{\theta}_k)$ 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,

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

1480

1481

With the probabilities $p(s_t | \mathbf{y}, \hat{\theta}_k)$ and $p(s_{t-1}, s_t | \mathbf{y}, \hat{\theta}_k)$ known, the expectation coefficients follow directly from **Equation 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 $\mathcal{L}(\theta | \mathbf{y}, \hat{\theta}_k)$ in **Equation 64** with respect to $\{\pi, \nu, \lambda, \mathbf{A}\}$, subject to the probability constraints $\sum_{k=1}^K \pi_k = 1$ and $\sum_{k=1}^K \mathbf{A}_{kl} = 1$, $1 \leq l \leq K$. The update equations for the model parameters are found as

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

fluorescence emission rates: $\hat{\boldsymbol{\nu}} = \boldsymbol{M}^{-1} \boldsymbol{b}$, where (79)

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

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

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

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

1489 Pooled inference on multiple traces

1490 Since the information available in a single MS2 fluorescence trace is not sufficient for the
1491 accurate inference of underlying model parameters, we perform a pooled EM inference
1492 assuming that the traces are statistically independent and governed by the same parameters.
1493 If $\mathbf{y}_{1:N}$ are N different fluorescence traces with corresponding trace lengths $T_{1:N}$, and $\mathbf{s}_{1:N}$
1494 are the hidden compound state sequences corresponding to each trace, we obtain

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

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

1495 Therefore, the objective function $\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\theta}}_k)$ maximized at each EM iterations takes the
1496 form
1497

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

1498 From the above equation, we recognize that the objective function for the pooled inference
1499 is the sum of objective functions written for each individual trace. Using the expression
1500 for the single-trace objective function obtained earlier (**Equation 64**), we find
1501

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

1502 where $\langle s_t^i(n) \rangle$ and $\langle s_t^i(n) s_{t-1}^j(n) \rangle$ are now the expectation coefficients obtained for the n^{th}
1503 fluorescence trace via the forward-backward algorithm, and $y_t(n)$ is the fluorescence at time

step t in the n^{th} trace. The update equations are then derived analogous to the single-trace case, with an additional summation performed over all traces, namely,

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

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

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

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

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

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

Execution of the mHMM method

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

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

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, \mathbf{r} , and the Gaussian precision parameter, $\lambda = 1/\sigma^2$, we first treat the fluorescence data $\mathbf{y}_{1:N}$ as identical and independently distributed (i.i.d.) and use a simplified time-independent EM algorithm to find their optimal values (cf. **Bishop (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 $\epsilon = 10^{-4}$ or the number of iterations exceeds 500. Because EM approaches typically infer locally optimal parameter values, the algorithm is run on the same dataset using multiple randomly chosen initial parameters (10-20 in our implementations), and the globally optimal set of values is chosen in the end. In the Matlab implementation of the EM algorithm, the variables are all stored in logarithmic forms to avoid overflow and underflow issues, which could occur when recursively evaluating the elements of the α and β matrices. Also, special care is taken when accounting for time points less than the elongation time, i.e. $t < w$, in which case the compound state is a collection of not w , but t promoter states, i.e. $s_t = \{z_t, z_{t-1}, \dots, z_1\}$.

1546 Because of the exponential scaling of the model complexity with the integer memory
 1547 window ($w = 7$ for the eve construct with $\Delta\tau = 20$ sec data sampling resolution), significant
 1548 computational resources were used when conducting inference on simulated and experi-
 1549 mental data. It took approximately 2 hours to conduct 25 mHMM inferences with different
 1550 initialization conditions on a machine with 24 CPU cores. Users of the mHMM method are
 1551 advised to have this metric as a reference when estimating the computational cost of their
 1552 inference.

1553 Windowed mHMM

1554 To investigate temporal trends in bursting parameters, we extended the mHMM method to
 1555 allow for a sliding window inference approach. From a technical perspective, this required a
 1556 revision of the inference formalism to be compatible with fragments of fluorescent traces in
 1557 which the beginning of the trace (initial rise in y_t from $t = 1$) was not included.

1558 To that end, we modified the first term in **Equation 59** to allow for all possible promoter
 1559 state sequences that could lead to the observation of the first fluorescence measurement in
 1560 the chosen time window ($[T_1, T_2]$), namely,

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

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

1561 Here $\boldsymbol{\pi}^{(T_1-w+1)}$ is the probability distribution of the earliest promoter state that still has an
 1562 impact on the observation of the first measurement in the sliding window, and D_{ni}^d is an
 1563 indicator variable which takes the value 1 only if the promoter state in the d^{th} position of the
 1564 i^{th} compound state is n .

1565 The modified expression for the joint probability distribution does not change the func-
 1566 tional form of the equations used for calculating the expectation coefficients. Maximization
 1567 equations for the emission rates and the noise also remain intact. Only the maximization
 1568 equation for the transition probabilities is revised from **Equation 83** into

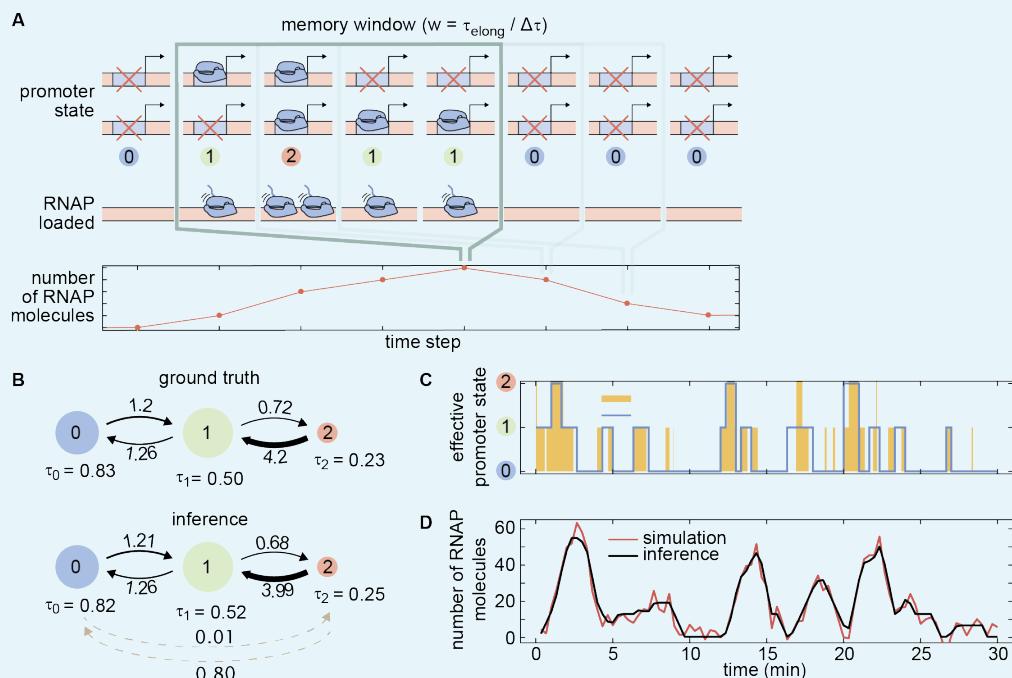
$$\hat{A}_{mn} = \frac{\sum_{i=1}^{T_2} \sum_{j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{mi} C_{nj} + \sum_{i=1}^{K^w} \sum_{d=2}^w \langle s_{T_1}^i \rangle D_{mi}^{d-1} D_{ni}^d \log A_{mn}}{\sum_{k=1}^K \sum_{i=1}^{T_2} \sum_{j=1}^{K^w} B_{ij} \langle s_t^i s_{t-1}^j \rangle C_{ki} C_{nj} + \sum_{k=1}^K \sum_{i=1}^{K^w} \sum_{d=2}^w \langle s_{T_1}^i \rangle D_{ki}^{d-1} D_{ni}^d \log A_{kn}}. \quad (97)$$

1569 We make a steady-state assumption within the sliding window and choose $\boldsymbol{\pi}^{(T_1-w+1)}$ to be the
 1570 stationary distribution of the current transition probability matrix, i.e. $\mathbf{A} \boldsymbol{\pi}^{(T_1-w+1)} = \boldsymbol{\pi}^{(T_1-w+1)}$.
 1571 We therefore use the current estimate of \mathbf{A} to evaluate $\boldsymbol{\pi}^{(T_1-w+1)}$ at each EM iteration, instead
 1572 of performing a maximization step.

1579 Appendix 4

1580 Statistical validation of mHMM

1581 We validated mHMM for the three-state ($K = 3$) architecture schematically illustrated in
1582 **Appendix 4-Figure 1A** by generating synthetic trajectories of effective promoter states using
1583 the Gillespie algorithm (Gillespie, 1976) and adding Gaussian noise to the resulting activity
1584 traces. Parameters in **Appendix 4-Table 1** were used for data generation. Pooled inferences
1585 were conducted on 20 independent datasets, each containing 9,000 data points,
1586 representative of the number of experimental data points in a central stripe region. The
1587 top panel of **Appendix 4-Figure 1B** shows the kinetic architecture used to simulate the
1588 promoter trajectory in **Appendix 4-Figure 1C** (yellow) as it switches through the multiple
1589 possible states. This promoter trajectory leads to the simulated trace of the number of RNA
1590 polymerase molecules actively transcribing the gene in **Appendix 4-Figure 1D** (red). Using
1591 mHMM, we found the best fitted path for our observable (**Appendix 4-Figure 1D**, black) and
1592 the corresponding most likely promoter state trajectory (**Appendix 4-Figure 1C**, blue).



1593 **Appendix 4 Figure 1. Statistical validation of mHMM.** **(A)** Three-state mHMM architecture where
1594 ON and OFF promoter states on each sister chromatid result in an effective three-state model. The
1595 trajectory of effective promoter states over the memory time window given by the elongation time
1596 dictates the number of RNA polymerase molecules loaded onto the gene. **(B)** Flow diagrams of
1597 promoter states and transition rates for the true parameters used to simulate trajectories (top) and
1598 corresponding average inference results obtained from 20 independent datasets (bottom). The area of
1599 each state circle is proportional to the relative state occupancy, and the thickness of the arrows is
1600 proportional to the transition rates. Dashed lines correspond to inferred transitions with very slow rates
1601 that were absent in the simulation. Rates are in min^{-1} and dwell times are in min. Error bars for the
1602 mean inferred parameters are shown in **Appendix 4-Figure 2**. **(C)** Sample simulated promoter activity
1603 trace (yellow) generated using the parameters in (B), overlaid with the best fitted trace (blue) obtained
1604 using the Viterbi algorithm (Viterbi, 1967). **(D)** Simulated and best fitted observable number of RNA
1605 polymerase molecules corresponding to the promoter trajectory shown in (C).

1608 **Appendix 4 Table 1.** Parameter values used for generating synthetic datasets in the statistical
1609 validation of the model. In order to perform this validation, we chose parameters that approximated
1610 those obtained through the mHMM inference on experimental data shown in **Figure 5**.

1612

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

1613

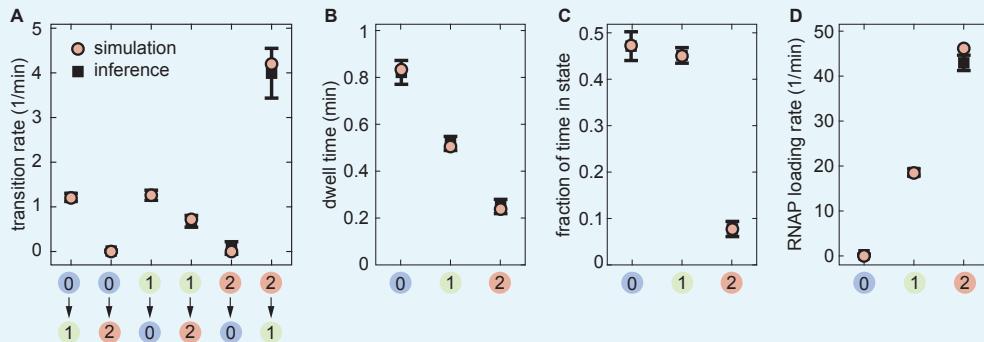
1614

1615

1616

1617

As shown in **Appendix 4–Figure 1B** 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.



1618

1619

1620

1621

1622

1623

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).

1625

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,

$$\mathbf{A} = e^{\mathbf{R}\Delta\tau}, \quad (98)$$

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

1628

1629

1630

1631

1632

1633

1634

1635

1636

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.

1637

Continuous vs. Poisson promoter loading

1638

1639

1640

1641

1642

1643

1644

1645

1646

1647

1648

1649

1650

1651

1652

1653

1654

1655

1656

1657

1658

1659

1660

1661

1662

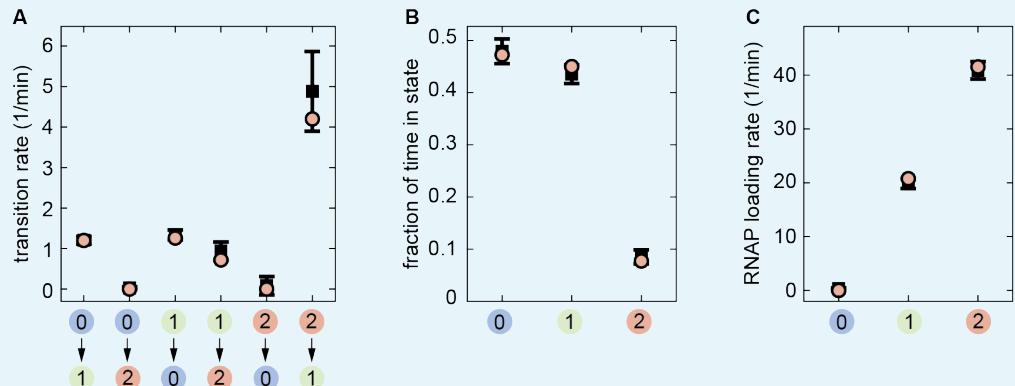
1663

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 ($v_{\text{elong}} = 46 \text{ bp/sec}$, *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 $l_{\text{RNAP}} = 50 \text{ bp}$ (*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,

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

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.



1664

1665

1666

1667

1668

1670

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

1673

1674

1675

1676

1677

1678

1679

1680

1681

1682

1683

1684

1685

1686

1687

1688

1689

1690

1691

1692

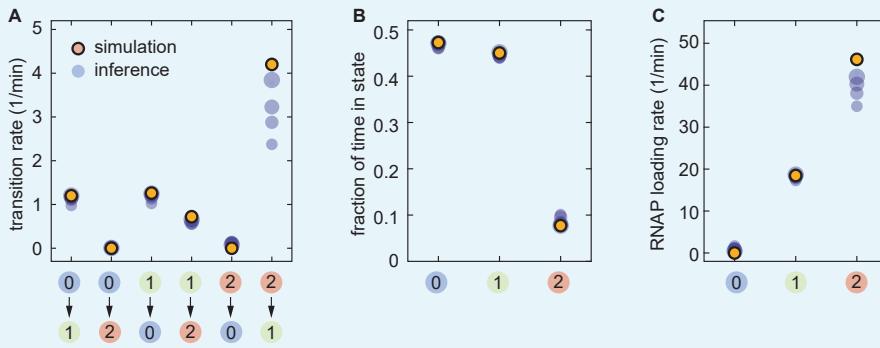
1693

1694

1695

model might be unable to capture all those transitions. To study this possible limitation of mHMM, we conducted inference on synthetic activity traces generated with varying sampling rates. Since the system memory ($w = \tau_{\text{elong}}/\Delta\tau$) needs to be an integer, we varied w in the [3, 7] range, correspondingly changing the sampling resolution from low ($\tau_{\text{elong}}/3 \approx 46\text{s}$) to high ($\tau_{\text{elong}}/7 = 20\text{s}$). We used the values in **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 4A**). The accuracy of inferred state occupancies is also remarkably high, making it robust to variations in the data sampling rate (**Appendix 4-Figure 4B**). 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 4C**). 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.



1696

1697

1698

1699

1700

1701

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).

1703

1704

1705

1706

1707

1708

1709

1710

1711

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 5A-C**), high ON rates and low OFF rates (**Appendix 4-Figure 5D-F**), and high ON rates and high OFF rates (**Appendix 4-Figure 5G-I**).

As expected, the inference is the most accurate when the data sampling frequency is greater than the transition rates (**Appendix 4-Figure 5A-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

1714

1715

1716

1717

1718

1719

1720

1721

1722

1723

1724

1725

1726

1727

1728

1729

1730

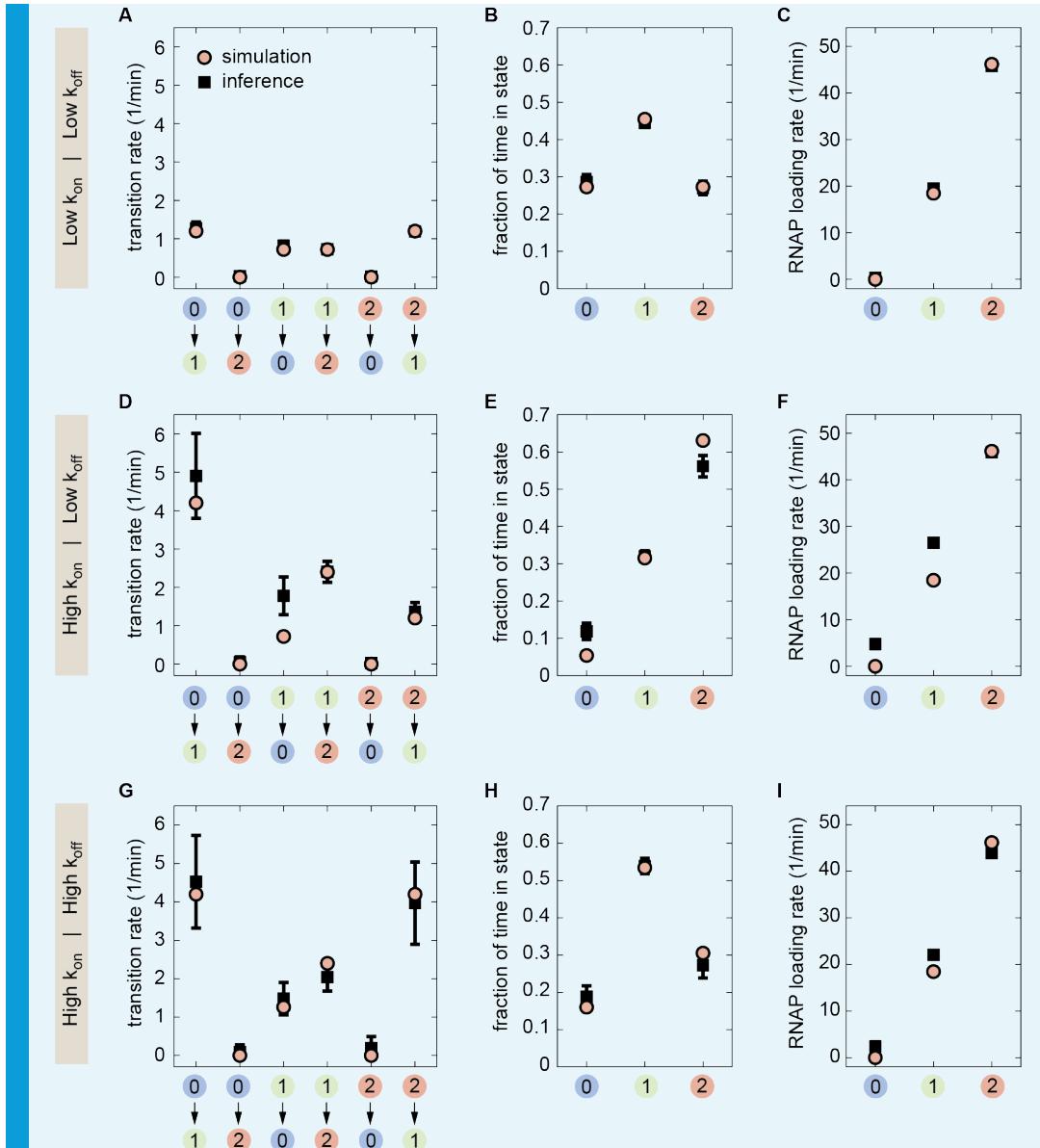
1731

1732

1733

parameters from their ground truth values occur when the ON rates are high and the OFF rates are low (**Appendix 4–Figure 5D–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 5F**). 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 5D**). Remarkably, the deviations caused by the high ON rates get substantially suppressed when the OFF rates are also made comparably high (**Appendix 3–Figure 5G–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.



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.)

1734
1735
1736
1737
1738
1739

1741
1742
1743
1744
1745
1746
1747
1748

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 exhibited the desired parameter trends. Consistent with our approach to the experimental

1749

1750

1751

1752

1753

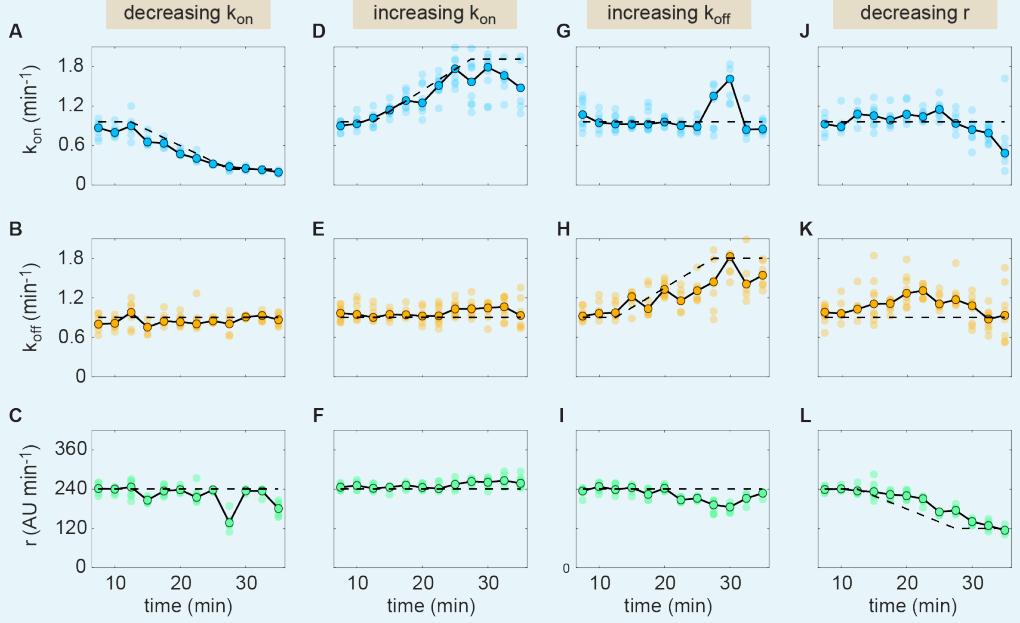
1754

1755

1756

1757

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 ($k_{\text{on}} = k_{01}/2$, $k_{\text{off}} = k_{10}$ and $r = r_1$ in **Appendix 5–Figure 2A**), as these were found to provide the most faithful indication of underlying system trends.



1758

1759

1760

1761

1762

1763

1764

1765

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.

1766

1767

1768

1769

1770

1771

1772

1773

1774

1775

1776

1777

1778

1779

1780

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 6A-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 6G-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 6J-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 $\tau_{\text{elong}} = 140$ 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 6J, 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

1781

1782

1783

1784

1785

1786

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.

1787 **Appendix 5**

1788

1789

1811

1812

1813

1814

1815

1816

1817

1818

1819

1820

1821

1822

1823

1824

1825

1826

1827

1828

1829

1830

1831

1832

1791

1833

1792

1834

1895

1896

1837

1797

1838

1898

1899

1800

1801

1802

1803

1804

1805

1806

1807

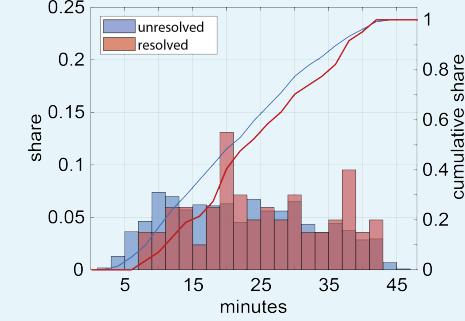
1808

1810

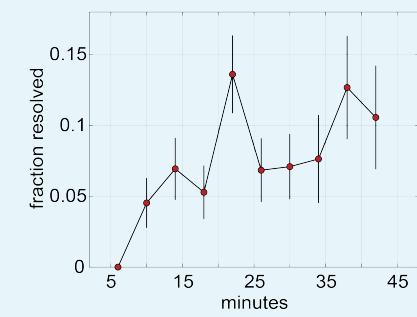
Sister chromatids

Detection of sister chromatid appearance

A



B



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

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 2B**). 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 1B** 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

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

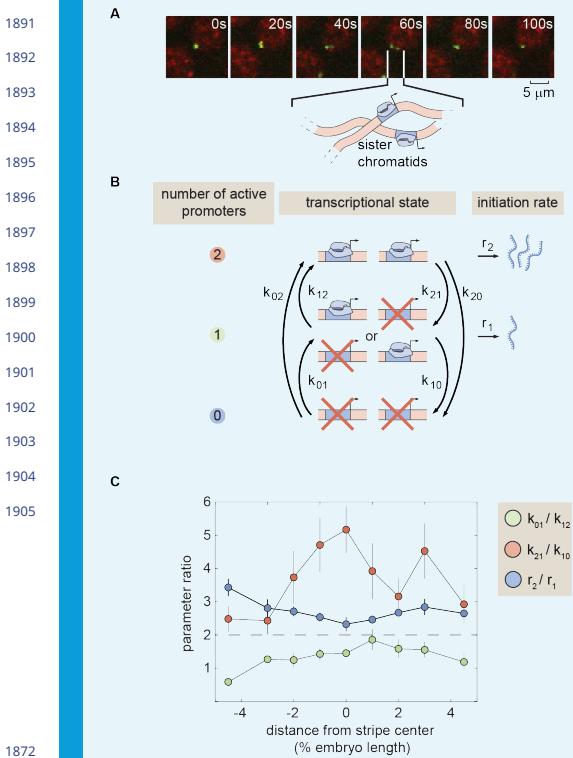
As indicated in **Appendix 5-Figure 1A**, we see evidence for resolved puncta by around 7 minutes into nuclear cycle 14. This is well within the average range for turn-on times observed throughout the stripe (see **Figure 3-Figure Supplement 2B**). Our results indicate that, at the very least, the genomic region containing our eve stripe 2 reporter is replicated within *some* nuclei by 6-8 minutes into nuclear cycle 14. **Appendix 5-Figure 1B** 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

1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870

to result in a progressive increase in this metric over time. However, such a trend is not evident. While we see *no* resolved sister loci between 4 and 8 minutes (first point in the plot in **Appendix 5-Figure 1B**), 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 1B** 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.

1871 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905

Probing for interactions between sister chromatids



1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889

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 4D**), then it is necessary to revisit the widely used two-state model of transcriptional bursting. In this revised scenario, each promoter on one of the sister chromatids undergoes fast ON/OFF switching. Therefore, each spot (encompassing two identical loci) can be in one of three states: (0) both promoters OFF, (1) one promoter ON and the other OFF, and (2) both promoters ON (**Appendix 5-Figure 2B**). 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 2A**. First, if these two promoters transcribe independently, then state (2) will have double the loading rate of state (1) such that $r_2 = 2r_1$. Second, the probability of both promoters transitioning simultaneously should be negligible; we expect no transitions between states (0) and (2) such that $k_{02} = k_{20} = 0$. Finally, if the promoters switch between their states in an independent manner, then there will be an extra constraint on their transition rates. For example, there are two paths to transition from (0) to (1) as either promoter can turn on in this case.

1910

1911

1912

1913

1914

1915

1916

1917

1918

1919

1920

1921

1922

1923

1924

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

1960

1961

However, there is only one possible trajectory from (1) to (2) because only one promoter has to turn on. This condition sets the constraint $k_{01} = 2k_{12}$. Similarly, $k_{10} = k_{21}/2$.

While the independence of sister chromatids is supported by recent single-molecule FISH experiments (*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 2B*, blue). 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 2C*, 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 2C*, 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.

1962 **Appendix 6**

1963

mHMM inference sensitivities

1964

Full three-state inference results

1965

1966 For the sake of simplicity, we presented our inference results in the main text using an effective two-state model in which two distinct active transcriptional states were combined into a single effective ON state (see **Figure 4E** and **F**). Here, for completeness, we include time-averaged and time-resolved inference results for the full three-state model where, as shown in **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.

1967

1968 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 1D**). 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 1E**) are strongly elevated in the stripe center.

1969

1970

1971

1972

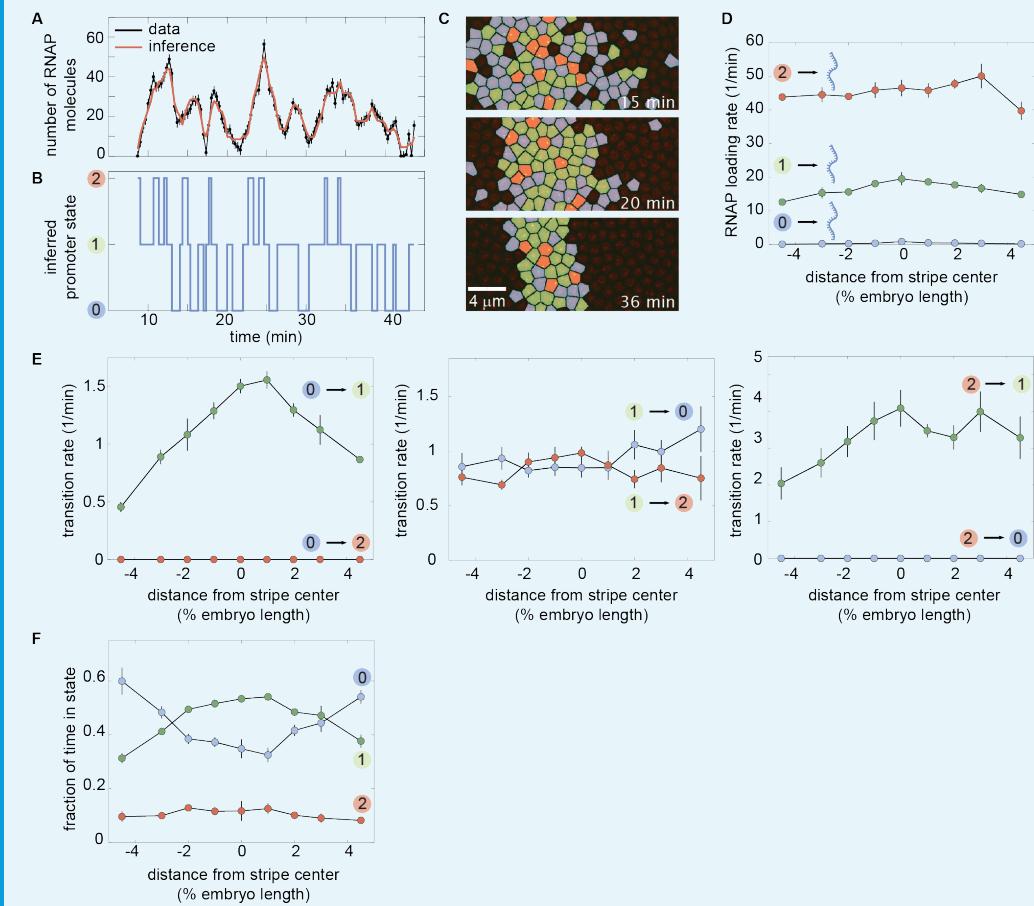
1973

1974

1975

1976

1977



1978

1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992

1993
1994
1995
1996

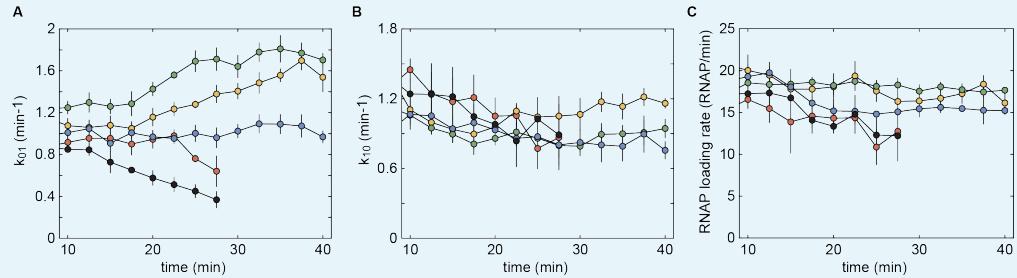
1997
1998
1999
2000
2001
2002
2003

2005
2006
2007
2008
2009
2010
2011
2012
2013

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 4D-F**). Due to a lack of statistics for state (2), we show only transition rates into and out of the first active state (middle state in **Figure 4E**).



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 4B**). Note that this approach is distinct from the *effective* two-state results presented in the main text. There, as outlined in **Figure 4D-F**, a three-state model was specified for inference and the results for the two active (ON) states were aggregated after the fact to simplify the presentation of the results. Conversely, here, we explicitly conducted inference using a two-state model.

Most of our findings remained unchanged in the context of the two-state model. Consistent with the three-state case, the two-state time-averaged 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 3A-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 3C**, 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 3C**, red). In general, this is not surprising, as our use of a simpler model likely means that multiple switching rates are being projected onto the k_{off} parameter. Specifically, if the *eve* stripe 2 system is indeed a true three-state system, then

2021

2022

2023

2024

2025

2026

2027

2028

2029

2030

2031

2032

2033

2034

2035

2036

2037

2038

2039

2040

2041

2042

2043

2044

2045

2046

2047

2048

2049

2050

2051

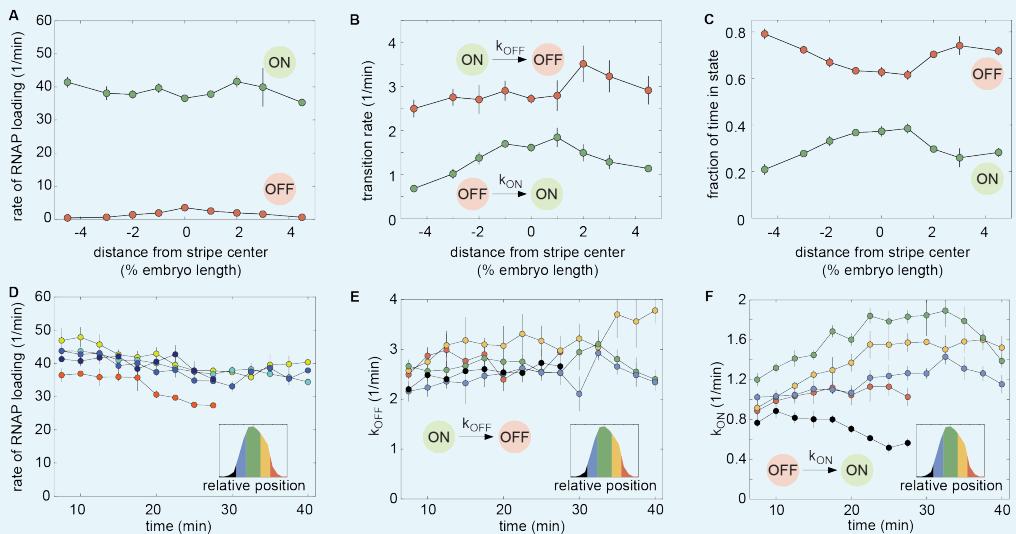
2052

2053

2054

2055

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} .

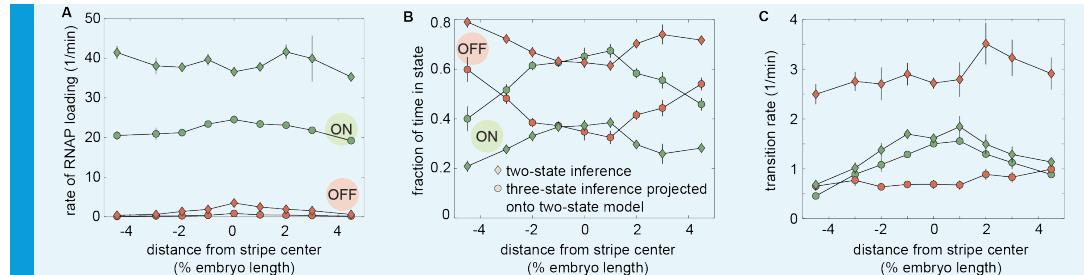


Appendix 6 Figure 3. Two-state mHMM inference. **(A-C)** Time-averaged 2-state inference results. **(A)** Consistent with three-state inference results, we observed no significant modulation in the rate of initiation along the axis of the embryo. Moreover, we found that k_{on} (green plot in **(B)**) 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 6E**, 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 3F**, red). This is not entirely surprising, as the trend returned by the three-state inference was relatively mild (**Figure 6E**, red), encompassing only the final two time points for which there was sufficient data to conduct inference. It is possible that the added complexity of the three-state model allowed it to register a subtle shift in the activation rate that was convolved with countervailing features in the two-state case. Future work will seek to elucidate the source of this discrepancy and further test the validity of the trend uncovered in the three-state case.

Comparing true and effective two-state inference results

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



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 4E* and F). Here, color schemes are consistent with those employed in **Appendix 6-Figure 3A-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 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 4C**, red squares and circles, respectively). The k_{off} value for the effective two-state model is defined as

$$k_{\text{off}} = \frac{k_{10}k_{21}}{k_{21} + k_{12}}. \quad (101)$$

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.

2101 **Appendix 7**

2102

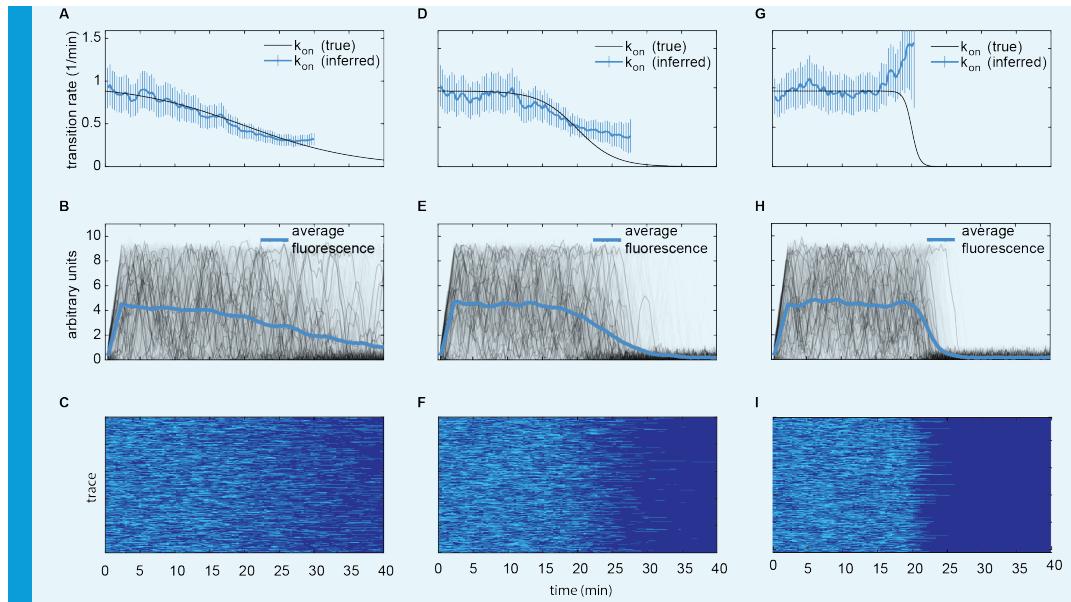
Inherent limits of bursting parameter inference

2103

2104 By definition, the onset of transcriptional quiescence coincides with the cessation of observ-
2105 able bursting activity. In the main text, we argue that this cessation appears to be driven
2106 by processes that are mechanistically distinct from those driving transcriptional bursting.
2107 It remains possible, however, that quiescence is instead driven by changes in the bursting
2108 machinery itself as illustrated in scenario (ii) in **Figure 6A**. If this is the case, it is important
2109 to note that fundamental limits exist to the time-scale of shifts in the bursting parameters
2110 that could be detected in any sort of time-dependent burst parameter inference (see, e.g.,
2111 **Figure 6**): changes of order-with or faster-than the time scale on which transcriptional bursts
2112 occur (1-3 min for eve stripe 2) cannot be detected. Notably, this is not a limit of the mHMM
2113 method, but rather reflects an inherent limitation set by system itself—in order to infer
2114 bursting parameters, we must observe bursts and, in order to infer a change in parameters,
2115 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.

2116

2117 To illustrate this limitation, we simulated three scenarios for a two-state transcriptional
2118 system in which k_{on} decreases to 0 s^{-1} over periods 15, 5, and 1 min in length. We then sought
2119 to recover the trend in k_{on} . To emphasize that the limitations are not specific to mHMM, we
2120 used the *true* promoter trajectories generated by our simulation algorithm to estimate k_{on} .
2121 These estimates thus represent the absolute best-case scenario for parameter inference, in
2122 which we recover the underlying behavior of the system exactly. The results indicate that, as
2123 expected, a transition in k_{on} that happens in the span of 1 minute is not detectable from a
2124 burst inference perspective (**Appendix 7–Figure 1A-C**). This indicates that, at this timescale,
2125 a shift in burst parameters (scenario ii in **Figure 6A**) would be indistinguishable from an
2126 abrupt, change in which the promoter entered a silent state outside of those considered
2127 by the bursting model (scenario (i) in **Figure 6A**). Interestingly, results for 5 and 15 minute
2128 k_{on} transitions (**Appendix 7–Figure 1D-I**) also indicate that even transitions that occur over
2129 longer periods of time cannot be fully recovered due to the fact that bursting behavior is
2130 observed over a limited window of time (40 minutes in our case). Thus, once the burst
2131 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.



2132

2133

2134

2135

2136

2137

2138

2139

2140

2141

2142

2143

2144

2145

2146

2147

2148

2149

2150

2151

2152

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}).

2153 Appendix 8

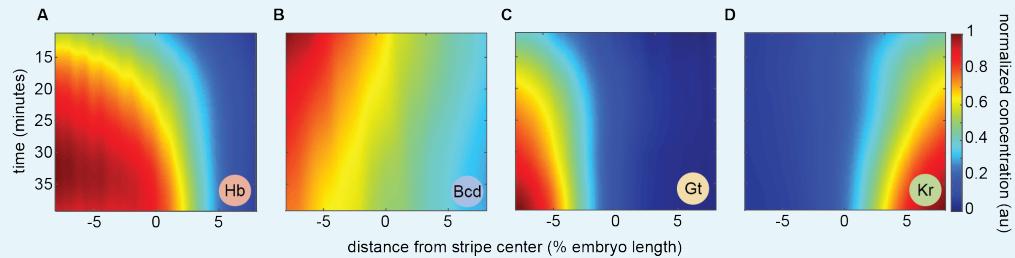
2154 Input-Output analysis details

2155
2156
2157
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.

2158 Input transcription factor data

2159 Data sources

2160 The input-output analysis presented in the main text made use of previously published data
2161 sets for the spatiotemporal concentration profiles of the gap genes Hunchback, Krüppel,
2162 and Giant (**Appendix 8-Figure 1A, C and D**). These data derive from elegant experiments in
2163 which individual embryos were co-immunostained for transcription factors of interest and
2164 precisely staged by measuring progressive cellularization over the course of nuclear cycle 14
2165 to generate a time series of protein concentration profiles spanning the course of this period
2166 of development (*Dubuis et al., 2013*). The Bicoid concentration data used for this analysis
2167 derives from live imaging experiments using a Bicoid-GFP fusion established by *Gregor et al.*
2168 (*2007*). These data come courtesy of Jonathan Liu and Elizabeth Eck (**Appendix 8-Figure 1B**).



2169
2170
2171
2172
2173
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.

2175 Data processing

2176 To prepare the Krüppel, Giant, and Hunchback profiles for use in our logistic regression
2177 analysis, we adopted an approach similar to that described in *Dubuis et al. (2013)*. Dorso-
2178 ventral orientation of embryos was found to have negligible effect on calculated intensity
2179 profiles and was ignored (i.e. all embryos were included, regardless of orientation). For
2180 each time point in nuclear cycle 14, a weighted temporal average was calculated using a
2181 sliding Gaussian kernel with $\sigma_t = 5$ min. For each time point, the minimum observed value
2182 across all anterior-posterior positions was then calculated and subtracted in order to remove
2183 background fluorescence. Normalized profiles were then calculated using the formula
2184

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

2185 An identical procedure was followed for processing the Bicoid-GFP data, with the addition
2186 of a spatial averaging step using a sliding Gaussian window of $\sigma_{\text{AP}} = .5$ % embryo length.
2187 This step was necessitated by the fact that, because individual embryos were imaged for
2188 the duration of nuclear cycle 14, multiple experiments contributed concentration data along
2189 the anterior-posterior axis for each time point. Thus averages in both space and time were
2190 needed in order to effectively aggregate these data into a single average spatiotemporal
2191 profile.
2192

2194 Finally, we discovered that the anterior-posterior axes in our live imaging data (both for
2195 eve stripe 2 and Bicoid-GFP) were inconsistent with the axes employed by the fixed data
2196 reported by the authors in *Dubuis et al. (2013)*. We addressed this issue by using eve stripe
2197 2 as a fiduciary mark to register the positions of the fixed and live data sets. Specifically, we
2198 aligned the mRNA peak predicted by our model at 40 minutes into nuclear cycle 14 with the
2199 peak in second stripe of the eve protein profile at 40 minutes, as reported in *Petkova et al.*
2200 (2019).

2201 **Logistic regression framework**

2202 The binomial logistic regression is a widely used statistical method for assessing the
2203 relationship between a set of predictor variables and a response variable of interest that is
2204 constrained to take on one of only two possible outcomes. In the context of our analysis, the
2205 predictor variables were the normalized transcription factor concentration profiles and the
2206 response variables were (i) the overall transcriptional state given by the transcriptional time
2207 window (active or silent?) and (ii) the bursting state amongst transcriptionally active loci (ON
2208 or OFF?). Inference was conducted at the level of individual gene loci. fmincon, a standard
2209 matlab function for constrained optimization, was used to fit all models discussed both in
2210 the main text and in this appendix.

2211 To prevent overfitting at the stripe centers, the selection of data sets for input-output
2212 inference were weighted to ensure equal representation of data points from across all
2213 regions of space and time included in the analysis. The data were divided into cells of size
2214 1% of the embryo length in width and 1 minute in duration for the purpose of calculating
2215 and assigning these weights. The number of data points in adjacent regions were factored
2216 into each region's weight score using a 2D Gaussian averaging kernel. Regions with fewer
2217 than 25 total data points were not included in the inference.

2218 **Inference details: transcriptional time window**

2219 For the time window input-output analysis, we considered only loci that were transcriptionally
2220 active for one or more time steps in nuclear cycle 14. Loci were classified as transcriptionally
2221 active for all time points between the first and last time points for which they exhibited
2222 detectable levels of transcriptional activity and silent for all time points following their final
2223 shut-off for which their nuclei were still present in the experimental field of view. Time points
2224 preceding the onset of activity were discarded. *Appendix 8-Figure 2A* illustrates how this
2225 quantity varies over space and time in our experimental data. We considered a class of
2226 logistic regression models in which each transcription factor was permitted to appear at
2227 most once, thus requiring that each factor act on eve2 in a uniform manner through space
2228 and time; i.e., the same protein could not activate expression on one stripe flank and repress
2229 on the other.

2230 **Inference details: transcriptional bursting**

2231 The bursting input-output analysis focused exclusively on transcriptionally engaged loci. The
2232 Viterbi algorithm was used to infer the instantaneous activity state (ON vs. OFF) for all loci.
2233 This activity state was taken as the response variable in our regression analysis. In all other
2234 respects, the inference procedure was identical to that conducted for the time window.

2235 **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

2237

Manuscript submitted to bioRxiv

2238

2239

2240

2241

2242

2243

2244

2245

2246

2247

2248

2249

2250

2251

2252

2253

2254

2255

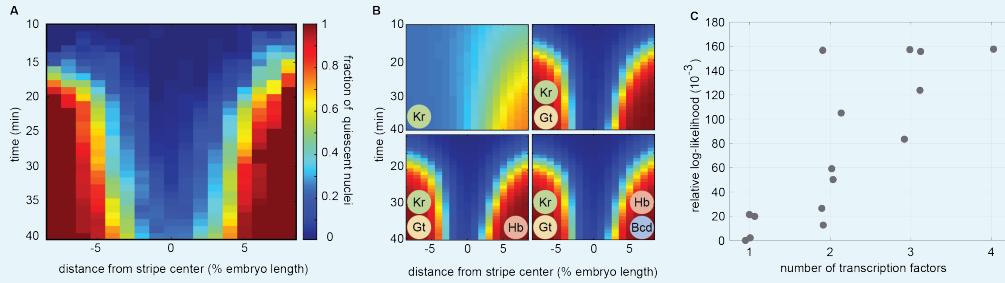
2256

2257

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 2B*). A quantitative comparison of model fit scores confirmed that the addition of Hunchback and Bicoid did nothing to improve model fit (*Appendix 8-Figure 2C*). 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 7A*. (B) Relaxing constraints on the functional nature of each transcription factor had no appreciable effect on the inference results. Profiles shown here are indistinguishable from those shown in *Figure 7D*. Once again, we find that the joint action of the repressors Giant and Krüppel is sufficient to explain the progressive onset of transcriptional quiescence in the stripe flanks. (C) A quantitative comparison of model fits reinforces the qualitative conclusions drawn from (B). Models including 3 and 4 transcription factors cannot improve on the fit achieved by the simpler double repressor model. Here blue dots indicate models for which only Giant and Krüppel make significant contributions to the model fit. This indicates that, while the 3 and 4 transcription factor models include additional parameters, these do not contribute appreciably to overall model fit, emphasizing the fact that these models behave, effectively, as double repressor models.

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 3C*, 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 3C*), comparison of this four-factor model's predicted activity profile with that of the Krüppel-Giant-Hunchback model revealed no material improvement

2280

2282

2283

2284

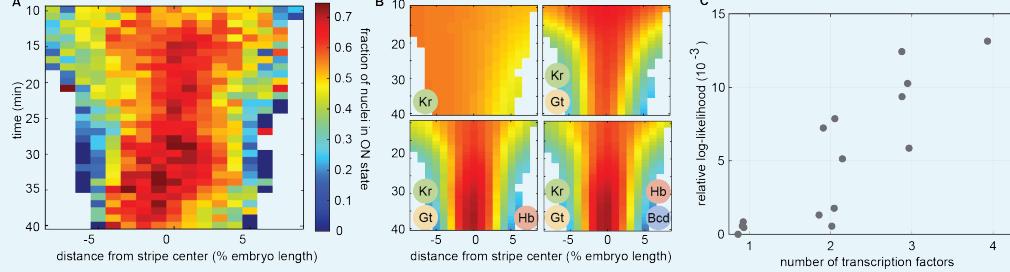
2285

2286

2287

2288

in the model's agreement with the experimental data (**Appendix 8-Figure 3B**, 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.



2289

2290

2291

2292

2293

2294

2295

2296

2297

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 7B**. **(B)** As with time window, relaxing the constraints on the functional nature of each transcription factor did little to alter the inference results presented in the main text (compare to **Figure 7E**). As with the constrained results, the joint action of Giant, Krüppel, and Hunchback appears sufficient to explain the spatiotemporal activity pattern revealed by mHMM inference. **(C)** A quantitative comparison of model fits.

2298 **Appendix 9**

2299 **Determining the RNA polymerase dwell time using autocorrelation**

2300 In order to conduct mHMM inference, it is necessary to specify the number of time steps w
2301 required for an RNA polymerase molecule to traverse the reporter gene,

2303
$$w = \frac{\tau_{\text{elong}}}{\Delta\tau}, \quad (103)$$

2304

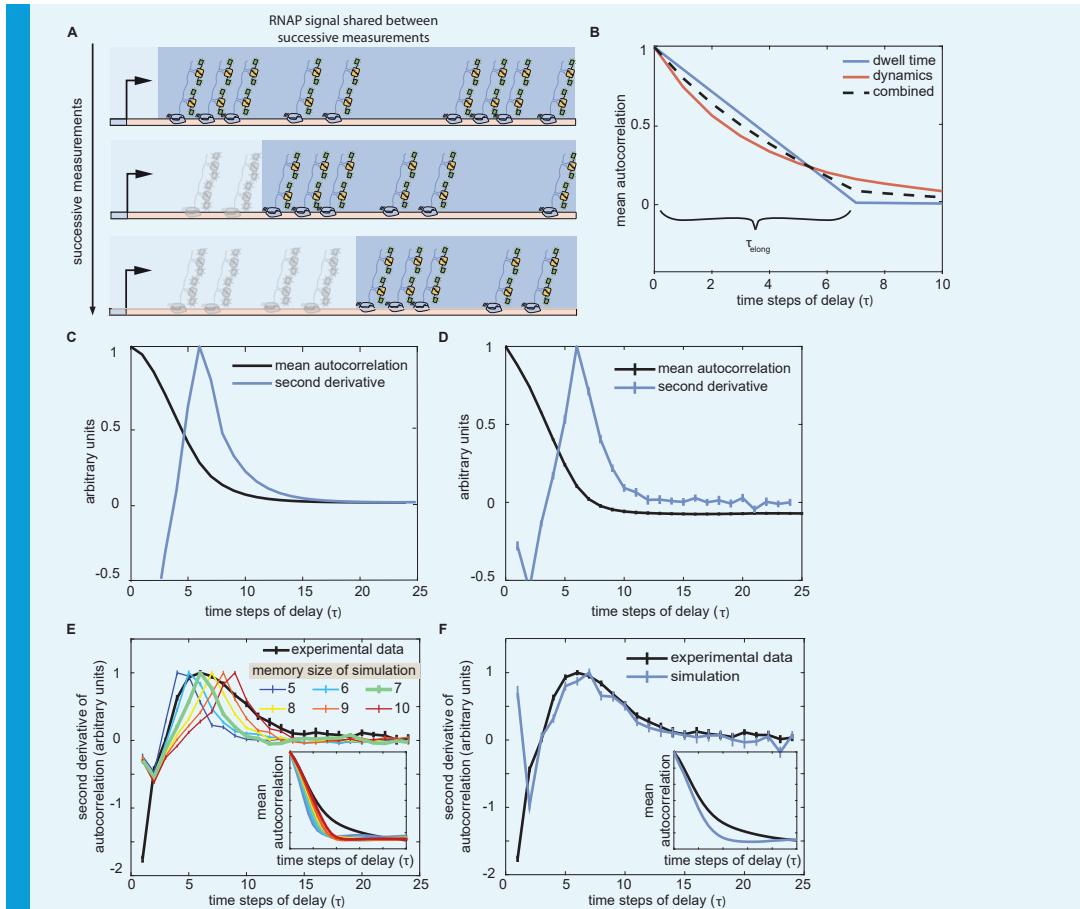
2305 where $\Delta\tau$ is set by the temporal resolution of our data acquisition and τ_{elong} is the elongation
2306 time which is unknown *a priori*. Past studies have estimated elongation rates for other
2307 systems involved in early patterning in the *Drosophila* embryo, but there is substantial
2308 disparity between the reported values. A live imaging study of transcriptional activity driven
2309 by the *hunchback* P2 enhancer reported an elongation rate of $1.4 - 1.7 \text{ kb min}^{-1}$ (**Garcia et al., 2013**).
2310 However, a recent study of the same regulatory element reported elongation rates of
2311 $2.4 - 3.0 \text{ kb min}^{-1}$ —nearly twice as fast (**Fukaya et al., 2017**). These results suggested that
2312 RNA polymerase elongation rates measured for other systems might not apply to our *eve*
2313 stripe 2 reporter. Thus, in order to ensure the validity of our inference, we developed an
2314 approach that uses the mean autocorrelation function of experimental fluorescence traces
2315 to estimate the elongation time directly from our data.

2316 The autocorrelation function $R_F(\tau)$ quantifies the degree to which a signal, $F(t)$, is corre-
2317 lated with a lagged version of itself, $F(t - \tau)$, and is given as a function of the time delay, τ ,
2318 between the two signal copies being compared such that

2319
$$R_F(\tau) = \frac{E[(F(t) - \mu_f)(F(t - \tau) - \mu_f)]}{\sigma_f^2}, \quad (104)$$

2320

2321 where μ_f is the average observed fluorescence, σ_f is the standard deviation of the fluo-
2322 rescence and E denotes the expectation value operator. As illustrated in **Appendix 9–Figure 1A**,
2323 the fact that it takes RNA polymerase molecules some finite amount of time to traverse
2324 the gene implies that the observed fluorescence at a transcriptional locus at some time
2325 t , $F(t)$, will be correlated with preceding fluorescence values $F(t - \tau)$ so long as $\tau < \tau_{\text{elong}}$
2326 because the two time points will share a subset of the same elongating RNA polymerase
2327 molecules. As τ increases, the correlation between $F(t)$ and $F(t - \tau)$ due to these shared RNA
2328 polymerase molecules will decay in a linear fashion (since the average number of shared
2329 RNA polymerase molecules decreases linearly with τ) until it reaches zero when $\tau = \tau_{\text{elong}}$
2330 (**Appendix 9–Figure 1B**, blue curve).



2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2352

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 time steps (140 s). **(E)** Simulated traces with elongation time of 7 time steps (green curve) exhibit a peak in the second derivative that coincides with the maximum of the experimental curve. Inset plots show corresponding mean autocorrelation curves for experimental data and simulations. **(F)** Stochastic simulations in which we allow for variation in elongation times distributed around a mean of 7 time steps qualitatively recapitulates the observed curve. (C-F, second derivative profiles depicted here are normalized relative to their maximum value for ease of depiction.)

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

2358

2359

2360

2361

2362

2363

2364

2365

2366

2367

2368

2369

2370

2371

2372

2373

2374

2375

2376

2377

2378

2379

2380

2381

2382

2383

2384

2385

2386

2387

2388

2389

2390

2391

2392

2393

2394

2395

2396

2397

2398

2399

2400

2401

2402

2403

2404

2405

2406

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 1B**, red curve)

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

Where λ_2 denotes the second largest eigenvalue of the transition 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 $R_P(\tau)$ indicated that the second derivative of the mean autocorrelation function reliably exhibits a peak that can be used to read out the value of τ_{elong} (**Medin et al., 2019**). **Appendix 9–Figure 1C** 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 1D**). These simulated traces included the expected contributions from both the Markov dynamics (red profile in **Appendix 9–Figure 1B**) and the finite RNA polymerase dwell time (blue profile in **Appendix 9–Figure 1C**). 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 1E** 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 $w = 7$ ($\tau_{\text{elong}} = 140$ s) best characterized our data (**Appendix 9–Figure 1E**, green curve). This implies that

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

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

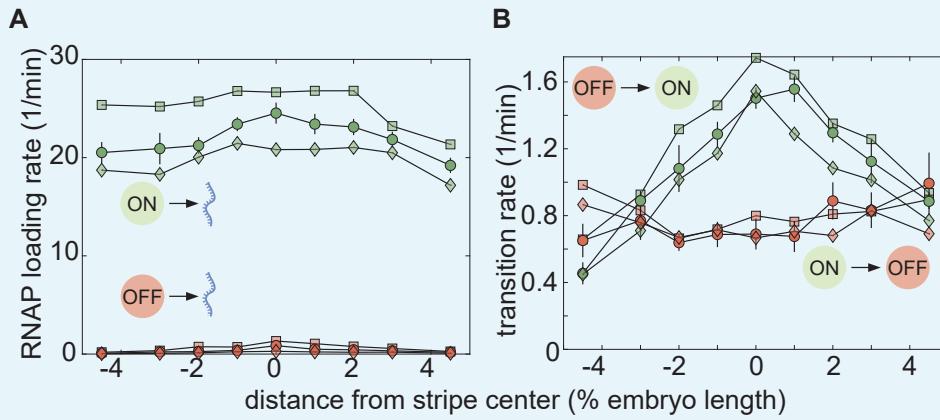
Appendix 9–Figure 1F shows how a simple adjustment to our simulation approach, wherein the elongation time steps w for individual RNA polymerase molecules were drawn from a Gaussian distribution with mean $\mu_w = 7$ and standard deviation $\sigma_w = 2.5$ time steps can qualitatively reproduce the wider 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 $\tau_{\text{elong}} = 120$ s and $\tau_{\text{elong}} = 160$ s (see below).

2407
2408
2409
2410
2411
2412
2413
2414
2415
2416

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 1E**). 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.



2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428

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 insensitive 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.)

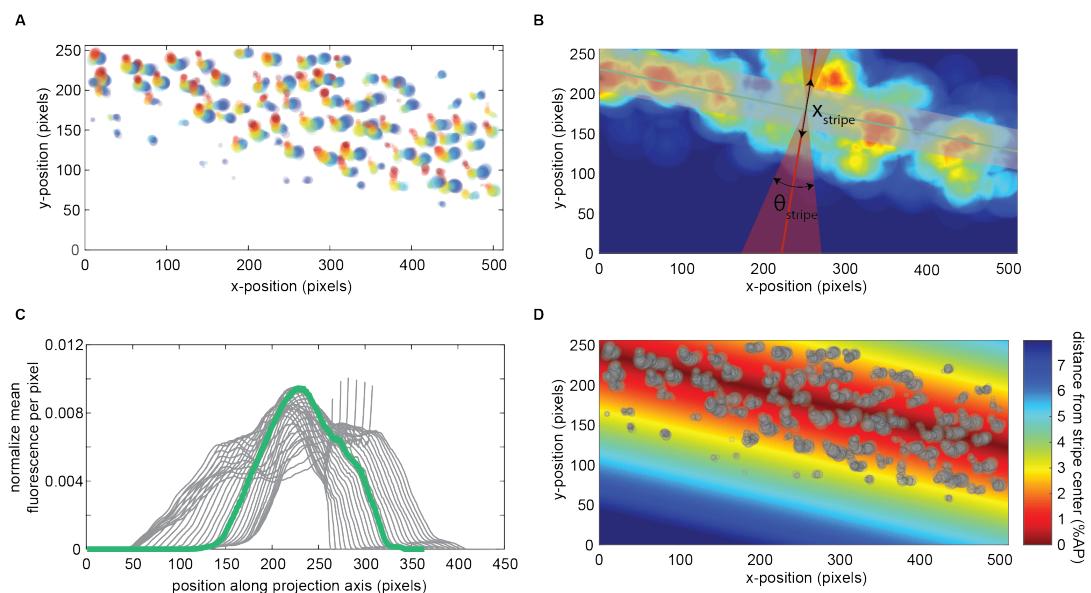


Figure 2-Figure supplement 1. Aligning stripes from multiple embryos. In order to minimize alignment errors when combining data from across multiple *Drosophila* embryos, an automated routine was employed to define a new experimental axis for each data set based upon the spatial distribution of transcriptional 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.

2430

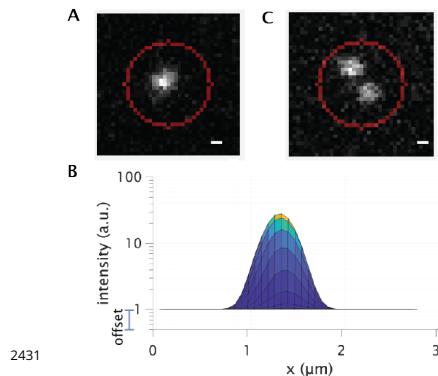


Figure 2-Figure supplement 2. Integrating MS2 Spots. **(A)** Fluorescence of a site of nascent transcript formation is measured by integrating raw pixel intensities in a circular region around the fluorescent MS2 spot of a predefined area (indicated by the red circle) and then subtracting off the background 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 of 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.

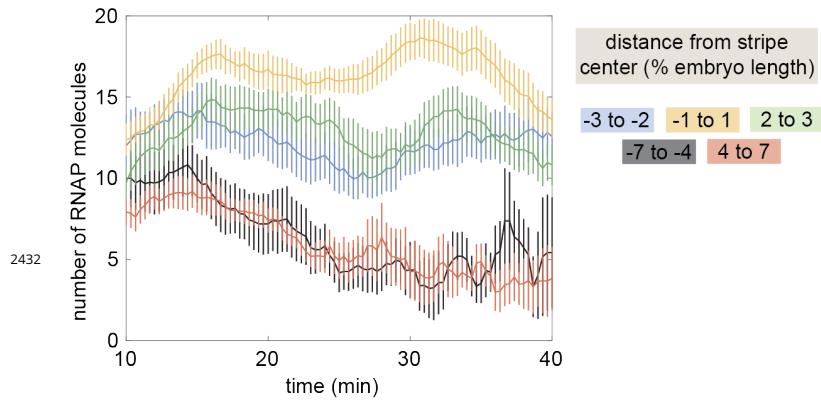


Figure 3-Figure supplement 1. Mean transcriptional activity. Mean transcriptional activity as a function of time for different positions along the stripe. (Average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

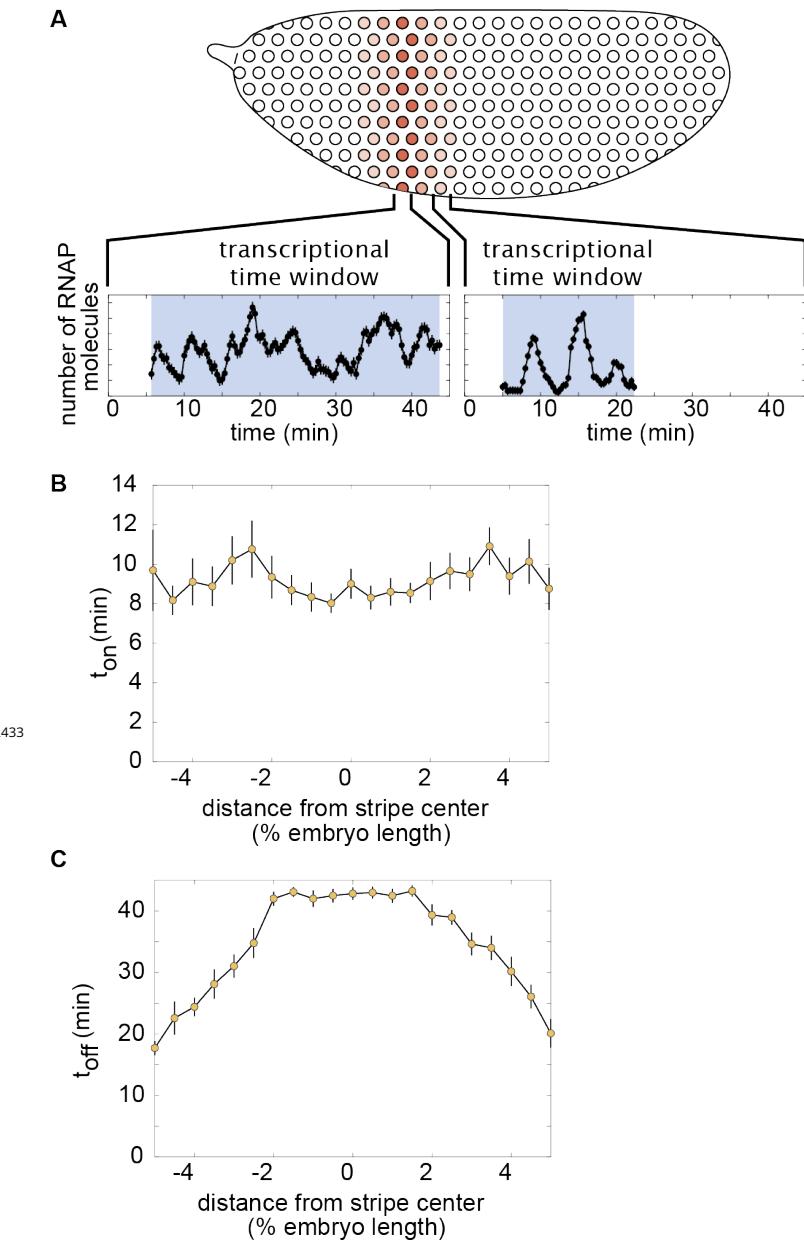


Figure 3-Figure supplement 2. Regulation of the transcriptional time window. (A) Single-nucleus measurements reveal that the duration of transcription is modulated along the stripe and that nuclei transcribe in a burst-like fashion. (B) Time for nuclei to activate transcription after mitosis, t_{on} , 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).

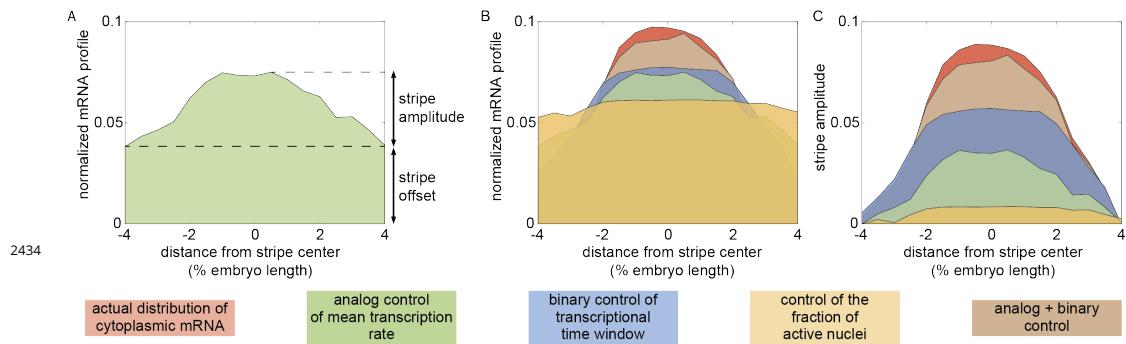


Figure 3-Figure supplement 3. Definition of stripe amplitude. (A) The normalized mRNA profile for the stripe can be separated into an offset and an amplitude. (B) Normalized mRNA profiles and (C) stripe amplitude for the cytoplasmic pattern of mRNA as well as for the contributions from the various regulatory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

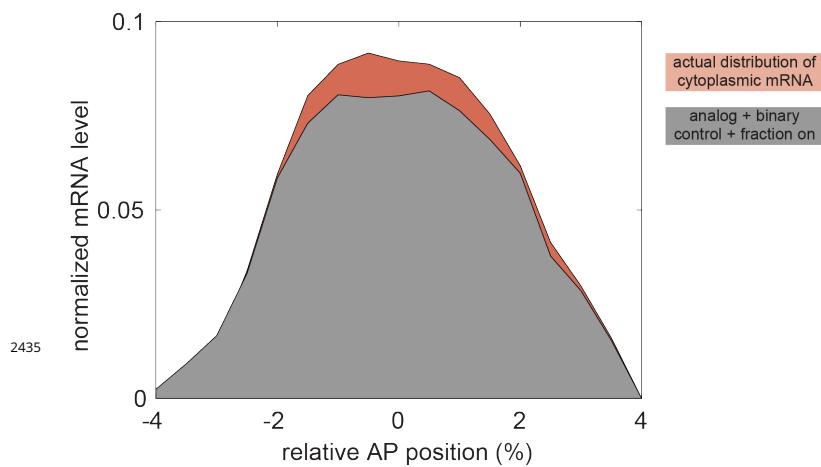


Figure 3-Figure supplement 4. Joint effect of mean rate, binary control, and fraction of active nuclei. Including of the predicted effect of anterior-posterior-dependent modulation of the fraction of active nuclei has little effect on the predicted cytoplasmic mRNA profile (compare brown profile in Figure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

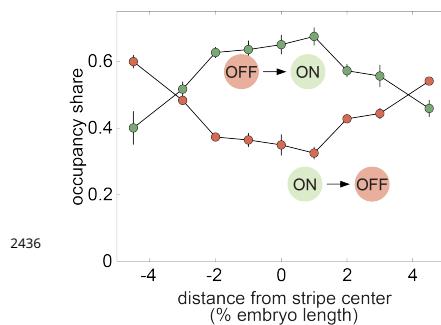


Figure 5-Figure supplement 1. Fraction of time spent in each transcriptional state. Fraction of time spent in the ON and OFF states as a function of the position along the stripe. (Error bars indicate the magnitude of the difference between the first and third quartiles of mHMM inference results for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