Title

Noise is the signal: uncovering the kinetic fingerprints of gene regulatory logic

Author

Lammers, Nicholas Christopher

Publication Date

2022

Peer reviewed|Thesis/dissertation

Noise is the signal: uncovering the kinetic fingerprints of gene regulatory logic

by

Nicholas C. Lammers

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Hernán G. Garcia, Chair
Professor Michael B. Eisen
Professor Rob Phillips
Professor Phillip L. Geissler

Spring 2022

Noise is the signal: uncovering the kinetic fingerprints of gene regulatory logic

Abstract

Noise is the signal: uncovering the kinetic fingerprints of gene regulatory logic

by

Nicholas C. Lammers

Doctor of Philosophy in Biophysics

University of California, Berkeley

Assistant Professor Hernán G. Garcia, Chair

Predicting how interactions between transcription factors and regulatory DNA sequence dictate rates of transcription and, ultimately, drive developmental outcomes remains an open challenge in quantitative biology. Indeed, despite decades of biochemical and genetic studies that have established a reasonably complete "parts list" of the molecular components driving eukaryotic transcription, the field nonetheless lacks a satisfactory understanding of how interactions between these molecular components unfold across space and time to give rise to gene regulatory logic. Recently, technical advancements have begun to provide glimpses into this molecular black box, dramatically improving our ability to trace how molecular pieces move, interact, and assemble. However, if we are to fully realize the immense potential of these exciting new technologies, then our ideas need to catch up with our experiments.

In this thesis, we argue that quantitative models have a central role to play in synthesizing the ever-increasing array of cutting-edge experimental measurements into a coherent theory for the molecular basis of transcriptional control. To this end, we seek to develop conceptual, theoretical and computational frameworks for dissecting how molecular reactions at individual gene loci give rise to the formation of dynamic patterns of gene expression and facilitate cellular decision-making. Chapters 2 and 3 describe previously published works that combine live imaging, statistical inference, and simple quantitative models to probe how transcription factor proteins regulate the dynamics of transcriptional bursting at target gene loci to give rise to stripes of gene expression early on in fruit fly development. Chapter 4 describes a series of analyses following-up on various results from these works. We also use this chapter to break new ground, however, examining how 2 spot experiments, which track the output of two identical gene loci in each cell, can be used to estimate rates of information transmission at individual gene loci from live imaging data.

In Chapters 5 and 6, we connect phenomenological models of transcriptional bursting employed in the preceding sections to truly molecular models that seek to understand how key transcriptional behaviors emerge from molecular interactions at the gene locus. Chapter 5 examines a puzzling gap revealed by recent live imaging studies between the rapid timescale

(seconds) of transcription factor binding and the slow timescale (minutes to hours) of transcriptional bursts, and proposes two simple theoretical frameworks for bridging this gap. Chapter 6 investigates how the presence of energy-dissipating processes within the eukaryotic transcriptional cycle can open the door to new kinds of gene regulatory logic that increase the rate at which gene loci transmit information. Finally, Chapter 7 describes ongoing work to develop a novel Bayesian framework, *burstMCMC*, that uses efficient inference techniques to examine how transcription factor proteins regulate transcriptional burst dynamics.

To my parents Chris and Linda,

for your constant love and support

Acknowledgments

It's surreal to find myself here, at the tail end of this journey. I can't stop puzzling over how something so long anticipated can feel so sudden. I didn't know what to expect when I started at UC Berkeley. It has certainly been a whirlwind. And a rollercoaster, and—not infrequently—a slog. Above all though, my time here has been an adventure unlike anything I've experienced before, and for which I am immeasurably grateful.

First and foremost, I want to thank my Ph.D. advisor, Hernán Garcia, for his support, insights, and guidance throughout this my graduate career at Berkeley. I couldn't have done it without him. Coming into grad school, I had no real sense for what kind of scientist I wished to become, or even for what sorts of paths were possible. Hernán, to a remarkable degree, gave me the freedom to find my own way, while ensuring that I still had the guidance I needed to do meaningful, impactful research. Moreover, his belief in collaborative science, his commitment to doing quantitative work amidst the messy world of biology, and, most importantly, his unwavering (sometimes unfathomable) optimism through the ups and downs of the scientific process have provided me with invaluable guideposts to strive for as I move forward with my own career. I am very grateful for—and proud of—the work that we have done together, and look forward to more adventures to come.

I also owe a great deal to my friends and colleagues in the Garcia Lab. I had the distinct pleasure of coming into the lab alongside three amazing scientists and people: Gabriella Martini, Meghan Turner, and Jonathan Liu. I have learned a great deal from each of them about how to think about scientific questions, but more importantly their myriad pursuits outside of lab reminded me that there is much more to life than computer screens and fruit flies (shocking, I know). To Gabriella: I couldn't have asked for a better "qual twin". I will dearly miss those afternoons at Kingfish watching basketball with you and (very occasionally) talking science. To Meghan: of all the up-and-down projects I've been on, ours easily had both the uppest ups and the downest downs. Somehow, it was still a joy to work (and commiserate) with you through it all. Let's go climbing again soon. I also want to thank Armando Reimer for his generosity in helping me to find my footing as I was first coming into the lab, and Jiaxi Zhao for keeping our various collaborations afloat with his curiosity, patience, and enthusiasm over the course of the past couple of years.

During my time at Berkeley, I have also been fortunate enough to engage in scientific collaborations outside of the Garcia Lab. My time working with Michael Eisen taught me a great deal about how to ask and answer meaningful questions by thinking creatively about data. Additionally, though I did not work with him directly, Rob Phillips has been incredibly generous with his time throughout my Ph.D. I am grateful for his help in making my manuscripts stronger and for his guidance at various stages along life's way. In addition, Vahe Galstyan, as much as anyone, has given me an amazing example of what it means to do interesting theoretical work on biological systems. His friendship and guidance were a boon to me throughout my Ph.D., both during the period of our direct collaboration and in the years thereafter.

Finally, this work would not have been possible without the support of my friends and family. Thank you to my parents, Chris and Linda, for the countless doors they have opened

for me, and for their patience as I've slowly figured out which ones to walk through. I also want to thank my friend and former roommate, Brian Goodman, for—year after year––patiently listening to me complain about how much work I had to do, and then dragging me out for beers anyway. Lastly, to my girlfriend, Kristi: sorry for being a hermit these past few weeks (months?). Thank you for making the bad days bearable, and for filling the good days with moments worth remembering. I couldn't have asked for a better puppy co-parent, and I can't wait to see what adventures lie in store for us in the months and years ahead.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ever since the discovery of the double helix, Biology has been recognized as an information science. And in few places is this connection between information and biological function more concrete than in the study of gene regulation, where gene loci function, quite literally, as communication channels, transducing input concentrations of transcription factor proteins into output levels of gene expression that then serve some further function within the cell. Over the past two decades, researchers have begun to capitalize on this connection, using concepts from information theory and statistical physics to interrogate both the theoretical limits of gene regulation [293, 294, 249, 35] and to search for hallmarks of optimal performance in real transcriptional networks experimentally [74, 237].

Yet, despite its utility, information theory's greatest strength—a complete agnosticism to the physical nature of a system's signaling components—holds it at a remove from the molecular mechanisms by which gene regulation is ultimately realized. The same formulas and bounds apply, regardless of whether the signal is being transmitted via protein patterns or fiber optics. As a result, describing biological systems in the language of channel capacities and bits can obscure as much as it reveals, masking the messy realities—the tangled hierarchies of molecular interactions between genes, the profound complexity of the stochastic molecular "hardware" [227], the as yet ill-explored horizons of gene regulation out of equilibrium [314]—that lie at the heart of most contemporary inquiries into the physical basis of gene regulatory function [166, 179].

Indeed, despite decades of biochemical and genetic studies that have established a reasonably complete "parts list" of the molecular components driving eukaryotic transcription [161], the field nonetheless lacks a predictive understanding of how transcriptional control emerges from molecular interactions at the gene locus. In recent years, significant technical advancements have lead to dramatic improvements in our ability to trace how molecular pieces assemble in space [227] and in time [166, 59, 179]. Yet even in the most sophisticated in vivo experiments to date (see, e.g., [213]), significant overlap between fluorescent protein emission spectra limits our ability to multiplex colors beyond 3. As a result, researchers are left—in many cases, quite literally [164]—to divine insights from isolated points of light set against a vast sea of hidden molecular machinations.

There is no single solution to this challenge, but over the course of this dissertation we

argue that quantitative models can and must play a central role in synthesizing the ever-accruing array of cutting-edge experimental measurements into a coherent theory for the molecular basis of transcriptional control. Along the way, we seek to illustrate that—far from being an irrelevant abstraction—information can serve as a guiding principle for navigating the complexity of biological systems, serving both to identify "useful" regulatory mechanisms from live imaging data and to place quantitative constraints on the performance of molecular models of transcription at and away from equilibrium. In the chapters that follow, we lay out a series of works that combine cutting-edge live imaging experiments, statistical physics, and computational modelling in an effort to bridge the gap the messy realities of transcriptional systems on the one hand, and (our as yet nascent) theoretical understanding on the other.

In the remainder of this introduction, we outline key themes running through these chapters, and summarize key findings from each work. Chapters 2 and 3 describe previously published works that combine live imaging, statistical inference, and simple quantitative models to probe how transcription factor proteins regulate the dynamics of transcriptional bursting at target gene loci to give rise to stripes of gene expression early on in fruit fly development. Chapter 4 describes a series of follow-up results to these works, and lays out a new vision for how so-called "2 spot" experiments, which track the output of two identical gene loci in each cell, can be used to estimate rates of information transmission from live imaging data. Chapter 5 surveys results from recent live imaging studies that point to a puzzling gap between rapid transcription factor binding and the slow timescales of the transcriptional bursts regulated by these binding events, and proposes two simple theoretical frameworks for bridging this gap. Chapter 6 describes ongoing work that examines how the presence of energy-dissipating processes within the eukaryotic transcriptional cycle can open the door to new kinds of gene regulatory logic and increase the rate at which gene loci transmit information. Finally, Chapter 7 describes ongoing work to develop a novel statistical framework, *burstMCMC*, for inferring how transcription factor proteins regulate transcriptional burst dynamics.

1.1 The many facets of transcriptional "noise"

From the eon-spanning march of natural selection to the hours-long sprint of *Drosophila melanogaster* (*D. mel*) development, random variation—noise—is a defining feature of biological systems [218], serving at once as the engine of life's possibilities and the arbiter of life's constraints. It is fitting, then (although, admittedly, somewhat serendipitous), that much of this dissertation revolves about different aspects of transcriptional noise: noise as a measure of regulatory potential, noise as a barrier to regulatory control, and—most of all—noise as a microscope, a window into the molecular origins of gene regulatory logic.

We can gain intuition for these three aspects of transcriptional noise by considering two spot experiments, which dissect transcriptional noise by comparing the transcriptional output of two identical copies of a gene expressed within the same cell [1, 81] (Figure 1.1A). Imagine we measure the total amount of mRNA, m, produced by some gene, g, expressed in a population of cells early on in fruit fly development and find that the total variance

of produced mRNA across all observed gene loci is equal to σ^2_{tot}. Two spot experiments allow us to decompose the σ_{tot} into two components: the intrinsic variability (σ^2_{int}), which is the variation between gene copies that are from the same cell (and, thus, exposed to the same regulatory inputs), and the extrinsic variability (σ^2_{ext}), which is defined as the cell-to-cell variability in m due to cell-specific differences, for instance, in input transcription factor concentrations (Figure 1.1B). The extrinsic variability captures the amount "useful" variation that conveys information about differences in regulatory inputs across different cells (Figure 1.1A), such that the expression $1 + \sigma^2_{ext}/\sigma^2_{int}$ can be used to estimate the capacity, in bits, for some gene g to convey biological information downstream cellular processes (Figure 1.1C(i)).

Yet, in all transcriptional system, the stochastic nature of the molecular interactions driving gene regulatory processes means that there will inevitably be some remainder, σ^2_{int}, that, depending on its size relative to the average rate of transcription, $\langle m \rangle$, will place limitations on the system's performance (Figure 1.1C(ii)). In Chapter 6, we investigate how σ^2_{int} interacts with other features of the transcriptional response to dictate limits on how rapidly gene loci can transmit information both at and away from thermodynamic equilibrium.

However, σ^2_{int} is more than merely an impediment to regulatory control: the structure of the intrinsic noise bears signatures—the kinetic fingerprints—that can be leveraged to learn about the nature of the molecular processes from which it arises (Figure 1.1C(iii)). One relatively simple way to do this is through analysis of the Fano factor, $\sigma^2_{int}/\langle m \rangle$, which we utilize in Chapter 4. However the concept of using noise to infer molecular structure is much broader than this, and features prominently (in various guises) in Chapters 2, 3, 5, and 7. Indeed, it was through the careful analysis of noise signatures that, even before the emergence of *in vivo* RNA fluorescence labeling technologies such as the MS2/MCP [14] and PP7/PCP systems [36], researchers were able to use single-cell distributions of nascent RNA and cytoplasmic mRNA molecules obtained with fluorescence *in situ* hybridization (smFISH) [88, 246] to predict that gene expression occurred in a discontinuous, "bursty" fashion [245, 316].

1.2 Using phenomenological models to capture the transcription factor control of transcriptional bursting

The inferences initially made based off of noise in fixed tissue distributions of nuclear mRNA counts were subsequently confirmed with new live imaging techniques [14, 36], which have directly revealed stochastic transcriptional bursts—periods of activity interspersed with periods of transcriptional silence—in cells in culture and within animals (Figure 1.2A-C) [111, 50, 170, 27, 290, 225, 217]. The simplest model that can capture these dynamics is the widely used two-state model of promoter switching (Figure 1.2D) , which posits that a gene promoter can exist in one of two states: a transcriptionally active ON state and a quiescent

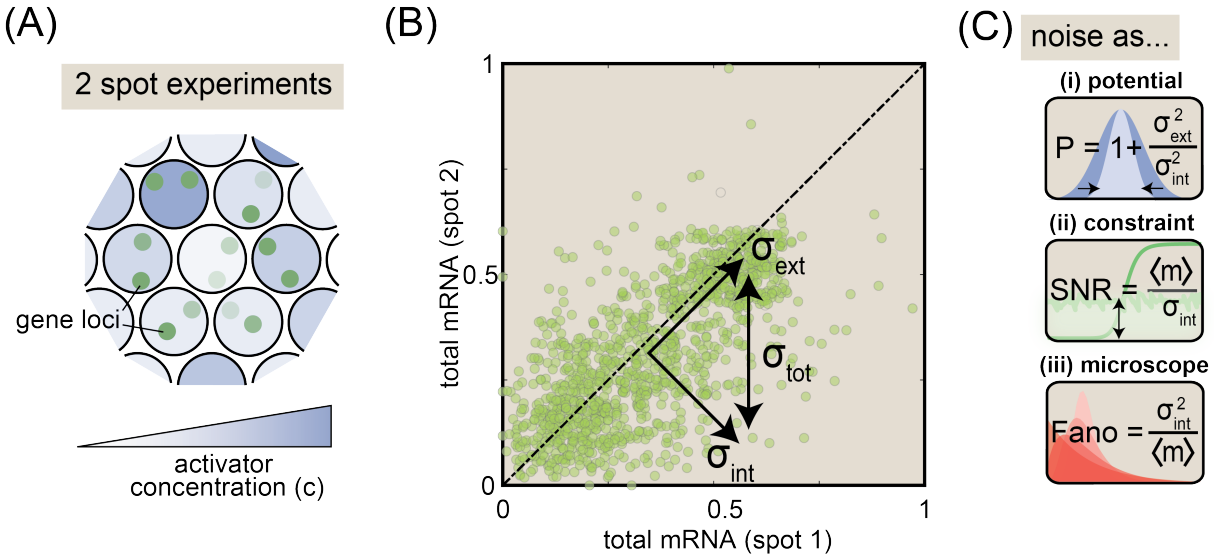

Figure 1.1: **Different faces of transcriptional noise. (A)** Cartoon illustrating two spot experiments, which measure the transcriptional output of two copies of the same gene in each cell. **(B)** Scatter plot showing hypothetical relationship total mRNA produced between sister spots within the same cell. **(C)** Panels illustrating the different facets of transcriptional noise discussed in the text.

OFF state. The promoter stochastically switches between these states with rates k_{on} and k_{off}, and loads new RNA polymerase II (RNAP) molecules at a rate r when in the ON state [307, 256, 21, 222]. The average burst duration, amplitude and separation are given by $1/k_{off}$, r and $1/k_{on}$, respectively (Figure 1.2E).

It is important here to point out that this two state model is, without a doubt, "wrong": a model with three parameters cannot possibly capture the full complexity of the mechanisms by which regulatory factors interact with the gene locus to give rise to bursty transcription [55, 332, 221]. And yet, I would argue, this model is useful nonetheless. Why? Because it provides a quantitative framework for decomposing the regulatory impact of different factors into distinct kinetic parameter. And while each bursting parameter does not necessarily map directly to a single molecular step in the transcriptional cycle, identifying which parameter(s) are subject to regulation can help narrow the set of possible molecular mechanisms. For instance, variation in r could indicate that transcription factors play an active role in the recruitment of RNAP to the promoter, or in the release of RNAP from promoter-proximal pausing [269].

Thus, by analyzing how transcription factor proteins alter the statistics of burst-like fluctuations at gene loci, we can move toward a truly molecular and mechanistic understanding of how transcriptional control emerges from its constituent molecular components. Yet, despite the potential utility of the two state formalism, few quantitative methods existed at the beginning of my graduate career that could reliably estimate burst parameters from MS2

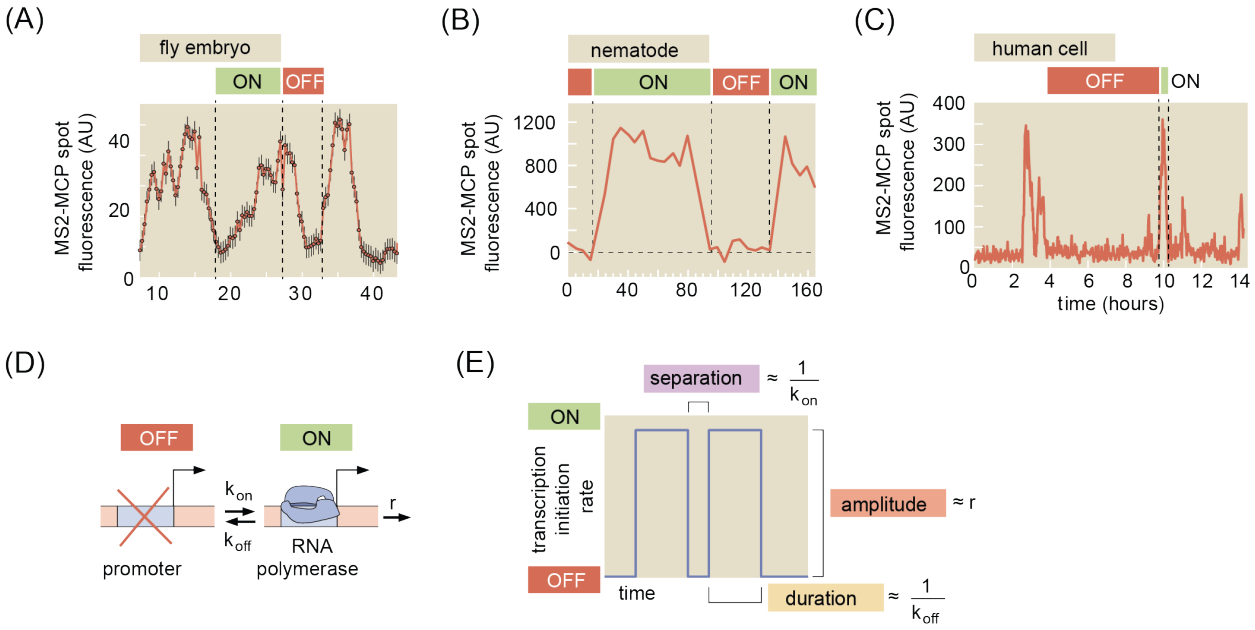

Figure 1.2: **Bursting is ubiquitous in eukaryotic transcription. (A-C)** Transcriptional bursting in (A) an embryo of the fruit fly *Drosophila melanogaster*, (B) the nematode *Caenorhabditis elegans*, and (C) human cells. **(D)** A two-state model of transcriptional bursting by a promoter switching between ON and OFF states. **(E)** Mapping the bursting parameters k_{on}, k_{off}, and r to burst duration, separation, and amplitude, respectively. (A, adapted from [165]; B, adapted from [177]; C, adapted from [251]).

and PP7 data in order to uncover how they were regulated by transcription factor proteins. To address this need, we developed a novel statistical method that uses compound-state Hidden-Markov models (cpHMM) to infer burst parameters from time traces of transcriptional activity [165]. In Chapters 2 and 3, as described below, we apply this tool to probe the kinetic fingerprints of transcription factor control of gene expression in early fruit fly development.

1.3 Different transcription factors employ the same kinetic mechanism to regulate transcriptional burst dynamics

Chapter 2 describes work—published in *PNAS* [165]—that combines cpHMM with theoretical modeling and live imaging to study the dynamic formation of stripe 2 of the widely studied *even-skipped* (*eve*) gene in the developing fruit fly embryo [278, 6]. We applied our cpHMM method to fit raw MS2 traces of *eve* stripe 2 activity (Figure 1.3A) and obtain the most likely sequence of underlying promoter states (Figure 1.3B) for all transcriptionally

engaged *eve* stripe 2 loci over time across the developing stripe (Figure 1.3C).

Previous works have established that the stripe is formed through the combined action of two transcriptional activators and two repressors [278] (though additional factors may play a role as well [302, 303]). Yet, despite this apparent regulatory complexity, our analysis reveals an emergent simplicity in the kinetics of transcription factor control. All four factors act upon the same kinetic parameter: the frequency of transcriptional bursts (Figure 1.3D and E). Moreover, while previous works largely focused on the regulation of transcriptional bursting across space, we find that the differential control of the *amount of time* for which *eve* stripe 2 loci engage in transcription plays a decisive role in driving stripe formation. Specifically, we find that rising repressor concentrations on the anterior and posterior flanks of the stripe induce gene loci to undergo an early transition into a long-lived period of transcriptional quiescence (Figure 1.3F), thereby concentrating mRNA production in the stripe center.

Taken together, this work elucidates how key regulatory factors interact with the *eve* stripe 2 enhancer over time to sculpt a sharp stripe of gene expression. It also raises several key questions that animate much of the remainder of this dissertation. First: we show that diverse regulatory factors interact with their target gene locus via the same kinetic step in the burst cycle (k_{on}), but what about different regulatory sequences? Does enhancer sequence, more so than regulatory factor identity, dictate how burst dynamics are manipulated to shape patterns of gene expression. We investigate this question next in Chapter 3. Second: the onset of transcriptional silencing described above *appears* irreversible in our live imaging data, but is it really? It is possible that quiescence is, in fact, reversible but simply not observed because repressor levels tend to rise monotonically over time along the stripe flanks. We address this question, in part, in Chapter 4. Lastly: what are the molecular mechanisms by which transcription factor binding ultimately leads to transcriptional bursting? We turn toward this mystery in Chapters 5 and 6.

1.4 A stripe is a stripe is a stripe: different enhancers sculpt stripes via identical kinetic mechanisms

Chapter 3 describes follow-up work—published in the journal *eLife* [13]—done in collaboration with the Eisen Lab at the University of California, Berkeley, in which we employ the same analytical methods to dissect the regulatory logic of the full even-skipped pattern, which consists of seven stripes of expression (Figure 1.4A). The *eve* stripes are produced via the largely independent activity of five discrete enhancer that drive individual stripes (the stripes 1,2, and 5 enhancers) or pairs of stripes (the stripe 3/7 and stripe 4/6 enhancers) [112, 125, 279] (Figure 1.4B). This modularity makes *eve-skipped* an ideal system for examining the connection between enhancer sequence and burst dynamics.

By conducting cpHMM inference on individual stripes, we found that, despite being created by five different regulatory enhancers, all seven stripes were sculpted using the same basic regulatory logic: enhancers up-regulated burst frequency and (to a lesser extent) the burst amplitude in the stripe centers and down-regulated them on the stripe flanks (Figure 1.4C-

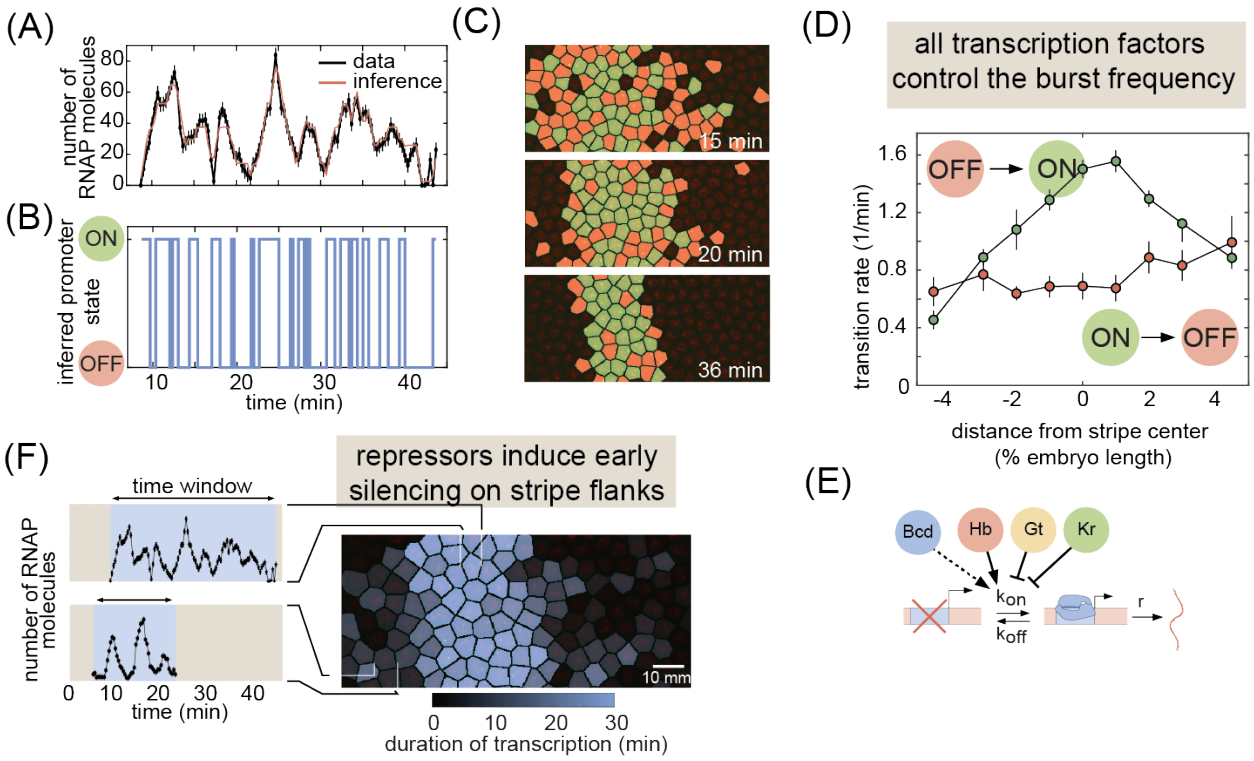

Figure 1.3: **Multimodal transcriptional control of pattern formation.(A)** Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (colors as in B). **(D)** Our cpHMM reveals that the transition rate between the OFF and ON state (equivalent to the burst frequency) is up-regulated in the stripe center, whereas k_{off} remains relatively constant. We also find r to be constant (data not shown; see Figure 2.5D). **(E)** This indicates that that all four key regulatory inputs of *even* skipped stripe 2, Bicoid (Bcd), Hunchback (Hb), Giant (Gt), and Krüppel (Kr) modulate burst dynamics via the same kinetic parameter: the burst frequency. We note that Bcd's role was found to be minimal, however (see Figure 2.7E-F). **(F)** Frame from a live imaging dataset, overlaid with patches indicating the observed duration of transcriptional activity in each cell across the stripe pattern. We find a clear spatial modulation in the duration of transcriptional engagement.

E). This result indicated a surprising degree of commonality in how different transcription factors and different enhancer sequences act to modulate gene expression. Consistent with our earlier experiments with *eve* stripe 2 (Chapter 2), we also found that the wholesale silencing of transcription in regions between stripes played a critical role in the emergence of the *even-skipped* pattern. Analysis in this chapter also uncovered the important role played by an additional regulatory "knob": the initial, digital decision of whether or not to engage in transcription *at all*. Our analysis revealed the presence of significant numbers of cells between the *eve* stripes along the anterior-posterior (AP) axis that never turned on in the first place, further amplifying the difference in mRNA levels between stripe and inter-stripe regions. We probe the molecular nature of this regulatory strategy further in Chapter 4.

Beyond the specific conclusions described above, the live imaging data underpinning this project serves as a beautiful illustration of the dynamism of developmental patterning. At one level, *eve* loci spanning most of the AP axis dance through dynamic patterns of quiescence and activity, darkness and light. Yet, even more strikingly, the *eve* stripes themselves split, sharpen, move, *change* in profound ways over a brief window of developmental time. These data thus serve as profound reminder of the fact that, in development, the spatial expression of genes such as *eve* consists not in a single, static "pattern", but rather in highly dynamic *trajectories* through a stereotyped series of spatial patterns.

We are at the dawn of a new period in the study of transcription, as new experimental techniques and advanced microscopy allow us to monitor transcriptional regulators, observe their behavior at the single-molecule level, and track the transcriptional output of a gene in living, developing animals. We have only barely begun to understand this new data and what it can tell us about biology. While the focus in Chapters 2 and 3 was on a single gene in a single species, we hope that these works will have a broader impact by beginning to establish rigorous frameworks for quantifying, characterizing and visualizing the dynamics of transcription at the single-cell level during development that will be required in the era of live imaging of transcription in development.

1.5 2 spot experiments shed light on the molecular basis of information transmission in *eve-skipped*

This chapter is intended both as a coda to the results put forward in the preceding two chapters, and as a bridge to the works that follow, in which we move from focusing on the control of bursting in and of itself towards thinking about its consequences for the molecular mechanisms of gene regulation and, more broadly, about how molecular noise impacts the capacity for gene circuits to transmit biologically useful information. Thought-provoking work out of the Gregor Lab at Princeton has demonstrated that spatial patterns of the regulatory inputs to *even-skipped* (the Gap Genes [141]) collectively encode approximately 4.1 bits of spatial information, on average [74]. More recently, a study from the same group demonstrated that the molecular processes that interpret Gap Gene inputs and dictate the spatial distribution of the *eve* stripes closely approximate predictions of an optimal statistical

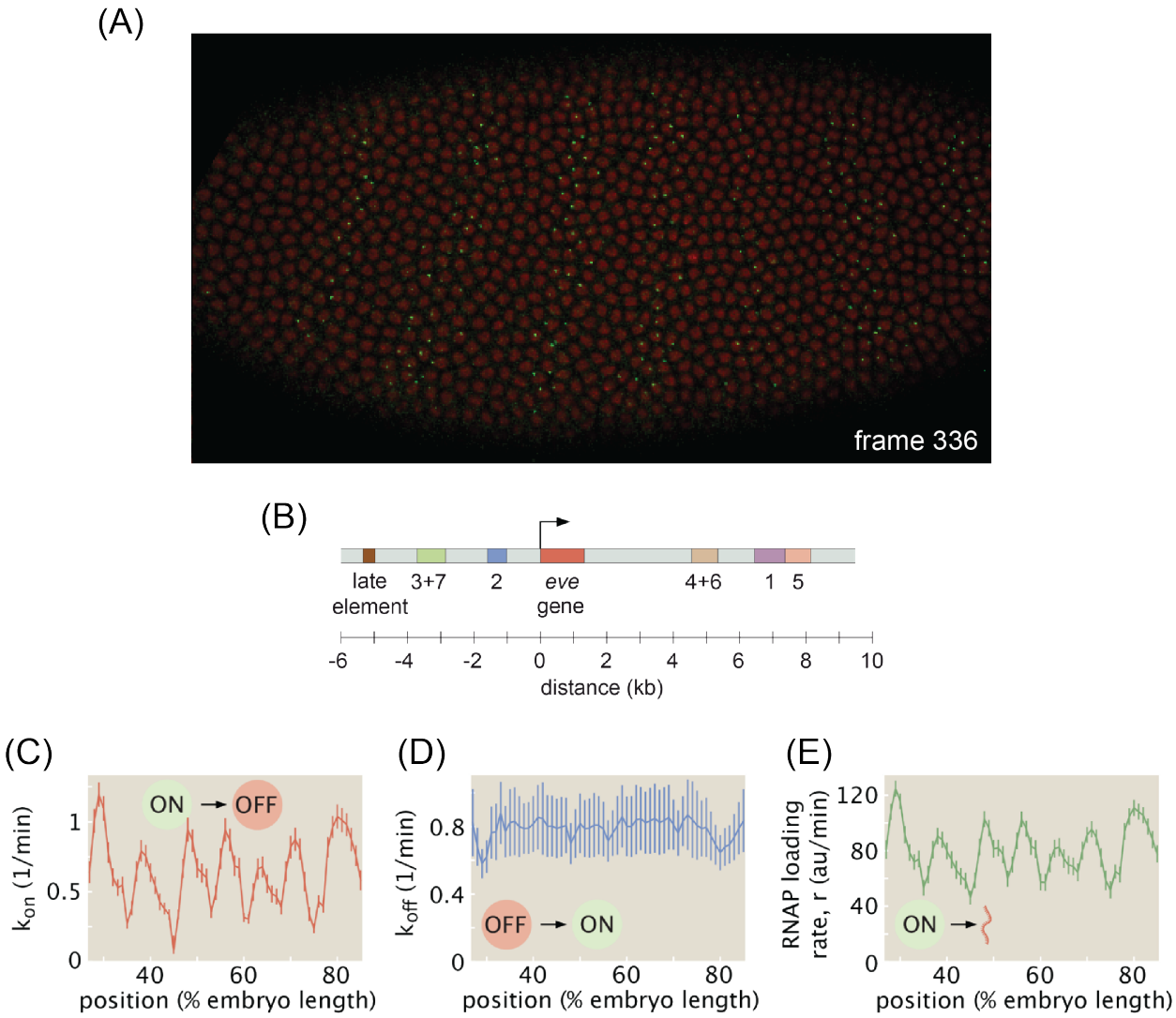

Figure 1.4: **A unified strategy for sculpting the seven stripes of *even-skipped*.** **(A)** Still from live imaging movie showing all seven stripes of *even skipped*. **(B)** Schematic depicting relative locations of the 5 stripe enhancers that regulate the *eve* stripes. Despite being driven by five distinct regulatory elements, we find that the *eve* stripes all exhibit the same burst parameter trends: **(C)** an up-regulation of k_{on} in stripe centers, **(D)** constant k_{off}, and **(E)** and up-regulation of the burst amplitude in stripe centers. Thus, each of the seven stripes is sculpted via the same set of kinetic mechanisms.

decoder, which the authors interpret as a signature of optimality [238]. Yet these results pertain to protein patterns in fixed tissue, and focus largely on the mature *even-skipped* pattern. How much information can a single *eve* locus actually transmit per unit time? How does this quantity change as a function of space and time? And what can this tell us about the molecular mechanisms that dictate the course of developmental processes?

Motivated by these questions, it was my ambition early on in my graduate research career to use multi-color live imaging experiments to track the regulatory inputs of (what else?) *eve-skipped* simultaneously with its transcriptional output in order to quantify the rate of information flow through individual transcriptional loci. Yet, technical hurdles made this ambition difficult to realize in practice. The crux of the difficulty, as first observed in [74, 73], lies in the fact that significant correlations between the cellular concentration of different gap genes makes it impossible to accurately estimate information flows without observing all relevant inputs simultaneously; a feat prohibited by the spectral overlap between fluorophores, which limits the total number of channels to ~ 3.

In Chapter 4, we argue that two spot experiments (Figure 1.1) provide a way around this constraint. The key insight is that comparing the transcriptional output of two gene loci from the same nucleus allows us to control for the set of regulatory inputs, no matter how complex, even when they are not directly observed (indeed, even when we do not know what the regulatory inputs *are*). Mathematically, this allows us to compare the prior uncertainty in the transcriptional output, $P(m)$, to the conditional uncertainty, given regulatory inputs, $P(m|\boldsymbol{c})$. We use this insight to assess the information content (regulatory potential) of different molecular "knobs" discussed in Chapters 2 and 3 (e.g. the average transcription rate, transcriptional silencing, and transcriptional engagement) and, more ambitiously, to track the rate of information transmission along the *even-skipped* pattern as a function of space and time. WE find that the timing of transcriptional silencing across the full *eve* pattern is highly informative, encoding about 1 bit of information on average (Figure 1.5A). We also find that the rate of information transmission by *eve* loci is highly variable, reaching its a peak of about 1 bit per minute in the gaps between stripes relatively early on in the process of pattern formation (20-35 minutes; see Figure 1.5B).

In addition, we show how the motion of the *eve* stripe pattern relative to underlying nuclei can be used as a natural input-output experiment, leveraging this feature to uncover evidence indicating that the repressor-induced silencing observed in Chapters 2 and 3 is, in fact, rapidly reversible (Figure 1.5C). This is consistent with findings from a separate work currently in preparation, undertaken in collaboration with Jiaxi Zhao, where we use optogenetic techniques to directly manipulate repressor concentrations, revealing rapid reversibility of transcriptional silencing in gene loci driven by the *even-skipped* stripe 4+6 enhancer.

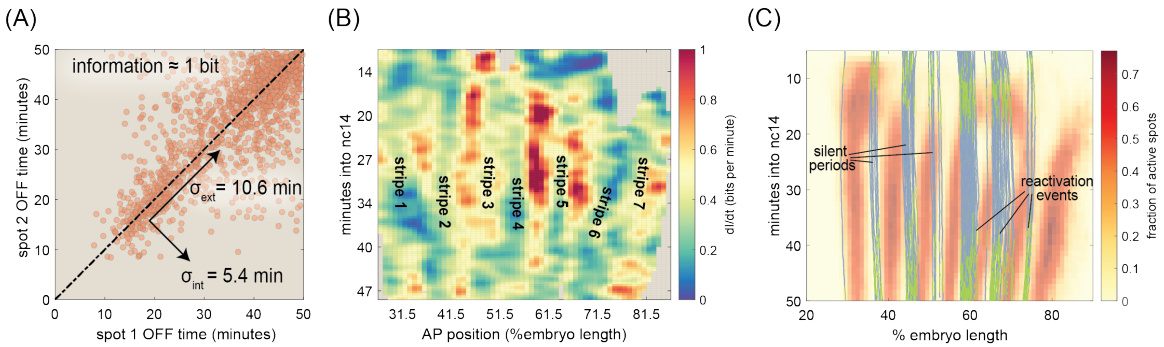

Figure 1.5: **Using two spot experiments to measure the information content of *eve* expression. (A)** Scatter plot comparing turn-off times between sister gene loci from the same cell. Information was calculated using Equation 4.10 from Chapter 4. **(B)** Heatmap of estimated rate of information transmission as a function of space and time. **(C)** Illustration of reactivation trajectories. Heatmap shading indicates fraction of active loci. Each line indicates AP position over time for an *eve* locus that was seen to undergo reactivation. Green segments indicate frames where transcription was observed. Blue segments indicate frames with no observed transcription. Note that reactivation events are clustered at the anterior stripe boundaries. This is because the net posterior-to-anterior shift in the *even-skipped* pattern is the primary driver of these events.

1.6 Using timescales as a lens to probe the molecular basis of transcriptional bursting

Bursting and its regulation are intimately tied to the molecular mechanisms that underlie transcriptional regulation as a whole. Yet, as discussed above, the models we and others have used to model bursty transcription are largely phenomenological in nature: they make no reference to the underlying molecular mechanisms that ultimately give rise to transcriptional control. Chapter 5 describes work, previously published in *Current Opinion in Biology* [167], where we argue that to make progress toward predicting transcriptional outcomes from underlying molecular processes, it is fruitful to start with the narrower question of how burst dynamics emerge from the kinetics of molecular transactions at the gene locus.

We review recent live imaging experiments that have revealed a significant temporal disconnect between transcription factor binding events, which generally last for seconds (see, e.g. Figure 1.6A), and the transcriptional bursts that these events control, which may last from a few minutes to multiple hours (Figure 1.6B). To explain this gap, we present two classes of molecular models for how slow bursting emerges from rapid transcription factor binding. The first posits the presence of one or more additional slow, rate-limiting molecular reactions (e.g. PIC assembly, mediator association, nucleosome removal) that

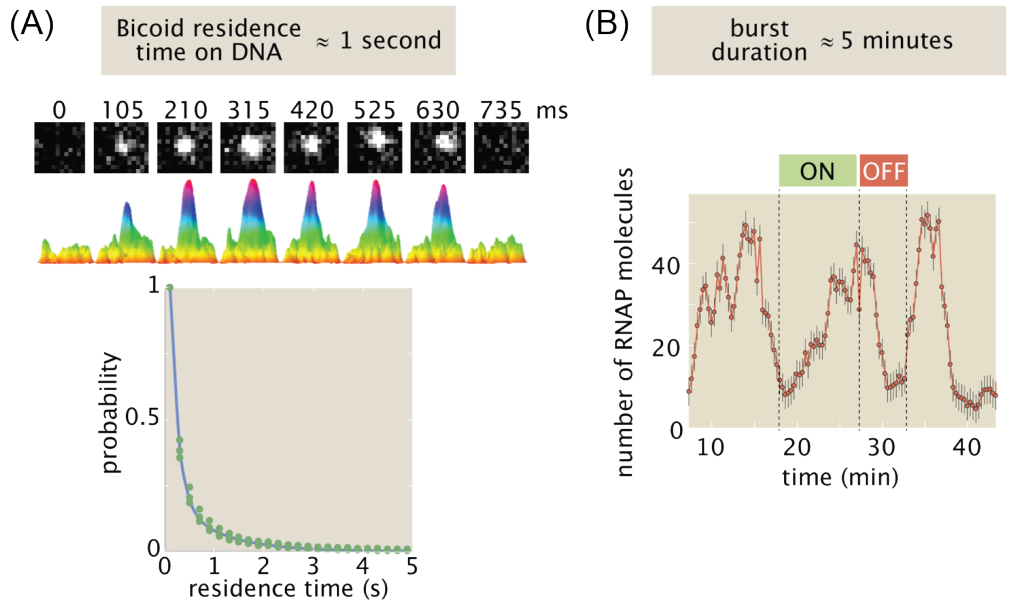

Figure 1.6: **A temporal disconnect between transcription factor binding and bursting. (A)** Data adapted from [212] that shows single-molecule measurements of binding dynamics for the transcriptional activator, Bicoid, in vivo. Mir and colleagues find an average residency time of approximately 1 second. **(B)** MS2 trace of transcription driven by the *hunchback* P2 enhancer, which is regulated by Bicoid. We find bursts that last not seconds but multiple minutes.

integrate activator binding while fluctuating on timescales in accordance with observed burst dynamics. Second, we show that slow dynamics can *emerge* as the collective result of many rapid binding events when cooperative binding interactions between regulatory factors are present. We conclude that a tighter discourse between theory and experiment is needed moving forward, if we are to synthesize the ever-growing deluge of in vivo measurements into a coherent model for how gene regulatory control emerges from stochastic molecular interactions at the gene locus.

1.7 Probing gene regulation away from equilibrium

Irrespective of its molecular origins, transcriptional bursting has implications for cellular decision-making, contributing significant levels of intrinsic noise to the transcriptional output of gene loci which, in turn, places strict constraints on how rapidly cells can harness that output to make decisions. In Chapter 6, we seek to draw together many of the themes discussed in earlier sections of this dissertation, using simple kinetic models of transcription to investigate how the molecular architecture of gene loci—the number and identity of biochemical steps in the transcriptional cycle, as well as the reaction rates that connect

these steps—dictates the rate at which bursty genes can transmit information to downstream cellular processes.

Thermodynamic models of transcription, which assume that gene circuits operate at equilibrium, have previously been employed with considerable success in the context of bacterial systems [241]. However, the presence of ATP-dependent processes—such as chromatin remodeling [329], PIC assembly [289], and Pol II initiation [321]—within the eukaryotic transcriptional cycle suggests that equilibrium models may be insufficient to capture how eukaryotic gene circuits sense and respond to input transcription factor concentrations. As a result, there is an urgent need for theoretical frameworks that can probe how non-equilibrium mechanisms reshape the transcriptional input-output function and, ultimately, redefine the limits of transcriptional control.

Thus, a key focus in this work in on how energy dissipation within the transcriptional cycle impacts the rate at which a gene locus transmit useful information and drive cellular decisions. Yet, rather than use information as a means to circumvent the molecular "hardware" that drives transcription, we instead seek to use the rate of information transmission as a quantitative lens to examine how energy-dependent changes to microscopic reaction rates at the gene locus lead to measurable changes in features of the transcriptional input-output function; namely, sharpness, precision, and specificity (Figure 1.7A). We draw from previous theoretical works [274, 69] to obtain quantitative estimates for how rapidly bursty gene expression driven by simple molecular models (Figure 1.7B) can transmit information and allow cells to distinguish between different input activator concentrations (c_1 and c_0; see Figure 1.7C and D).

We find that biologically plausible rates of energy-dissipation can drive significant gains in information transmission (Figure 1.7E), but that the molecular mechanisms underlying these gains change depending on the level of interference from the off-target binding of non-cognate factors at the gene locus. When interference is low, information is maximized by harnessing energy to push transcriptional sharpness beyond its equilibrium limits. Conversely, once non-cognate factor interference is high, information-maximization begins to instead demand that genes harness energy to increase transcriptional specificity at the price of decreased sharpness in order to buffer against non-cognate factor interference.

1.8 Fast, flexible burst parameter inference with *burstMCMC*

In this chapter, we describe ongoing efforts to develop *burstMCMC*, a novel computational framework that uses Markov Chain Monte Carlo (MCMC) methods [107] to uncover how transcription factor proteins modulate the dynamics of transcriptional bursting at target gene loci. Despite its utility, the cpHMM method developed in 2 has notable limitations. Most significantly, the model's computational complexity (and, hence, the time required for inference) scales as K^w, where K is the number of activity states (typically 2 or 3) and $w = t_{elong}/\Delta\tau$ is the number of time steps required for RNA Polymerase (RNAP)

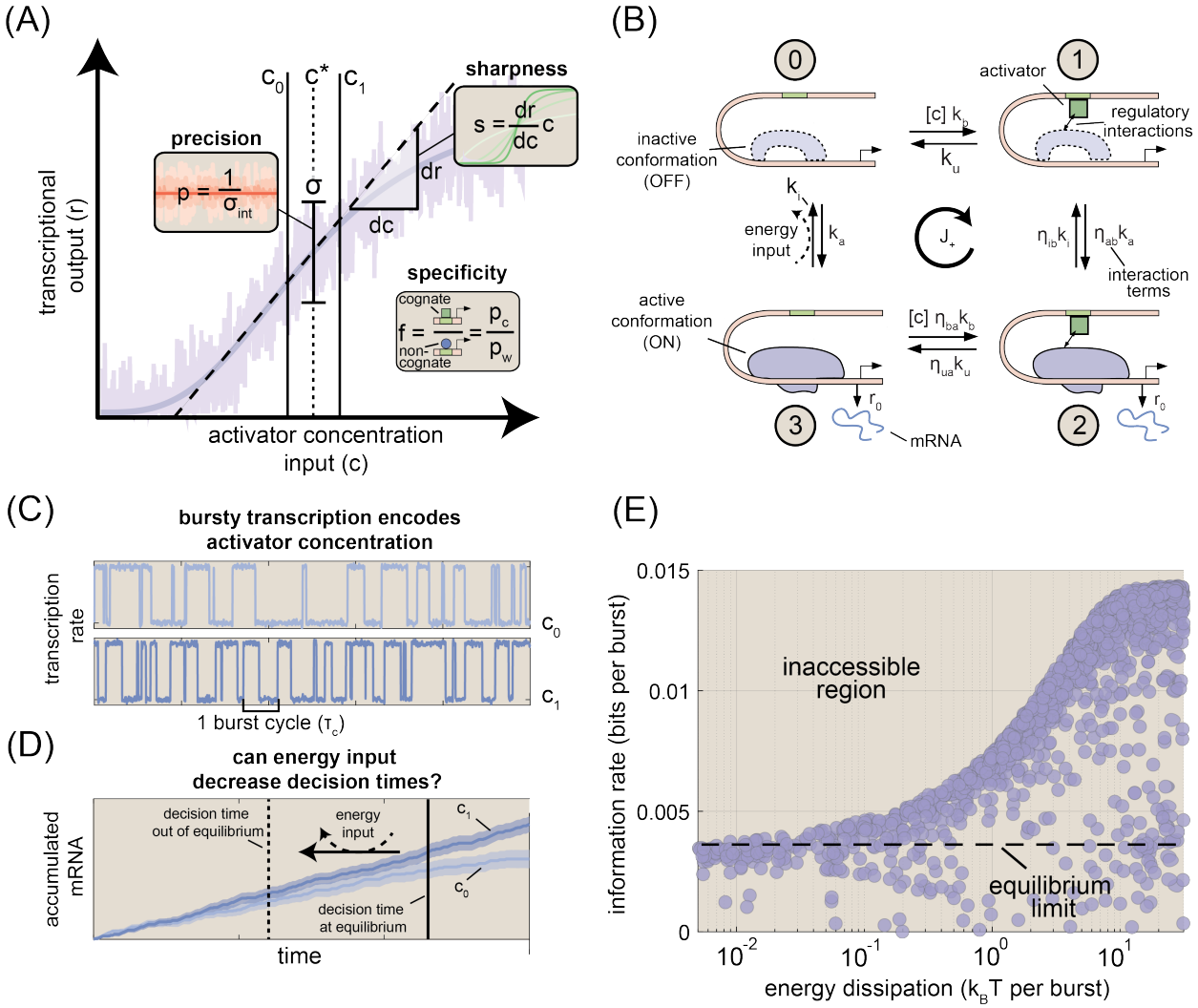

Figure 1.7: **Probing gene regulation away from equilibrium.** **(A)** Gene regulatory input-output function illustrating the basic biological problem considered in this work. Here a cell must distinguish between two activator concentrations, c_0 and c_1, based off of the transcriptional output of a gene locus (purple curve). We examine how three aspects of the transcriptional input-output function—sharpness, precision, and specificity—combine to dictate the rate at which the transcriptional output can drive biological decisions. **(B)** Four state MWC-like model of transcription used as the foundation of our investigations throughout. Here a single activator (green square) may bind to a specific site at the gene locus, and mRNA production occurs when the gene locus switches into its active (ON) conformation. A hypothetical energy input is depicted along the rate from state 3 to state 0. **(C)** Simulated burst dynamics for one realization of the model shown in (B). **(D)** Illustrative simulation results for accumulated mRNA levels driven by c_1 and c_0. Solid lines show trajectories for a single locus and shaded regions indicate the standard deviation in levels taken across 100 simulated trajectories. **(E)** Rate of information transmission as a function of energy dissipation rate for a parameter sweep exploring all possible model realizations. Modest rates of energy dissipation can lead to a significant increase in the maximum amount of information that can be transmitted per burst cycle.

molecules to traverse the gene. In practice, this scaling means that access to powerful servers or computing clusters is required to run the model for many systems biological systems. In addition, cpHMM requires w as an input, which can be problematic since the RNAP elongation time is gene-specific and typically unknown *a priori* for most systems. With *burstMCMC*, we aim first and foremost to address these limitations of the cpHMM method. Our ultimate aim is broader than this, however: we envision *burstMCMC* as an accessible, comprehensive computational platform for extracting quantitative insights from MS2/PP7 measurements of transcriptional systems. The focus here lies on progress toward the former task; however, we briefly outline next steps and eventual goals for model development both here and in Chapter 7.

The crux of the exponential scaling issue noted above lies in the fact that, as a part of its inference, cpHMM must keep track of the likelihood of *all possible* promoter states at each time point along an MS2 trace, which scales as K^w (where K is the number of promoter activity states). Recently, however, Bowles and colleagues [28] showed that the number of models states that actually have a non-negligible likelihood at any given point in time can be several orders of magnitude smaller than K^w. As a result, only a few hundreds or thousands of states need to be tracked at any point in time to achieve accurate inference results. The authors use this insight to develop *burstInfer*, an inference approach that reduces burst inference times by truncating the space of possible model states at each experimental time point [28]. Yet, whereas cpHMM yields accurate inference results for systems with up to 3 distinct activity states and continuous promoter transitions—i.e. switching can occur at any point on or between experimental time steps—*burstInfer* only applies to systems with up to two activity states and, further, makes the assumption that promoter switching is discrete (i.e. happens only at experimental time points); an assumption that likely does not comport the true nature of transcriptional systems.

With *burstMCMC*, we take an alternative approach. Rather than algorithmically truncating state space to reduce model complexity, we use MCMC methods—originally developed to efficiently sample high-dimensional ensembles in Statistical Physics [210]—to efficiently explore the high-dimensional state space of potential burst model states. As with all MCMC methods, burst MCMC works by drawing repeated samples of each parameter in such a way that these samples are guaranteed (in theory, at least) to converge to the true values implied by the experimental data (Figure 1.8A). We show that *burstMCMC* approaches cpHMM performance for both discrete and continuous three state systems, all while reducing computational complexity by orders of magnitude for biologically salient values of the elongation parameter, w (Figure 1.8B). In addition, we demonstrate that *burstMCMC* can reliably infer the elongation time, w, thus removing a significant barrier to the application of burst inference to a wider array of genes and providing a new computational lens into a biological parameter that is of interest in its own right (Figure 1.8C; see, e.g., [99, 190]).

We close by sketching out key next steps for model development. First, we lay out model improvements (many of which are already under way) that are aimed at increasing the speed, accuracy, and robustness of *burstMCMC* inference. Second, we discuss extensions to the model that probe new facets of the regulation of bursting dynamics, including hierarchical models for single cell burst inference and models with concentration-dependent parameters

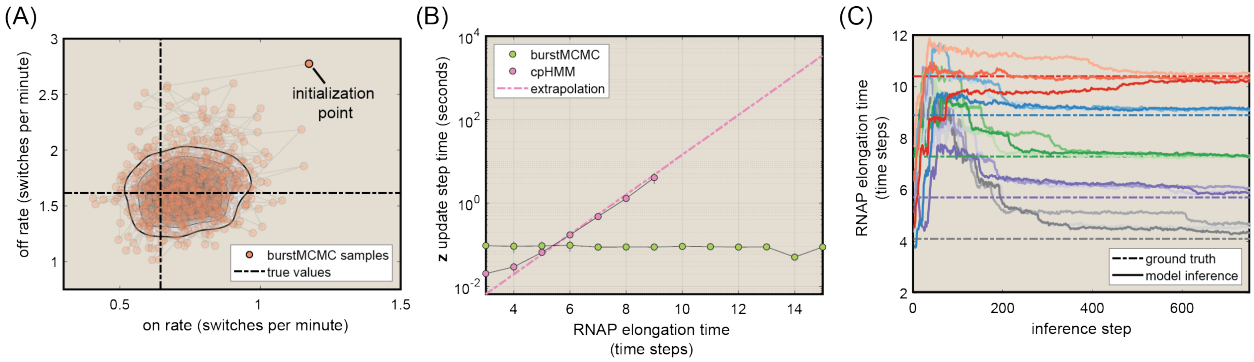

Figure 1.8: ***burstMCMC*: a fast, flexible framework for burst parameter inference.** **(A)** Plot showing a series of 1000 MCMC samples. Each circle indicates an individual sample, dashed lines indicate true parameter values, and gray shaded regions indicate posterior bivariate density inferred by *burstMCMC*. **(B)** Plot comparing the average time required for *burstMCMC* to resample $\boldsymbol{z}$ (green circles) to the time needed for a single expectation step in cpHMM inference (pink circles) as a function of the RNA polymerase dwell time on the reporter gene (w). Dashed pink line extrapolates cpHMM performance to regimes that could not be calculated directly given computational constraints. Whereas cpHMM computational time scales exponentially with w, we see that the trend for *burstMCMC* is essentially flat. **(C)** Inference results for systems with different RNAP elongation times. Solid lines indicate results for different MCMC runs. Dashed lines of same color indicate the correct w values. We see that the model consistently converges to the correct value for a wide range of elongation times.

Finally, we draw upon concepts from Chapters 5 and 6 to speculate about how mean first passage times could serve as a computationally tractable link between phenomenological bursting models and bona fide molecular models that can shed direct light on the molecular mechanisms underlying the transcription factor control of burst dynamics.

Chapter 2

Multimodal transcriptional control of pattern formation in embryonic development

This chapter is a reproduction of reference [165]

Abstract

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

2.1 Introduction

During embryonic development, tightly choreographed patterns of gene expression—shallow gradients, sharp steps, narrow stripes—specify cell fates. The correct positioning, sharpness, and amplitude of these patterns of cytoplasmic mRNA and protein ensure the reliable determination of animal body plans [236]. Yet, despite decades of work mapping the gene

regulatory networks that drive development, and despite extensive efforts to dissect the regulatory logic of the enhancer elements that dictate the behavior of these networks, a predictive understanding of how gene expression patterns and developmental outcomes are driven by transcription factor concentrations remains a central challenge in the field [301].

Predicting developmental outcomes demands a quantitative understanding of the flow of information along the central dogma: how input transcription factors dictate the output rate of mRNA production, how this rate of mRNA production dictates cytoplasmic patterns of mRNA, and how these mRNA patterns lead to protein patterns that feed back into the gene regulatory network. While the connection between transcription factor concentration and output mRNA production rate has been the subject of active research over the last three decades [172, 72, 280, 285, 147, 115, 142, 263, 301, 259], the connection between this output rate and the resulting cytoplasmic patterns of mRNA has remained largely unexplored. For example, a graded stripe of cytoplasmic mRNA within an embryo could arise as a result of radically different transcriptional dynamics at the single-nucleus level (Figure 2.1A). Specifically, if individual nuclei along this stripe modulate their average RNA polymerase (RNAP) loading rate, then graded control of the mean rate of transcription results: nuclei in the middle of the stripe transcribe at a higher average rate than nuclei on the stripe boundaries (Figure 2.1B). We identify this graded transcriptional control strategy with the analog control of gene expression. Alternatively, transcription factors could exert control over the length of time a nucleus is transcriptionally active (Figure 2.1C). In this binary control scheme—akin to an on/off switch that dictates whether a nucleus is transcriptionally active or quiescent—individual nuclei transcribe at the same average rate regardless of their position along the stripe, but for different lengths of time. Finally, some nuclei might not engage in transcription at all during the formation of the pattern (Figure 2.1D). Here, a larger fraction of nuclei engage in mRNA production in the stripe center than in the boundaries. Any of these scenarios, or some combination thereof, can explain the formation of a cytoplasmic mRNA pattern.

In order to quantify the contribution of each regulatory strategy to pattern formation, and thereby move towards a deeper understanding of the molecular processes at play, it is necessary to measure the rate of RNAP loading in individual nuclei, in real time, in a living embryo. However, to date, most studies have relied on fixed-tissue techniques such as mRNA FISH and immunofluorescence in order to obtain snapshots of the cytoplasmic distributions of mRNA and protein as development progresses [142, 84, 232, 62, 231]. Such techniques are virtually silent regarding the regulation of single-cell gene expression over time, and are thus ill-suited to the study of how spatiotemporal variations in transcriptional dynamics give rise to patterns of cytoplasmic mRNA.

In this work, we investigated how single-cell transcriptional activity leads to the formation of stripe 2 of the widely studied *even-skipped* (*eve*) gene in the developing fruit fly embryo [278, 6]. Previous work has established that the stripe is formed through the interplay of transcriptional activators and repressors [278]. In addition, recent studies have indicated that the *eve* stripe mRNA profiles are graded and highly reproducible [27, 286, 195, 238], suggesting that the detailed cytoplasmic distribution of mRNA that makes these stripes is key to the transmission of spatial information along the gene regulatory network

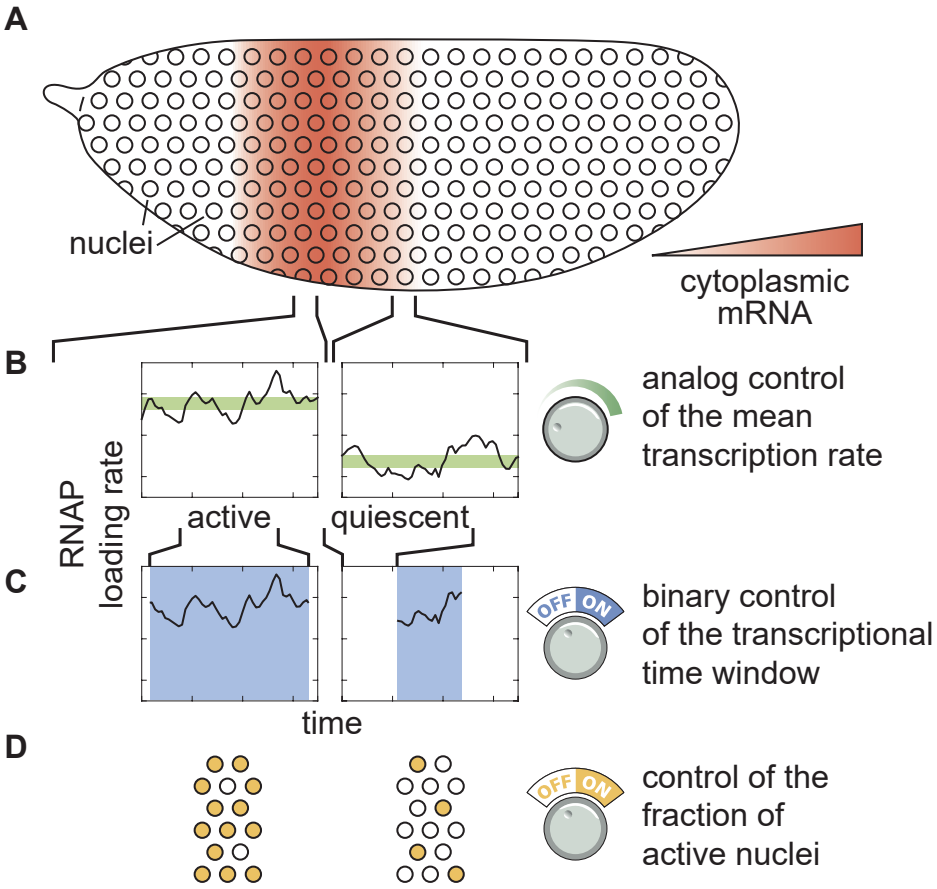

Figure 2.1: **Multiple modes of pattern formation by single-cell transcriptional activity.** **(A)** Cytoplasmic mRNA patterns could arise from transcription factors exerting control over **(B)** the mean transcription rate, **(C)** the transcriptional time window dictating when a nucleus is transcriptionally active or quiescent, **(D)** the fraction of active nuclei, or some combination thereof.

that drives *Drosophila* development and reinforcing the need to develop models of gene regulation capable of connecting quantitative variations in input transcription factor patterns to graded output rates of transcription. To do this, we combined live imaging with theoretical modeling in order to study transcription at the single-cell level in real time, seeking a quantitative connection between the spatiotemporal variations in input transcription factor concentrations, the control of *eve* transcription, and the formation of cytoplasmic patterns of mRNA.

We found that all three regulatory strategies outlined in Figure 2.1 quantitatively contribute to the formation of *eve* stripe 2. First, a smaller fraction of nuclei become active and engage in transcription in the periphery of the stripe than in the center, though this regulation of the fraction of active nuclei makes only a minor contribution to stripe formation. Second, consistent with previous studies, we found that the rate of mRNA production is significantly elevated in the center of the stripe [27]. Strikingly, however, we discovered that this analog control of the transcription rate is insufficient to quantitatively recapitulate the cytoplasmic mRNA stripe pattern. In addition to the control of the rate of mRNA production among nuclei, we also observed a pronounced regulation of the *window of time* during which *eve* loci were engaged in transcription across the stripe, with those in the stripe center expressing for approximately three times longer than those in the flanks. While it is widely appreciated that genes are transcriptionally competent for limited windows of time during development, we found that—in the case of *eve* stripe 2—this binary transcriptionally engaged/disengaged logic is not merely a necessary precondition for pattern formation—it is the *primary driver* thereof. Thus, we conclude that the regulation of *eve* stripe 2 is multimodal in nature, with contributions from three distinct regulatory strategies (Figure 2.1B-D). Nonetheless, stripe formation can be quantitatively explained almost entirely through the interplay between two distinct control strategies: binary control of the duration of transcriptional engagement (Figure 2.1C) and control of the mean rate of transcription (Figure 2.1B).

Building upon this result, we developed novel computational approaches to uncover the mechanistic underpinning of each regulatory strategy. We employed a compound-state hidden Markov model (cpHMM) to uncover variations in transcriptional bursting dynamics in individual nuclei across space and time [287, 217, 55]. We uncovered that, consistent with previous results, transcription factors control the rate of transcription by altering the frequency of transcriptional bursts [98, 330]. In addition, we utilized logistic regressions to correlate *eve* stripe 2 transcriptional dynamics with changes in input transcription factor concentrations. This analysis revealed that the transcriptional time window adheres to different regulatory logic than transcriptional bursting: while repressor levels alone were sufficient to explain the early silencing of nuclei in the anterior and posterior stripe flanks, the control of bursting among transcriptionally engaged nuclei depends upon the input concentrations of both activators and repressors. Thus, our findings point to the presence of two distinct regulatory mechanisms that control transcription and gene expression patterns in early development, showcasing the potential for theoretical modelling and biological numeracy to yield novel biological insights when coupled with precise and quantitative experimental observation.

2.2 Results

2.2.1 Predicting cytoplasmic mRNA distributions from transcriptional activity

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

$$\frac{\mathrm{dmRNA}}{\mathrm{d}t}(x,t) = \underbrace{p_{\mathrm{on}}(x)}_{\substack{\text{fraction of}\\ \text{active nuclei}}} \underbrace{R(x,t)}_{\text{synthesis}} - \underbrace{\gamma\,\mathrm{mRNA}(x,t)}_{\text{degradation}}, \tag{2.1}$$

where $\mathrm{mRNA}(x,t)$ indicates the mRNA concentration at position x along the embryo at time t, $R(x,t)$ corresponds to the mRNA synthesis rate averaged over multiple nuclei within the same position x, $p_{\mathrm{on}}(x)$ is the fraction of active nuclei (corresponding to the regulatory strategy shown in Figure 2.1D) and γ is the degradation rate (see SI Appendix, section A.3.1 for details of this derivation).

In order to examine the quantitative consequences of the three potential regulatory strategies (Figure 2.1B-D), we adopted widespread assumptions in the modeling of transcriptional regulation. First, we assumed that the degradation rate γ is a constant and not under any kind of spatiotemporal control. Comparisons between model predictions and empirically measured levels of cytoplasmic mRNA suggest that this assumption is reasonable (see SI Appendix, section A.3.2). Second, we posited that at each position throughout the embryo the synthesis rate $R(x,t)$ does not vary significantly in time such that it can be approximated by its time average $R(x) = \langle R(x,t) \rangle$. This assumption will be revised later in the text in order to account for the time-dependent regulation of the mean rate of transcription. Finally, we assumed that nuclei along the axis of the embryo start transcribing at time $t_{\mathrm{on}}(x)$, and stop transcribing and enter a state of transcriptional quiescence at time $t_{\mathrm{off}}(x)$. Under these assumptions, Equation 2.1 can be solved analytically, resulting in

$$\mathrm{mRNA}(x,t) = \underbrace{\frac{R(x)}{\gamma}}_{\substack{\text{mean transcription rate}}} \times \underbrace{\left(e^{-\gamma(t-\min\{t_{\mathrm{off}}(x),t\})} - e^{-\gamma(t-t_{\mathrm{on}}(x))}\right)}_{\text{transcriptional time window}} \times \underbrace{p_{\mathrm{on}}(x)}_{\text{fraction active}}. \tag{2.2}$$

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

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

In order to test the simple model of pattern formation put forward in Equation 2.2, we quantified transcription of stripe 2 of *eve* in the fruit fly. We imaged the transcription of an *eve* stripe 2 reporter using the MS2 system [104, 193, 27]. Transcripts of a reporter gene driven by the *eve* stripe 2 enhancer and the *eve* promoter contain repeats of a DNA sequence that, when transcribed, form stem loops [14]. These stem loops are recognized by maternally provided MS2 coat protein fused to GFP (Figure 2.2A). As a result, sites of nascent transcript formation appear as fluorescent puncta within individual nuclei (Figure 2.2B and Video A.4). As described in SI Appendix Figure A.2, the intensity of these fluorescent puncta is proportional to the number of RNAP molecules actively transcribing the gene. These resulting fluorescence values could then be calibrated using single-molecule FISH in order to estimate the number of RNAP molecules actively transcribing the gene (see Materials and Methods and [104]). By aligning multiple embryos (see SI Appendix, Figure A.1), we obtained the average number of actively transcribing RNAP molecules as a function of time and position throughout the embryo (Figure 2.2C).

Using the MS2 system, we quantified each potential regulatory strategy and determined its predicted contribution to pattern formation according to our model in Equation 2.2. We first used the average fluorescence intensities of our MS2 traces to estimate the time-averaged rate of RNAP loading, $R(x)$ as described in SI Appendix, section A.3.2. We found that this rate is modulated along the axis of the embryo (Figure 2.3A and B; see also Video A.4, SI Appendix Figure A.3, and Materials and Methods): whereas in the center of the stripe RNAP molecules are loaded at a rate of approximately 16 molecules/min, this loading rate decreases to about 8 molecules/min at the boundaries.

We next used our MS2 data to examine spatial trends in the transcriptional time window. Our data revealed that the transcriptional time window is modulated along the stripe (see SI Appendix, Figure A.4A). Whereas the time at which each nucleus becomes transcriptionally active, $t_{\text{on}}(x)$, was constant across the stripe, with all nuclei becoming active 8 ± 4 min after the previous anaphase (SI Appendix, Figure A.4B), the time at which nuclei stop transcribing and become quiescent, $t_{\text{off}}(x)$, showed a strong modulation along the embryo's axis (SI Appendix, Figure A.4C). As a result, the time window over which each transcriptional locus is engaged in transcription, $\Delta t = t_{\text{off}} - t_{\text{on}}$, is sharply modulated along the stripe (Figure 2.3C and D and Video A.4), with nuclei in the stripe center transcribing for > 30 min and nuclei on the boundaries only transcribing for approximately 10 min. We note that, in order to derive these results, it was necessary to account for potential effects of the detection limit in our experiments of approximately 4 RNAP molecules per locus on estimates of the timing of the appearance and disappearance of fluorescent puncta. This procedure is outlined in detail in the SI Appendix in section A.3.3, as well as in figures A.12 and A.13.

Finally, our analysis also revealed the magnitude of the modulation of the fraction of active nuclei along the stripe. Most nuclei along the stripe were engaged in transcription. In the stripe center, nearly 100% of nuclei transcribed at some point during the nuclei cycle. This number reduced to about 80% at the boundaries (Figure 2.3E and F and Video A.4).

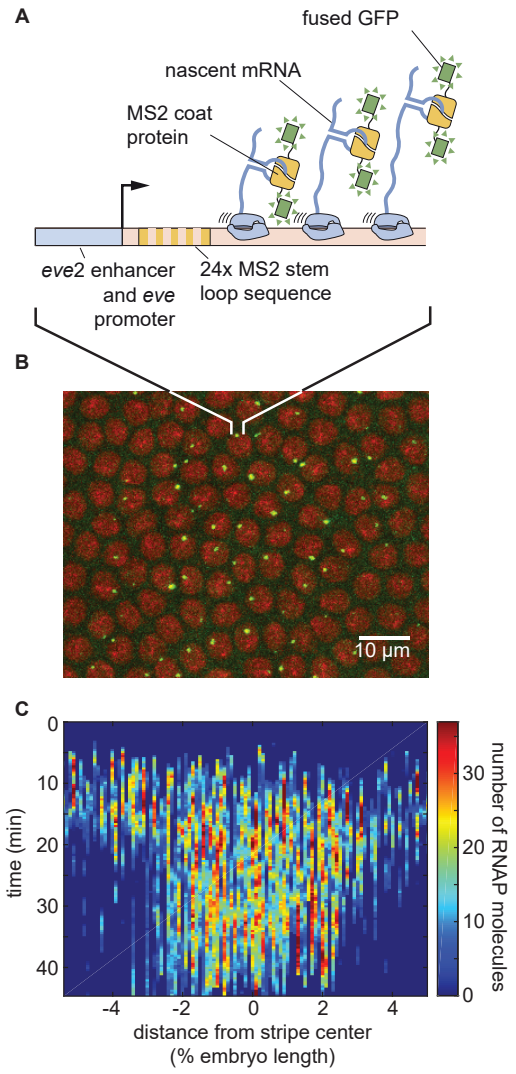

Figure 2.2: **Measuring transcriptional dynamics of *eve* stripe 2 formation using the MS2 system. (A)** MS2 stem loops introduced in an *eve* stripe 2 reporter gene are bound by MS2 coat protein fused to GFP. **(B)** Sites of nascent transcript formation appear as green fluorescent puncta whose intensity reports on the number of actively transcribing RNAP molecules. Nuclei are visualized through a fusion of RFP to Histone. **(C)** Mean number of RNAP molecules actively transcribing the gene as a function of space and time. (C, data averaged over 11 embryos).

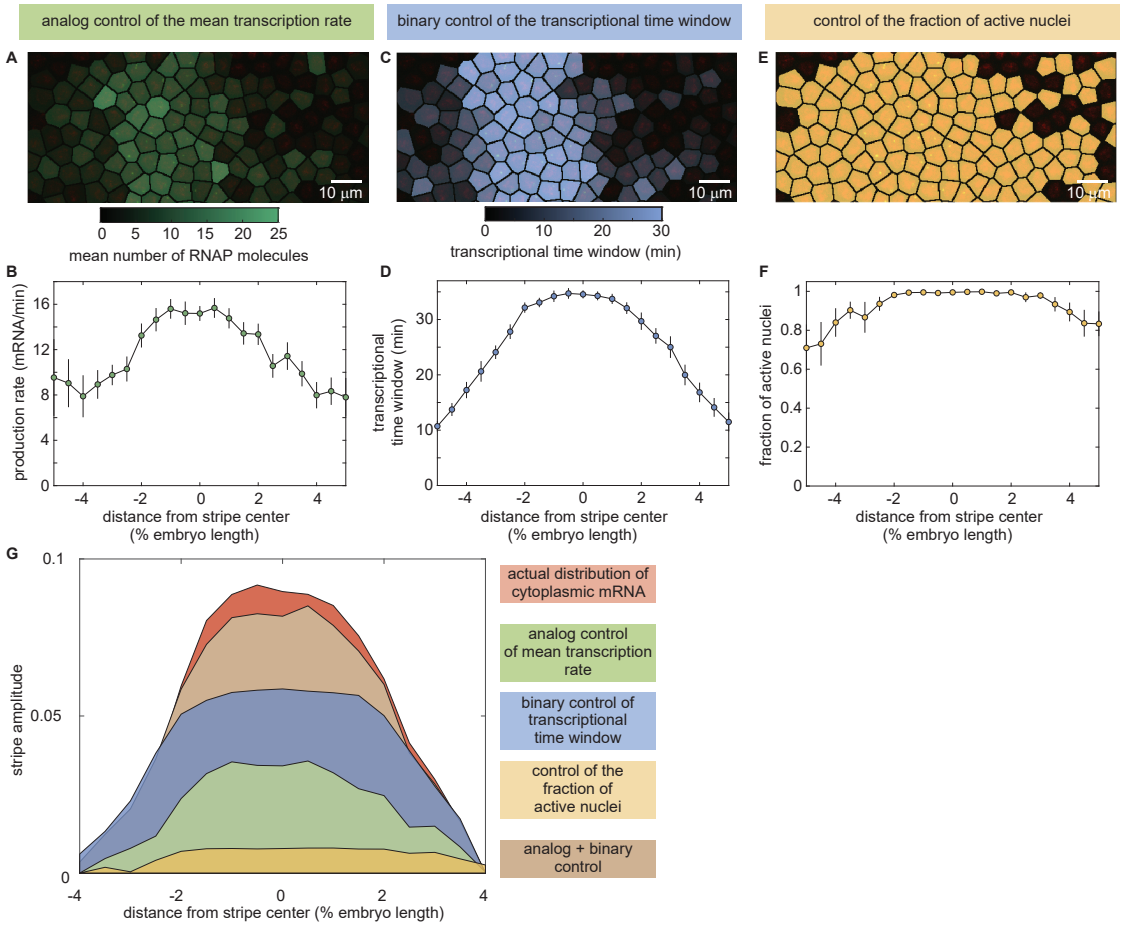

Figure 2.3: **Regulatory strategies for pattern formation in** *eve* **stripe 2. (A,B)** Time-averaged rate of mRNA production, **(C,D)** transcriptional time window, and **(E,F)** fraction of active nuclei as a function of position along the embryo. **(G)** Amplitude of the cytoplasmic mRNA distribution 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 SI Appendix, Figure A.5 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).

The analysis in Figure 2.3A-F reveals that each of the three regulatory strategies identified in Figure 2.1 is at play in the embryo, and that they all have the potential to contribute to pattern formation. However, these measurements alone cannot inform us on *how much* each of these strategies contributes to the cytoplasmic mRNA pattern. To quantify the degree to which each regulatory strategy contributes to the formation of *eve* stripe 2, we employed

the model described in Equation 2.2.

Figure 2.3G indicates the quantitative contribution of each regulatory strategy (each term on the right-hand side of Equation 2.2) to the formation of this cytoplasmic pattern. The cytoplasmic pattern of accumulated mRNA, corresponding to the left-hand side of Equation 2.2, was obtained by integrating from our live-imaging data (see SI Appendix, section A.3.2 for details). Regulation of the fraction of active nuclei along the embryo (Figure 2.3G, yellow) contributes negligibly to this mRNA pattern. In contrast, both the analog regulation of the mean rate (Figure 2.3G, green) and the binary control of the transcriptional time window (Figure 2.3G, 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 2.3G, brown) is sufficient to quantitatively recapitulate the stripe of cytoplasmic mRNA from single-cell transcriptional activity.

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

Are the binary and analog control strategies driven by distinct molecular mechanisms, or are they different manifestations of the same underlying process? To uncover the molecular mechanism behind the analog control of the mean rate of transcription, we analyzed the transcriptional activity of individual nuclei. Previous work demonstrated that the rate of gene expression at individual loci within the *eve* stripe 2 pattern is highly stochastic [27]. Indeed, as shown in Figure 2.4A, our data revealed punctuated peaks and troughs in the number of active RNAP molecules. These features have been related to the rate of RNAP initiation at the *eve* promoter by assuming that transcriptional activity is "burst-like", with the promoter rapidly loading mutliple RNAP molecules onto the gene at a constant rate during discrete "bursts" of activity interspersed with periods of inactivity [27]. This and other evidence from live imaging [27, 98, 68], as well as data from fixed-tissue approaches [230, 187, 317, 330], support a minimal two-state model of promoter switching (Figure 2.4B): promoters switch stochastically between ON and OFF states with rates k_{on} and k_{off}. In this model, promoters in the ON state engage in the loading of RNAP (and, correspondingly, mRNA production) at rate r. Thus we find that, in order to describe *eve* stripe 2 transcriptional dynamics, we need to account for both the short, transient ON periods dictated by transcriptional bursts, and a longer transcriptional time window that describes the period of time over which loci engage in this transcriptional bursting.

In the bursting model, the mean rate of transcription is given by the product of the fraction of time spent in the ON state with the transcription rate in this active state [233, 156, 258, 319]

$$\underbrace{R(x)}_{\substack{\text{mean}\\ \text{transcription rate}}} = \underbrace{r(x)}_{\substack{\text{RNAP loading}\\ \text{rate}}} \times \underbrace{\frac{k_{\text{on}}(x)}{k_{\text{on}}(x)+k_{\text{off}}(x)}}_{\substack{\text{fraction of time}\\ \text{in ON state}}}, \tag{2.3}$$

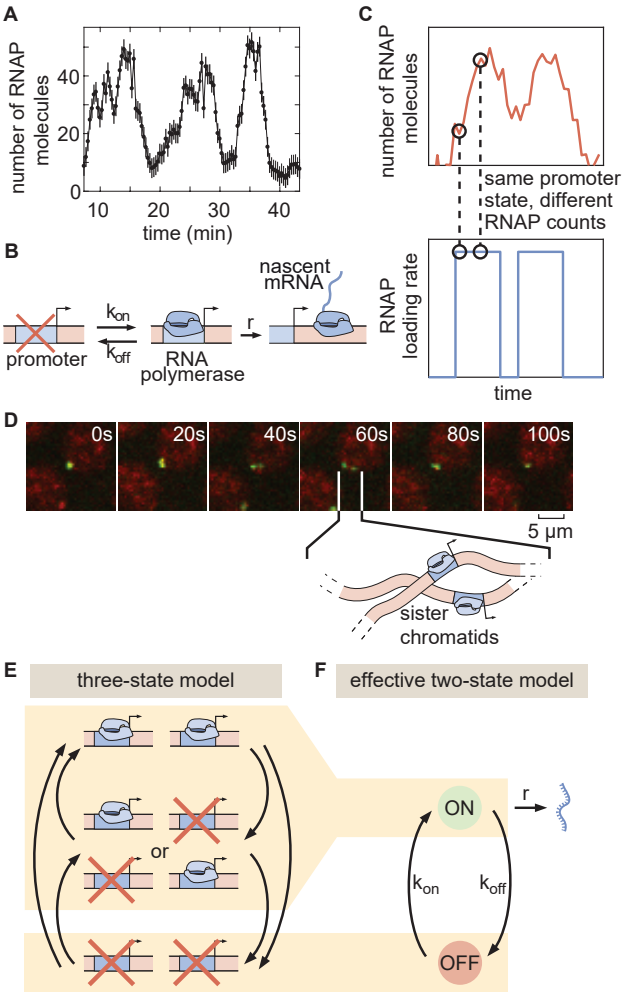

Figure 2.4: **Transcriptional bursting in *eve* stripe 2. (A)** Single-nucleus measurements reveal that nuclei transcribe in bursts. **(B)** Two-state model of bursting of a single promoter. **(C)** The same hidden rate of RNAP loading (bottom) can correspond to different observable numbers of RNAP molecules 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 [104].)

where all parameters are allowed to vary as a function of position along the embryo, x (see SI Appendix, section A.3.1 for details of this derivation). Thus, within this framework, the observed modulation of the mean rate of transcription across the stripe (Figure 2.3G,

green) implies that one or more of these bursting parameters is subject to spatially controlled regulation. However, the mean rate trend alone is not sufficient to identify *which* of the three bursting parameters (k_{on}, k_{off}, and r) is being regulated by the input transcription factors in order to control the average transcription rate. While each bursting parameter does not necessarily map directly to a single molecular step in the transcriptional cycle, identifying which parameter(s) are subject to regulation can help narrow the set of possible molecular mechanisms. For instance, variation in r could indicate that transcription factors play an active role in the recruitment of RNAP to the promoter, or in the release of RNAP from promoter-proximal pausing [269].

Typically, the *in vivo* molecular mechanism of transcription factor action is inferred from measurements of transcriptional noise obtained through snapshots of dead and fixed embryos or cells using theoretical models [22, 230, 327, 281, 187, 149, 265, 317, 229, 277, 7, 330, 128]. In contrast, MS2-based live imaging can directly inform on the dynamics of transcriptional bursting in real time. The MS2 approach, however, reports on the total number of actively transcribing RNAP molecules and not on the instantaneous rate of RNAP loading at the promoter, which is the relevant quantity for estimating k_{on}, k_{off}, and r. To date, approaches for extracting bursting parameters from such data in multicellular organisms have mainly relied on the manual analysis of single-nucleus transcriptional dynamics [27, 98] or autocorrelation-based methods that infer mean bursting parameters across ensembles of traces [170, 57, 68]. A computational method for inferring the rates of RNAP loading (Figure 2.4C, bottom) from the total number of actively transcribing RNAP molecules in single cells (Figure 2.4C, top) is thus needed to obtain the bursting parameters.

Hidden Markov models (HMMs) are widely used to uncover the dynamics of a system as it transitions through states that are not directly accessible to the observer [18]. However, our observable (the MS2 signal) does not correspond to the hidden variable of interest (the promoter state) in a one-to-one fashion (compare Figure 2.4C top and bottom). Instead, the observable MS2 signal reflects the net effect of promoter switching over a period equal to the time that an RNAP molecule takes to transcribe the whole gene. Thus, instantaneous fluorescence does not just depend on the current promoter state; it exhibits a dependence on how active the promoter has been over a preceding window of time, which effectively constitutes a memory for recent promoter states [46, 319, 55, 45]. Classic HMM approaches cannot account for this kind of system memory.

In order to model the process of transcription and extract the kinetic parameters of promoter switching, we augmented classic HMMs to account for memory (details about implementation of the method are given in SI Appendix, section A.3.4). Similar approaches were recently introduced to study transcriptional dynamics in cell culture and tissue samples [287, 217, 325, 332, 132, 31, 55, 87]. We used simulated data to establish that cpHMM reliably extracts the kinetic parameters of transcriptional bursting from live-imaging data (SI Appendix, section A.3.5), providing an ideal tool for dissecting the contributions from individual bursting parameters to observed patterns of transcriptional activity across space and time.

Before applying our model to real-time transcriptional data, we had to account for the rapid replication of the *D. melanogaster* genome at the beginning of each nuclear cycle [273],

which leads to the presence of two distinct *eve* loci within each fluorescent spot (Figure 2.4D and Video A.4). The first evidence of resolved chromatids appears as early as 8 minutes into nuclear cycle 14 (SI Appendix, Figure A.26)—coincident with the average onset time of transcription (SI Appendix, Figure A.4B). Moreover, our analysis indicates that replication of the relevant portion of the genome likely occurs in all *eve*-expressing nuclei by no later than 10 minutes following mitosis (SI Appendix, Figure A.26). Thus, we conclude that the vast majority of our data feature two distinct *eve* loci within each diffraction-limited transcription spot. Moreover, while the distance between sister loci varies over time (see, *e.g.* Figure 2.4D), they nonetheless stay in relatively close proximity to ensure their proper segregation from each other at the next mitosis [264] such that the fluorescent intensity signals extracted from our data reflect the integral over both loci (SI Appendix, Figure A.2). As a result, if we assume that each locus can be well-represented by a two-state model (OFF/ON) of transcriptional bursting, then an effective three-state model (OFF+OFF/ON+OFF/ON+ON) is needed to capture *eve* dynamics (Figure 2.4E). Thus, we elected to employ such a three-state model in our analysis. Due to conflicting evidence from previous studies [208, 187, 330], we made no prior assumptions about the nature or degree of cooperativity between sister chromatids either in transitions between activity states or in the rates of initiation in each state (see SI Appendix E for details). While these assumptions increased the complexity of our model, we believed that a conservative approach that left the model free to infer the presence or absence of sister interactions was warranted, given our ignorance regarding the nature and strength of interactions between adjacent gene loci. For ease of exposition, we present our main results in the context of an effective two-state model, in which, as detailed in SI Appendix, section A.3.1, the system is considered to be in the ON state so long as *either* chromatid is bursting (Figure 2.4F). Note that none of our conclusions below are affected by this choice of an effective model as shown in SI Appendix, section A.3.7, where we present full results for the three-state model.

A typical experimental trace for a nucleus in the core of the stripe is shown in Figure 2.5A, along with its best fit, which corresponds to the cpHMM-inferred promoter trajectory in Figure 2.5B. Our ability to infer the instantaneous promoter state in individual nuclei throughout development is further illustrated in Figure 2.5C and Video A.4. These data revealed that, as development progresses and the stripe sharpens, the *eve* promoter continuously fluctuates between the ON and OFF states on a time scale of approximately 1-2 minutes.

In order to infer time-averaged bursting parameter values, we grouped traces by position along the anterior-posterior axis. The rate of RNAP loading, r, remained constant throughout the stripe (Figure 2.5D), suggesting that none of the transcription factors regulating *eve* stripe 2 act directly on the rapid series of molecular steps involved in the initiation of transcription by RNAP. Similarly, we noted no significant spatial modulation of the rate of switching out of the ON state, k_{off} (Figure 2.5E). In contrast, the rate of switching into the ON state (also known as burst frequency), k_{on}, was strongly up-regulated in the stripe center (Figure 2.5E). These observations suggested that, control the mean transcription rate, transcription factors act primarily on the rate of promoter turning on, consistent with previous results both in embryos [317, 68, 98] and in single cells [281, 265, 229, 7]. This regulatory

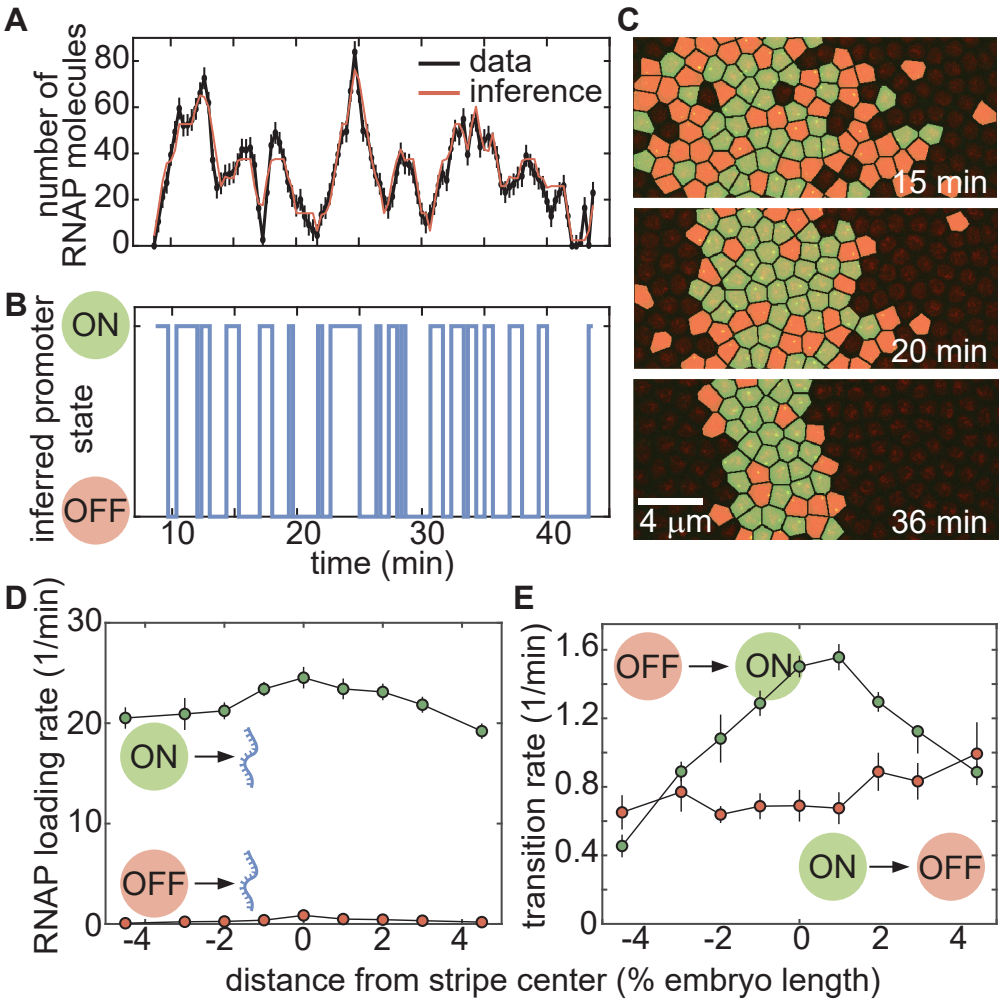

Figure 2.5: **Inferring bursting dynamics using a memory-adjusted Hidden Markov model.** **(A)** Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout 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 cpHMM 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 [104]; D,E, error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

modality increases the fraction of time that loci near the stripe center spend in the ON state (SI Appendix, Figure A.7 and [330]).

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

Having determined that the analog control of the mean transcriptional rate is realized by the modulation of the burst frequency, k_{on}, we next sought to uncover the molecular mechanism by which the binary regulation of the transcriptional time window is implemented. In one possible scenario, the onset of transcriptional quiescence at the end of the transcriptional time window would reflect a fundamental change to the molecular character of the transcriptional locus such that the bursting framework no longer applies. For instance, repressing transcription factors could alter in the local chromatin landscape by repositioning promoter or enhancer nucleosomes [4]; changes that could block the binding of activators at the stripe 2 enhancer or of general transcription factors at the promoter, and thus abolish further activator-mediated bursting (Figure 2.6A, top). Alternatively, if the rates of promoter switching vary in time, then the time window could be explained without invoking an extra silenced state that is mechanistically distinct from the processes driving transcriptional bursting. Specifically, transcriptional quiescence could be achieved by progressively reducing the frequency (k_{on}), intensity (r), and/or duration ($1/k_{\text{off}}$) of transcriptional bursts. For example, it is possible that increasing repressor levels in the stripe flanks could disrupt the capacity for activators to initiate transcription bursts via short-range quenching interactions [60]; a mechanism that would manifest as a decrease in k_{on} over time.

In order to determine whether quiescence can be explained within the bursting framework, we divided the stripe into the five regions shown in Figure 2.6B. For each region, we sought to determine whether the bursting dynamics varied over time in a manner that could explain the dynamics of entry into quiescence of individual nuclei (Figure 2.6C). To probe for this time-dependence in transcriptional bursting, we extended our cpHMM method to obtain promoter-bursting parameters over discrete periods of time by performing inference on our live-imaging data using a sliding window (see SI Appendix, section A.3.4 for details). Our inference revealed that the rate of promoter turn on, k_{on}, varied significantly in time (Figure 2.6D). Specifically, k_{on} decreased in both the anterior and posterior stripe boundaries (black and red curves) as development progressed and the fraction of active nuclei decreased (grey shaded region), while loci in the stripe center (green and yellow curves) exhibited a significant *increase* in k_{on}. Further, while relatively constant at most positions along the stripe, both the rate of RNAP loading when in the ON state, r, and the rate of promoter turn off, k_{off}, decreased slightly in (Figure 2.6E and F).

These findings confirmed our time-averaged inference results (Figure 2.5D and E) indicating that k_{on} was the primary kinetic pathway through which transcription factors influence *eve* stripe 2 transcription dynamics. Moreover, the coincidence of the decrease in k_{on} in flank nuclei with the onset of transcriptional quiescence (grey shaded region in Figure 2.6D) seemed to suggest that, at least in part, quiescence in the stripe flanks could be driven by the

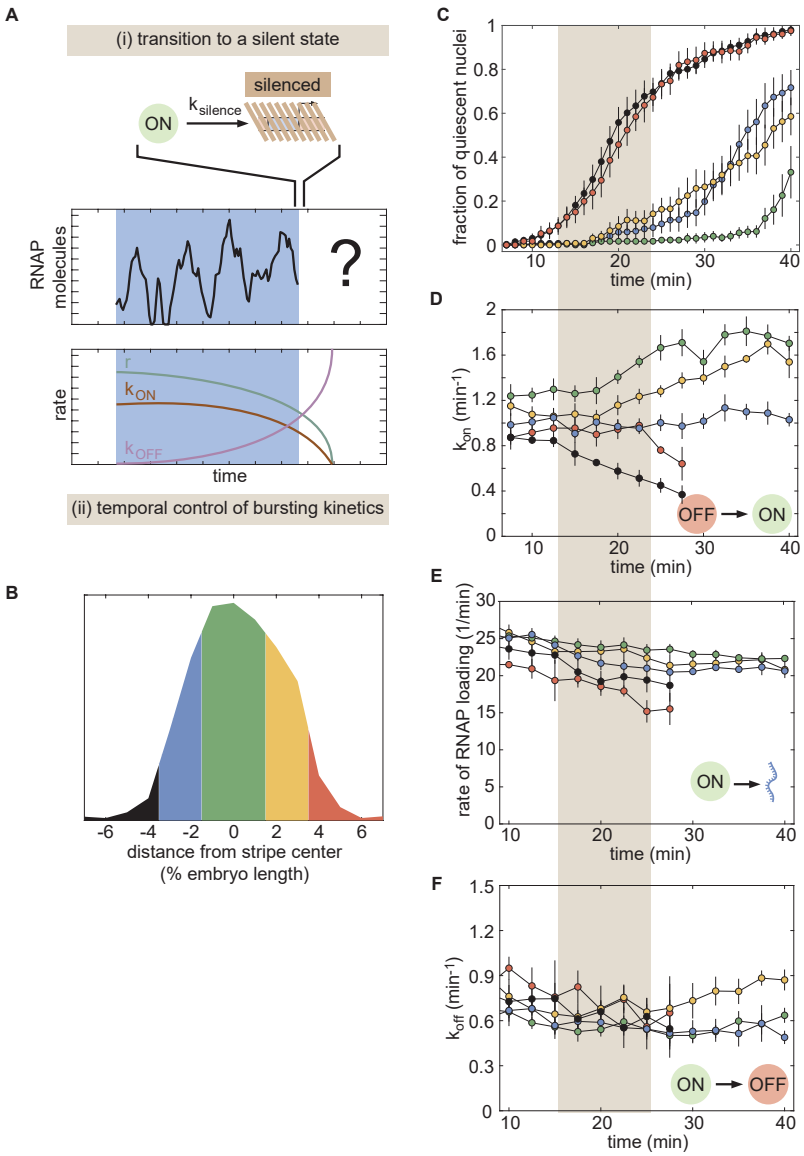

Figure 2.6: **Investigating the molecular character of transcriptional quiescence. (A)** Two hypotheses explaining promoter quiescence onset: (i) a transition into an alternative, long-lived transcriptionally silent state and (ii) the modulation of one or more bursting parameters over time. **(B)** Division 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 RNAP 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 cpHMM inference results for bootstrapped samples of experimental data. See Materials and Methods for details)

temporal modulation of bursting parameters (Figure 2.6A, bottom). However, other trends in our data were not consistent with the view that a decrease in k_{on} drives transcriptional quiescence.

Although 70% and 50% of nuclei in the regions directly anterior and posterior of the stripe center were quiescent by 40 min into the nuclear cycle (blue and yellow curves in Figure 2.6C), we detected no corresponding decrease in k_{on}. In fact, k_{on} actually *increased* in some inner regions of the stripe (Figure 2.6D)—a trend that would increase overall transcriptional activity and would therefore go against the establishment of transcriptional quiescence.

The divergent outcomes observed in the central stripe regions, with the rate of transcriptional bursting remaining constant or increasing at *eve* loci within the engaged population of nuclei even as loci in neighboring nuclei turn off for good, runs counter to the hypothesis that quiescence is driven by the temporal modulation of the promoter switching parameters. It is conceivable that temporal changes in bursting parameters associated with the onset of quiescence occur too rapidly to be captured by our model. However, as discussed in SI Appendix, section A.3.9 , these changes would need to occur on the same time scale as bursting itself (1 to 3 min). Given that both the other temporal trends detected by our inference (Figure 2.6) and the shifts in the input transcription factors themselves (SI Appendix, section A.3.8) unfold on significantly slower timescales (5-15 min), we concluded that while possible, a scenario where bursting dynamics are changing too quickly to detect is unlikely.

The contradictory trends observed in the stripe center and flanks indicated that entry into transcriptional quiescence might involve processes not captured within the bursting model (Figure 2.6A, top), thus suggesting that binary control of the transcriptional time window and the transcriptional bursting driving the analog control of the mean transcription rate may arise from distinct molecular processes.

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

eve stripe 2 is mainly established by the combined action of two activators, Bicoid and Hunchback, and two repressors, Giant and Krüppel [93, 278, 6]. If transcriptional bursting and the transcriptional time window are controlled by distinct molecular processes, then distinct forms of regulatory logic may be at play. For example, the Bicoid and Hunchback activators could control transcriptional bursting, while the Giant and Krüppel repressors could dictate the entry into the quiescent state. In order to reveal the molecular logic controlling each regulatory strategy, we sought to correlate the fraction of nuclei that have entered the quiescent state (Figure 2.7A) and the fraction of nuclei in the bursting ON state (Figure 2.7B) with the corresponding spatiotemporal patterns in the input concentrations of these four transcription factors.

We measured Bicoid concentration profiles using a a well-established Bicoid-GFP fusion [116] and obtained spatiotemporal concentration profiles for Krüppel, Giant, and Hunchback from published immunofluorescence data [74]. We combined these data with our live-imaging data of *eve* stripe 2 transcriptional activity to generate an "average embryo" in which the

concentration of all relevant inputs and the output transcriptional activity at each point in time and space were known (Figure 2.7C and Video A.4). Building upon previous work [139], we utilized logistic regressions to probe the regulatory role played by each of these four factors in the spatiotemporal control of transcriptional bursting and the transcriptional time window. The logistic regression is a widely used method of inferring predictive models in processes with binary outcomes. For example, in order to query the regulatory logic behind the control of the transcriptional time window, the model probes the impact of each transcription factor on the relative likelihood of a locus entering the quiescent state versus the likelihood of remaining transcriptionally engaged such that

$$\log\left(\frac{P_{\text{quiescent}}}{P_{\text{engaged}}}\right) = \beta_0 + \beta_1\,[\text{Bcd}] + \beta_2\,[\text{Hb}] + \beta_3\,[\text{Gt}] + \beta_4\,[\text{Kr}]\,, \tag{2.4}$$

where the coefficients β_n indicate the magnitude and nature (activating or repressing) of the transcription factor's regulatory function. In estimating these coefficients, we used prior knowledge about the function of each transcription factor, requiring Bicoid and Hunchback to play activating roles, and Krüppel and Giant to play repressing roles [280, 278]. We used an analogous model to investigate the regulatory logic controlling transcriptional bursting by inferring the factors that determine the relative likelihood that nuclei are in the bursting ON versus the OFF state, P_{ON}/P_{OFF}.

Our analysis of the fraction of nuclei in the quiescent state revealed that no single transcription factor can explain quiescence dynamics (Figure 2.7D and E). However, a simple model in which increasing levels of the repressors Giant and Krüppel drive the onset of transcriptional quiescence in the anterior and posterior stripe flanks, respectively, recapitulated experimentally observed trends. The further addition of Hunchback and/or Bicoid had no impact on the model's predictive power, suggesting that activator concentrations have no influence over the molecular processes responsible for silencing. Relaxing constraints on the functional role of each transcription factor–for instance, allowing the presumed activators to function as repressors–also provided no significant improvement over models presented here as shown in SI Appendix, section A.3.8.

We next turned our attention to the relationship between transcription factor levels and the fraction of nuclei in the ON state (Figure 2.7B). Unlike the transcriptional time window, repressor levels alone could not recapitulate the observed bursting profile; Hunchback levels were also necessary in order to fully capture the spatiotemporal bursting dynamics (Figure 2.7E and G). Specifically, we linked a rise in Hunchback concentration to an observed rise in the fraction of nuclei in the ON state in the stripe center between 30 and 35 min into the nuclear cycle (Figure 2.7B and F).

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

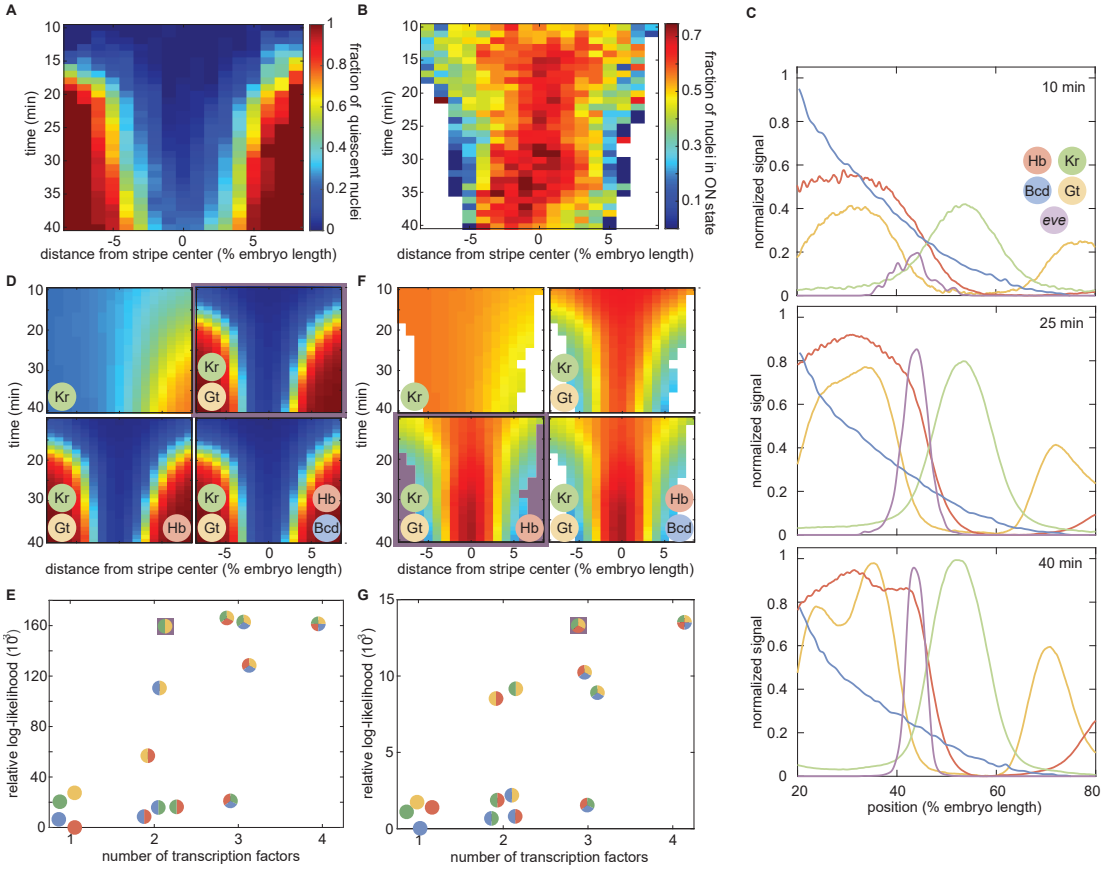

Figure 2.7: **Probing the regulatory logic of bursting and the transcriptional time window.** **(A)** Fraction of nuclei in the transcriptionally quiescent state and **(B)** fraction of nuclei in the bursting ON state as a function of time and position along the embryo. **(C)** Snapshots of input transcription factor levels and predicted *eve* mRNA levels of our "average" embryo at 10, 25, and 40 minutes into nuclear cycle 14. **(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).

in which the long-lived trancriptionally silent state observed in flank nuclei constitutes a distinct molecular state outside of the bursting model.

Discussion

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

A priori, there are several distinct regulatory strategies at the single-cell level capable of generating spatially differentiated patterns of cytoplasmic mRNA (Figure 2.1), each with distinct implications for the nature of the underlying molecular processes at play. Several recent studies have revealed that the average rate of transcription is mainly modulated across the embryo by tuning the frequency of transcriptional bursting [186, 27, 317, 68, 98, 330]. Yet it has remained unclear whether this modulation of the rate of transcription (and thereby mRNA production) is the dominant modality by which input concentrations of transcription factors drive the formation of patterns of gene expression, or if, instead, it is simply the most readily apparent mechanism among multiple distinct control strategies.

In this work, we derived a simple theoretical model that predicts how the interplay between regulatory strategies at the single-cell level dictates the formation of a cytoplasmic gene expression pattern (2.2). We applied this model to single-cell live-imaging measurements of an MS2 reporter driven by the *eve* stripe 2 enhancer; an approach that allowed us to dissect the regulatory logic of a well-characterized regulatory element free from the confounding influences of other enhancers located in the endogenous *eve* locus. We demonstrated–quantitatively–that the modulation of the mean rate of transcription is alone insufficient to account for the formation of a sharp stripe of gene expression (Figure 2.3F, green). We discovered that the window of time over which promoters engage in transcription is sharply controlled along the axis of the embryo (Figure 2.3C and D) and that the joint action of the analog control of the rate of transcription and the binary control of the duration of transcription is necessary and sufficient to quantitatively recapitulate most of the full stripe profile (Figure 2.3F, brown). While this work focused on dissecting the regulatory logic of the *eve* stripe 2 enhancer in the context of a minimal construct, it is important to note that our conclusions are not limited to this reporter construct and also apply to the endogenous regulation of *eve*. As shown in SI Appendix Figure A.11, an analogous analysis performed on a the expression dynamics of a reporter BAC containing the full endogenous *eve* locus [11] indicated that stripe formation in this endogenous context is also dominated by the interplay between the regulation of the mean rate of transcription and of the transcriptional time window.

Here, we contribute to a growing body of work that illustrates the utility of using simple quantitative models to extract insights into the workings of complex biological phenomena [139, 202, 317]. Our discovery of the key role of the differential duration of the transcriptional time window in pattern formation was only made possible by biological numeracy; that is, by going beyond the *qualitative* description of pattern formation and demanding a *quantitative* agreement between our theoretical predictions and the experimental data [239]. While it is widely appreciated that genes are expressed for discrete windows of time over the course of development [104, 193, 192], we have demonstrated that—in the case of this *eve* stripe 2 reporter—this binary transcriptionally engaged/quiescent logic is actively regulated by transcription factors to drive pattern formation. Thus the differential duration of transcriptional activity comprises a necessary element of any quantitative description of pattern formation.

Our work contributes to an increasingly diverse and exciting discourse in quantitative developmental biology regarding the importance of the temporal component of transcriptional regulation in specifying developmental outcomes. For example, one recent study has demonstrated that the limited readout time imposed by short nuclear cycles in early *Drosophila* development places strict constraints on the kinds of regulatory architecture that could be responsible for driving observed patterns of *hunchback* gene expression [296]. Other recent work has indicated that the pioneer factor Zelda plays a key role in regulating both the timing and probability of transcriptional activation following mitosis [320, 75]. Our work complements these previous observations by exploring yet another aspect of the interplay between timing and transcriptional regulation. We have shown that, in the case of *eve* stripe 2, transcription factors regulate the onset of transcriptional quiescence, t_{off}, across the stripe, thus demonstrating that the embryo actively leverages the differential duration of transcriptional engagement as a strategy to generate patterns of gene expression. Together, these recent findings suggest that, if the field is make progress toward a predictive picture of pattern formation in development, it will be necessary to go beyond the widespread steady-state, *static* picture of pattern formation in development put forward by previous single-cell transcriptional activity studies that focused on the study of snapshots of fixed embryos [230, 187, 317, 330] and embrace a *dynamical* description that acknowledges that development is a process that occurs outside of steady state [11].

To determine whether this binary control of the transcriptional time window and the analog control of the mean transcription rate share a common molecular mechanisms, we utilized a variety of theoretical and computational tools in conjunction with our live-imaging data. Specifically, to uncover how the mean rate of transcription is regulated across the stripe, we developed a cpHMM that is capable of inferring the instantaneous activity state of individual gene loci from MS2 traces. We used this cpHMM to infer average promoter-switching parameters across the stripe (Figure 2.5). In agreement with previous measurements of various gene expression patterns [317, 68, 98, 330], our results revealed that the burst frequency (k_{on}) is the main bursting parameter regulated by the input transcription factors across *eve* stripe 2. This increase in k_{on} in the stripe center functions to increase the fraction of time that nuclei spend in the active transcriptional state.

Importantly, our cpHMM algorithm is not limited to the *eve* stripe 2 system and should

prove useful to infer the underlying regulatory kinetics of any gene that is tagged using approaches such as the MS2 or PP7 systems in any organism [170, 98]. Further, the method could be used to infer the state of the ribosome as mRNA is being translated into protein in novel single-molecule *in vivo* translation assays [219, 306, 322, 315]. Thus, we envision that our method will be useful for the broader biophysical analysis of *in vivo* cellular processes at the single-molecule level.

Having identified k_{on} as the primary kinetic mode by which transcription factors modulate the mean rate of expression across *eve* stripe 2, we next sought to probe the relationship between bursting and the transcriptional time window (Figure 2.6A). We adapted our cpHMM to go beyond time-independent models of promoter switching to infer the regulation of these rates across both space and time. We observed striking temporal trends indicating that the burst frequency responds dynamically to time-varying transcription factor inputs. However, we noted a significant disconnect between temporal trends in the burst frequency and the onset of transcriptional quiescence. In particular, k_{on} either increased or remained constant near the stripe center even as a significant fraction of *eve* nuclei transitioned into quiescence (Figure 2.6C and D). We reasoned that the onset of transcriptional quiescence is likely not the result of a progressive reduction in burst frequency, amplitude, or duration, and is instead driven by molecular processes that are distinct from those that regulate transcriptional bursting, such as a repressor-induced shift in nucleosome position that prevents activating transcription factors from binding to the stripe 2 enhancer.

To test this hypothesis, we utilized a logistic regression framework and time-resolved data for the primary regulators of *eve* stripe 2 to query the regulatory logic exhibited by the time window and bursting, respectively (SI Appendix, section A.3.8). In this context, the logistic regressions served as a robust statistical tool for drawing inferences from existing data that were not obvious (or verifiable) by simple visual inspection. Consistent with our time-resolved cpHMM results, the two regulatory strategies responded to transcription factor concentrations in different ways. On the one hand, increasing levels of Giant and Krüppel were sufficient to explain the onset of transcriptional quiescence in the stripe flanks (Figure 2.7A and D). This observation points to a model in which repressor levels act unilaterally—without respect to coincident levels of activator proteins—to shut off transcription at loci in an (at least effectively) irreversible fashion. Conversely, the joint action of Giant, Krüppel, and Hunchback was necessary to recapitulate the observed pattern of transcriptional bursting (Figure 2.7B and F).

This difference in the regulatory logic observed for the two strategies dissected in this work suggests that control of the transcriptional time window and the modulation of the average transcription rate arise from two distinct, orthogonal molecular mechanisms. It is also notable that our model finds that Hunchback activation is necessary in order to fully explain the observed pattern of transcriptional bursting in *eve* stripe 2. A recent study has suggested that Hunchback actually functions as a repressor of *eve* stripe 2, and that indirect activation occurs via counter-repression of Hunchback by the maternal factor Caudal [303]. While we cannot rule out the possibility that Hunchback acts indirectly, the strong link between rising Hunchback levels and the increase in *eve* 2 activity in the stripe center is most consistent with Hunchback playing a traditional activating role. Additional work will

be necessary to determine whether this correlation between rising Hunchback levels and increased stripe activity can be reconciled with the counter-repression hypothesis proposed in [303]. Lastly, we note that the striking absence of a direct functional role for Bicoid in the regulation of either phenomenon suggests that, while Bicoid is almost certainly necessary for the expression of *eve* stripe 2 [278], it does not play a direct role in dictating the magnitude or duration of *eve* stripe 2 transcription. In this interpretation, Bicoid functions like a general transcription factor, facilitating the transcription of *eve* 2 without directly conferring spatiotemporal information.

In addition to gleaning valuable insights into the mechanisms driving the regulation of transcription of the *eve* stripe 2 enhancer, our logistic regression framework makes quantitative and falsifiable predictions about the regulation of this stripe for combinations of input transcription factors concentrations that the embryo does not encounter in the wild-type setting. For instance, our finding that repressors alone drive the onset of transcriptional quiescence predicts that this onset should be unaltered in mutated *eve* stripe 2 enhancers where some or all Hunchback binding sites have been disrupted. In this scenario, transcriptional activity, initially arising due to permissive levels of Bicoid, would shut off in precisely the same manner as observed for the full enhancer (compare upper right panel in Figure 2.7D to lower left panel). In the absence of Hunchback activation, the model also predicts reduced levels of transcriptional bursting, particularly late in nuclear cycle 14 (compare upper right panel of Figure 2.7F to lower left panel). Similarly, our model could be used to predict the expected stripe profile in mutant embryos, where the expression of one or more gap genes has been altered or abolished. We note, however, that the interconnected nature of the gap gene network [142] means that it would be necessary to re-image *all* three gap genes that regulate *eve* stripe 2 in order to generate data such as shown in Figure 2.7C, since any change to one will affect the expression patterns of all. Thus, additional binding site mutation studies similar to the one described above likely represent the most direct path to testing our model's predictions. Taken together, we anticipate that the approaches outlined in this work will serve as a tool both for extracting additional insights from experimental data *and* for motivating additional experiments aimed at answering meaningful questions about the mechanistic underpinnings of gene regulation.

We also observe that certain aspects of the system remain beyond the scope of our model. Most notably, while loci engaged in transcriptional bursting appear to continuously sense changes in transcription factor concentrations, it remains an open question whether loci continue to read out transcription factor concentrations following the onset of transcriptional quiescence. While the transition *appears* irreversible in our data, it is possible that quiescence is, in fact, reversible but simply not observed because repressor levels increase over time in our region of interest. The temporally resolved manipulation of repressor concentration through, for example, optogenetics [137], could make it possible to deplete repressors from the nucleus after transcriptional quiescence in order to determine whether this quiescent state is reversible.

In order to further test these and other hypotheses, it will be critical to move beyond spatiotemporal averages for transcription factor inputs (Figure 2.7C) and, instead, use live single-nucleus measurements to directly correlate input transcription factor concentration

dynamics with the corresponding transcriptional activity at the single-cell level [135]. Experimentally, we recently demonstrated the simultaneous measurement of inputs and outputs in single nuclei of a living fly embryo using novel genetically encoded LlamaTags [24]. We believe that utilizing this novel technique, in conjunction with the theoretical methods presented here, to query the effects of targeted disruptions to transcription factor binding domains on regulatory enhancers will constitute a powerful assay for querying transcription factor function at the molecular level. Thus, there are clear experimental and theoretical paths to uncovering the detailed quantitative mechanisms behind the molecular control of transcriptional bursting and quiescence in development. Such a quantitative description is a necessary step toward a predictive understanding of developmental decision-making that makes it possible to calculate developmental outcomes from knowledge of the nature of the transcription factor interactions within gene regulatory networks.

2.3 Materials and methods

2.3.1 Reporter construct

This work employed the same *eve* stripe 2 reporter construct developed by [27]. This construct contains the *even-skipped* (*eve*) stripe 2 enhancer and promoter region (spanning -1.7 kbp to +50 bp) upstream of the *yellow* reporter gene. 24 repeats of the MS2 stem loop sequence were incorporated into the 5' end of the reporter gene.

2.3.2 Sample preparation and data collection

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

2.3.3 Image analysis

Image analysis of live embryo movies was performed based on the protocol found in [104] with modifications to the identification of transcriptional spots, which were segmented using

the Trainable Weka Segmentation plugin for FIJI using the FastRandomForest algorithm [5]. In comparison with a previous algorithm based on Difference of Gaussians [187, 104, 27], this alternative spot segmentation approach was found to be superior for the detection of dim transcription spots—a feature critical to establishing the precise timing of the cessation of activity at transcriptional loci.

2.3.4 cpHMM inference code

All scripts relating to the cpHMM inference methodology developed in this work are available at the GarciaLab/cpHMM GitHub repository. See the extended SI Materials and Methods section, as well as SI Appendix, section A.3.4 for additional details.

2.3.5 Acknowledgements

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

Chapter 3

Kinetic sculpting of the seven stripes of the Drosophila *even-skipped* gene

This chapter is a reproduction of reference [13]

3.1 Abstract

We used live imaging to visualize the expression dynamics of the Drosophila melanogaster even-skipped gene at single-cell and high temporal resolution as its seven stripe expression pattern forms, and developed tools to characterize and visualize how transcriptional bursting varies over time and space. We find that despite being created by the independent activity of five enhancers, even-skipped stripes are sculpted by the same kinetic phenomena: a coupled increase of burst frequency and amplitude. By tracking the position and activity of individual nuclei, we show that stripe movement is driven by the exchange of bursting nuclei from the posterior to anterior stripe flanks. Our work provides a conceptual, theoretical and computational framework for dissecting pattern formation in space and time, and reveals how the coordinated transcriptional activity of individual nuclei shape complex developmental patterns.

3.2 Introduction

The patterns of gene expression that choreograph animal development are formed dynamically by an interplay between processes - transcription, mRNA decay and degradation, diffusion, directional transport and the migration of cells and tissues - that vary in both space and time. However the spatial aspects of transcription have dominated the study of developmental gene expression, with the role of temporal processes in shaping patterns receiving comparably little attention [25, 104]. Gene expression patterns are dynamic on many levels. They form, change and disappear over time, often as cells, tissues and organs are forming and moving in the developing embryo [173]. Furthermore the transcriptional process that creates these patterns is itself highly dynamic. The classical view of transcription as

a switch or a tunable rheostat has been replaced in recent years by the recognition that mRNA synthesis occurs in bursts, with promoters switching stochastically between an ON state where polymerases are loaded and begin elongating, and an OFF state where few or no new transcripts are initiated (Figure 3.1A) [327, 111, 19, 144, 50, 324, 247, 186, 223, 187, 317, 180, 98, 68, 128, 165].

A slew of studies, from theoretical models [160, 233, 255, 257, 319, 46, 47, 268, 156, 258] to imaging-based analyses [317, 98, 265, 149, 111, 217, 287, 281, 229, 7, 128, 330] , have shown that overall rates of mRNA synthesis can be adjusted by controlling the bursting process. Changing the duration or bursts, the separation between bursts, or the rate at which polymerases are loaded during a burst (Figure 3.1B) will affect mRNA production, and modulating any or all of these parameters over space and time could, in principle, produce arbitrarily complex output patterns. However, it remains unclear how diverse the kinetic strategies employed by different regulatory sequences actually are, and what, if anything, constrains how these different kinetic parameters are used by evolution to shape patterns of expression.

In this paper we set out to compare the ways in which different enhancers that drive similar types of spatiotemporal patterns during animal development deploy transcriptional bursting to produce their outputs by examining transcription at the single-cell level in living embryos. We use as our model the Drosophila melanogaster even-skipped (eve) gene whose seven stripes ring the embryo in the cellularizing blastoderm (nuclear cycle 14; nc14) in the hour preceding gastrulation [286, 92, 148, 195, 198, 93] .

The eve stripes are produced through the largely independent activity of five discrete enhancers (Figure 3.1C) that drive individual stripes (the stripe 1, stripe 2, and stripe 5 enhancers) or pairs of stripes (the stripe 3/7 and stripe 4/6 enhancers) [125, 112, 279] . These enhancers respond in different ways to canonical maternal factors Bicoid (Bcd) and Caudal (Cad), and gap genes Hunchback (Hb), Giant (Gt), Krüppel (Kr), Knirps (Kni) and Sloppy Paired 1 (Slp1), among others, balancing activating and repressive inputs to generate novel output patterns [201] . For example, the eve stripe 2 enhancer is activated in the anterior by Bcd and Hb, and repressed by Gt and Kr, ultimately expressing in a stripe of nuclei that fall between the domains occupied by these two repressors [93, 278] .

Transcriptional bursting is widespread during D. melanogaster development [187, 317, 27, 98, 22, 135, 330, 230, 165] . For example, [27] utilized the MS2 system, which exploits the interaction between the phage MS2 coat protein (MCP) and a short RNA stem loop to fluorescently label nascent transcripts as they are being synthesized [14, 104, 91] , to directly visualize and quantify transcription from an eve stripe 2 transgene at single nucleus resolution. They showed that the stripe is generated by bursts of transcriptional activity in the nuclei that form it, and that the aggregate pattern is highly dynamic, forming and dissipating rapidly during nc14.

Our objective in carrying out this work was twofold: first, to characterize the detailed dynamics of this classical and well-studied pattern as a means to reveal how multiple enhancers dictate potentially distinct bursting dynamics to shape a gene expression in the embryo, and second, to establish a rigorous systematic framework for analyzing such data, and conceptual paradigms for characterizing what we observe from this new type of experiment. Indeed,

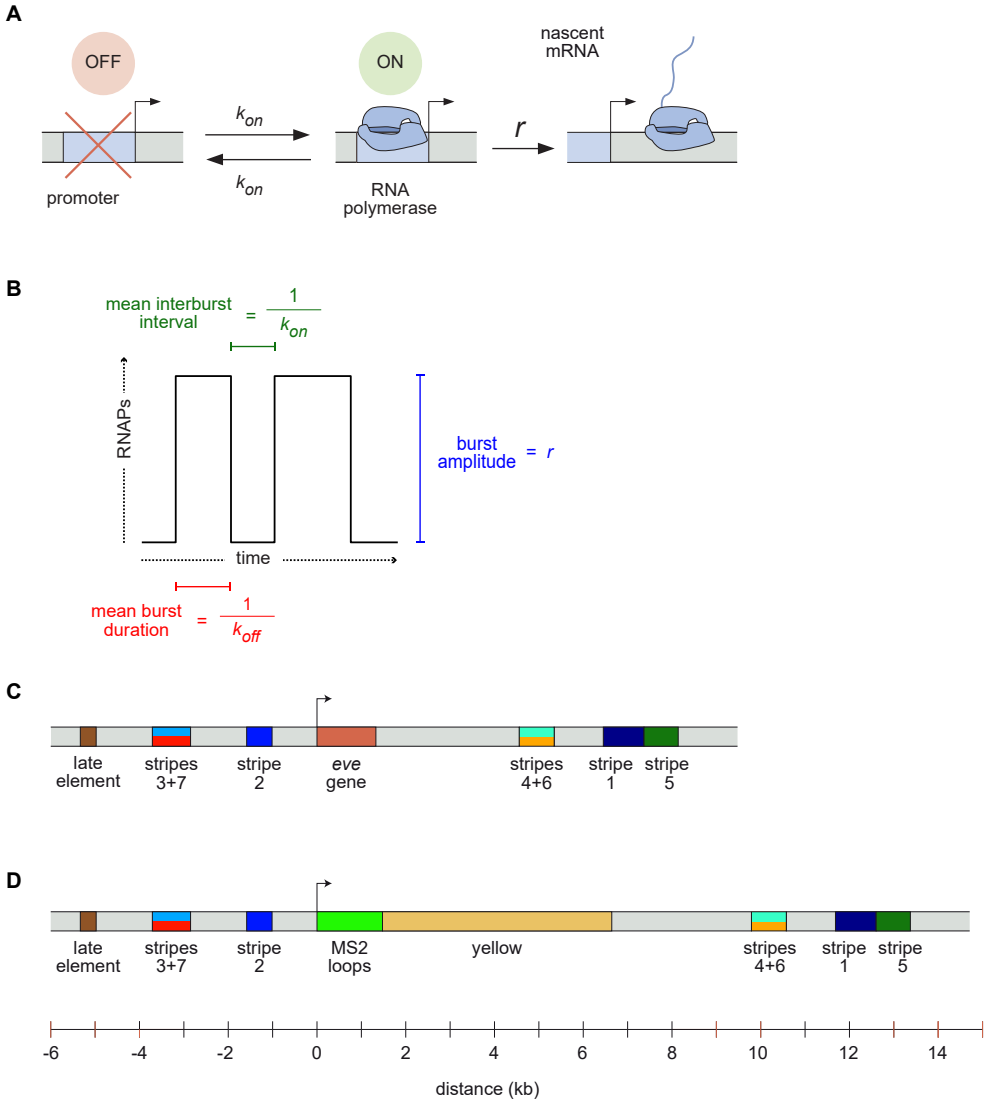

Figure 3.1: **Visualizing live transcription from the seven stripes of D. melanogaster *even-skipped*.** **(A)** Simple model of transcriptional bursting by promoter switching between ON and OFF states. **(B)** The promoter switching parameters define the burst duration, the duration between bursts, and amplitude. **(C)** Wild-type eve locus showing the five stripe enhancers (1,2,3+7,4+6,5) and the late enhancer element. Colors for individual stripes are used throughout figures. **(D)** Layout of the engineered eve BAC showing the locations of the MS2 stem loop array and yellow gene.

the advent of live imaging in the context of development calls for the establishment of a new language and new metrics for characterizing the formation of gene expression patterns in space and time.

We use a variety of new analyses to generate a kinetic fingerprint of eve transcription during stripe formation - a record of temporal and spatial variation in the bursting state of the promoters of 3,000 nuclei covering all seven stripes throughout nc14 - and to visualize different aspects of eve regulation. We find that all seven eve stripes are sculpted by the same regulatory strategies: the elimination of new bursts between stripes; the enhancement of bursting across stripes through a coupled increase in k_{on} and r; and the refinement and movement of stripe positions by the addition of bursting nuclei along the anterior edge of the stripes and the loss of bursting along their posterior edge. Thus, in this experiment and with our new set of analytical tools, we capture not only how single cell transcriptional activity encodes the formation of the stripes, but also how this activity is modulated in space and time in order to create and displace a complex pattern of gene activity across the embryo.

3.3 Results

3.3.1 Live imaging of eve expression

We used recombineering [308] to modify a bacterial artificial chromosome (BAC) [299] containing the D. melanogaster eve gene and all of its enhancers and regulatory elements [300] , replacing the coding region with an array of MS2 stem loops followed by the D. melanogaster yellow (y) gene (Figure 3.1D; [27]). The 4,329 base pair y gene, which is longer than the endogenous eve transcript, is phenotypically neutral and provides a means to increase the number of RNA Polymerase II (Pol II) molecules loaded onto the gene in order to amplify the signal (see Methods for a discussion of how the structure of the reporter genes affects the fluorescence signal, analyses and inferences performed throughout this work). We inserted the engineered BAC into a targeted site on chromosome 3L using ΦC31 integrase-mediated recombination [89] , and homozygosed the line, which showed no signs of adverse effects of the transgene.

We crossed males from this line with females carrying transgenes that express in embryos an RFP-labeled histone to visualize nuclei, and an MCP-GFP fusion that binds the MS2 RNA stem loops. The result is the direct visualization and quantification of nascent transcripts at the transgene locus as fluorescent puncta [104]. The temporal and spatial pattern of eve transgene transcription recapitulates the well-characterized dynamics of eve expression, most notably formation of the characteristic seven stripes in the late blastoderm (Figure 3.2; [286, 148, 195]. Further, as recently demonstrated in [165] this BAC reporter construct quantitatively recapitulates the cytoplasmic eve mRNA pattern as measured by FISH [165, 196, 184].

We used laser-scanning confocal microscopy to record, with high temporal resolution and high magnification, two color (MCP-GFP and histone RFP) movies of embryos from before nc14 through gastrulation. We optimized our data collection strategy to sample

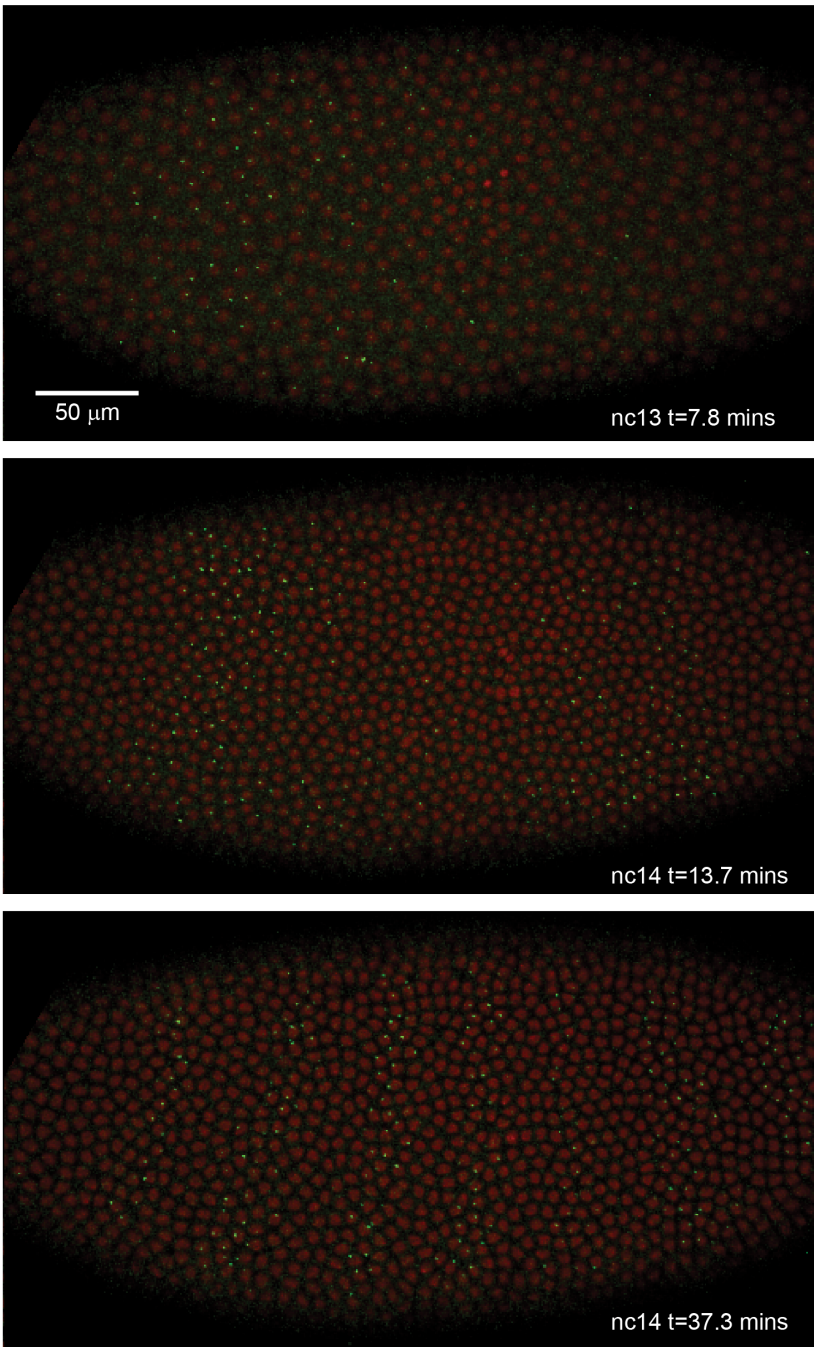

Figure 3.2: **Live expression of *even-skipped*.** Stills from maximum projection renderings of image stacks of an embryo spanning all seven stripes. This movie was collected with a 40x objective for illustration purposes only. Movies used for data analysis were collected at higher resolution as described in the text.

multiple stripes (three to five) in each movie, to obtain high temporal resolution (one Z-stack, corresponding to each time point of our movies, every 16.8 seconds) and to have optimal signal to noise with minimal bleaching. In total, we collected 11 movies , with every stripe imaged at least five times (see Table 1).

We used a custom image processing pipeline [104, 165] to identify nuclei, track fluorescent puncta and extract fluorescence intensities in each nucleus over time. The resulting data (File S1) contains fluorescence traces from 2,961 nuclei at an interpolated time interval of 20s, representative examples of which are shown in Figure 3.3A.

We first sought to reexamine the previously characterized temporal dynamics of stripe formation [286, 92, 148, 195] using the increased temporal resolution (relative to earlier analyses of fixed embryos and of slowly maturing fluorescent protein reporters) of these data (Figure 3.3B). Early imaging studies described eve as being expressed broadly in nc13 and early nc14 embryos before refining sequentially into four, then seven stripes [93, 198] . Subsequent work with improved labeling and imaging techniques [286, 92] revealed an initial phase with broad domains in the anterior and posterior, followed by the formation of stripes from within these broad domains and, eventually, amplification of the stripe pattern.

During nc14, we first observe eve transcription beginning approximately five minutes after the onset of anaphase. The initial transcription covers a broad swath of the embryo, from around 25% to 75% egg-length, with the highest activity in two domains roughly centered in the anterior and posterior halves of the embryo respectively. The greatest fluorescence signal during the first 25 minutes of nc14, when stripes are not yet fully formed, is in the most anterior region of eve transcription, in an area in which stripe 1 will eventually form.

Although the full seven stripe pattern is not fully formed until around 25 minutes, the three anterior-most stripes are already apparent as locally high areas of fluorescence intensity as early as 10 minutes. By 20 minutes stripes 1, 2 and 3 have clearly separated from background, stripes 4 and 6 appear to split off from a large posterior domain, and stripe 7 forms de novo in the posterior. Stripe 5 appears as a distinct stripe last, emerging in an area of low transcriptional activity left behind following the splitting of stripes 4 and 6. The stripes persist for the remainder of nc14, gradually increasing in fluorescence intensity until they reach a peak at around 35 minutes into nc14.

The positions of stripes 1-3 along the anterior-posterior (AP) axis are largely stable after they form, while stripes 4-6 show small anterior shifts. Stripe 7 makes a more dramatic movement towards the anterior, moving approximately 8% of egg-length, or around 40 μm from its initial location. The quantitative characterization of this stripe movement, the decoupling between stripes and nuclei, and the quantification of transcriptional bursting dynamics in each nucleus necessitated the development of a method, described below, to dynamically define the position of stripes throughout each movie.

3.3.2 Modeling and inference of promoter state

As expected, the fluorescence traces from individual nuclei show clear hallmarks of transcriptional bursting, with apparent stochastic fluctuations between states with low and high fluorescence output (Figure 3.3A). Following previous work in the field [111, 50, 327, 186,

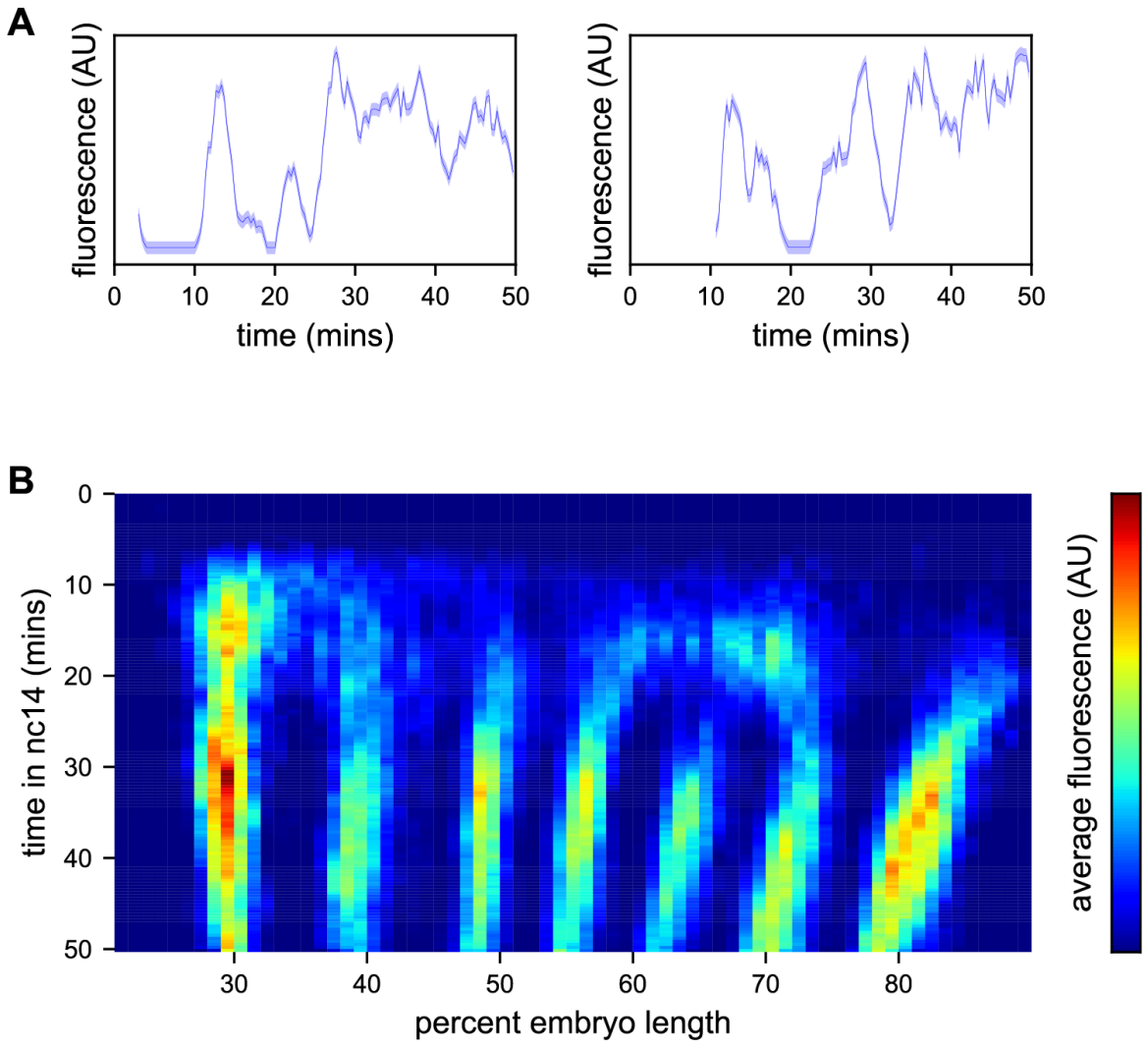

Figure 3.3: **Spatiotemporal dynamics of *even-skipped* expression. (A)** Fluorescence traces from two representative nuclei (particle ID = 1.0163 and 11.0448). **(B)** Average fluorescence over space and time showing stripe formation, modulation and movement. The time resolution along the y-axis is 20s. The positions of nuclei along the x-axis were registered across movies based on the inferred position of stripe centers, and placed into bins of one percent embryo length, with the average fluorescence of all nuclei in each bin plotted. (A, shading corresponds to the error estimated based on the background fluorescence fluctuations as described in [104])

223, 187, 317, 180, 98, 68, 128, 330, 27, 230, 184] , we model bursting as a simple Markovian process in which a promoter switches stochastically between an OFF and an ON state with rates k_{on} and k_{off}. When the promoter is in the ON state, we assume it loads polymerases continuously with a constant rate r (Figure 3.1A).

In our implementation of the MS2 system, once a polymerase molecule transcribes the stem loops located at the 5' end of the gene, the MCP-GFP molecules bound to the stem loops produce a constant fluorescent signal at the locus that persists until this polymerase completes its traversal of the gene. Building off of the method presented in [165], we estimated this polymerase transit time as the displacement that gives the minimum value in the autocorrelation of the single frame differences in the fluorescent signal (see Methods). The rationale for this approach was that every increase in signal due to polymerase loading at time t should be accompanied by a corresponding decrease in signal at time t+telong due to the completion of a transcriptional elongation cycle with a delay equal to the elongation time [56, 68]. We arrived at an estimate of 140s (Figure 3.4A), consistent with a direct measurement of the rate of polymerase elongation of 2,700 bp/min from [99] and the length of the construct (6,563bp).

We model the bursting process at each promoter in discrete time steps of $\Delta t = 20$s, set by the time resolution of our imaging. Under our model, in each time window a promoter is either OFF and not loading polymerases, or ON and loading polymerases at a fixed rate. A promoter that is in the ON state loads $\Delta t \times r$ polymerases, producing a single pulse of fluorescence proportional to $\Delta t \times r$ (with the proportionality factor determined by the fluorescence of GFP and the fraction of MS2 loops bound by MCP-GFP). This pulse lasts at the locus for 140s, at which point all polymerase molecules loaded during the original time window have terminated transcribing (Figure 3.4B). Since we do not calibrate the fluorescence signal to the number of polymerase molecules for this construct, in practice we fold the proportionality factor into r altering its units from polymerases loaded per unit time to fluorescence signal produced per unit time. Since many transcriptional bursts last for longer than 20s, the fluorescence output of a single burst is a sum of the pulses generated during each time window.

In the embryos we imaged here, the MS2 BAC is heterozygous, contributed only by the father, while the mother contributes the MCP-GFP. However DNA replication occurs within an average of 10 mins for loci in nc14 [207, 243, 273], meaning that there are actually two sister chromatids with the MS2 containing transgene in every nucleus. Because of sister chromatid cohesion, we can not, in general, discriminate both copies [100, 312, 187]. As it is still unclear how the sister chromatids influence each other's transcription [207, 165, 330], we model the locus as having two independent promoters that operate independently but governed by the same bursting parameters. Thus the system can be in one of three states: OFF, one promoter ON, and two promoters ON (Figure 3.4B).

If we know the state of the promoter over time, we can reconstruct its expected fluorescence output by summing 140s pulses beginning at each point where the promoter is ON and having height r if one promoter is ON or height 2 x r if two promoters are ON (Figure 3.4C). Traces modeled from hypothetical promoter state sequences (Figure 3.4C) have the features of the observed fluorescence signal: linear increases in intensity (corresponding to periods

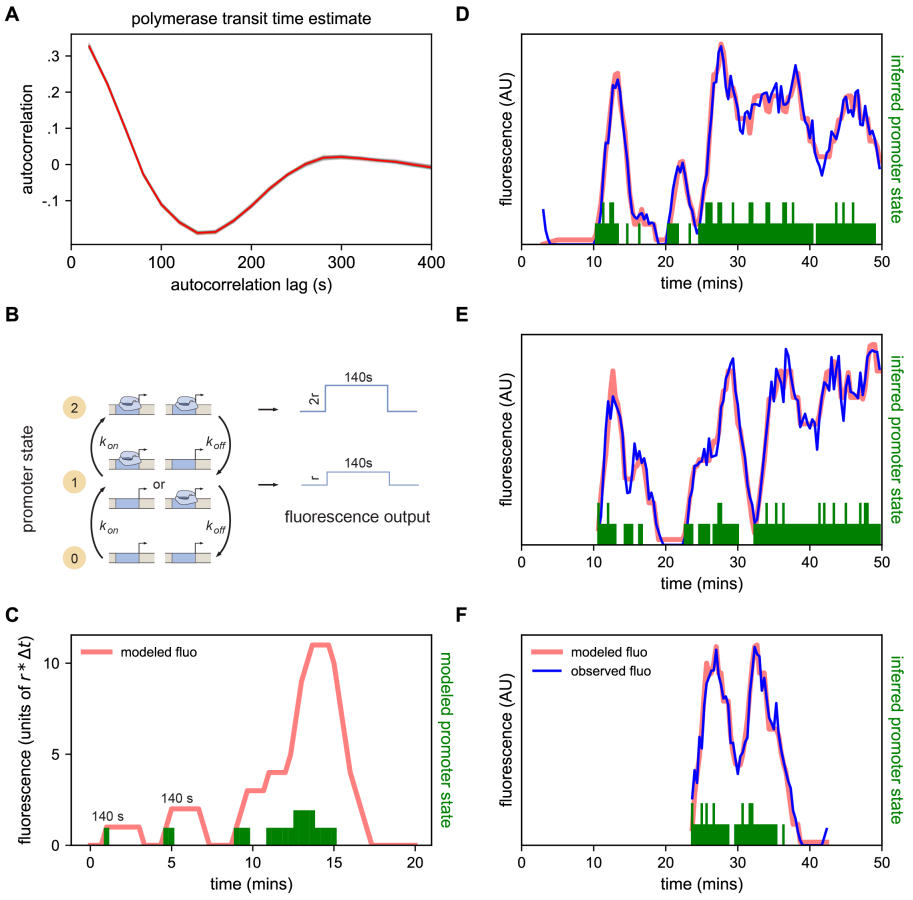

Figure 3.4: **Modeling bursting in individual nuclei. (A)** A key parameter in relating fluorescence output to the bursting state of a promoter is the time it takes for a polymerase to transit the gene, which we determined as approximately 140s by examining the autocorrelation (red line) of the change in fluorescence. Gray lines show 100 bootstraps over randomly selected sets of 80% of nuclei; note they almost perfectly overlap the red line. **(B)** Three state model accounting for post-replication presence of sister chromatids. When either promoter is ON for a short time period Δt, loads polymerases at a constant rate contributing a pulse of polymerase that persists for 140s. **(C)** Simplified example of the expected observed fluorescence (red line) produced from a hypothetical promoter state sequence. The fluorescence is the sum of the fluorescence pulses produced when one or both promoters are ON (given by the height of the green bars). **(D-F)** Representative fluorescence traces from individual nuclei (blue lines), inferred bursting pattern (green bars) and fluorescence imputed by cpHMM (red line) for particles 1.0163 (D), 11.0448 (E) and 5.0231 (F).

when the promoter is ON); plateaus (corresponding to periods when transcriptional initiation is matched with previously initiated polymerases completing their transit of the gene); and linear signal decays (corresponding to periods when the promoter is OFF but previously initiated polymerases are still transiting the gene) [27, 104].

However, when given a fluorescence trace, it is not trivial to infer the promoter state sequence that generated it, owing to the time convolution between promoter state and fluorescence output. To solve this problem, we developed a compound state hidden Markov model (cpHMM, described in [165]) that estimates global parameters k_{on}, k_{off}, and r for a set of traces, and allows us to identify the maximum-likelihood promoter state sequence under these parameters for every trace via the Viterbi algorithm.

The cpHMM thus accomplishes two aims central to treating these data in a more rigorous and biologically meaningful manner. First, it allows us to describe the bursting behavior of any set of nuclei in quantitative terms. Across all seven stripes, the model infers approximate k_{on} k_{off} values of 0.60 events per minute and an r of 67 AU per minute. And second, by providing a means to fit a sequence of ON and OFF states to the data from each nucleus, the cpHMM allows us to shift the focus in the analysis of individual traces from fluorescence, which only indirectly reflects the temporal behavior of a promoter, to the instantaneous promoter state (Figure 3.4D-F).

3.3.3 Dynamic determination of stripe positions

Before analyzing the data further we had to solve two practical problems. To compare the kinetic behavior of individual stripes, we had to determine which nuclei were in each stripe at every time point, a process complicated by the movement of stripes relative to both the embryo and nuclei. Further, to analyze the data in toto, we also had to register the 11 movies relative to each other and to the embryo.

To address these challenges, we used a Gaussian mixture model to cluster bursting nuclei in each movie in a series of overlapping six-minute time windows based on their x and y positions in the image (Figure 3.5A). This clustering reliably separates nuclei into individual stripes. We next determined the orientation of each stripe to the AP and imaging axes by fitting a line to coordinates of all nuclei assigned to that stripe in each movie (Figure 3.5B). We fit a line with this slope to bursting nuclei from each time window (Figures 5C and 5D), and use these fits to generate a linear model of the position of each stripe in each image over time, which we use to reorient the stripe so that it is perpendicular to the image x-axis (Figure 3.5E).

We next use the known coordinates of the anterior and posterior poles of the embryo to convert the image x-axis to AP position, and register the examples of each stripe from different movies by setting the AP position of the center of each stripe at 35 min in nc14 to the mean AP position of all examples of that stripe at 35 min, adjusting the position of the stripe at other time points by the same correction (Figure 3.5F). As the stripes are not all present until after 25 min in nc14, we assign and register nuclei before that point based on the stripe position at 25 min. The stripe assignment is invariant over bootstrapping of

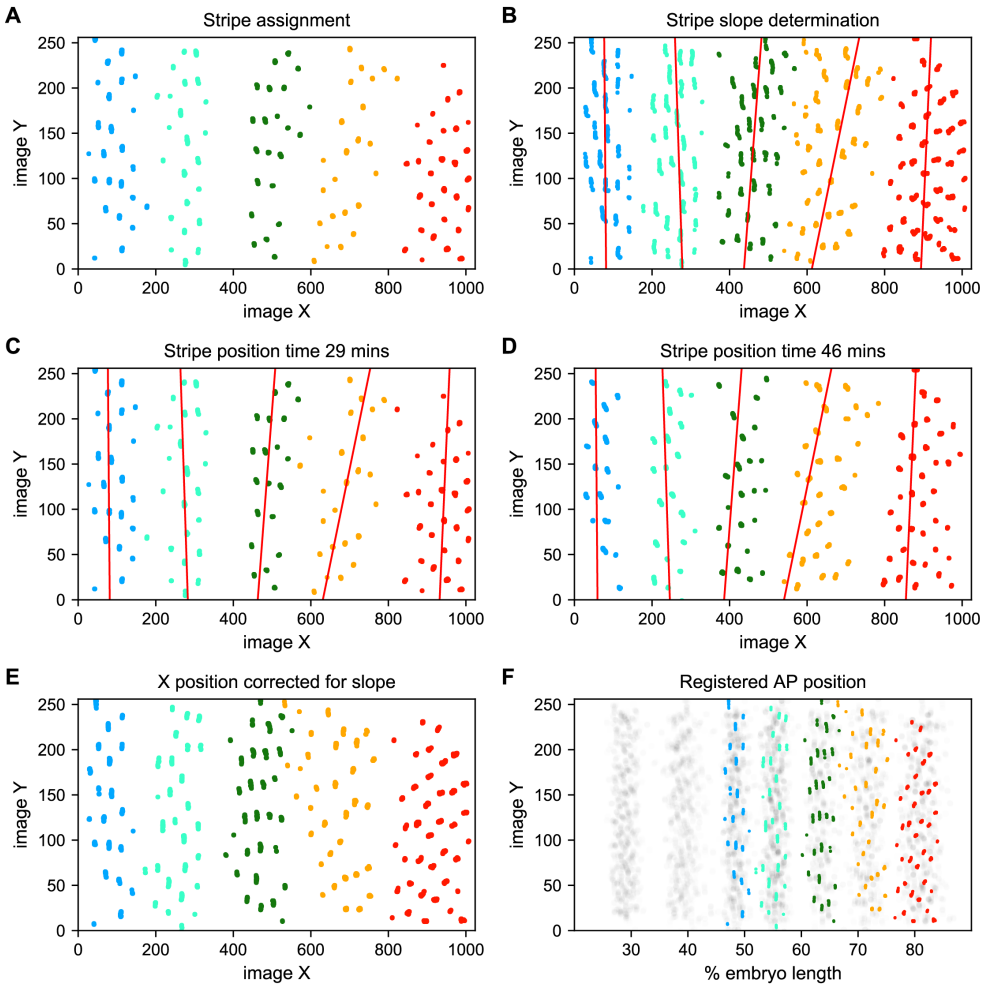

Figure 3.5: **Stripe assignment and alignment.** **(A)** We preliminarily assign bursting nuclei to stripes by applying a Gaussian mixture model to each movie independently in overlapping six-minute time windows with the number of Gaussians equal to the number of stripes captured in that movie. **(B-D)** We next determine the orientation of each stripe to the imaging axes by fitting a line to coordinates of all nuclei from t ¿ 25 min assigned to that stripe in each movie and time window. **(E)** We next determine the orientation of each stripe to the imaging axes by fitting a line to coordinates of all nuclei from t ¿ 25 min assigned to that stripe in each movie and time window. **(F)** The known coordinates of the anterior and posterior poles of the embryo are used to convert the image x-axis to AP position and register the stripes from different movies to each other, as shown here for nuclei from Movie 2 colored by stripe and nuclei corresponding to all other movies drawn in grey.

movies, and the standard deviation of the AP displacement over bootstrapping of movies is 0.0016.

Collectively these data represent an easy to visualize and interpret kinetic fingerprint of stripe formation: a record of every transcriptional burst that occurred during the formation of eve stripes in these embryos (Figure 3.6A).

3.3.4 Bursting dynamics of individual nuclei

We used the output of the cpHMM and registration process to examine the locations of transcriptional bursts along the AP axis and over time (Figure 3.6). The most striking feature is the almost complete lack of observable transcriptional bursts in the regions between stripes from 25 minutes into nc14, with the exception of the 5-6 interstripe which is discussed below (note that this is not an artifact of the movie alignment and orientation process, as this effect is seen clearly in individual movies). We took advantage of the fact that we were tracking bursts in individual nuclei in order to analyze the relationship between this absence of bursting in interstripe regions and the single-nucleus bursting behavior within stripes.

Stripes are defined by sharp spatial boundaries, with the transition between the low bursting (quiescent) state and the frequently bursting (active) state occurring from one column of nuclei to the next (Figure 3.6), consistent with the classical descriptions of eve stripe patterns [279, 93, 96, 278, 309, 52]. They also have sharp temporal boundaries: all of the interstripe regions, save that between stripes 6 and 7, form in regions where there was appreciable bursting early in nc14 that disappears at around 25 minutes into the nuclear cycle (Figure 3.6).

To better understand how the low-bursting state in interstripes is established, we looked at the bursting history of the nuclei in these regions (Figure 3.7). The first feature we noticed was that most of the nuclei that ultimately form the interstripe were never detected to burst at any point in nc14 (Figure 3.7A,B). With the exception of the 5-6 interstripe, these never-ON nuclei effectively form the boundaries between stripes, as essentially every nucleus within each stripe bursts at some point during nc14 (Figure 3.7A,B).

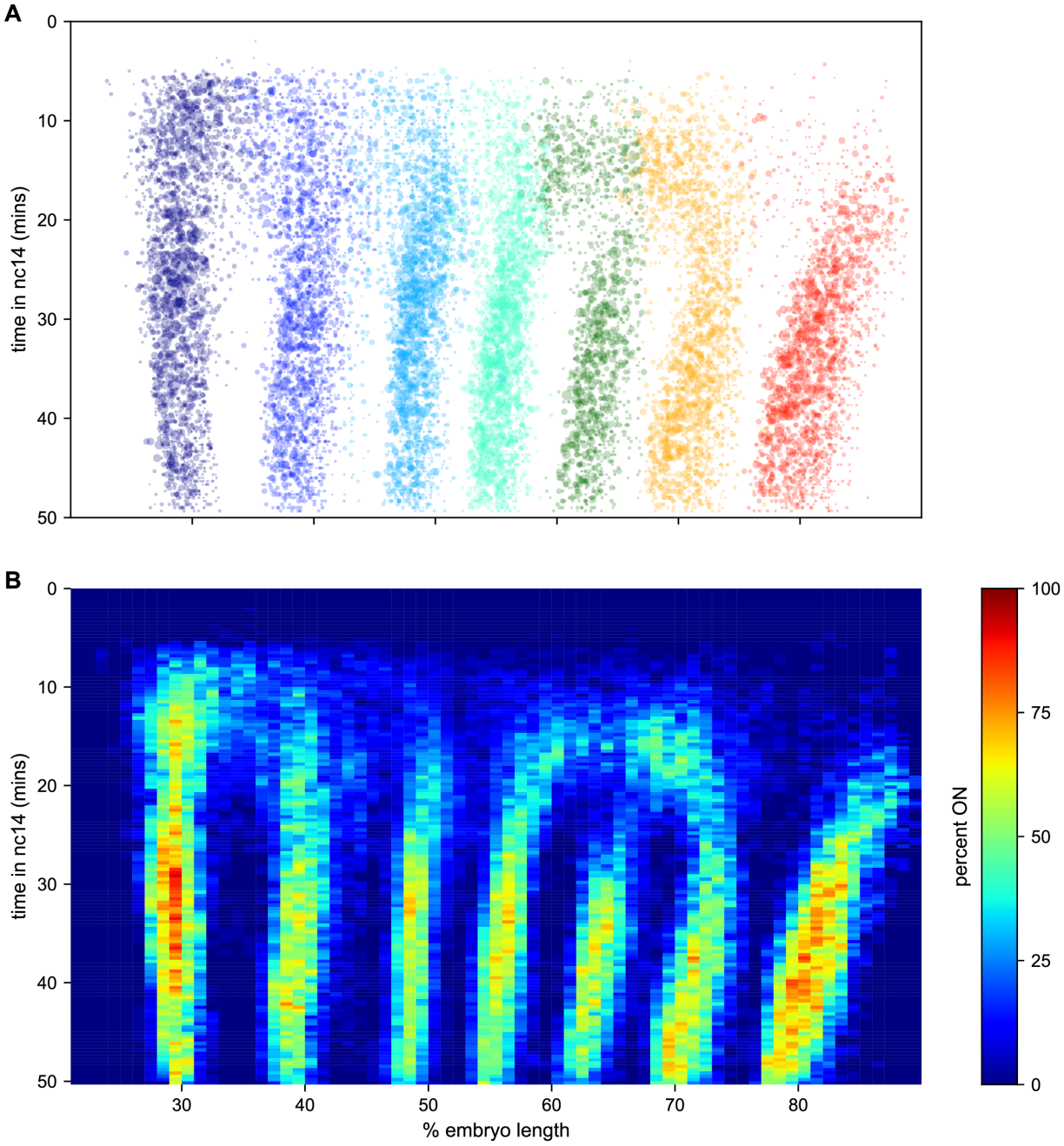

Figure 3.6: **The kinetic fingerprint of even-skipped stripe formation. (A)** Inferred location of every transcriptional burst in all 11 movies as a function of time and where along the anterior-posterior axis (plotted as fraction of embryo length) each burst occurred. The size of the dot represents the duration of the burst. Collectively the data create a kinetic fingerprint of eve stripe formation. **(B)** Instantaneous fraction of nuclei in the transcriptionally active ON state as a function of time and position along the embryo.

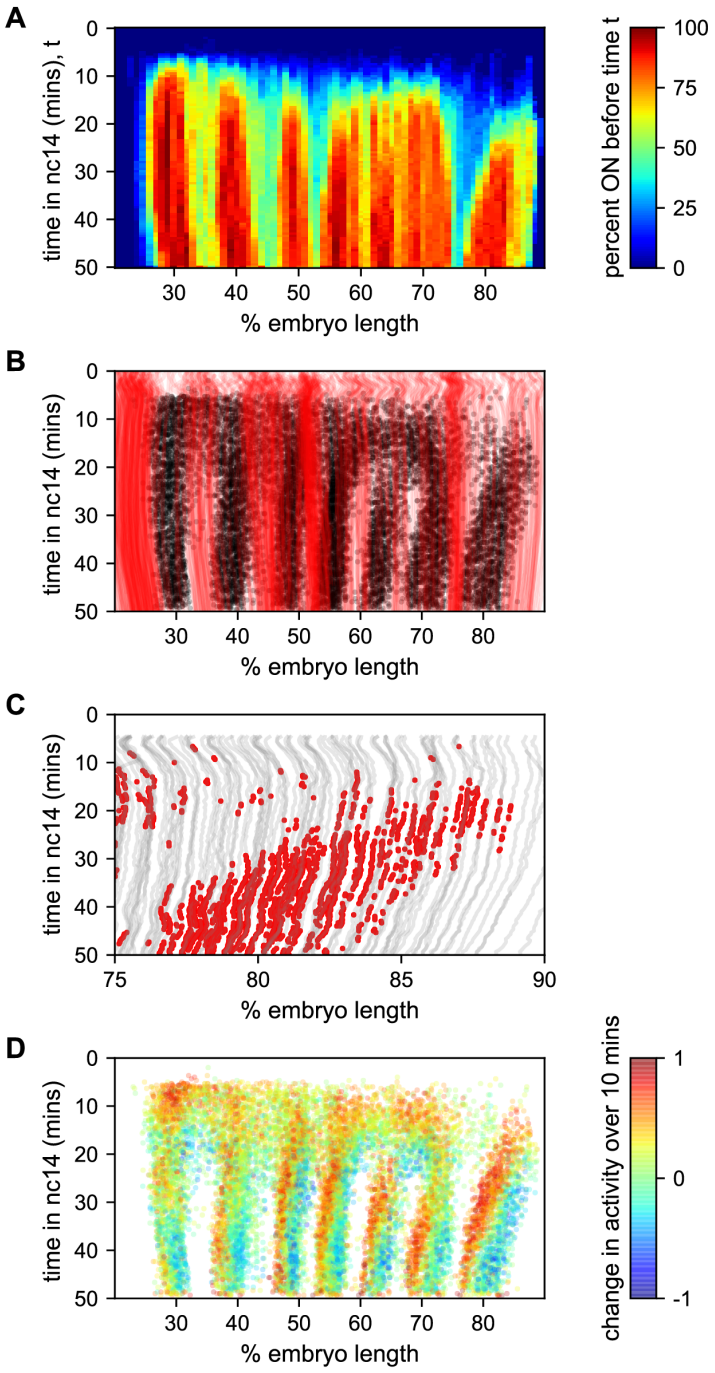
A
time in nc14 (mins), t
% embryo length
percent ON before time t
100
75
50
25
0
B
time in nc14 (mins)
% embryo length
C
time in nc14 (mins)
% embryo length
D
time in nc14 (mins)
% embryo length
change in activity over 10 mins
1
0
-1

Figure 3.7: **Stripe formation and movement. (A)** Fraction of nuclei bursting before time t as a function of position along the embryo. **(B)** Locations of new bursts (black dots) in space and time along with spatiotemporal traces of nuclei that are in the OFF state throughout nc14 (red lines). **(C)** Traces of nuclei positions over time (gray lines) from stripe 7 region of movie EVED6 with timepoints where new bursts initiated colored red to illustrate stripe movement relative to nuclei. **(D)** Difference in transcriptional activity as defined as the difference between the fraction of the time each nucleus is in the ON state in the subsequent 10 min minus the fraction of time the nucleus is in the ON state in the preceding 10 min. Positive values represent a nucleus turning on or increasing activity, while blue values indicate a nucleus turning off or decreasing activity.

The contrast in bursting history between stripes and interstripes is less pronounced in the posterior, where there are fewer such never-ON nuclei in the interstripe region (Figure 3.7B, notice the lower density of red single-nuclei tracks corresponding to never-ON nuclei). In order to reveal the source of this reduced number of never-ON nuclei in posterior interstripes, we analyzed their bursting history. Figure 3.7C shows the AP positions of the nuclei in one movie covering stripe 7 as a function of time, with the period in which they are part of the stripe highlighted. Although the stripe is clearly present throughout this period, no nuclei remain a part of the stripe for the entirety of this 25 minute period. As time progresses, nuclei at the posterior edge of stripe 7 shift from an active state, in which the promoter stochastically alternates between the ON and OFF transcriptional states, to a quiescent state in which we observe no subsequent bursting. In contrast, nuclei just off the anterior edge of the stripe switch from a quiescent to an active state at roughly the same rate. This leads to a net overall anterior movement of the stripe, akin to treadmilling, at a velocity of approximately one percent of embryo length every three minutes.

Consistent with (Lim et al. 2018), the other stripes exhibit smaller and varied anterior shifts (Figure 3.7-S1), but in every case the shift is associated with a similar coupled gain of active nuclei along the anterior edge and loss along the posterior edge. This effect is most clearly seen in Figure 3.7D, which shows, for each time point where a nucleus initiates a new burst, the difference in activity (defined as the difference between the fraction of the time the nucleus is in the ON state in the subsequent 10 min minus the fraction of the time the nucleus is in the ON state in the preceding 10 min). For all seven stripes there is a clear spatial pattern, with nuclei along the anterior edge of the stripe entering a bursting state and blue nuclei along the posterior edge becoming quiescent, indicating a movement of stripes relative to nuclei. Hence, stripe movement is associated with the dynamic switch of nuclei between active and quiescent states and not just with the movement of nuclei themselves.

3.3.5 All seven eve stripes are created by the same regulation of bursting kinetics

We next turned to the questions of how the spatial pattern of nuclear transcriptional activity described above is produced by regulating bursting kinetics, and whether this regulation differs among the seven eve stripes. In principle, any pattern of transcriptional activity could

be achieved by modulating the duration, separation and/or amplitude of bursts across space and time. For example, a stripe could be created by varying burst separation along the anterior-posterior axis, with nuclei in the stripe center having lower burst separation, and those outside the stripe having long periods without bursts, or no burst at all. Alternatively, the same stripe could be created with uniform burst separation across nuclei, but elevated burst duration or amplitude within the stripe, or by modulating multiple parameters simultaneously.

Ideally, we would like to have a measure of the bursting parameters governing the behavior of every nucleus. However, individual MS2 traces have too few time points to allow for accurate cpHMM inference of burst parameters at the single trace level. We therefore used the cpHMM to infer k_{on}, k_{off}, and r for groups of nuclei binned on their mean fluorescence output and stripe. The logic of the fluorescence binning was that, given that $\langle fluorescence \rangle \propto r\frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}$ [165], nuclei that have similar k_{on}, k_{off}, and r will have similar fluorescence outputs. Our inference shows that k_{on} is very strongly regulated as a function of average fluorescence in a consistent manner across stripes (Figure 3.8A). In contrast, only a weak drop in k_{off} is observed (Figure 3.8B). Finally, r also featured a strong upregulation as a function of average fluorescence across stripes (Figure 3.8C).

As shown in Figure 3.8D, each stripe contains nuclei with a relatively wide range of average fluorescence values. In order to reveal the bursting parameters across the AP axis for each stripe, we averaged the single-cell bursting parameters determined in each stripe (Figure 3.8A-C) weighted by the relative number of nuclei in each fluorescence bin present at each position along the AP axis (Figure 3.8D). We find that the variation in bursting parameters observed as a function of average fluorescence largely echoes the modulation of fluorescence in space (Figure 3.8E). Specifically, while there is a subtle downregulation of koff within stripes, kon and r are significantly upregulated in the center of each stripe.

Thus, not only do the five eve enhancers employ a common regulatory strategy for modulating the fluorescence output of nuclei to create a stripe, decreasing burst separation and increasing burst amplitude with a constant burst duration, the precise quantitative relationship among these bursting parameters is maintained across a wide range of molecular inputs and fluorescence outputs.

3.4 Discussion

The most remarkable aspect of eve regulation is that what appears to be a regular, repeating pattern of nearly identical stripes is created by the largely independent activity of five separate enhancers responding to different combinations of activators and repressors [**Fujioka1995** , 96, 279, 6, 278]. We have now shown that the connection between the stripe enhancers is more than just that they produce the same kind of pattern: they realize these patterns through the same control of transcriptional bursting.

Although, in principle, complex patterns of transcription could be generated by the independent regulation of k_{on}, k_{off} or r, many of the key features of eve stripe regulation we observe here involve the modulation of k_{on} and r in concert. The most straightforward

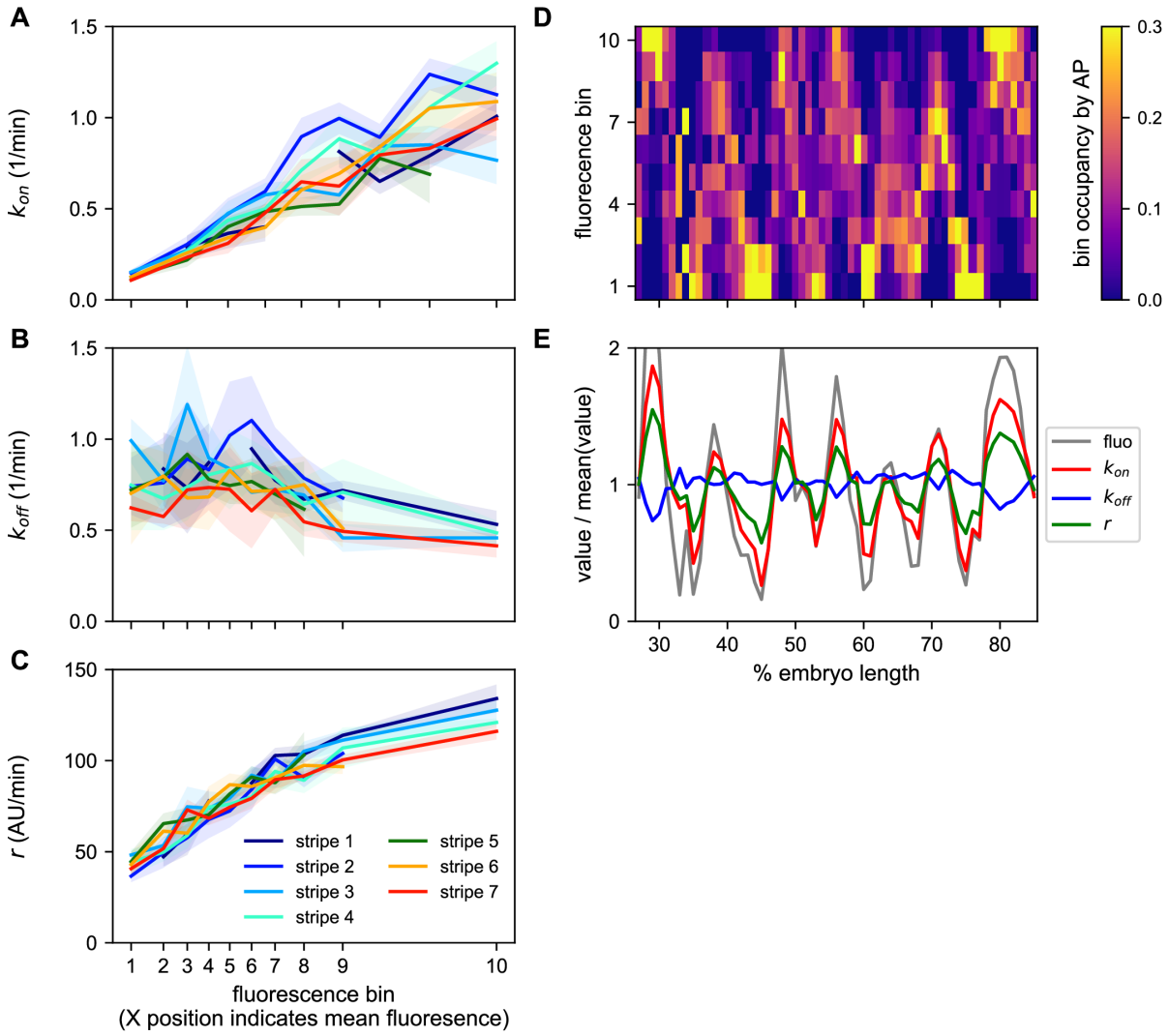

Figure 3.8: **A common bursting control mechanism across all *even-skipped* stripes. (A-C)** cpHMM inference was carried over nuclei binned according to their average fluorescence value indicating that while (A) k_{on} and (C) r are subject to the same regulation along all stripes, (B) k_{off} remains unchanged. Error bars are calculated by taking the standard deviation across cpHMM inference results for multiple bootstrapped samples of experimental data. **(D)** Distribution of average nuclear fluorescence values along the AP axis. **(E)** Mean nuclear fluorescence values for each AP position together with the corresponding averaged and weighted bursting parameters.

explanation for this shared mode of bursting control is that there is a single molecular pathway via which eve transcriptional bursting is regulated, with enhancers essentially having access to only a single tunable parameter. Whether this parameter is determined by the gene through, for example, the promoter sequence, or whether this single molecular pathway reflects some broad common property of gene regulation, such as constraints on the general transcriptional machinery, remains an open question. The limited data on bursting control available for other genes in the fly [85, 98, 330] suggests that control mechanisms are not ubiquitously the same and that they might be unique to different classes of genes.

An alternative explanation for the observed commonalities in the control of bursting is that there is a functional reason to use this strategy. Namely, that this is not the result of a common molecular mechanism, but rather of common selective pressures acting on the five enhancers independently. The particular bursting control strategy uncovered here might, for example, be more robust to fluctuations in transcription factor concentrations or temperature, or provide more precise spatiotemporal gene expression control [271, 114]. New experiments and theoretical work will be necessary in order to uncover the specific molecular pathways by which bursting is controlled and to understand the functional consequences of different bursting strategies that create the same mRNA levels.

In addition to this modulation of bursting, the fraction of nuclei that engage in transcription at any point in the nuclear cycle is higher in stripe centers than in interstripes. This regulation of the fraction of active nuclei, also seen in other genes [104], seems to reside outside of the bursting framework. Such regulation, as well as the spatial modulation of the window of time over which bursting ensues, suggests the presence of multiple and overlapping modes of regulation that go beyond the control of bursting parameters and that can be as relevant for pattern formation [165].

3.4.1 Stripe movement is driven primarily by expression flow

Just as gene expression patterns are dynamic in time [27], they are dynamic in space, resulting in the movement of expression domains throughout the embryo during development [142, 157]. The anterior movement of eve stripes during nc14 has been previously described [157, 184] , and proposed to arise from a combination of nuclear movement (nuclear flow) and movement in the pattern of regulators (expression flow), especially repressors, which are known to shift anteriorly during nc14 as well [142, 143]. While Keränen et al. [157] concluded that the relative contributions of these two forces were roughly equal, our data suggest that, especially in the posterior, expression flow dominates the anterior shift of the eve stripes.

A typical nucleus in stripe 7 moves around one percent of embryo length in the final 25 min of nc14. The stripe, however, moves around five percent of embryo length during that time (see Figure 3.7C). Because we are tracking both the position and activity of individual nuclei, we can visualize expression flow in action. We see nuclei transition from low activity in the anterior interstripe to high activity in the stripe, from high activity in the stripe to low activity off the posterior flank of eve expression, and in some cases both.

This effect is most pronounced for the posterior stripes, but is observed for the more anterior stripes as well, although the magnitude of the shift decreases for more anterior stripes (Figure 3.7-S1). The difference in the amount of the effect we and Keränen et al. attribute to expression flow is likely an effect of differences in the data used. Because we are looking at instantaneous transcription rates while they looked at accumulated mRNA, there is a considerable temporal lag and integration of the transcriptional activity over the life time of eve mRNA in their data, which has the effect of underestimating the extent to which the stripes actually move.

We also note that the extent to which nuclear flow by itself would be expected to shift output patterns measured as a function of position in the embryo is unclear, as it would depend on the extent to which the repositioning of regulators drives movement of nuclei (which it is believed to do [20], and the corresponding effect that nuclear movement has on the positioning of regulators, which remains largely unknown. One open question relates to the temporal relationship between changes in the position of the repressor array that drives stripe position and the transcriptional output of the stripes. For example, the anterior shift of the stripes of eve as well as fushi tarazu has been proposed to originate, in part, from cross-repression between these two genes [184]. Recent advances in the simultaneous monitoring of protein concentration and transcriptional output in living embryos should help answer this question in the near future [24, 165].

3.4.2 Characterizing dynamics patterns demands dynamics measurements

That gene expression is a fundamentally dynamic process is not new information. However, the tools we have had at our disposal to study gene expression so far have tended to emphasize its static features, down to the language we use to describe the transcriptional output of a gene. In textbooks and the scientific literature, eve has a gene expression pattern consisting of seven stripes. But, as some earlier work emphasized [145], and we have directly visualized here, the transcriptional output of eve, rather than a single "pattern" is a rapidly changing as a function of time and space: it is dynamic at many time scales and across space and nuclear positions. Indeed, at no point does eve approach anything even remotely like a steady state.

We are at the dawn of a new period in the study of transcription, as new experimental techniques and advanced microscopy allow us to monitor transcriptional regulators, observe their behavior at the single-molecule level, and track the transcriptional output of a gene in living, developing animals. We have only barely begun to understand this new data and what it can tell us about biology. While the focus in this paper was on a single gene in a single species, we hope that this and our accompanying work [165] will have a broader impact by beginning to establish rigorous frameworks for quantifying, characterizing and visualizing the dynamics of transcription at the single-cell level during development that will be required in the era of live imaging of transcription in development.

3.5 Methods

3.5.1 Generation of MS2 tagged eve BAC

We used bacterial recombineering [308] to modify a bacterial artificial chromosome (BAC) [299] containing the D. melanogaster eve gene and all of its enhancers and regulatory elements (BAC CH322-103K22) [300]. We replaced the coding region with an array of 24 MS2 stem loops fused to the D. melanogaster yellow gene (Figure 3.1B; [27] as described below. We inserted our eve::MS2::yellow BAC-based construct in the D. melanogaster genome at chromosome 3L through C31 integrase-mediated recombination (see Generation of fly lines), and generated a viable homozygous fly line (w-; +; eve::MS2::yellow) as detailed below.

3.5.2 Reporter design

In principle the length of the reporter should not limit our ability to estimate burst parameters. However, in practice a reporter construct that is too short will have insufficient signal. Further, one that is too long will increase the dwell time of each RNA polymerase molecule on the gene and, as a result, our cpHMM inference will require too many computational resources. Our choice of reporter construct structure strikes a balance between these two limitations and is ideally suited for inferring bursting parameters in the time range where eve resides, as well as for boosting the signal-to-noise ratio. See [165] for a more detailed discussion of reporter length-related tradeoffs.

3.5.3 Specifics of recombineering

We modified a CHORI BAC CH322-103K22 derived from [300], which contained the entire eve locus and a GFP reporter instead of the eve coding sequence (CH322-103K22-GFP). We replaced the GFP reporter with MS2::yellow (6665 bp) through a two step, scarless, galK cassette-mediated bacterial recombineering [308]. Briefly, we transformed our starting CH322-103K22-GFP BAC into E.coli recombineering strain SW102. We then electroporated the strain with a galK cassette flanked by 50bp-long DNA homology arms homologous to the MS2::yellow (6665 bp) reporter. Upon electroporation, we selected transformants on M63 minimal media plates with galactose as a single carbon source. We achieved a correct replacement of GFP sequence by galK cassette in the BAC context (CH322-103K22-galK), validated by observing the digestion patterns produced by ApaLI restriction enzyme.

We next purified the CH322-103K22-galK BAC and transformed it into fresh E. coli SW102 cells. We electroporated these cells with the purified MS2::yellow insert and used M63 minimal media plates with 2-deoxy-galactose to select against bacteria with a functional galK gene. We used colony PCR to screen for colonies with a correct MS2::yellow insertion (CH322-103K22-MS2) replacing the galK cassette. We validated this insertion by observing ApaLI, XhoI, SmaI, and EcoRI restriction digestion patterns and through PCR and Sanger sequencing of the insertion junctions. We transformed our CH322-103K22-MS2 BAC in

E.coli EPI300 cells to induce high copy numbers and purified it with a Qiagen plasmid Midiprep kit.

3.5.4 Generation of fly lines

We sent a sample of our purified CH322-103K22-MS2 BAC to Rainbow Transgenic Flies, Inc. for injection in D. melanogaster embryos bearing a C31 AttP insertion site in chromosome 3L (Bloomington stock 24871; landing site VK00033; cytological location 65B2). We received the flies that resulted from that injection and used a balancer fly line (w- ; + ; +/TM3sb) to obtain a viable MS2 homozygous line (w- ; + ; MS2::yellow). We used line (yw; His::RFP; MCP::GFP) as the maternal source of Histone-RFP and MCP-GFP [104].

3.5.5 Embryo Collection and Mounting

Embryo collection and mounting was done as specified in [103]. In short, we set fly crosses between 30 males (w-; +; eve::MS2::yellow) and 80 females (yw; His::RFP; MCP::GFP) in a plastic cage capped with a grape juice agar plate. We collected embryos from cages two to ten days old by adding a fresh plate for 30 minutes and aging for 60 minutes to target embryos 90 min or younger.

Embryos were mounted on a gas-permeable Lumox Film (Sarstedt - Catalog 94.6077.317) embedded on a microscope slide hollowed on the center. Then, we coated the hydrophobic side of the Lumox film with heptane glue and let it dry. The film allows oxygenation of embryos during the 2-3h long imaging sessions while heptane immobilizes them.

We soaked an agar plate with Halocarbon 27 oil, picked embryos with forceps, and laid them down on a 3 x 3 cm piece of paper tissue. We dechorionated embryos by adding 2 drops of bleach diluted in water (5.25%) on the paper tissue and incubating for 1.5 minute. We removed bleach with a clean tissue and rinsed with 4 drops of distilled water. We then placed the tissue paper with dechorionated embryos in water, and picked buoyant embryos with a brush.

We lined 30 apparently healthy embryos on the Lumox film slide and added 2-3 drops of Halocarbon 27 oil to avoid desiccation, and covered the embryos with a cover slip (Corning® Cover Glass, No.1, 18 x 18mm) for live imaging.

3.5.6 Imaging and Optimization of Data Collection

Movies of embryonic development were recorded on a Zeiss-800 confocal laser scanning microscope in two channels, (EGFP: 488 nm; TagRFP: 561 nm). We imaged embryos on a wide field of view, along their anterior-posterior axis, of 1024 x 256 pixels (202.8µm x 50.7µm), encompassing 3-5 stripes per movie. We tuned laser power, scanning parameters, master gain, pinhole size and laser power to optimize signal to noise ratio without significant photobleaching and phototoxicity.

For imaging, the following microscope settings were used: 63x oil-objective, scan mode 'frame', pixel size of 0.2µm, 16 bits per pixel, bidirectional scanning at a speed of 7, line step

of 1, laser scanner dwelling per pixel of 1.03µs, laser scanner averaging of 2, averaging method Mean, averaging mode Line, 488 nm laser power of 30µW (EGFP), 561 nm laser power of 7.5µW (TagRFP) (both powers were measured with a 10x air-objective), Master Gain in EGFP detector of 550V, Master Gain in TagRFP detector of 650V, Digital Offset in both detectors of 0, Digital Gain in both detectors of 1.0, and a pinhole size of 1 airy unit under the imaging conditions mentioned above (44µm, 0.7µm/section), laser filters EGFP:SP545 and TagRFP:LBF640. This resulted in an imaging time of 633 ms per frame and a full Z-stack of 21 frames in intervals of 0.5µm every 16.8s. Following [27, 23, 24, 165] , the imaging conditions were determined not to affect normal development as reported by the timing of the nuclear cycles in early development. We stopped imaging after 50 min into nuclear cycle 14, and took mid-sagittal and surface pictures of the whole embryo for localization of the recorded field of view along the embryo's AP axis.

3.5.7 Image processing

We used a Matlab computational pipeline based on [104, 165] to segment and extract numeric data from our raw movies. Briefly, this software segments and processes the images from the two channels (channel 1: MCP::GFP, channel 2: Histone::RFP) on which we collected our data. For segmentation of channel 1, we used Fiji-Weka Segmentation 3D software; this machine-learning-based method relies on the manual segmentation of a variety of MCP::GFP labeled transcriptional foci in a given 21 frame Z-stack from a single dataset (EVE_D11) to produce a model for the segmentation of all datasets recorded under the same imaging conditions. Next, we segmented and tracked the Histone::RFP labeled nuclei on channel 2. Subsequently, we assigned MCP::GFP labeled transcriptional foci to their corresponding Histone::RFP labeled nuclei. Since we collected whole embryo pictures of each of our datasets, we were able to match and locate the recorded fields of view to their right position along the body of their corresponding embryos. Finally, we extracted position and fluorescence values over time of all transcriptional foci to generate data structures ready to use in further analyses.

3.5.8 Estimation of polymerase transit time

To estimate the transit time of the polymerase along the construct (which is used to determine the persistence of the fluorescence signal from a single transcript at the locus) we first calculated, for each nucleus, the difference in fluorescence signal between adjacent timepoints Dn,t = Fn,t+1 - Fn,t where Fn,t is the fluorescence signal for nucleus n at time point t and then calculated the Pearson correlation coefficient of the vectors [..., Dn,t , Dn,t+1, Dn,t+2, ...] and [..., Dn,t+d , Dn,t+d+1, Dn,t+d+2, ...] over values of d from 1 to 20 representing time displacements of 20 to 400 seconds. The minimum correlation occurred at 140 seconds.

3.5.9 Compound-state Hidden Markov Model

For this work we employed a statistical method that utilizes a compound-state Hidden Markov Model to infer bursting parameters from experimental fluorescence traces. The theory and implementation of this method are described in detail in [165]. Briefly, parameters were inferred using a standard version of the Expectation Maximization Algorithm implemented using custom-written scripts in Matlab. Our inference is carried over the full duration of activity of each active nucleus during nuclear cycle 14. Bootstrap sampling was used to estimate the standard error in our parameter estimates. Subsets of 3,000 data points were used to generate time-averaged parameter estimates. Inference was not conducted for groups for which fewer than 1,000 time points were available.

3.5.10 Data Analysis and Figures

All data were analyzed in Python using a Jupyter notebook with custom code to process raw data and generate figures. The Jupyter notebook and all data required to run it is available in at `https://github.com/mbeisen/Berrocal_2020`.

3.5.11 Data filtering

We first filtered the raw data to remove data with observations spanning less than 2,000 seconds, as well as nuclei that were poorly tracked over time (defined as nuclei that moved across the movies at an average rate of over 4.5 pixels per minute. This left 430,073 observations from 2,959 nuclei.

3.5.12 Stripe assignment and registration

We used the Gaussian mixture model module of the Python library scikit-learn [234] to cluster all nuclei time points in each movie in each of a series of overlapping 428 second time windows beginning at 25 min in nc14, specifying the number of components equal to the number of stripes captured in the movie and using the setting $\text{covariance}_{\text{type}}$ = 'tied'. We preliminarily assigned nuclei time points to a stripe if they were consistently clustered in that stripe in the relevant time windows. We then pooled all nuclei time points assigned to the same stripe and fit a line to the median x and y positions in the bottom ($y < 128$) and top ($y > 128$) halves of the image. We considered the slope of this line to represent the orientation of the stripe to the image x axis. We then went back to each time window and fit the nuclei assigned to the stripe with a line with the previously computed slope fixed. This produced an association of time with stripe position, from which we derived a linear model that describes the position of each stripe in each movie at every time point.

We assigned all nuclei time points (not just bursting ones) to stripes by identifying the stripe whose predicted position at the relevant time was closest (along the x axis) to the nucleus being analyzed, and assigned a nucleus to the most common stripe assignment for its individual time points. We then corrected the reorientation of the stripe at each time point

to be perpendicular to the image x-axis (to enable projection along the AP axis) by setting its new image x-axis position to be the x position of the stripe in the middle of the y-axis (y = 128) plus the offset of the nucleus to the unoriented stripe along the x-axis. Finally, we used the positions of the anterior and posterior poles of the embryo to map image x coordinates to AP position. We then adjusted the AP position of each stripe in each movie such that the center of the stripe at 35 min in nc14 had the same AP position.

Author contributions

AB designed and generated the labeled eve transgene, collected all of the imaging data and ran the initial image processing. NL and MBE performed all of the higher level analyses and generated figures. HGG and MBE wrote the paper. HGG and MBE conceived of the experiments. HGG and MBE provided funding and supervised every aspect of the project.

3.5.13 Data availability

All of the raw and processed data described in this paper are available on Data Dryad at doi:10.6078/D1XX33 and computational notebooks with necessary data to regenerate analyses and figures is available at `https://github.com/mbeisen/Berrocal_2020`.

3.5.14 Acknowledgements

This work was supported by an HHMI Investigator award to MBE. HGG was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl & Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236). NL was supported by NIH Genomics and Computational Biology training grant 5T32HG000047-18. AB was supported by a Doctoral Fellowship from The University of California Institute for Mexico and the United States (UC MEXUS) and CONACyT.

Chapter 4

Probing the limits of information transmission in *even-skipped* with 2 spot experiments

Abstract

The emergence of robust developmental programs from noisy molecular processes is one of the enduring mysteries of developmental biology. Here, we use 2 spot experiments, which track the transcriptional output of two identical gene loci in each cell, to dissect observed variability in *even-skipped* (*eve*) transcription in early fruit fly development. We demonstrate that having two replicate loci from each nucleus allows us to obtain estimates for the rate of information transmission at *eve* loci, in bits, by comparing the relative magnitudes of the total (σ^2_{tot}) and intrinsic (σ^2_{int}) noise components. We use this insight to quantify the information content of different regulatory "knobs," finding for instance, that the turn-off times of *eve* loci convey 1 bit of information, on average. We also track the rate of information transmission across space and time as the *even-skipped* pattern forms, finding that information transmission peaks midway through nuclear cycle 14, and is concentrated along stripe boundaries, which achieve rates of up to approximately 1 bit per minute.

4.1 Introduction

This chapter is intended both as a coda to the results put forward in the preceding two chapters, and as a bridge to the works that follow, in which we move from focusing on the control of bursting in and of itself towards thinking about its consequences for the molecular mechanisms of gene regulation and to how molecular noise impacts the capacity for gene circuits to transmit biologically useful information. Here, I lay out a series results and vignettes arising from my analysis of (as yet) unpublished live imaging datasets taken and processed by Augusto Berrocal, my coathor in the work from Chapter 3. These datasets feature the

even-skipped BAC reporter used in Chapter 3 (see Figure 3.1D) expressed homozygously, such that there can be up to two active *eve* loci per nucleus.

We begin by tying up a few loose ends from Chapters 2 and 3, using two spot data to probe the molecular logic of two single-cell regulatory events found to play a role in the formation of the *even-skipped* pattern: the initial decision to engage in transcription (at all, ever; see Figure 2.1E and F), and the early onset of long-lived transcriptional quiescence in the regions between *eve* stripes (Figures Figure 2.1C and D). We find that the two *eve* loci within each nucleus make initial decision to enter into transcription (or not) independently; that is, there is no nucleus-specific variable that leads to coordination in transcriptional engagement between loci within the same nucleus.We also show that *eve* loci that have been transcriptionally silenced by high repressor concentrations in the regions between stripes can, likewise, re-engage in transcriptional activity if they enter regions that favor transcriptional activation, suggesting that the molecular mechanisms responsible for long-lived repression at this stage in development are reversible on rapid timescales.

Next, we use different facets transcriptional noise to probe the nature and limits of the regulatory control of *even-skipped* by spatiotemporal patterns of transcription factor proteins. Building off of previous works [46, 317, 331], we show that the level of intrinsic noise in transcription can be used to infer how transcription factors control different burst parameters to modulate the rate of transcription. Our results are consistent with findings in Chapter 3, which indicated a model in which transcription factors increase *eve* transcription by up-regulating the burst frequency (k_{on}) and the burst amplitude (r).

To close, we propose that two spot experiments can serve as more than just a microscope into the kinetic levers of gene regulation: they can be harnessed to probe the fundamental limits of transcriptional control. We demonstrate how the unique nature of these data—two gene "replicates" exposed to the same set of transcriptional inputs—can be used to marginalize over the (partly or wholly unobservable) set of transcriptional inputs ($\boldsymbol{c}$) to estimate the "regulatory potential" of the *eve* locus, defined as

$$P_{reg} = \frac{\sigma^2_{tot}}{\sigma^2_{int}} = 1 + \frac{\sigma^2_{ext}}{\sigma^2_{int}}, \tag{4.1}$$

where σ^2_{tot} is the total variability in transcriptional output, σ^2_{int} is the intrinsic variability between loci from the same nucleus, and $\sigma^2_{ext} = \sigma^2_{tot} - \sigma^2_{int}$ is the extrinsic component of variability that indicates the degree to which sister loci from the same locus are correlated (or anticorrelated) with one another. We show that Equation 4.1 can be related to the information content, in bits, of a transcriptional signal, and use this insight to dissect the information content of different regulatory strategies uncovered in previous chapters. Furthermore, we use cpHMM decoding to estimate the rate of information transmission by *eve* loci as a function of space an time. We find that *eve* loci transmit information at a maximum rate about 1.2 bits per minute, and that, allowing for mRNA accumulation, the total predicted information content of the *eve* pattern reaches a maximum of about 4.2 bits for the *even-skipped* pattern (allowing for accumulation of mRNA over time). This is strikingly similar to the 4.1 bits found by [74] to be encoded collectively by the Gap Genes that regulate *eve*.

4.2 Probing the molecular nature of transcriptional engagement and transcriptional silencing

4.2.1 *eve* loci within each nucleus decide to engage in transcription independently

In Chapter 3, we found that a significant (but variable) fraction of *eve* loci located in regions between stripes never engage in transcription across the full duration of nuclear cycle 14 (see Figure 3.7B). Thus the initial decision of whether or not to engage in transcription constitutes a significant regulatory event that is subject to spatial control, presumably by patterns of transcription factor proteins. Consistent with this past work, we also see a spatial modulation in the probability, p, that *eve* loci engage in transcription (dashed line in Figure 4.1B), with fewer loci located between stripes ever engaging while nearly all *eve* loci in stripe centers enter into a transcriptionally active state.

A simple first question to ask about transcriptional engagement is whether, as illustrated in Figure 4.1A, the decision to engage in transcription unfolds in a nucleus-wide fashion, such that both loci within the same nucleus tend to engage (or not) together, or if, instead, each locus decides independently. Following unpublished work by Albert Lin and colleagues, we observe that we can use the spatial modulation of the engagement probability, $p(x)$, to formulate testable predictions for the case when each locus engages in transcription independently. This model predicts that the fraction of nuclei with two actively transcribing loci will be given by

$$p_2(x) = (p(x))^2 . \tag{4.2}$$

Similarly, the fraction of nuclei with only one engaged locus will correspond to

$$p_1(x) = 2p(x)\,(1 - p(x))\,, \tag{4.3}$$

and the fraction of nuclei with no engaged loci will be

$$p_0(x) = (1 - p(x))^2 . \tag{4.4}$$

Figure 4.1B compares our measurements for each of the three quantities above (colored circles) with the predictions of the "independent engagement" model. We find excellent agreement between predicted and measured frequencies of zero, one, and two loci being engaged across the full expanse of the *even-skipped* pattern. This indicates that the regulatory processes that dictate whether or not a locus engages in transcription are intrinsic to each *eve* locus. We note that conclusion is consistent with the findigs of Lin et. al. in the context of transcription driven by *hunchback* P2, another widely studied regulatory element in early fly development.

4.2.2 Long-lived transcriptional silencing in inter-stripe regions is reversible

A striking feature of the *even-skipped* pattern that we an others have noted is its dynamism [13, 184]. For instance, in Chapter 3, we find that *eve* stripe seven moves towards the anterior at a rate of about one percent of embryo length every three minutes. Importantly, this movement largely occurs independent of the motion of nuclei within the embryo, which means that the same nucleus may (for instance) be part of an active stripe region for one period, and part of an inactive interstripe region for another.

This pattern motion relative to the nuclei provides a unique opportunity to probe the dynamic nature of transcriptional activation and transcriptional silencing. Specifically, in Chapter 2, we found that repressor-induced silencing of transcription on the flanks of *eve* stripe 2 constituted a major regulatory driver of stripe formation. In the context of the stripe 2 reporter used in those experiments, however, it was impossible to ascertain whether the mechanism inducing this long-lived silencing was truly irreversible, or whether it merely *appeared* irreversible because *eve* loci on the stripe flanks were only ever exposed to repressor concentrations that increased over time. However, in the context of the full *even-skipped* pattern relative to individual nuclei means that, if silencing is reversible, we should see cases where silenced *eve* loci reactivate upon being overtaken by a dynamic domain of stripe expression.

To test this hypothesis, we searched for unambiguous events in our data where individual *eve* loci initially engage in transcription, undergo transcriptional silencing, and then re-engage at some later time (Engaged $\rightarrow$ Silent $\rightarrow$ Engaged). We found that such events are quite commonplace. Out of 6,142 qualifying loci that were detected to have engaged in transcription, we see 575 that were silent for at least 10 minutes, and then re-engaged subsequently. Indeed, even if we are stricter, and demand that loci be silent for a full 20 minutes to qualify, we still find 151 that meet this standard.

Figure 4.1C shows the trajectory of these 151 loci in space and time, overlaid on top of a heat map indicating the fraction of transcriptionally active *eve* loci. Green line segments indicate active periods and blue segments indicate inactive periods. This heat map shows where each of the seven stripes are located over time. In these regions, the relative concentrations of activators and repressors favors active *eve* transcription. In most cases, we see that the reactivated loci begin on the posterior flank of one stripe, end up in an inter-stripe region (where they are silenced), and, finally, are overtaken by the anterior edge of the neighboring stripe, which leads to reactivation. It is conceivable that, in these cases, the original enhancer driving activity (e.g. the stripe 2 enhancer) remains permanently silenced while the second enhancer (e.g. stripe 3) takes over; however, reactivation trajectories in the stripe 1 and stripe 4+6 region suggest that the same enhancer can be reactivated. Thus, we interpret these data as providing strong evidence that transcriptional quiescence is transient and readily reversible.

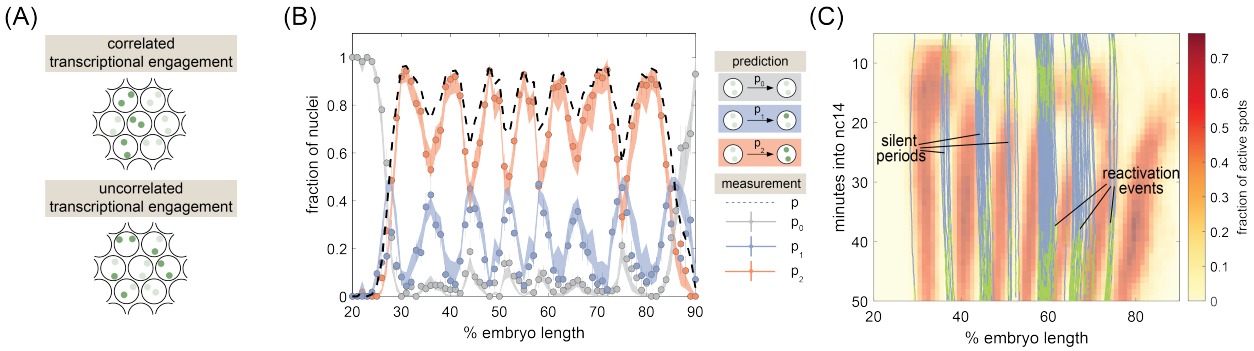

Figure 4.1: **Probing the molecular nature of transcriptional engagement and transcriptional silencing. (A)** Two distinct scenarios of transcriptional engagement where the decision to engage in transcription is nucleus-wide or independent for each locus within a given nucleus. **(B)** Plot comparing observed frequencies of nuclei with 0, 1, and 2 engaged *eve* transcriptional loci to predictions assuming that loci decide to engage independently. **(C)** Illustration of reactivation trajectories. Heatmap shading indicates fraction of active loci. Each line indicates AP position over time for an *eve* locus that was seen to undergo reactivation. Green segments indicate frames where transcription was observed. Blue segments indicate frames with no observed transcription.

4.3 Intrinsic noise trends are consistent with hypothesis that *eve* is regulated via burst amplitude and burst frequency

A central finding in Chapter 3 is that the five enhancers regulating the seven stripes of the *even skipped* pattern modulate *eve* activity in the same way: by controlling k_{on} and r. These results were obtained using cpHMM inference. At the time, we had no complementary method to independently confirm this finding; however, it has previously been established that the intrinsic noise in transcriptional output can be used to shed light on the regulation of transcriptional bursting [317, 330, 46]. To do this, we calculated the intrinsic variance in the accumulated fluorescence, defined as

$$\sigma^2_{int} f = \frac{1}{2}\langle (f_1 - f_2)^2 \rangle, \tag{4.5}$$

where f_1 and f_2 are the total accumulated fluorescence produced by spots 1 and 2 from the same nucleus. We performed this calculation for all nuclei featuring two active spots that were concurrently active for at least 25 minutes (a total of 894 unique nuclei).

As in Chapter 3, we then grouped spot pairs according to their mean fluorescence, and calculated the average intrinsic noise for each group. As illustrated in Figure 4.2A, we found that the intrinsic noise increases approximately linearly with average spot fluorescence. To determine whether this trend is consistent with our previous findings using cpHMM burst

inference, we made use of formulas from [46] that relate the mean activity and intrinsic noise for a gene with elongation time T to underlying burst parameters. Specifically, we have that

$$\langle f \rangle = \frac{rTk_{on}}{k_{on} + k_{off}} \tag{4.6}$$

for the mean and

$$\sigma^2_{int} f = \langle m \rangle \left[1 + \frac{2rk_{off}}{(k_{on} + k_{off})^2} + \frac{2rk_{off}}{(k_{on} + k_{off})^3} \left(\frac{e^{-T(k_{on}+k_{off})} - 1}{T} \right) \right] \tag{4.7}$$

for the intrinsic variance. These expression allow us predict how the intrinsic variance scales with the mean for a known set of burst parameters. Plugging in cpHMM inference results from Chapter 3 predicts the thrend shown as a solid black line in Figure 4.2A. We see that the prediction captures the monotonic increase of $\sigma^2 f$ with $\langle f \rangle$; however, the predicted magnitudes do not agree. One likely explanation for this is that equations 4.6 and 4.7 pertain to a *true* two state system, whereas those presented in Chapter 3 are for an *effective* two state model, derived from three state model (see Figure 2.4).

In fact, as shown in the Appendix for Chapter 2 (Figure A.31), fits to a true 2 state model tend to return k_{off} and r that are approximately four and two times larger than effective 2 state parameters derived from 3 state model fits. If we apply these correction factors, we find that our predictions approach the observed intrinsic noise trend (see dashed line in Figure 4.2). Overall, we conclude that the intrinsic noise vs. mean fluorescence trend is consistent with the model put forward in Chapter 3; however, more work is needed to establish whether noise-based and cpHMM-based inference results are truly in quantitative agreement with one another.

There are two routes to achieving this: first, true two state inference could be conducted on the *eve* data used in the previous chapter to permit a direct comparison to the two state noise-based model presented by [46]. Alternatively, expressions for the intrinsic noise for systems with 3+ states that are developed in Chapter 6 (and taken originally from [311]) could be adapted to incorporate the effects of the RNAP elongation time, T, on the noise signatures. We favor the second route, since the three state model is generally a more defensible framework for conducting burst inference in the fruit fly embryo. In addition, future work should seek to compare intrinsic noise trends across individual stripes to assess whether these are consistent with the claim that all stripes exhibit the same regulatory signatures. Data limitations may prohibit this with the current set, but the question certainly bears further investigation. smFISH would also provide a powerful complementary experimental lens on this question.

4.4 Using 2 spot experiments to probe the information capacity of gene loci

In this section I sketch out some thoughts on connections between two spot experiments, the different regulatory knobs we've explored in Chapters 2 and 3, information transfer in

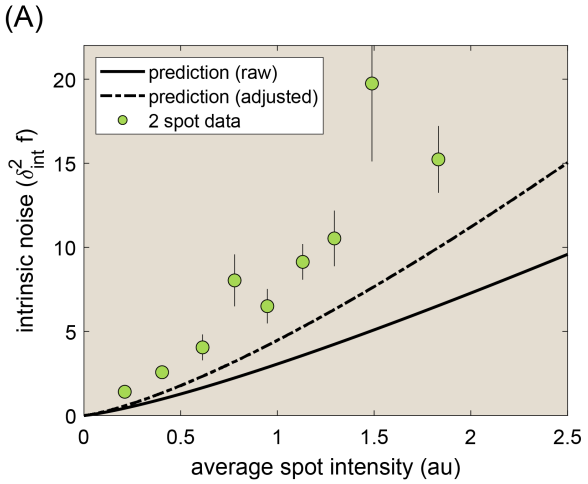

Figure 4.2: **Comparing intrinsic noise signatures to cpHMM inference predictions. (A)** Plot comparing observed intrinsic noise vs. mean fluorescent trends to predictions based off of cpHMM inference in Chapter 3. Green circles indicate noise signatures extracted from 2 spot data. Solid line indicates predictions using parameters for the effective 2 state model inferred in Chapter 3. Dashed line indicates approximate prediction when parameters are adjusted to simulate true 2 state model inference results.

gene circuits. Imagine a "perfect" experiment in which we can track protein concentrations for all four Gap Genes (and whatever additional factors are deemed relevant) over time in developing embryos while also tracking the transcriptional output of the *eve* stripes. Forget MS2 for a moment and imagine instead that we can measure the instantaneous rate of transcription, r, at individual gene loci directly. What could we do with data from such an experiment?

One obvious question that we could answer is how accurately the transcription rate r at gene i encodes information from the vector of transcription factor inputs experienced by that gene, $\boldsymbol{c_i}$. An intuitive way to approach this question would be to ask how much larger the initial variance in transcription rates is than the variance given a fixed set of the transcription factor inputs:

$$P_{reg} = \frac{Var(r)}{Var(r|\boldsymbol{c})}, \tag{4.8}$$

where we have defined this ratio as the regulatory potential of the gene locus. Note that, if the correspondence between r and $\boldsymbol{c}$ is perfectly precise, $Var(r|\boldsymbol{c}) \to 0$ and (assuming $Var(r) \neq 0$), P_{reg} becomes infinitely large. Conversely, if the regulatory factors in $\boldsymbol{c}$ have no impact on transcription, we will simply have that $P_{reg} = 1$. Thus, Equation 4.8 provides an intuitive measure of the strength of the regulatory relationship.

Of course, in practice, it is (for the moment) impossible to measure all relevant inputs to a gene like *eve* simultaneously. Yet it turns out that P_{reg} is still within our reach. Indeed, as we've already given away in the intro, the intrinsic noise, which captures variability between

sister gene loci from the same nucleus, is *precisely* $Var(r|\boldsymbol{c})$, such that

$$P_{reg} = \frac{\sigma^2_{tot}}{\sigma^2_{int}}. \tag{4.9}$$

In fact, if we take $p(r)$ and $p(r|\boldsymbol{c})$ to be approximately Gaussian, then we have that the mutual information between r and c is simply:

$$I(r;\boldsymbol{c}) = \log_2\left(\frac{\sigma^2_{tot}}{\sigma^2_{int}}\right)\log_2(e) = \log_2\Big(P_{reg}\Big)\log_2(e), \tag{4.10}$$

where e is the base of the natural logarithm.

Thus, we start to see the power of two spot experiments: the fact that we have access to two gene loci exposed to the same (potentially complex) set of transcription factor inputs allows us to sidestep the challenge of measuring regulatory inputs entirely and probe the information content of transcriptional outputs directly. We note that an analogous procedure could be performed with nuclei binned by relevant biological parameters; space and time, for instance. However, this procedure would inexorably underestimate $I(r;\boldsymbol{c})$, since it would factor in nucleus-to-nucleus variability in $\boldsymbol{c}$. Thus, two spot experiments provide a unique window into how much information can be transmitted by different regulatory outputs. We note that, for the purposes of the exploratory analyses that follow, we assume that relevant variables are Gaussian distributed for convenience. Testing this assumption and correcting for deviations will represent an important step in follow-up analyses.

4.4.1 Quantifying the information content of different regulatory strategies

In Chapter 2, we found that both the average rate of transcription and the duration of transcriptional engagement played a significant role in driving the formation of *even-skipped* stripe 2, but how much capacity does each of these regulatory strategies actually hold to transmit biologically salient information? Here, as illustrated in Figure 4.3A, we use our 2 spot data to calculate the amount of extrinsic and intrinsic variation in the average spot intensity (a proxy for the average transcription rate), and two components of the transcriptional time window: locus turn-on times and locus turn-off times. For this simple exercise, we look at these noise metrics taken across the full *even-skipped* pattern; however, we note that it would be informative to also take a more granular approach and look at how the regulatory impact of each knob varies as a function of position along the AP axis.

We visualize σ_{ext} and σ_{int} for each regulatory parameter by generating scatter plots where each point captures the relative parameter value for spot 1 (x axis) and spot 2 (y axis) for a single nucleus. Figure 4.3B shows this plot for the average spot fluorescence. By eye, it is clear that we see a non-trivial amount of correlation between sister spots, indicating that (unsurprisingly) the mean transcription rate is subject to some degree of transcription factor control. Our calculations confirm this, indicating that the total variation in spot intensity is

composed of roughly equivalent intrinsic and extrinsic components, such that $\sigma^2_{tot} f \approx 2\sigma^2_{int} f$. Plugging this in to Equation 4.10, we find that this equates to approximately 2/3 of a bit.

Moving on to the locus turn-on time, Figure 4.3C indicates a markedly lower degree of correlation between sister spots. Indeed, we find that the ON time conveys only about 1/5 of a bit's worth of information on average across the *eve* expression pattern, suggesting that the timing of the initial onset of transcription at gene loci contains little useful regulatory information. The on time scatters in Figure 4.3C also reveal interesting structures: we see a sparse central component of highly correlated late turn-on times, as well as two corresponding vertical and horizontal components (indicated by dashed ovals). We attribute to late turn-on events previously noted in Chapter 3 (Figure 3.7C), which occur as a result of the dynamic drifting of the *even-skipped* pattern over time.

Lastly, we find that a large degree of the total variability we observe in the transcriptional turn-off time is attributable to "useful" extrinsic variability (Figure 4.3D). Indeed, we find that intrinsic variability between sister spots comprises only about one fourth of the total variability in off times that we observe across the *eve* pattern. As a result, the transcriptional off time is highly informative, conveying precisely 1 bit of information on average across the *even-skipped* profile. Here especially it would be interesting to look at the information content of transcriptional OFF times as a function of AP position. Our expectation is that this regulatory event could convey significantly more than 1 bit of information at stripe boundaries, and significantly less at stripe centers.

Thus, we have used 2 spot experiments to build off of analyses presented in Chapters 2 and 3, moving beyond average spatial and temporal trends to use the relative magnitudes of σ^2_{tot} and σ^2_{int} to calculate the actual information content of key regulatory strategies shaping gene expression and pattern formation early on in fruit fly development. Yet, so far, we've focused on static measures of transcriptional control. Next we will apply this same framework to assess the rate of information transmission by *eve* across space and time as the 7 stripe pattern takes form over the course of nuclear cycle 14.

4.4.2 Active *eve* loci transmit information at up to 1 bit per minute

We can apply the same approach presented in the previous section to measure how much information actively transcribing *eve* loci transmit per unit time. The first step in this analysis was to apply the cpHMM method to obtain an estimate of the instantaneous rate of transcription, r, over time for individual *eve* loci (Figure 4.4A). Figure 4.4B shows the average instantaneous rate of transcription calculated from these decoded traces as a function of AP position and time. Note that the positions of the stripes are clearly visible in these plots as regions of elevated transcription. In similar fashion, we can likewise calculate the intrinsic and extrinsic noise components.

We can plug these two noise components into Equation 4.10 to calculate the information content of *eve* transcription. Our ultimate aim, however, is to calculate an Information *rate*, with units of bits per unit time. And here we encounter a subtlety that must be reckoned

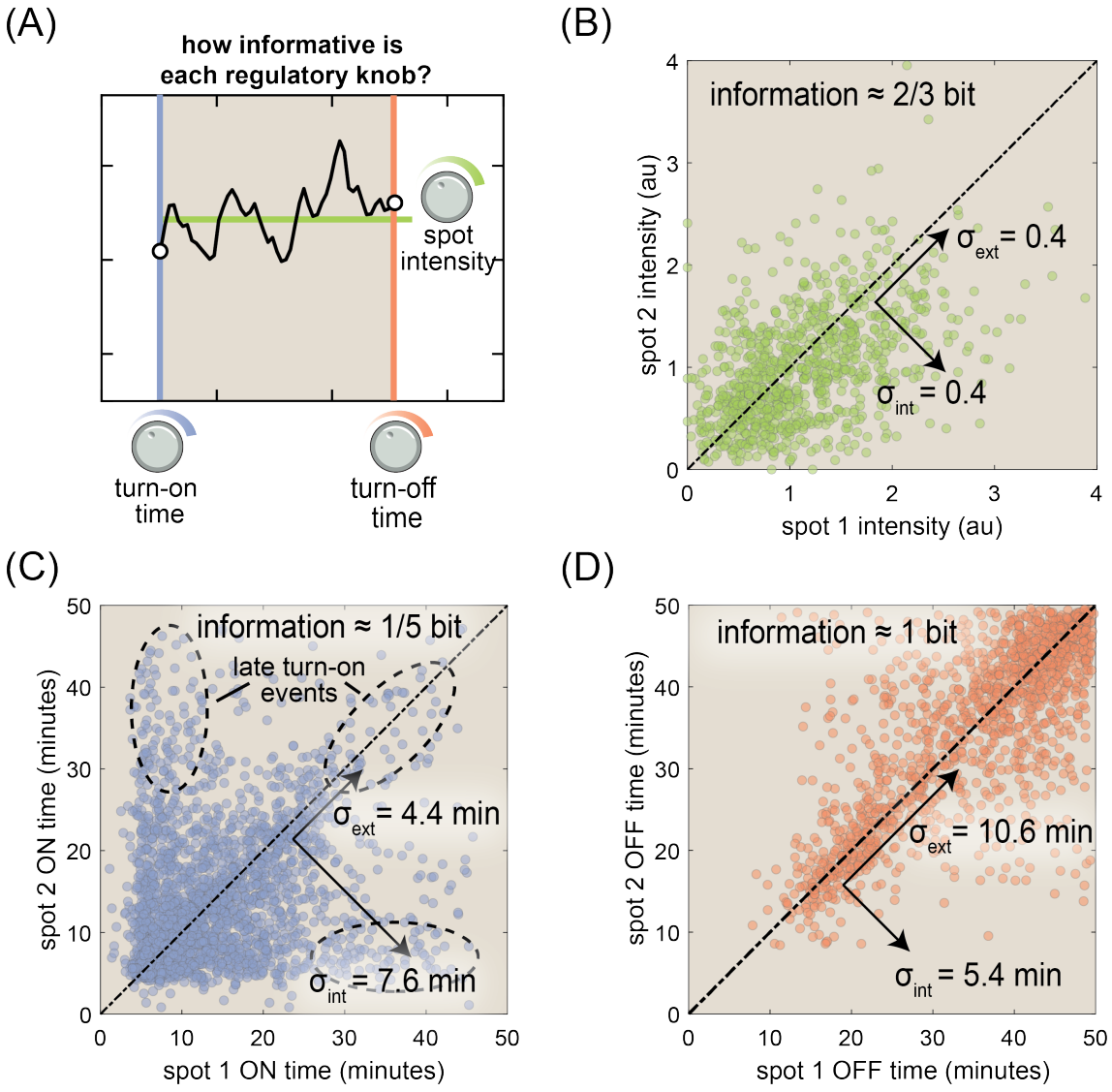

Figure 4.3: **Quantifying the information content of different regulatory strategies.** **(A)** We can use 2 spot experiments to measure the information content of three key parameters in *eve* transcription: the transcriptional turn-on time, the turn-off time, and the average rate of expression while active. **(B)** Scatter plot showing relationship between average spot intensities for sister spots within the same nucleus. **(C)** Scatter plot comparing turn-on times between sister spots. Dashed ovals indicate regions corresponding to instances where one or both *eve* loci turn on late in nuclear cycle 14, likely due to the movements of the *even-skipped* pattern over time. **(D)** Scatter plot depicting relative turn-off times for sister spots.

with: the question of averaging. On the one hand, using a single frame to calculate dI/dt yields noisy, uninformative estimates. The natural solution is to calculate the average value of this quantity across multiple time points; however, we find that the information content of the transcriptional signal saturates as the size of the averaging window, Δt, grows larger (Figure 4.4C), such that a naive rate estimate of $dI/dt = I(\Delta t)/\Delta t$ will yield a quantity that decreases as Δt increases.

We solve this problem by noting that the function $I(\Delta t)$ is well fit by the CDF of an exponential distribution times some constant, such that:

$$I(\Delta t) \approx I_0(1 - e^{-\Delta t/\tau_I}). \tag{4.11}$$

Differentiating this expression with respect to time and solving for the instantaneous limit ($\Delta t = 0$) gives us a robust estimate for the rate of information transmission:

$$\frac{I(\Delta t)}{dt} = \frac{I_0}{\tau_I}. \tag{4.12}$$

The red line in Figure 4.4D indicates the fit of this function to the empirical estimates of $I(\Delta t)$ for the entire dataset, which yields a characteristic time constant of $\tau_I = 3.4$ minutes. We estimate the maximum information content, I_0, separately for each time-AP group, but assume that this time constant holds for the full dataset.

Using these parameter estimates, we can now estimate the rate of information transmission by *eve* loci as a function of space and time. Figure 4.4D shows these estimates, where we see that the information rate varies substantially for different parts of the *eve* pattern. Comparing Figure 4.4D to the stripe positions reveals that stripe boundaries tend to be information rich, while stripe centers tend to be information *poor*. For instance, we see that *eve* loci in the regions between stripes 2 and 3, 4 and 5, and 5 and 6 all approach information rates of 1 bit per minute. On the other hand, the stripe centers can reliably be picked out as vertical stripes of blue and green, which corresponds to information rates of between 0 and 0.3 bits per minute. We also see that the information rate is not constant in time, with the highest rates of transmission occurring between 20 and 35 minutes into nuclear cycle 14.

Do these rates make sense? A previous study has established that the Gap Gene network, which constitutes the primary set of regulatory inputs for *even-skipped*, encodes about 4.1 bits of information on average [74]. Thus, we reasoned that I_0, which indicates the asymptotic value that I approaches as we average the transcriptional signal for longer and longer, should not significantly exceed this value. From Equation 4.12, we see that I_0 can be calculated by multiplying the information rate by τ_i (so, 3.4 minutes in this case). In agreement with the above reasoning, we find that the upper limit of the information that can be gleaned from *eve* transcription (I_0) is approximately 4 bits, with the maximum value observed equal to 4.21 bits. It is notable, however, that many regions fall far below this upper limit, with I_0 values typically falling between 0 and 2 bits in stripe centers. We also note that, although I_0 technically represents an asymptotic value, our calculations indicate that these values can be achieved with as little as 10 minutes of averaging (or accumulation), a timescale in agreement with the 7 minute half-life measured for a different pair rule gene, *fushi tarazu*, during nuclear cycle 14 [79].

Thus, our results appear largely to be consistent with previous works, although it is notable that we find maximum rates of transcriptional information to occur during a period of time (20-35 minutes) that significantly precedes the window of time (37-50 minutes) for which these works have found the information content of *eve*'s Gap Gene inputs to be highest [74, 73, 238]. A potential explanation for this is that we have limited ourselves here to looking at gene loci that have (i) already turned on and (b) not yet been silenced. Thus, it is possible that the high information content in the stripe flank regions persists, and even increases once transcriptionally silenced gene loci are factored in.

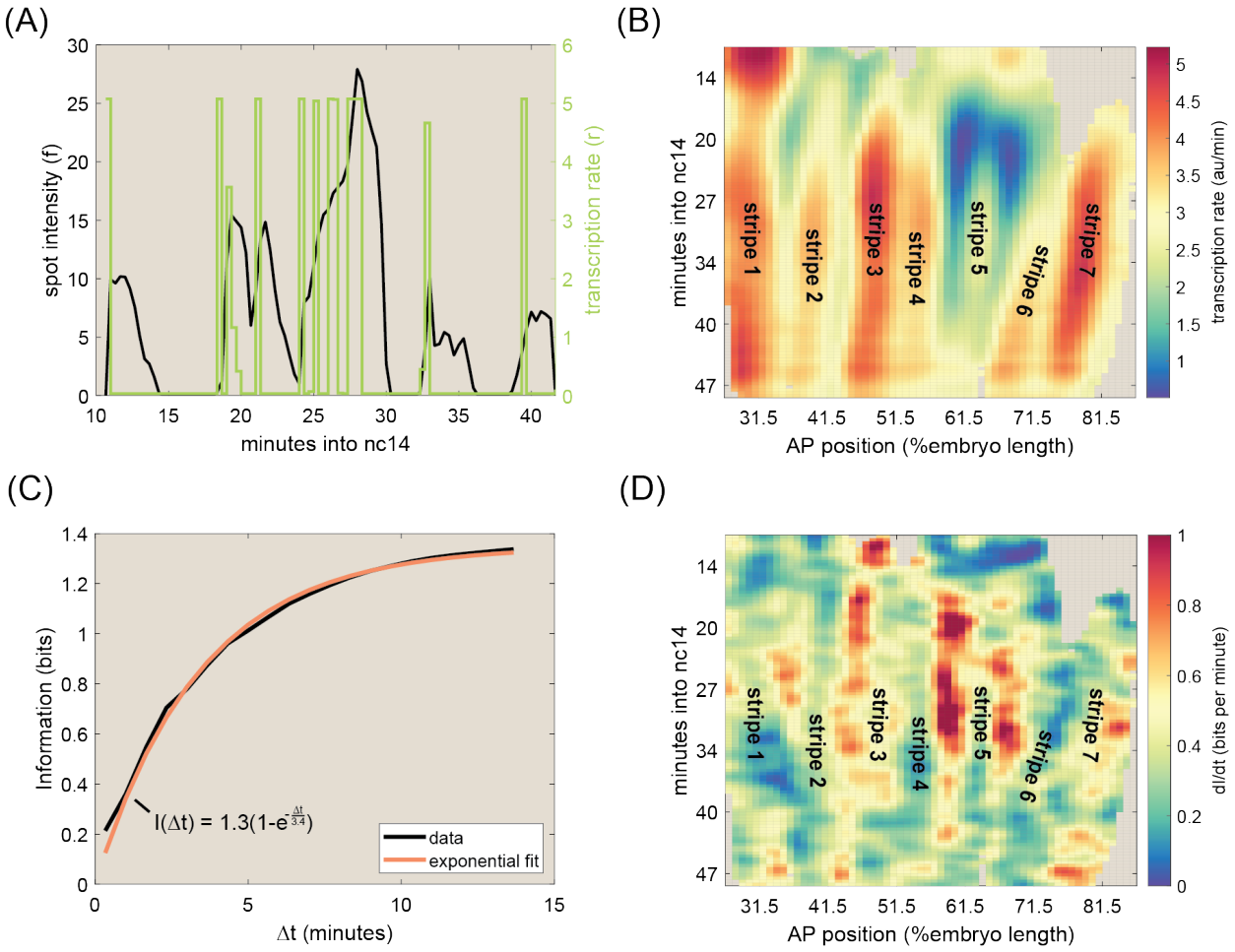

Figure 4.4: **Quantifying the rate of *eve* information transmission over time. (A)** Plot showing a raw MS2 trace (black) overlaid with the cpHMM inference of the instantaneous transcription rate (green). We use these decoded traces to generate estimates of **(B)** the mean rate pf transcription as a function of AP position and time into nuclear cycle 14. **(C)** Plot indicating the estimated information content of the *eve* transcriptional output as a function of the length of the averaging window used (in black) alongside the best-fit function (red). **(D)** shows the estimated rate of information transmission across the *eve* pattern at different points in time.

Chapter 5

A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting

This chapter is a reproduction of reference [166]

Abstract

Eukaryotic transcription generally occurs in bursts of activity lasting minutes to hours; however, state-of-the-art measurements have revealed that many of the molecular processes that underlie bursting, such as transcription factor binding to DNA, unfold on timescales of seconds. This temporal disconnect lies at the heart of a broader challenge in physical biology of predicting transcriptional outcomes and cellular decision-making from the dynamics of underlying molecular processes. Here, we review how new dynamical information about the processes underlying transcriptional control can be combined with theoretical models that predict not only averaged transcriptional dynamics, but also their variability, to formulate testable hypotheses about the molecular mechanisms underlying transcriptional bursting and control.

5.1 A disconnect between transcriptional bursting and its underlying molecular processes

Over the past two decades, new technologies have revealed that transcription is a fundamentally discontinuous process characterized by transient bursts of transcriptional activity interspersed with periods of quiescence. Although electron microscopy provided early hints of bursty transcription [208], the advent of single-molecule fluorescence *in situ* hybridization (smFISH) [88, 246], was key to establishing its central role in transcription. The single-cell distributions of nascent RNA and cytoplasmic mRNA molecules obtained using this

technology provided compelling, if indirect, evidence for the existence and ubiquity of gene expression bursts, and indicated that their dynamics were subject to regulation by transcription factors [245, 316]. These fixed-tissue inferences have been confirmed with new *in vivo* RNA fluorescence labeling technologies such as the MS2/MCP [14] and PP7/PCP systems [36], which directly reveal stochastic bursts of transcriptional activity in living cells in culture and within animals (Figure 5.1A-C) [111, 50, 170, 27].

What is the role of transcriptional bursting in cellular decision-making? One possibility is that bursty gene expression is intrinsically beneficial, helping (for instance) to coordinate gene expression or to facilitate cell-fate decision-making [80]. Alternatively, bursting may not itself be functional, but might instead be a consequence of key underlying transcriptional processes, such as proofreading transcription factor identity [114, 271].

Bursting and its regulation are intimately tied to the molecular mechanisms that underlie transcriptional regulation as a whole. In this Review we argue that, to make progress toward predicting transcriptional outcomes from underlying molecular processes, we can start with the narrower question of how the burst dynamics emerge from the kinetics of molecular transactions at the gene locus. To illustrate the importance and challenge of taking kinetics into account, we highlight two interrelated molecular puzzles that arise from new measurements of the dynamics of key transcriptional processes *in vivo*.

First, as illustrated in Figure 5.1D and reviewed in detail in Appendix Table B.1, despite qualitatively similar bursty traces from different organisms, bursts unfold across markedly distinct timescales ranging from several minutes [12, 165], to tens of minutes [177, 251], all the way to multiple hours [287]. Is this wide range of bursting timescales across organisms reflective of distinct molecular mechanisms or is it the result of a common set of highly malleable molecular processes?

Second, recent live imaging experiments have revealed a significant temporal disconnect between transcription factor binding events, which generally last for seconds, and the transcriptional bursts that these events control, which may last from a few minutes to multiple hours. The majority of the molecular processes underlying transcriptional control are highly transient (Figure 5.1E), with timescales ranging from milliseconds to seconds (see Appendix Table B.2 for a detailed tabulation and discussion of these findings).

In this Review, we seek to address this second puzzle by surveying key theoretical and experimental advances that, together, should shed light on the molecular origins of transcriptional bursting and transcriptional regulation. We leverage this framework to examine two kinds of molecular-level models that explain how slow burst dynamics could arise from fast molecular processes. Finally, we present concrete experimental strategies based on measuring variability in the timing of bursts that can be used to discern between molecular models of transcriptional bursting.

Overall, we seek to illustrate how iterative discourse between theory and experiment sharpens our molecular understanding of transcriptional bursting by reformulating cartoon models as concrete mathematical statements. Throughout this Review, we focus on illustrative recent experimental and theoretical efforts; we therefore do not attempt to provide a comprehensive review of the current literature (see [58, 181, 307, 250, 313, 242] for excellent reviews).

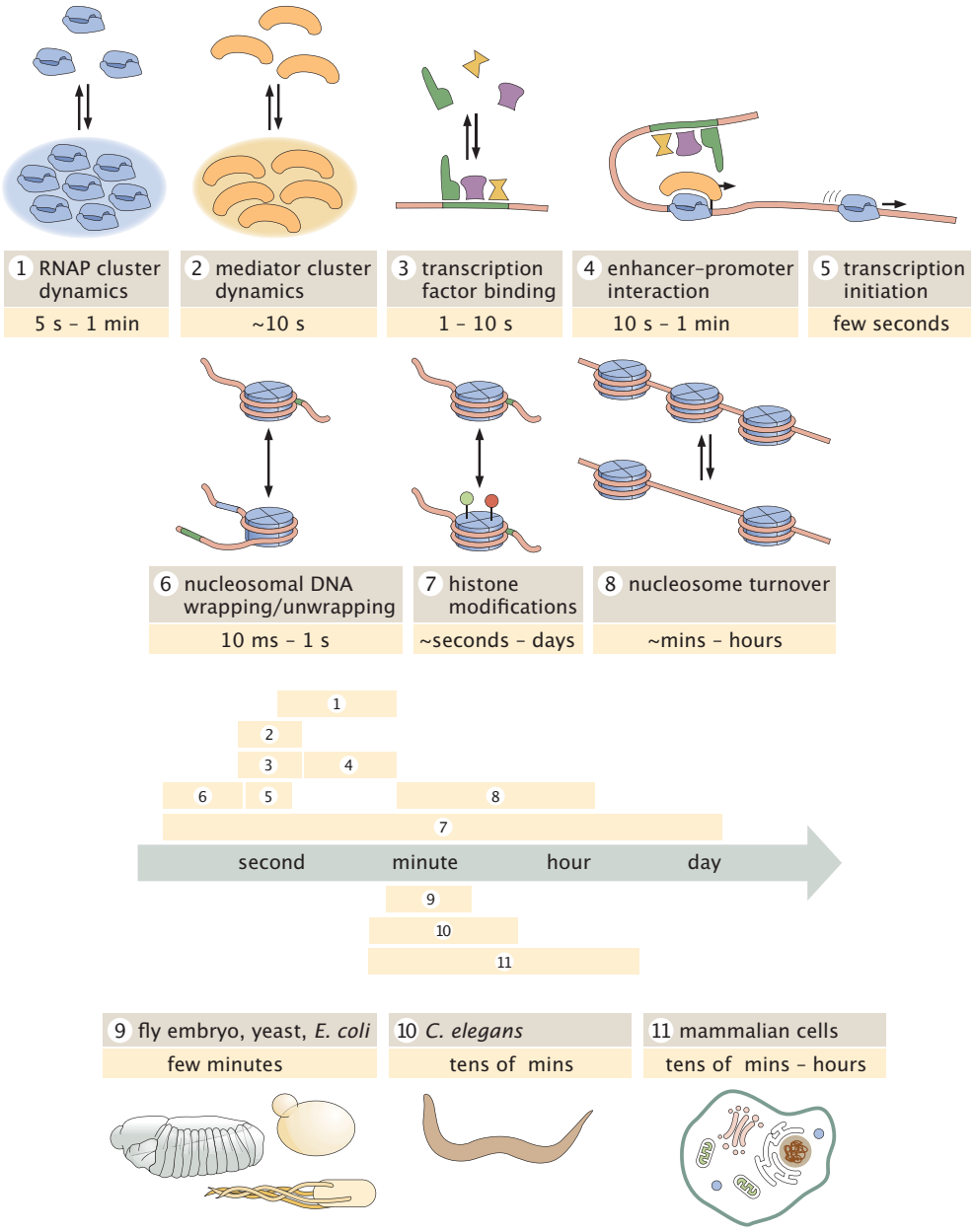

Figure 5.1: **Separation of timescales between transcriptional bursting and its underlying molecular processes. (A,B,C)** Transcriptional bursting in (A) an embryo of the fruit fly *Drosophila melanogaster*, (B) the nematode *Caenorhabditis elegans*, and (C) human cells. **(D)** In these and other organisms, bursting dynamics (average period of ON and OFF) span a wide range of timescales from a few minutes to tens of hours. **(E)** Timescales of the molecular processes behind transcription range from fast seconds-long transcription factor binding to slower histone modifications, which may unfold across multiple hours or days. A detailed summary of measurements leading to these numbers, including references, is provided in Appendix Table B.1 and Appendix Table B.2. (A, adapted from [165]; B, adapted from [177]; C, adapted from [251]).

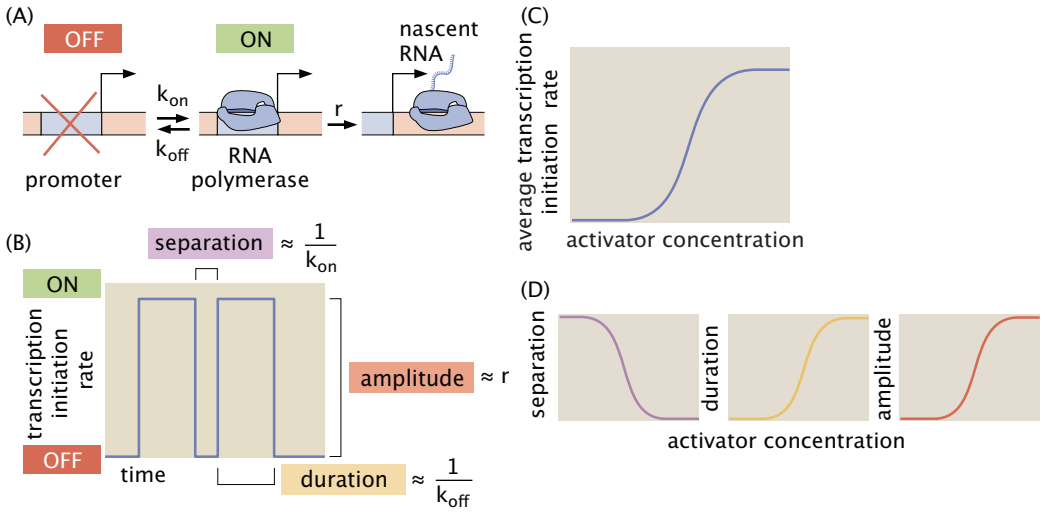

Figure 5.2: **The two-state model of transcriptional bursting**. **(A)** A two-state model of transcriptional bursting by a promoter switching between ON and OFF states. **(B)** Mapping the bursting parameters k_{on}, k_{off}, and r to burst duration, separation, and amplitude, respectively. **(C)** The action of an activator results in an increase in the average rate of transcription initiation. **(D)** In the two-state model, this upregulation can be realized by decreasing burst separation, increasing burst duration, increasing burst amplitude, or any combination thereof.

5.2 The two-state model: a simple quantitative framework for bursting dynamics

To elucidate the disconnect between molecular timescales and transcriptional bursting, we invoke a simple and widely used model of bursting dynamics: the two-state model of promoter switching. While the molecular reality of bursting is likely more complex than the two-state model suggests [55, 332, 221], there is value in examining where this simple model breaks down. This model posits that the promoter can exist in two states: a transcriptionally active ON state and a quiescent OFF state (Figure 5.2A). The promoter stochastically switches between these states with rates k_{on} and k_{off}, and loads new RNA polymerase II (RNAP) molecules at a rate r when in the ON state [307, 256, 21, 222]. Figure 5.2B shows a hypothetical activity trace for a gene undergoing bursty expression, where a burst corresponds to a period of time during which the promoter is in the ON state. The average burst duration, amplitude and separation are given by $1/k_{off}$, r and $1/k_{on}$, respectively.

Because the *instantaneous* transcription initiation rate during a burst is r and zero otherwise, the *average* initiation rate is equal to r times the fraction of time the promoter spends in this ON state p_{on},

$$\left\langle \text{initiation rate} \right\rangle = r\, p_{on}, \tag{5.1}$$

where brackets indicate time-averaging. As shown in B.2, in steady state, p_{on} can be expressed as a function of the transition rates k_{on} and k_{off}:

$$p_{on} = \frac{k_{on}}{k_{on} + k_{off}}. \tag{5.2}$$

Plugging this solution into Equation 5.1 results in the average rate of mRNA production as a function of the bursting parameters given by

$$\left\langle \text{initiation rate} \right\rangle = \underbrace{r}_{\substack{\text{transcription rate} \\ \text{in ON state}}} \quad \underbrace{\frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}}}_{\substack{\text{probability of} \\ \text{ON state}}}. \tag{5.3}$$

Equation 5.3 shows that, within the two-state model, transcription factors can influence the mean transcription rate by modulating any one of the three burst parameters (or a combination thereof). For example, consider an activator that can increase the mean transcription rate (Figure 5.2C) by decreasing k_{off}, increasing k_{on} or r, or any combination thereof (Figure 5.2D). Both live-imaging measurements and smFISH have revealed that the vast majority of transcription factors predominantly modulate burst separation by tuning k_{on} [97, 165, 12, 27, 330, 70, 316, 186]. There are also examples of the control of burst amplitude and duration [330, 85, 177].

Yet although experiments have identified *which* bursting parameters are under regulatory control, the question of *how* this regulation is realized at the molecular level remains open (with one notable exception in bacteria [44]). This is because the two-state model is a *phenomenological* model: we can use it to quantify burst dynamics without making any statements about the molecular identity of the burst parameters. Nonetheless, by putting hard numbers to bursting and identifying which parameter(s) are subject to regulation, this framework constitutes a useful quantitative tool to formulate and test hypotheses about the molecular mechanisms underlying transcriptional control.

For instance, consider the observation that many activators modulate burst separation. This observation can be explained if transitions between the ON and OFF states reflect the binding and unbinding of individual factors to regulatory DNA. Here, k_{off} would be the activator DNA-unbinding rate and k_{on} would be a function of activator concentration $[A]$,

$$k_{on}([A]) = [A]k_0^b, \tag{5.4}$$

where k_0^b is the rate constant for activator binding.

The two-state model highlights the absurdity of this proposition: if k_{off} were an activator unbinding rate, then it would be on the order of 1 s^{-1} (Figure 5.1D and E, box 7). However, measurements of burst duration reveal that k_{off} must be orders of magnitude smaller ($\lesssim 0.01$ s^{-1}, Figure 5.1D). Thus, the two-state model lends a quantitative edge to the disconnect in Figure 5.1, confirming that transcriptional bursting cannot be solely determined by the binding kinetics of the transcription factors that regulate it. We must therefore extend our simple two-state framework to incorporate molecular mechanisms that allow rapid transcription factor binding and transcriptional bursts that are orders of magnitude slower.

5.3 Bridging the timescale gap: kinetic traps and rate-limiting steps

Recent works have considered kinetic models of transcription that describe transition dynamics between distinct microscopic transcription factor binding configurations. These models make it possible to investigate how molecular interactions facilitate important behaviors such as combinatorial regulatory logic, sensitivity to changes in transcription factor concentrations, the specificity of interactions between transcription factors and their targets, and transcriptional noise reduction [271, 114, 296, 182, 262, 82].

We illustrate how these kinetic models can shed light on the disconnect between the timescales of transcription factor binding and bursting using the activation of the *hunchback* minimal enhancer by Bicoid in the early fruit fly embryo as a case study [116, 82, 231, 296, 76, 70]. Recent *in vivo* single-molecule studies have revealed that Bicoid specifically binds DNA in a highly transient fashion ($\sim 1 - 2$ s) [212, 214], suggesting that Bicoid binding cannot dictate the initiation and termination of *hunchback* transcriptional bursts, which happen over minutes [70]. We seek molecular models that recapitulate two key aspects of bursting: (1) the emergence of effective ON and OFF transcriptional states, and (2) "slow" (>1 min) fluctuations between these states. We sketch out the mathematical basis of these efforts and review key results below; more detailed calculations can be found in B.3.

Following [296], we consider a simple activation model featuring an enhancer with identical activator binding sites. While the full model for the *hunchback* minimal enhancer consists of six binding sites, we first use a simpler version with three binding sites to introduce key features of our binding model before transitioning to the more realistic six binding sites version when discussing our results. We capture the dynamics of activator binding and unbinding at the enhancer by accounting for the transitions between all possible binding configurations (Figure 5.3A). Our assumption of identical activator binding sites leads to two simplifications: (1) the same rate, $k_{i,j}$ governs the switching from any configuration with i activators bound to any configuration with j bound and, (2) all binding configurations with the same number n of activators bound have the same rate of transcription, $r_n = r_0\, n$, which we posit to be proportional to the number of bound activators. As a result we need not track specific binding configurations and may condense the full molecular representation in Figure 5.3A into a simpler four-state chain-like model with one state for each possible value of n (Figure 5.3B).

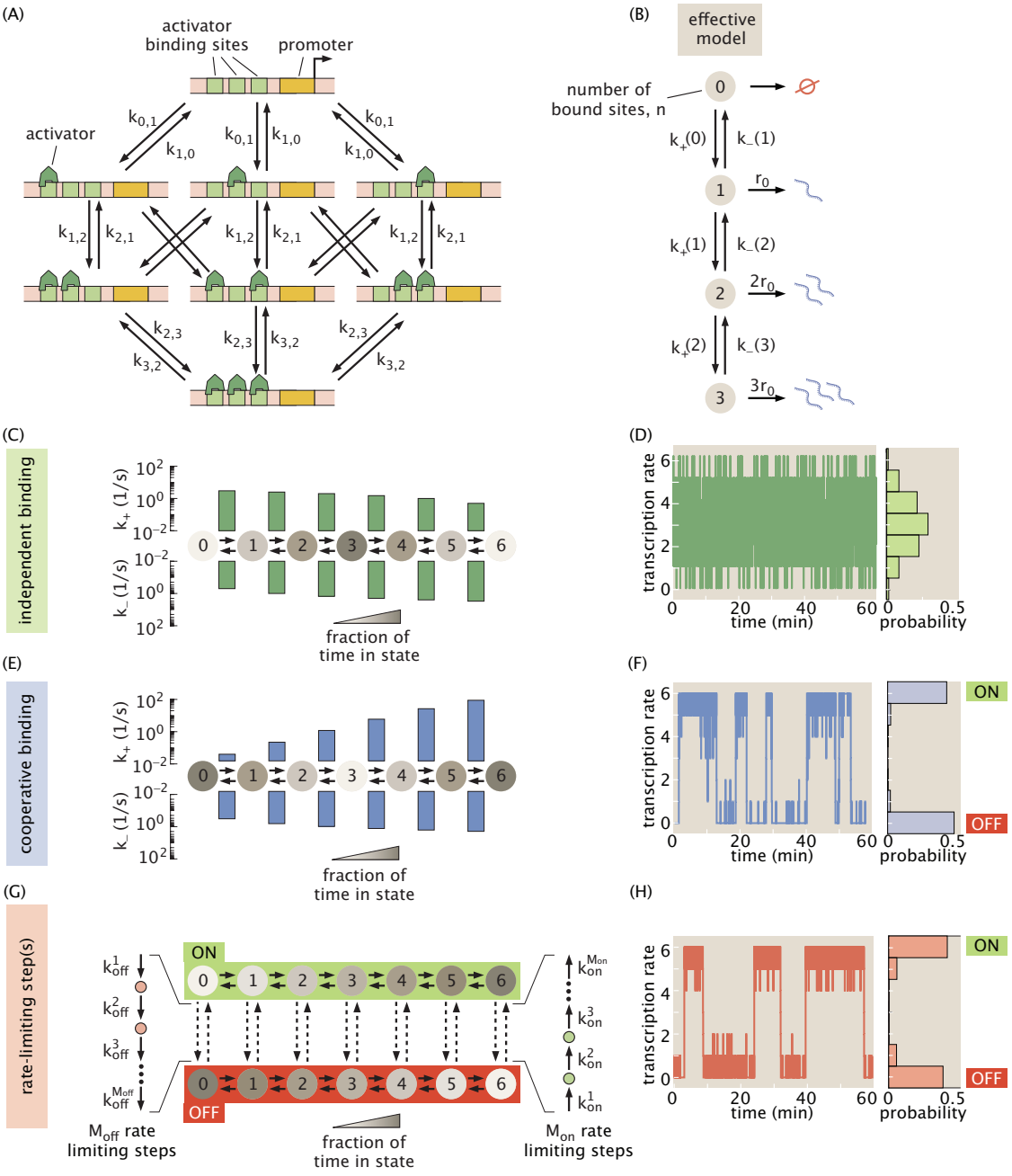
(A)
activator binding sites
promoter
activator
(B)
effective model
number of bound sites, n
(C)
independent binding
fraction of time in state
(D)
transcription rate
time (min)
probability
(E)
cooperative binding
fraction of time in state
(F)
ON
OFF
(G)
rate-limiting step(s)
ON
OFF
M_off rate limiting steps
M_on rate limiting steps
fraction of time in state
(H)
ON
OFF

Figure 5.3: **Using theoretical models to understand the origin of ON/OFF bursting dynamic**. **(A)** Model with three activator binding sites. The transition rates between states with i and j activators are given by $k_{i,j}$. **(B)** The model in (A) can be simplified to an effective four-state chain model in which each state corresponds to a certain number of bound molecules and the transcription rate is proportional to the number of bound activators. **(C)** Independent activator binding model with effective binding and unbinding rates plotted above and below, respectively. Shading indicates the fraction of time that the system spends in each state. **(D)** Stochastic simulations indicate that rapid activator binding alone drives fast fluctuations about a single transcription rate. **(E)** Cooperative binding model in which already-bound activators enhance the binding rate of further molecules. **(F)** Simulation reveals that cooperativity can cause the system to exhibit bimodal rates of transcription and slow fluctuations between effective ON and OFF states. **(G)** Rate-limiting step model in which several molecular steps can connect a regime where binding is favored (ON) and a realization where binding is disfavored (OFF). **(H)** Simulations demonstrate that rate-limiting steps can lead to bimodal transcriptional activity reminiscent of transcriptional bursting. Simulation results were down-sampled to a resolution of 0.5 s to ensure plot clarity in D, F, and G. (Parameters: C, D, $k^b = k^u = 0.5\text{s}^{-1}$; E, F, $k^b = 0.004$ s^{-1}, $k^u = 0.5$ s^{-1}; and $\omega = 6.7$; G,H, $k^u_{on} = k^u_{off} = 0.5$ s^{-1}, $k^b_{off} = 0.01$ s^{-1}, $k^b_{on} = 21$ s^{-1}, $M_{off} = 1$, $M_{on} = 2$, $k^1_{off} = 0.0023$, $k^1_{on} = k^2_{on} = 0.0046$ s^{-1}.)

Transitions up and down the chain in Figure 5.3B are governed by the effective binding and unbinding rates $k_+(n)$ and $k_-(n)$. To calculate these rates from the microscopic transition rates $k_{i,j}$, consider, for example, that there are three possible ways of transitioning from the 0 state to the 1 state, each with rate $k_{0,1}$. Thus, the effective transition rate between states 0 and 1 is given by $3k_{0,1}$. More generally, in the effective model, activator binding rates are

$$k_+(n) = (N - n)k_{n,n+1}, \tag{5.5}$$

where n indicates the current number of bound activators and N is the total number of binding sites. Similarly, activator unbinding rates are given by

$$k_-(n) = nk_{n,n-1}. \tag{5.6}$$

These transition rates allow us to generalize to the more realistic enhancer with six binding sites.

We first examine a system in which activator molecules bind and unbind independently from each other (Figure 5.3C). There are only two unique microscopic rates in this system: activator molecules bind at a rate $k_{i,i+1} = k^b = k^b_0[A]$, with $[A]$ being the activator concentration and k^b_0 the binding rate constant, and unbind at a rate $k_{i,i-1} = k^u$. We fix the unbinding rate $k^u = 0.5\ s^{-1}$ to ensure consistency with recent experimental measurements of Bicoid in [212, 214]. For simplicity, we also set $k^b = 0.5$ s^{-1} (see B.3.2.3 for details).

To gain insight into the model's transcriptional dynamics, we employ stochastic simulations based on the Gillespie Algorithm [110]; however a variety of alternative analytic and numerical approaches exist [82, 76, 262]. Our simulations reveal that independent binding

leads to a unimodal output behavior in which the transcription rate fluctuates rapidly about a single average (Figure 5.3D). This result is robust to our choices of k^b or k^u, as well as the number of binding sites in the enhancer (B.3.2.2). The observed lack of slow, bimodal fluctuations leads us to conclude that the independent binding model fails to recapitulate transcriptional bursting.

A thought-provoking study recently suggested that protein-protein interactions between transcription factors near gene loci could generate burst-like behavior [134]. Inspired by this work, we extend the independent binding model to consider cooperative protein-protein interactions between activator molecules [298] that catalyze the binding of additional activators. Here, the activator binding rate is increased by a factor ω for every activator already bound, leading to

$$k_{i,i+1} = k^b \omega^i. \tag{5.7}$$

Because we assume that activator unbinding still occurs independently, the effective unbinding rates remain unchanged (Equation 5.6).

Stochastic simulations of the cooperative binding model in Figure 5.3F reveal that the output transcription rate now takes on an all-or-nothing character, fluctuating between high and low values that act as effective ON and OFF states. Further, our simulation indicates that these emergent fluctuations are quite slow (0.13 transitions/min for the system shown), despite fast activator binding kinetics. Both of these phenomena result from large imbalances between $k_+(n)$ and $k_-(n)$ that act as "kinetic traps".

Consider the case with five bound activators. If $k_+(5) \gg k_-(5)$, then the enhancer is much more likely to bind one more activator molecule and move to state six than to lose an activator and drop to state four. For instance, if $k_+(5)/k_-(5) = 23$ (Figure 5.3F), then the system will on average oscillate back and forth between states five and six 23 times before it finally passes to state four. While it is possible to generate this kind of trap without cooperativity at one end of the chain or the other by tuning k^b, cooperative interactions are needed to simultaneously achieve traps at both ends. Finally, it is important to note that phenomenon is not limited to activator binding: cooperative interactions in fast molecular reactions elsewhere in the transcriptional cycle, such as in the dynamics of pre-initiation complex assembly, could, in principle, also induce slow fluctuations.

Inspired by the MWC model of protein allostery [76, 204], a second way to bridge the timescale gap between activator binding and transcriptional bursting is to posit two distinct system configurations: an ON configuration where binding is favored ($k^b \gg k^u$) and an OFF configuration that is less conducive to binding ($k^b \ll k^u$). From any of the seven binding states, this system can transition from OFF to ON by traversing M_{on} slow steps, each with rate $k^i_{on} \ll k^u$, where i is the step number (Figure 5.3G). Similarly, transitions from ON to OFF are mediated by M_{off} steps with rates given by k^i_{off}. Stochastic simulations indicate that this system yields bimodal transcription that fluctuates between high and low activity regimes on timescales set by the rate-limiting molecular steps (Figure 5.3H). Thus, as long as these steps induce a sufficiently large shift in activator binding (k^b), the rate-limiting step model reconciles rapid activator binding with transcriptional bursting.

Figure 5.1B suggests candidates for these slow molecular steps. For example, the ON state in Figure 5.3G could correspond to an open chromatin state that favors binding while the OFF state could indicate that a nucleosome attenuates binding such that $M_{on} = M_{off} = 1$. Our model also allows multiple distinct rate-limiting steps. For instance, chromatin opening could require multiple histone modifications ($M_{on} \geq 2$, $M_{off} = 1$), or chromatin opening may need to be followed by enhancer-promoter co-localization to achieve a high rate of transcription ($M_{on} = 2$, $M_{off} = 1$).

Although they are not the only possible models, the cooperativity and rate-limiting step scenarios discussed above represent two distinct frameworks for thinking about how slow processes like bursting can coincide with, and even arise from, rapid processes like activator binding. The next challenge in identifying the molecular processes that drive transcriptional bursting is to establish whether these models make experimentally distinguishable predictions.

5.4 Using bursting dynamics to probe different models of transcription

While we cannot yet directly observe the microscopic reactions responsible for bursting in real time, these processes leave signatures in transcriptional dynamics that may distinguish molecular realizations of bursting such as those of our cooperative binding (Figure 5.3E) and rate-limiting step (Figure 5.3G) models. Inspired by [323, 288, 332, 75], we examine whether the distribution of observed burst separation times (Figure 5.4A) distinguishes between these two models. In keeping with literature convention, we refer to these separation times as *first-passage times* from OFF to ON.

The variability in reactivation times provides clues into the number of hidden steps in a molecular pathway. For instance, suppose that bursts are separated by an average time $\tau_{off} = 1/k_{\text{on}}$, as defined in the two-state model in Figure 5.2A and B. If there is only a single rate-limiting molecular step in the reactivation pathway ($M_{on} = 1$ in Figure 5.3G), then the first-passage times will follow an exponential distribution (Figure 5.4B) such that the variability, defined as the standard deviation (σ_{off}), will simply be equal to the mean (τ_{off}). Now, consider the case where two distinct molecular steps, each taking an average $\tau_{off}/2$, connect the OFF and ON states ($M_{on} = 2$). To calculate the variability in the time to complete *both* steps and reactivate, we need to add the variability of each step in quadrature:

$$\sigma_{off} = \sqrt{\left(\frac{\tau_{off}}{2}\right)^2 + \left(\frac{\tau_{off}}{2}\right)^2} = \frac{\tau_{off}}{\sqrt{2}}. \tag{5.8}$$

More generally, in the simple case in which each step has the same rate, given an average first-passage time of τ_{off}, the variability in the distribution of measured first-passage times will decrease as the number of rate-limiting steps, M_{on}, increases following

$$\sigma_{off}(M_{on}) = \frac{\tau_{off}}{\sqrt{M_{on}}}. \tag{5.9}$$

As predicted by Equation 5.9, increasing the rate-limiting step number reduces the width of the distribution for the rate-limiting step model obtained from stochastic simulations, shifting passage times from an exponential distribution when $M_{on} = 1$ to increasingly peaked gamma distributions when $M_{on} > 1$ (Figure 5.4B).

Based on these results, since the fluctuations between high- and low-activity regimes reflect transitions through many individual binding states in the cooperative binding model (Figure 5.3E), we might also expect this model to exhibit non-exponential first passage times. Instead, the first-passage times are exponentially distributed (Figure 5.4C). This result is consistent with earlier theoretical work that examined a chain model similar to ours and found that sufficiently large reverse rates (k^u in our case) cause first-passage time distributions to exhibit approximately exponential behavior [9].

The coefficient of variation ($CV = \sigma_{off}/\tau_{off}$) provides a succinct way to summarize the shape of passage time distributions for a wide range of parameter values. Figure 5.4D plots σ_{off} against τ_{off} for each of the model architectures considered in Figure 5.4B and C for a range of different τ_{off} values. Points representing distributions with $CV = 1$ will fall on the line with slope one and points for distributions with $CV < 1$ will fall below it. We see that both the cooperative binding model and the single rate-limiting step model have CV values of approximately one for a wide range of τ_{off} values, consistent with exponential behavior. Conversely, all models with multiple rate-limiting steps have slopes that are significantly less than one.

Thus, by moving beyond experimentally measuring average first-passage time for a given gene and examining its *distribution*, it is possible to rule out certain molecular mechanisms. For example, a non-exponential distribution would be evidence against the cooperative binding and single rate-limiting step models (see B.3.1 and B.3.4 for details about stochastic simulations and first-passage time calculations). While these conclusions are specific to the models considered here, the general approach of invoking the distributions rather than means and using stochastic simulations to derive expectations for different models can be employed to discriminate between molecular hypotheses in a wide variety of contexts. Indeed, the examination of distributions has been revolutionary throughout biology by making it possible to, for example, reveal the nature of mutations [197], uncover mechanisms of control of transcriptional initiation [257] and elongation [266, 3], measure translational dynamics [33], and even count molecules [252].

Note that, while appropriate for *qualitatively* estimating the order of magnitude of bursting timescales, raw fluorescence measurements from MS2 and PP7 experiments such as those in Figure 5.1A-C do not directly report on the promoter state. Rather, the signal from these experiments is a convolution of the promoter state and the dwell time of each nascent RNA molecule on the gene body [165]. As a result, inference techniques like those developed in [165, 55] are often required to infer underlying burst parameters and promoter states that can be used to estimate first-passage time distributions. Other techniques, such as measuring the short-lived luminescent signal from reporters [332], have also successfully estimated first-passage times.

The first-passage time analyses discussed here are just one of an expansive set of approaches to determining the best model to describe experimental data. For instance, direct

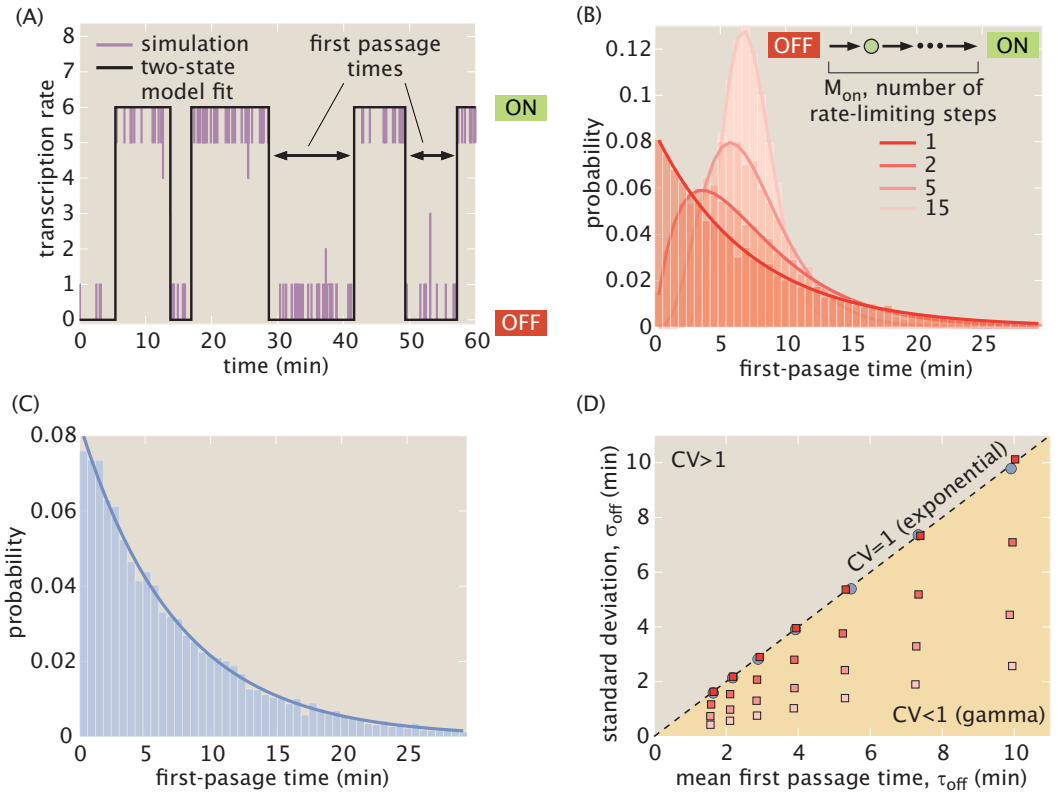

Figure 5.4: **Using first-passage time distributions to discriminate between models of transcriptional bursting**. **(A)** The outcome of stochastic simulations like those in Figure 5.3D, F, and G (purple) is fit to a two-state model (black) and the first-passage times out of the OFF state are measured. **(B)** First-passage times for the rate-limiting step model as a function of the number of rate-limiting steps M_{on} calculated using stochastic simulations. A single step results in an exponential distribution, but distributions break from this behavior when more steps are added, yielding increasingly peaked gamma distributions. **(C)** In contrast, first-passage times for the cooperative binding model follow an exponential distribution. **(D)** Standard deviation as a function of mean first-passage time for various parameters choices of the cooperative binding (blue) and the rate-limiting step models (red, with color shading indicating the M_{on} values considered in (B)). Distributions with $CV = 1$, such as the exponential distribution, fall on the line of slope one while gamma distributions, with $CV < 1$, fall in the region below this line. (Parameters: B, $k^u_{on} = k^u_{off} = 0.5$ s^{-1}, $k^b_{off} = 0.01$ s^{-1}, $k^b_{on} = 21$ s^{-1}, $M_{off} = 1$, $k^1_{off} = 0.0023$ s^{-1}, $k^i_{on} = M_{on}0.0023$ s^{-1}; C, $k^b = 0.004$s^{-1}, $k^u = k_{off}$, and $\omega = 6.7$; D, see simulation scripts on GitHub for exact parameter values.)

fits of models to experimental time traces could be used to identify the most appropriate model (see, e.g. [55, 275]). A discussion of this and other approaches falls beyond the scope of this work, but we direct the reader to several excellent introductions to elements of this field [275, 276, 209, 71].

5.5 Conclusions

The rapid development of live-imaging technologies has opened unprecedented windows into *in vivo* transcriptional dynamics and the kinetics of the underlying molecular processes. We increasingly see that transcription is complex, emergent, and—above all—highly dynamic, but experiments alone still fail to reveal how individual molecular players come together to realize processes that span a wide range of temporal scales, such as transcriptional bursting.

Here we have argued that theoretical models can help bridge this crucial disconnect between single-molecule dynamics and emergent transcriptional dynamics. By committing to mathematical formulations rather than qualitative cartoon models, theoretical models make concrete quantitative predictions that can be used to generate and test hypotheses about the molecular underpinnings of transcriptional control. We have also shown how, although different models of biological phenomena might be indistinguishable in their averaged behavior, these same models often make discernible predictions at the level of the distribution of such behaviors.

Moving forward, it will be critical to continue developing models that are explicit about the kinetics of their constituent molecular pieces, as well as statistical methods for connecting these models to *in vivo* measurements in an iterative dialogue between theory and experiment. In particular, robust model selection frameworks are needed to navigate the enormous space of possible molecular models for transcriptional control. Such theoretical advancements will be key if we are to synthesize the remarkable experimental findings from recent years into a truly mechanistic understanding of how transcriptional control emerges from the joint action of its molecular components.

Chapter 6

Competing constraints shape the limits of gene regulation out of thermodynamic equilibrium

6.1 Abstract

Gene regulation is central to cellular function. Yet, despite decades of work, we nonetheless lack quantitative models that can predict how transcriptional control emerges from molecular interactions at the gene locus. Thermodynamic models of transcription, which assume that gene circuits operate at equilibrium, have previously been employed with considerable success in the context of bacterial systems. However, the presence of ATP-dependent processes within the eukaryotic transcriptional cycle suggests that equilibrium models may be insufficient to capture how eukaryotic gene circuits sense and respond to input transcription factor concentrations. Here, we employ simple kinetic models of transcription to investigate how energy dissipation within the transcriptional cycle impacts the rate at which genes can transmit information and drive cellular decisions. We find that biologically plausible levels of energy input can lead to significant gains in how rapidly gene loci transmit information, but discover that the regulatory mechanisms underlying these gains change depending on the level of interference from off-target activator binding. When interference is low, information is maximized by harnessing energy to push the sharpness of the transcriptional response to input transcription factors beyond its equilibrium limits. Conversely, when interference is high, conditions instead favor genes that harness energy to increase transcriptional specificity. Our analysis further reveals that equilibrium gene regulatory mechanisms break down as transcriptional interference increases, suggesting that energy dissipation may be indispensable in systems where non-cognate factor interference is sufficiently large.

6.2 Introduction

Throughout biology, systems must make accurate decisions under time constraints using noisy molecular machinery. Eukaryotic gene regulation exemplifies this challenge: here genes must read input concentrations of transcription factor proteins and respond by producing appropriate levels of gene product (mRNA and eventually protein) in order to drive downstream cellular decisions. Interestingly, the gene activity underlying cellular decision-making is often subject to large amounts of noise. Indeed, experiments across a wide range of organisms have revealed that eukaryotic transcription is highly stochastic, occurring in episodic bursts [26, 290, 225, 185]—periods of activity interspersed with periods of transcriptional silence—that unfold over timescales ranging from a few minutes to multiple hours [166]. This stochasticity means that the rate of transcription is, invariably, a noisy reflection of transcription factor concentration. Given enough time, the accumulation of gene product will tend to average out this noise, but biological processes, by their very nature, must operate under time constraints: cells in developing fruit fly embryos have only minutes to determine their developmental fates [2, 69], antigen recognition in T-cells unfolds over the span of a single day [228], and even cells in adult tissues are constrained by mRNA half lives that range from minutes to days [235].

A key question, therefore, is how the molecular architecture of gene loci—the number and identity of biochemical steps in the transcriptional cycle, as well as the reaction rates that connect these steps—dictates the amount of time needed for bursty gene expression to drive accurate cellular decisions. In particular, while it is widely accepted that processes within the eukaryotic transcriptional cycle consume biochemical energy [59, 314], we do not yet know what non-equilibrium should "look like" in the context of transcriptional systems. Indeed, it remains challenging not only to predict unambiguous signatures of energy expenditure that can be detected experimentally [123, 231, 76], but also whether energy consumption is actually harnessed to improve gene regulatory performance and, thereby, improve the accuracy of cellular decision-making.

Here, we use concepts from information theory and statistical physics as a lens to investigate how energy dissipation impacts the timescale on which gene circuits can drive cellular decisions. We consider a simple binary choice scenario wherein a cell must decide, as rapidly as possible, whether it is subjected to a high or low concentration of a transcriptional activator (c_1 or c_0?) based on the transcriptional output of a gene locus. The basis for this decision is the gene's input-output function (Figure 6.1A), which emerges from microscopic interactions between input activator molecules and the basal transcriptional machinery at the gene locus (Figure 6.1B) that induce differences in the output dynamics of transcriptional bursting (Figure 6.1C) for high (c_1) and low (c_0) activator concentrations. These differences in the burst dynamics, in turn, drive different rates of mRNA accumulation (Figure 6.1D). Since each ON/OFF fluctuation is stochastic, the resulting levels of gene expression are noisy, and the cell must wait some time T before it is possible to accurately distinguish between c_1 and c_0. Our central question in this work is whether energy dissipation allows gene loci to decrease the decision time, T, and, if so, how this performance gain manifests in terms of measurable features of the transcriptional input-output function.

There are multiple distinct ways in which energy dissipation could alter the input-output behavior of a gene locus to improve cellular decision-making. As illustrated in Figure 6.1A, non-equilibrium processes could function to increase sensitivity to differences in input transcription factor concentration ("sharpness") or to suppress transcriptional noise ("precision"). Since our model assumes that, in addition to the concentration of the cognate activator, C, the gene locus is subject to some level of non-cognate factors, W, energy dissipation could also serve to buffer against interference from off-target activation ("specificity").

Recent works have begun to uncover a complex space of trade-offs between these three aspects of transcriptional performance both at and away from thermodynamic equilibrium. A recent work found that systems operating at thermodynamic equilibrium suffer from strict trade-offs between transcriptional specificity and transcriptional precision, but that this trade-off could be overcome by gene circuits that spend energy to enhance specificity through a scheme reminiscent of classical kinetic proofreading [272, 226, 136]. A separate study demonstrated that energy dissipation could likewise function to enhance transcriptional sharpness [82]. Yet, interestingly, while energy can increase sharpness and specificity *separately*, another recent study has indicated that non-equilibrium levels of specificity come at the cost of sub-optimal sharpness [113]. The same work found that energy dissipation tends to *decrease* transcriptional precision, though this conclusion likely hinges on specific modelling assumptions made by the authors [113]. Thus, the field has begun to uncover tantalizing insights into the interplay between energy dissipation and different aspects of transcriptional performance. However, it remains unclear how these non-equilibrium gains and trade-offs ultimately fit together to impact how effectively gene circuits can harness transcription factor concentrations to drive cellular decisions.

In this work, we identify a key quantity, the rate of information transmission (IR) from input transcription factor concentrations to the output rates of transcription, as the quantitative link between energy-dependent changes to different aspects of the transcriptional input-output function (Figure 6.1A) and the amount of time required for gene loci to drive accurate biological decisions [274, 69]. As a starting point, we focus on the case where interference from non-cognate factor binding is negligible. We demonstrate that energy dissipation increases IR, both for the simple model with one binding site ($N_B = 1$) and two locus conformations ($N_{LC} = 2$) in Figure 6.1B, and for more complex architectures that feature multiple binding sites or multiple locus conformations. We uncover a novel tradeoff between transcriptional sharpness and transcriptional precision that dictates the upper IR limit in non-equilibrium gene circuits and show that, while both sharpness and precision can be improved by energy dissipation, increasing sharpness is the key to maximizing the rate of information transmission away from equilibrium. In addition, we find that, in gene loci operating out of equilibrium, activator binding at a single site can drive increased sharpness by regulating multiple successive steps in the assembly of the transcriptional machinery. This non-equilibrium "kinetic" cooperativity matches the sharpness achieved by "classical" cooperative interactions between activators bound at multiple sites in gene loci operating at equilibrium.

Next, we extend our model to consider the effects of interference due to the binding of non-cognate transcriptional activators. In contrast to recent modeling efforts [272, 113], we

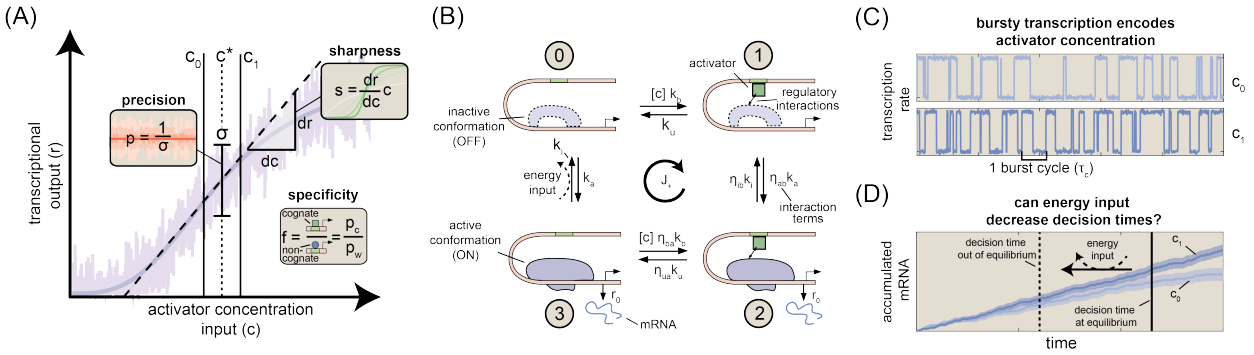

Figure 6.1: **Three factors shaping transcriptional information transmission. (A)** Gene regulatory input-output function illustrating the basic biological problem considered in this work. Here, a cell must distinguish between two activator concentrations, c_0 and c_1, based off of the transcriptional output of a gene locus (purple curve). We examine how three aspects of the transcriptional input-output function—sharpness, precision, and specificity—combine to dictate the rate at which the transcriptional output can drive biological decisions. **(B)** Four state MWC-like model of transcription used as the foundation of our investigations throughout. Here a single activator (green square) may bind to a specific site at the gene locus, and mRNA production occurs when the gene locus switches into its active (ON) conformation. A hypothetical energy input is depicted along the rate from state 3 to state 0. In practice, our framework permits non-equilibrium driving to occur along any of the 8 transition rates in the model. **(C)** Simulated burst dynamics for one realization of the model shown in (B). Activator binding drives different burst dynamics at loci exposed to high and low activator concentrations. The burst cycle time is defined as the average time required to complete one ON $\rightarrow$ OFF $\rightarrow$ ON, and sets the timescale over which biological decision unfold. **(D)** Illustrative simulation results for accumulated mRNA levels driven by c_1 and c_0. Solid lines show trajectories for a single locus and shaded regions indicate the standard deviation in levels taken across 100 simulated trajectories.

build off of [35] to consider a model in which cognate and non-cognate factors must compete with one another to activate a single gene locus. We find that the relative concentration of wrong-to-right activator species, w/c, defines a shifting tradeoff between transcriptional sharpness and transcriptional specificity. When w/c is small, (e.g. in the fruit fly embryo), non-equilibrium gene circuits that maximize sharpness drive the fastest decisions. On the other hand, when w/c is large (e.g. in cells in mammalian tissues), we find that gene circuits must instead prioritize transcriptional specificity.

In closing, we identify hallmarks of non-equilibrium gene regulation that may be amenable to experimental detection. We use our model to illustrate how simple experiments involving point mutations to activator binding sites could lead to robust signatures of non-equilibrium regulatory processes. In addition, we discuss findings that emphasize the importance of using theoretical models that account from non-cognate factor binding when interpreting experimental measurements of gene expression.

6.3 Results

6.3.1 A simple model for probing the connection between energy and information in transcription

To examine the interplay between energy and information transmission, we sought to establish models of gene circuits that capture two essential characteristics of eukaryotic transcription: the interplay between specific and general transcription factors, and the widespread presence of transcriptional bursting. Gene regulation hinges upon interactions between specific and general transcription factors. Even though salient regulatory information tends to reside exclusively in the concentration a handful of specific transcription factors targeted to binding sites within enhancers [302], these proteins are not alone sufficient to give rise to transcription. Instead, transcription and transcriptional control depend on the interplay between specific regulatory factors and general transcription factors, such as mediator [113, 200, 254, 153], RNA polymerase [290], nucleosomes [272, 215], and the various sub-units of the pre-initiation complex [227]. While these factors do not themselves carry biological information, they nonetheless constitute key molecular steps within the transcriptional cycle. Moreover, some of these processes entail the dissipation of biochemical energy, opening the door to non-equilibrium behaviors. This multiplicity of molecular players implies that gene loci may exist in multiple (and perhaps many) distinct molecular states corresponding different binding configurations of specific and general molecules (see, e.g., [15]).

Additionally, in recent years it has become apparent that transcriptional dynamics is characterized by stochastic, episodic bursts of activity interspersed with periods of transcriptional silence [26, 98, 188, 331, 290, 168] that unfold over timescales ranging from a few minutes to multiple hours [166]. Since the concentration of specific transcription factors has been shown to regulate burst dynamics [168, 331, 318], a simple model would suggest that transcriptional bursts originate from the binding and unbinding of specific transcription factors. However, recent *in vivo* measurements have revealed that activators and repressors typically bind DNA for seconds, rather than minutes or hours [168, 185]. This temporal disconnect between bursting and transcription factor binding suggests a model in which transcriptional burst cycles—corresponding to OFF $\rightarrow$ ON $\rightarrow$ OFF fluctuations at the locus (Figure 6.1C)—are determined not by transcription factor binding but, instead, by molecular reactions involving one or more general transcription factors.

Together, these observations—the interplay between specific and general regulatory factors and the disconnect between timescales of binding and bursting—support an MWC-like framework [240, 113, 272, 215] for modelling transcription wherein specific transcription factors act as effector molecules, conditioning the frequency with which the gene locus fluctuates between active and inactive transcriptional conformations. The simplest model that meets this description is one where a transcriptional activator binds to single a binding site at the gene locus and where the locus can exist in one of two conformations: and inactive (OFF) state where no mRNA is produced, and a transcriptionally active (ON) state where mRNA is produced at rate r_0.

If we neglect (for now) the binding of non-cognate transcription factors, this leads to

the four state model shown in Figure 6.1B. This model contains four basal reaction rates: the transcription factor binding and unbinding rates (k_+ and k_- respectively), and the locus activation and deactivation rates (k_a and k_i). We leave the molecular identity of the activation step unspecified, but note that it may in principle be any of the elements of the general transcriptional machinery mentioned above. In addition to these basal rates, the η terms in Figure 6.1B capture interactions between the transcription factor and the activation step. For instance, η_ab encodes the degree to which having an activator bound alters the rate of locus activation ($\eta_\mathrm{ab} > 1$ corresponds to activating activity). Without these interaction terms, activator binding would have no impact on locus activation, and, thus, no information transfer would be possible. Lastly, we note that the average rate of mRNA production in this model is simply equal to $r_0(\pi_2 + \pi_3)$, where π_i is the steady-state probability of finding the system in state i.

6.3.2 Calculating energy dissipation rates and decision times

At equilibrium, all state transitions in our model must obey the law of microscopic reversibility. The dissipation of energy along one or more of the microscopic transitions shown in Figure 6.1B lifts this strict equilibrium constraint and opens the door to novel forms of non-equilibrium gene regulatory logic. For the model shown in Figure 6.1B, the amount of energy dissipated per unit time (Φ) can be expressed as

$$\Phi = \boldsymbol{J} \ln \frac{\eta_\mathrm{ab}\eta_\mathrm{ua}}{\eta_\mathrm{ib}\eta_\mathrm{ba}}, \tag{6.1}$$

where the η terms are defined in Figure 6.1B and the net cycle flux, $\boldsymbol{J}$, encodes the expected number of *extra* clockwise ($\boldsymbol{J} > 0$) or counterclockwise ($\boldsymbol{J} < 0$) cycles the system will complete per unit time; i.e. the degree to which the microscopic transitions in the system are biased in a particular direction [133]. Φ, is a strictly positive quantity with units of $\mathrm{k_BT}$ per unit time that indicates how "near" or "far" a system is from thermodynamic equilibrium [133, 169]. For ease of comparison across different realizations of our model gene circuit, we express Φ in units of $\mathrm{k_BT}$ per burst cycle ("energy per burst").

Our central aim in this text is to understand how energy dissipation impacts rate at which gene loci transmit information and drive cellular decisions. For simplicity, we assume that c_0 and c_1 are constant in time. We also stipulate that the difference between these two concentrations (δc) is relatively small, such that $\delta c = c_1 - c_0 = 0.1c^*$, where c^* is the midpoint concentration $c^* = (c_1 + c_0)/2$. This value of δc is equivalent, for example, to concentration differences for the activator Bicoid between adjacent nuclei in early fruit fly development [117]. Figure 6.1D shows trends indicating the predicted transcriptional output of a gene locus when it is exposed to a high concentration of activator (c_1) and a low concentration (c_0). Intuitively, it should be easier to distinguish between these two scenarios when (i) the difference between average rates of transcript production (slope of the lines in Figure 6.1D) is large or (ii) the noise (shaded regions) in the accumulated output is small.

The information rate, IR, codifies this intuition, providing a quantitative measure of a gene's ability to read out and respond to different input activator concentrations. Formally,

IR is defined as the rate of change in the Kullback-Leibler divergence between our two hypotheses ($C = c_0$ and $C = c_1$) given the expected transcriptional output of our model gene circuit (see [61] for a formal definition of the Kullback-Leibler divergence). If we take the noise in the transcriptional output to be approximately Gaussian (see Appendix C.2.1), then the information rate can be expressed as

$$\mathrm{IR} = \frac{1}{2} \underbrace{\left(\frac{\delta c}{c^*}\right)^2}_{\text{input}} \times \underbrace{s^2 p^2}_{\text{output}}, \tag{6.2}$$

where IR is strictly positive and has units of information per unit time, and where s and p are the sharpness and precision of the transcriptional response, respectively, as defined in Figure 6.1A. See Appendix C.2.2 for a full derivation of this expression. We note that the native units of Equation 6.2 are natural log units ("nats"), but that, for ease of exposition, all we will give all informational quantities throughout the text in the more familiar "bits", such that IR has units of buts per burst cycles ("bits per burst"). Note also that the precision term, p, pertains solely to noise from intrinsic fluctuations between microscopic states at the gene locus and does not account for Poisson noise resulting from mRNA synthesis. In general, this shot noise is expected to be small relative to the noise from locus fluctuations for the parameter regimes considered (see Appendix C.2.3 for details).

Equation 6.2 is composed of two terms: an input component that encodes the size of the activator concentration gradient, and an output component that depends on two measurable features of the transcriptional input-output function: the sharpness and the precision (Figure 6.1A). This expression provides quantitative support for intuitions outlined above: IR can be increased both by increasing the difference between the transcription rates driven by c_1 and c_0 (by increasing the sharpness) and by decreasing the amount of noise (by increasing precision). Moreover, since both s and p can be calculated analytically from the microscopic reaction rates in model our gene circuit (see Appendix C.2.4), Equation 6.2 makes it possible to calculate and compare information rates for gene circuits with different microscopic reaction rates.

The information rate, in turn, dictates how rapidly cells can distinguish between the two activator concentrations, c_0 and c_1, based off of the accumulated transcriptional output of a gene circuit. Previous works have established that the theoretical lower limit for the time required to distinguish between c_0 and c_1 is given by

$$\langle T \rangle = \ln\left(\frac{1-\epsilon}{\epsilon}\right) \frac{1-2\epsilon}{\mathrm{IR}}, \tag{6.3}$$

where ϵ is an additional parameter that sets the error tolerance in the decision-making process representing the probability of deciding incorrectly—i.e. choosing c_1 when the true value is c_0 (or vice versa) (see Appendix C.2.5 and [69] for further details). We note the error-tolerance ϵ in Equation 6.3 is extrinsic to the gene circuit model, and depends on the nature of the cellular processes that are responding to the accumulated transcriptional output. Unless otherwise noted we will follow [69] and set $\epsilon = 0.32$, which is equivalent to an error level of "1 sigma."

6.3.3 Energy dissipation increases the rate of information transmission

With our framework in place, we sought to determine whether increasing Φ could improve the rate at which our model gene circuit drives cellular decisions between c_0 and c_1. To examine this question, we built off of methods employed in [82, 76] to develop a numerical algorithm capable of searching the space of permissible values for the transition rates that connect the four molecular states in our model (Figure 6.1B) to identify boundaries in bivariate parameter space. The method employs an evolutionary approach, iteratively updating microscopic transition rates until model realizations are found that that lie near the boundary of achievable information rates (IR from Equation 6.2) and energy dissipation rates (Φ from Equation 6.1). This approach will play an important role throughout in helping to unravel complex relationships between key transcriptional parameters. See Appendices C.2.6 & C.2.7 for further details regarding its implementation. For this analysis (and all following) we placed minimal constraints on the systems' production rate, excluding only gene circuits that were near saturating ($\overline{r} > 0.98r_0$) or basal ($\overline{r} < 0.02r_0$) activity levels.

Figure 6.2A shows the relation between IR and Φ resulting from our numerical analysis. Here, each circle represents the IR and Φ values for a single realization of our model gene circuit (Figure 6.1B), as defined by its compliment of transition rate values. Near equilibrium, our analysis reveals that gene circuits can transmit information no faster than 0.035 bits per burst (far left-hand-side of Figure 6.2A). According to Equation 6.3, this means that the best equilibrium gene circuits require at least 110 burst cycles to drive a decision between the c_1 and c_0 with an error probability of 32% (Figure 6.2B). In the developing fruit fly embryo (*D. melanogaster*), where the burst timescale (τ_b) is approximately 2 minutes [166], this translates to a decision time of 3.7 hours; far too slow to meet the time constraints imposed by the duration of early nuclear cleavage cycles (between 8 and 60 minutes [2]). Our equilibrium gene circuit would need even longer in adult nematode (*C. elegans*) and mouse (*M. musculus*) cells, where τ_b is much slower, with measurements ranging from 61 to 105 minutes ($T \geq 112$ hours) and 30 minutes to multiple hours ($T \geq 55$ hours) for worms [178] and mice [168], respectively. In each case, these timescales likely exceed decision time limits imposed by cell cycle times and mRNA half decay timescales (horizontal lines in Figure 6.2B; see Appendix C.2.8 for further details).

Can gene circuits that dissipate energy do better? Our analysis indicates that energy dissipation does indeed open the door to improved information transmission, leading to a fourfold increase in the upper IR limit from 0.0035 to 0.014 bits per burst cycle (Figure 6.2A). Moreover, we find that this performance gain is realized at biologically plausible levels of energy consumption: IR reaches its maximum non-equilibrium value at $\Phi \approx 20$ k_BT per cycle, which is approximately equivalent to the hydrolysis of one to two ATP molecules [211]. This corresponds to an energy-dependent decrease in decision time from 110 to just 29 burst cycles (red shaded region in Figure 6.2B). This is enough to meet the upper decision limit for mouse cells (Figure 6.2B). Yet there remains an absolute speed limit that no amount of energy dissipation can overcome, as shown by the empty space below the red non-equilibrium boundary in Figure 6.2B.

How can gene circuits do better? Real transcriptional systems are typically far more complex than the simple four state model from Figure 6.1B: gene enhancers typically feature multiple transcription factor binding sites [302] and, just as importantly, transcriptional activation depends on the combined action of multiple distinct general transcription factors [166]. These considerations suggests that, to overcome this speed limit, we must examine the impact of tuning two molecular "knobs": the number of specific activator binding sites in our model (N_B), and the number molecular conformations available to the gene locus (N_{LC}). For simplicity, we focus on systems in which all binding sites are identical, and likewise assume the kinetics of all molecular transitions between locus conformations to be identical. While restrictive, we will see that this simple approach is sufficient to give rise to rich, biologically salient behaviors. We note also that, while in the main text we explore the effects of varying N_{LC} and N_B separately, these two mechanisms are mutually compatible and, indeed, may act jointly in real biological systems. See Appendix C.2.9 for further details regarding the implementation of these higher-order models.

Adding binding sites improves information-energy tradeoffs

We first examined the performance of gene circuit models with multiple binding sites. In these models (as with the four state model described above), activator binding does not directly dictate transitions into and out of transcriptionally active molecular states, but, instead, functions increase to or decrease the likelihood of these transitions. Concretely, this means that each bound activator contributes an extra factor of η_{ab} and η_{ib} to impact locus activation dynamics. Models with multiple binding sites also permit cooperative interactions between activator molecules (see Appendix C.2.9 for details). With these assumptions in place, we employed our parameter sweep algorithm to explore tradeoffs between the rate of energy dissipation (Φ) and the rate of information transmission (IR) for systems with 1 to 5 activator binding sites. In all cases, we held the number of locus conformations constant at $N_{LC} = 1$ (as in Figure 6.1B).

As illustrated in Figure 6.2C, we find that adding activator binding sites has the effect of shifting the IR vs. Φ tradeoff boundary from Figure 6.2A upwards, which allows for higher rates of information transmission for the same rate of energy dissipation. This leads to significant IR gains, even in gene circuits operating near the equilibrium limit (as indicated by the vertical dashed line in Figure 6.2C), with the upper equilibrium limit increasing by a factor of twenty five from 0.0035 bits per burst cycle for $N_B = 1$ to 0.088 bits per cycle for $N_B = 5$. This means that equilibrium gene circuits with 5 binding sites need as little as 5 burst cycles to distinguish between c_1 and c_0; easily satisfying the decision time constraints of the biological systems shown in Figure 6.2B. More generally, we find that the lower decision time limit scales as the inverse of the number of binding sites squared ($T \sim N_B^{-2}$, see Figure C.1A).

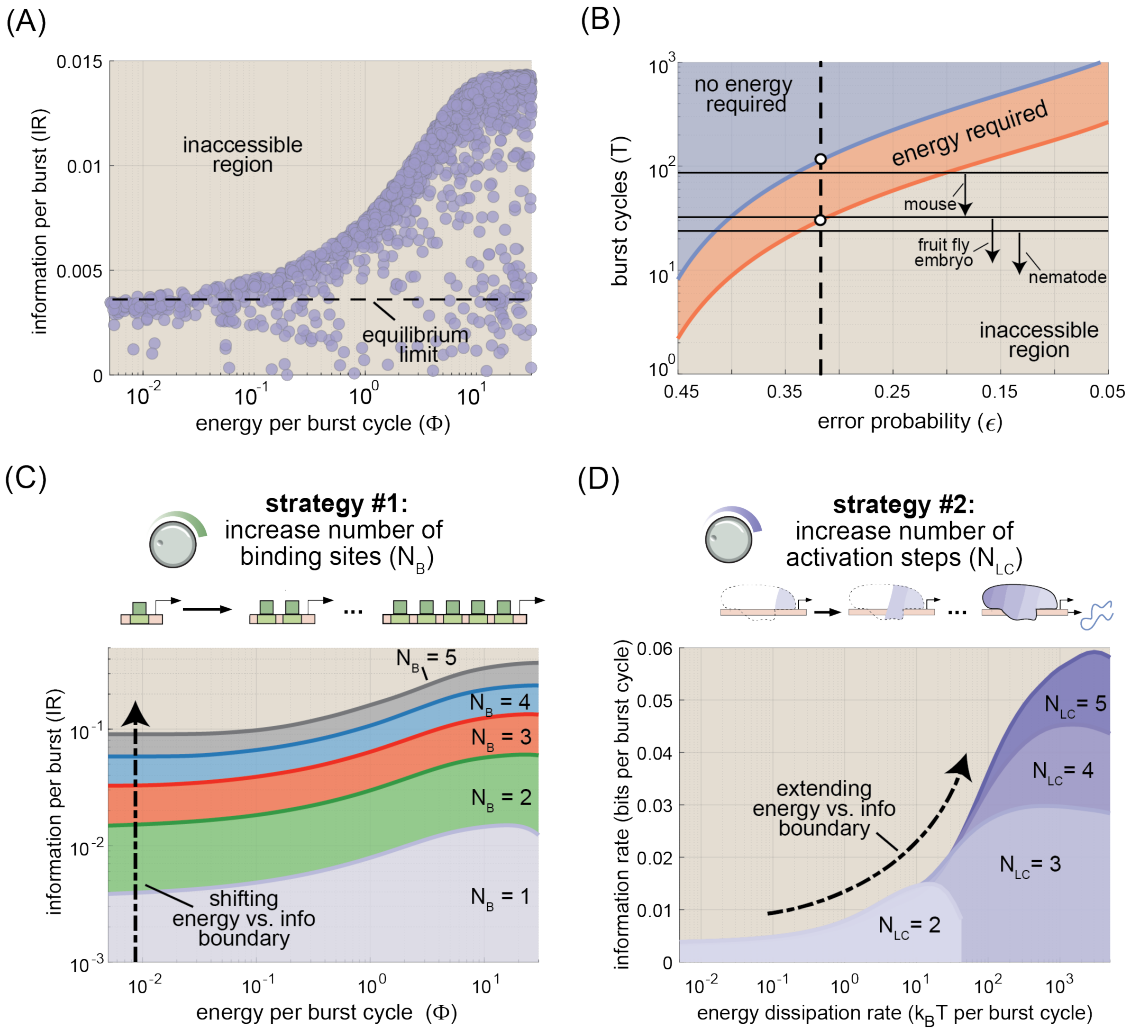

Figure 6.2: **Energy dissipation increases the rate of information transmission in gene circuits.** **(A)** Information rate (IR from Equation 6.2) as a function of energy dissipation rate (Φ from Equation 6.1) for a parameter sweep exploring all possible model realizations. Modest rates of energy dissipation can lead to a significant increase in the maximum amount of information that can be transmitted per burst cycle. **(B)** The amount of time needed to distinguish between c_0 and c_1 as a function of the probability of deciding incorrectly for equilibrium and non-equilibrium gene circuits. The decision time is given in terms of the number of transcriptional burst cycles required for a decision to be made. Note that the x-axis is arranged in order of *decreasing* error probability (i.e. increasing accuracy) from left to right. Horizontal lines indicate approximate decision times (in burst cycles) for different biological systems. **(C)** Parameter sweep results for achievable IR and Φ values for gene circuits with one to five activator binding sites. Achievable regimes for each molecular architecture are indicated as color-coded shaded regions. **(D)** Sweep results illustrating achievable IR vs. Φ regimes for gene circuits featuring two to five locus conformations.

Adding locus conformations allows gene circuits to harness higher rates of energy dissipation

What about multiple locus conformations? To pursue this question, we once again built off of the four state model shown in Figure 6.1B, this time changing the number of conformations ($2 \leq N_{LC} \leq 5$) while holding the number of binding sites fixed at $N_B = 1$ (top panel of Figure 6.2D). In all cases, we assume that only one locus conformation is transcriptionally active (ON) and that the remaining $N_{LC} - 1$ are inactive (OFF) conformations, reflecting a scenario in which the assembly of multiple distinct regulatory factors is required for productive transcription.

We conducted parameter sweeps to examine the interplay between energy dissipation and information transmission for these systems. As with adding binding sites, we find that the addition of locus conformations leads to increased rates of information. Unlike increasing N_B, however, these IR gains do not come for free. Instead, the addition of locus conformations extends the Φ-IR boundary into higher-energy regimes, allowing non-equilibrium gene circuits to achieve larger gains in IR at the expense of increased rates of energy dissipation (Figure 6.2D).

This increased IR gain means that systems with multiple locus conformations can drive decisions between c_1 and c_0 more rapidly than the simple four state gene circuit shown in Figure 6.1B. For example, we find that non-equilibrium gene circuits with five locus conformations can drive decisions nearly four times as rapidly as systems with a single step (8 burst cycles vs. 29 burst cycles; see Figure 6.2B). This 8 burst cycle limit approaches what can be achieved by an equilibrium gene circuit with five activator binding sites (5 burst cycles; see Figure 6.2E), underlining the similarity between adding binding sites at equilibrium and adding conformational states out of equilibrium. However, this parity has an energetic cost: to approach the performance of the five binding site model, the one binding site system with five conformations must dissipate at least $180 k_B T$ per burst.

6.3.4 Non-equilibrium sharpness increase drives improved information transmission

According to Equation 6.2, the energy-dependent increases in IR uncovered in Figure 6.2 must result from increased sharpness, increased precision, or some combination thereof. Thus, to uncover how energy reshapes the transcriptional input-output function to increase IR, we used our numerical sweep algorithm to examine the space of achievable sharpness and precision values for our baseline four state model (Figure 6.1B) both at and away from thermodynamic equilibrium. One challenge in comparing sharpness and precision levels across different gene circuits is that the upper bounds on both s and p on the fraction of time, $\overline{a}$, the system spends in the transcriptionally active conformation, which we allow to vary between different realizations of our model gene circuit. Thus, for ease of comparison across different model realizations, we give all results in terms of normalized sharpness and precision measures: $S = s/b$ and $P = pb$, where $b = \overline{a}(1 - \overline{a})$. See Appendix C.2.10 further details.

Figure 6.3A shows the results of our analysis, with each circle representing the S and P values for a single gene circuit realization. For systems operating at equilibrium (blue dots in Figure 6.3A), we find that both S and P are bounded by "Hopfield barriers" (dashed lines) [136, 82] with values of 1 and $1/\sqrt{2}$, respectively. These bounds place strict limits on information transmission at equilibrium, and have a straightforward interpretation: they are precisely equal to the sharpness and precision of a simple two state gene circuit with a single activator binding site, where the ON rate is concentration-dependent ($k_{\text{on}} \propto [c]$). Indeed, a detailed examination of the microscopic rates defining IR-optimized equilibrium systems—the maximally sharp and precise equilibrium gene circuits in the upper right-hand corner of the blue equilibrium region in Figure 6.3A—reveals that this apparent equivalence is literally true. Optimal equilibrium gene circuits behave as effective two state systems at the microscopic level, spending the vast majority of their time in one of two states: OFF with activator unbound (state 0 in Figure 6.1B) and ON with activator bound (state 2). Thus, at equilibrium, the added molecular complexity introduced by the conformational shift that characterizes our four-state gene circuit is more curse than blessing: the *very best* that the four-state network can do at equilibrium is to match the performance of a simpler two-state system (see Appendix C.2.11 for further details).

Energy dissipation permits gene circuits to overcome equilibrium performance bounds, increasing S by up to a factor of 2 and P by up to a factor of $\sqrt{2}$ with respect to their equilibrium limits (Figure 6.3A and B). Previous works have found that energy dissipation can increase sharpness [82]; however, to our knowledge, this is the first time that a non-equilibrium improvement in transcriptional precision has been uncovered at biologically plausible levels of energy dissipation. Yet, while energy can improve sharpness and precision individually, the absence of realizable gene circuits in the upper right hand corner of Figure 6.3A indicates that genes cannot maximize both simultaneously. This tradeoff places inexorable limits on the degree to which energy can boost IR, and stems from the fact that maximally sharp and maximally precise gene circuits require distinct and incompatible underlying molecular architectures (see Appendix C.2.12 for details).

By examining the microscopic transition rates of high-performing gene circuits returned by the parameter sweep algorithm, we can identify molecular motifs that correspond to different performance characteristics. In maximally precise gene circuits, we find that all states are (nearly) equiprobable, all clockwise rates are nearly identical in magnitude, and all counter-clockwise rates are negligible. This leads to a clock-like system with four equal steps per cycle, a design which renders the microscopic transitions as deterministic as possible and, as a result, minimizes noise from microscopic fluctuations (Figure 6.3B). In contrast, maximally sharp non-equilibrium gene circuits exhibit an all or nothing architecture similar to that observed in IR-maximized equilibrium systems. This time, however, non-equilibrium driving activator binding to regulate *both* the locus activation step and the locus inactivation step; thereby doubling transcriptional sharpness (Figure 6.3B; see Appendix C.2.13 for details).

Because sharpness and precision cannot be maximized simultaneously, gene circuits that dissipate energy must "choose" which aspect of transcriptional performance to maximize. And from the perspective of IR maximization the choice is clear: Figure 6.3A shows the

location 100 gene circuits within 1% of the maximum of 0.014 bits per cycle maximum (Figure 6.2A) in S − P phase space (gray circles). This reveals that the most informative gene circuits maximize transcriptional sharpness (S = 2) at the cost of retaining equilibrium levels of precision (P = $1/\sqrt{2}$); a result that makes sense, given that non-equilibrium systems can boost S by up to a factor of 2, while P is limited to a maximum gain of $\sqrt{2}$. As with the equilibrium case, these S and P values have an intuitive interpretation: they are simply equal to the expected sharpness and precision of a two state system, this time one in which both the ON and OFF rates are concentration-dependent. Thus, even though spending energy to overcome the constraints of detailed balance opens up a vast new space of possible regulatory schemes, we find that maximally informative non-equilibrium gene circuits exhibit an emergent simplicity, converging upon architectures in which their many molecular degrees of freedom collapse into a few effective parameters that define system behavior.

Non-equilibrium gains in sharpness drive IR increases in more complex regulatory architectures

To assess the generality of our results, we used our parameter sweep algorithm to examine equilibrium and non-equilibrium tradeoffs between sharpness and precision for more complex gene circuits with 2-5 activator binding sites and 3-5 locus conformations . In all cases, we found that energy dissipation increases the upper limits of S and P, and that—as with our simple the state model—these non-equilibrium performance gains cannot be realized simultaneously (Figure C.15A and B). For all models considered, we likewise found that the gains in information rate uncovered in Figure 6.2 are maximized by spending energy to increase sharpness, rather than precision (see Appendix C.2.14 for further details). For the case of multiple activator binding sites ($N_B > 1$), the N_B-dependent increases in the information rate shown in Figure 6.2C are a straightforward consequence of the fact that increasing the number of binding sites increases the upper sharpness limit both at and away from equilibrium (Appendix C.2.14; [113, 82]).

More surprisingly, we find that increasing the number of molecular conformations (N_{LC}) while holding the number of activator binding sites constant can, likewise, increase transcriptional sharpness in systems operating out of equilibrium. Figure 6.3C shows the range of achievable S values for non-equilibrium systems as a function of N_{LC}. We see that the upper S limit scales linearly with N_{LC}, such that $S_{neq} \leq N_{LC}$. Remarkably, this scaling is identical to the effect of adding activator binding sites at equilibrium, where $S_{eq} \leq N_B$ (Figure C.15C). This equivalence provides an intuitive explanation for our finding that systems with multiple molecular steps can drive faster decisions between c_1 and c_0 (Figure 6.2D and Figure C.1B), underscoring the fact that, when away from equilibrium, it becomes possible for activator binding at a single site to regulate multiple distinct steps of the transcriptional cycle in a way that matches the effect of multiple activator binding sites in equilibrium systems (Figure 6.3D).

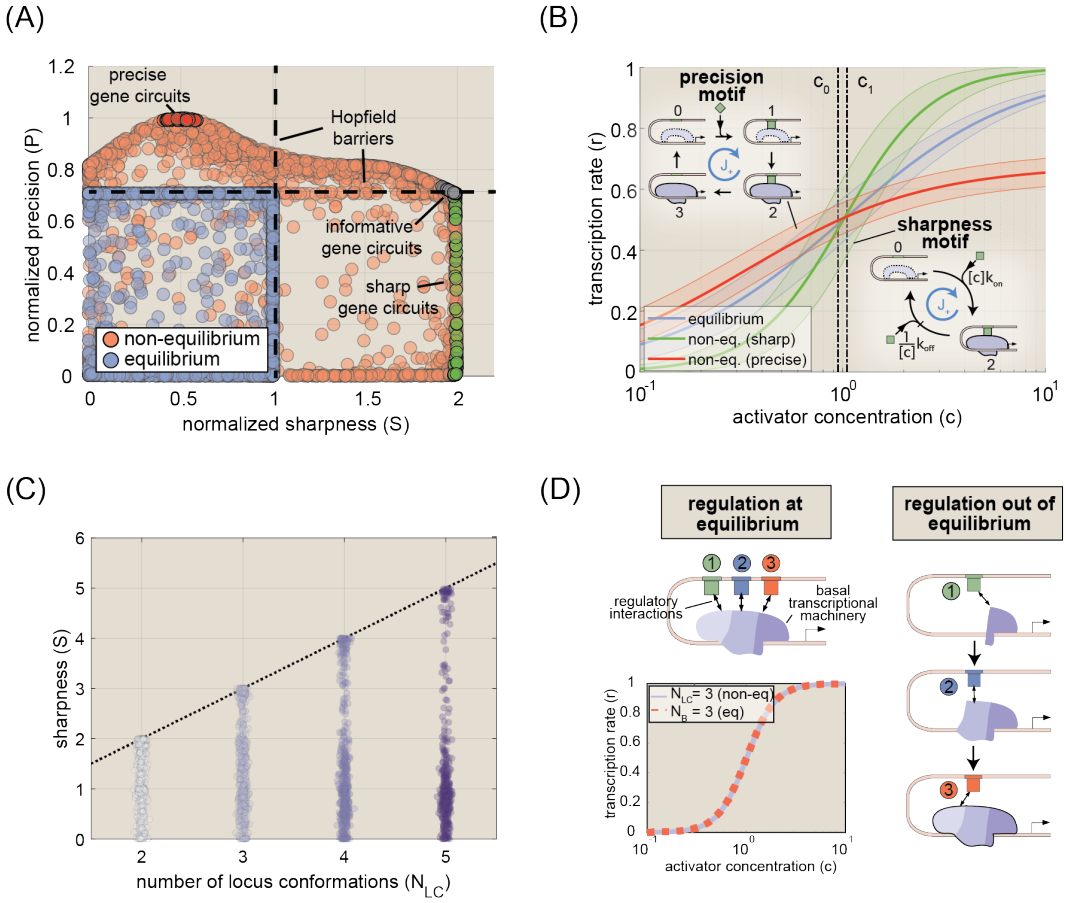

Figure 6.3: **Increased transcriptional sharpness drives increased information transmission away from equilibrium.** **(A)** Scatter plot of parameter sweep results showing the normalized sharpness and precision of 3,000 simulated gene circuits with and without equilibrium constraints. **(B)** Illustrative input-output functions for a maximally informative equilibrium gene circuit (blue) from the parameter sweeps shown in (A), as well as maximally sharp and precise non-equilibrium gene circuits (green and red, respectively). Shaded region indicates predicted noise levels in the gene expression patterns after 25 transcriptional cycles. Cartoons illustrate the molecular motifs for maximally precise and sharp non-equilibrium gene circuits. **(C)** Plot of achievable non-equilibrium sharpness levels for with 2-5 locus conformations and one activator binding site. Each circle represents a single gene circuit model. We find that normalized sharpness is bounded by the number of locus conformations. **(D)** Cartoon illustrating functional equivalence between three binding sites at equilibrium and three molecular conformations out of equilibrium. Plot shows predicted input-output functions for each case, demonstrating the equivalent levels of sharpness driven by the two strategies.

6.3.5 Energy dissipation is required for rapid cellular decisions at high non-cognate factor concentrations

So far, we have focused on the determinants of information transmission for the case when non-cognate factor binding is negligible. Yet, in any real biological setting, cells will not contain a single species of transcription factor, but many. To drive timely biological decisions, therefore, a gene circuit must not only sense and respond to its cognate transcription factor, but also efficiently filter out "irrelevant" signals from non-cognate factors. This process is inherently challenging in eukaryotes, where short DNA binding footprints lead to modest energetic differences between specific (correct) and nonspecific (incorrect) transcription factor binding events on the order of $4.6\mathrm{k_BT}$ [199], which means that non-cognate transcription factors unbind from gene loci approximately 100 times faster than cognate factors ($\alpha = k_-^w/k_- \approx 100$).

Is this hundredfold difference in binding kinetics enough to drive decisions in real biological systems? We can gain intuition for this question by examining a stripped down scenario in which a cognate and non-cognate activators must compete to bind a single binding site (Figure 6.4A). We can quantify the severity of non-cognate factor interference by dividing the fraction of time the site is bound by a cognate factor (n_c) by total fraction of time it spends bound by *either* the cognate or the non-cognate species ($n_c + n_w$). If we assume equal basal binding rates (k_+) for cognate and non-cognate species, then this fraction given by

$$p_c = \frac{n_c}{n_c + n_w} = \frac{f}{f + \frac{w}{c}}, \tag{6.4}$$

where in the second ratio on the right hand side we have introduced a new quantity, the transcriptional specificity (f), which is defined as the ratio of the (average) number of cognate and non-cognate factors bound, normalized by concentration:

$$f = \frac{w}{c}\frac{n_c}{n_w}. \tag{6.5}$$

We note that Equation 6.5, which considers competition between two activator species to bind and activate a single gene, is distinct from (and, in some sense, the complement of) specificity definitions employed in previous works, which examine the problem from the perspective of a single activator species that regulates two genes: a cognate and a non-cognate locus [272, 113] (see Appendix C.2.15 for details).

From Equation 6.4 we see that f sets the scale the severity of non-cognate factor interference. At equilibrium, f is simply equal to the affinity factor α (which we will set to 100 for concreteness), such that cognate factor binding dominates when $w/c < \alpha$ and non-cognate factors dominate when w/c exceeds α. Where do actual biological systems fall? A recent study pursuing synthetic enhancer design in the early fly embryo cited a list of 47 pertinent regulatory factors that it controlled for in order to avoid off target binding [302, 83], leading to an estimate of $w/c = 47$. Plugging this into Equation 6.4, we find that the cognate factor is predicted to be bound about 2/3 of the time in the fly embryo. At the other end

of the spectrum, we can use the genomic abundance of transcription factor proteins to estimate upper bounds on w/c values for adult nematode and mouse cells, yielding estimates of $w/c \leq 698$ and $w/c \leq 1,426$, respectively [37]. This time, Equation 6.4 predicts that cognate binding accounts for only a small fraction of total binding interactions—as little as 1/8 in worms and 1/15 in mice—suggesting that equilibrium affinity difference alone may be insufficient to cope in these cases. To examine how these high levels of interference impact the timescale of biological decisions, and to determine whether energy dissipation can improve upon this equilibrium baseline, we must extend our gene circuit model to incorporate interference from non-cognate activator binding.

To do this, we draw inspiration from [35], adding a second "wrong" activation cycle to our original four state model (Figure 6.1A) wherein the binding of a non-cognate factor to the gene locus can also induce transitions into the active conformation. This leads to the six state model shown in Figure 6.4B, where, for simplicity, we have grouped all non-cognate activators into a single concentration term: W. Here, states 4 and 5 are identical to states 1 and 2, save for the fact that a non-cognate activator species (blue circle) is bound, rather than the cognate activator (green square). For notational convenience, the unbinding rates of the non-cognate activator k_-^w are written as the specific unbinding rate k_- multiplied by an affinity factor $\alpha = k_-^w/k_-$. As stipulated above, we take $\alpha = 100$ throughout the main text; however, we note that our results are robust to the choice of α.

With our model defined, we employed parameter sweeps to examine the upper limits on information transmission as a function of the ratio of wrong-to-right activator concentrations (w/c). Throughout this process, the cognate factor concentration was held fixed at $C = c^*$, such that w was the only variable concentration parameter. Figure 6.4C illustrates the results of our analysis, plotting the range of achievable information rates as a function of the relative wrong factor concentration. In keeping with our calculations above, our results reveal that the rate of information transmission at equilibrium drops precipitously once w/c exceeds α, with an upper limit equal to the maximum rate of information transmission when $w = 0$ times the cognate fraction squared: $\text{IR} \leq \text{IR}_0^{\text{eq}} p_c^2$ (dashed blue line in Figure 6.4C). Away from equilibrium, the picture is more complex. When $w/c < \alpha$, the upper IR scales as p_c^2 (dashed red line); however, the scaling becomes less severe once w/c exceeds α, approaching a linear relationship. At a practical level, this shift to linear scaling has the effect of making non-equilibrium gene circuits more robust to high non-cognate factor concentrations, with the relative IR from energy dissipation increasing from a factor of 4 when $w \approx 0$ all the way to a factor of 1000 when $w/c = 10^5$ (Figure 6.4C, inset). More fundamentally, it suggests that a qualitative change occurs in the way that energy is used once $w/c > \alpha$; an observation that we will return to in Section 6.3.6.

We next used Equation 6.3 to calculate the amount of time required for a cell to decide between concentrations c_0 and c_1 of the cognate activator species for different values of w/c, starting with gene circuits constrained to operate at equilibrium. As in Figure 6.2B, we compared our model's performance to the decision time limits for different biological systems, this time placing each system in its appropriate place along the w/c axis. In all systems considered, gene circuits generally have a few tens of burst cycles over which to transmit information, with no system exceeding 100 bursts (black error bars in Figure 6.4D). This is

significantly faster than can be achieved by our simple six state model with one binding site and two locus conformations at equilibrium, even when $w = 0$ (see purple shaded region in Figure 6.4D).

Next, we sought to determine if equilibrium gene circuits with multiple sites could do better. As illustrated in Figure 6.4D, these results indicated that equilibrium gene circuits with 3 or more activator binding sites (red, blue, and gray regions) are sufficient to drive timely decisions in "low intereference" systems such as the early fruit fly embryo. However, as with the one binding site case, we once again observe a precipitous decline in performance once $w/c > \alpha$. Indeed, the best equilibrium model ($N_B = 5$) can drive decisions in no fewer than 1,060 burst cycles—the equivalent of at least 530 hours (3 weeks) for mouse cells—when $w/c \approx 1400$ (the upper limit for mice). This is over an order of magnitude too slow for the mouse system's decision time limit of 86 bursts (Figure 6.4D). Moreover, meeting this time limit is not merely a matter of adding a handful of additional sites: our analysis suggests that at least 17 activator binding sites are needed at equilibrium (see FigureC.2C). Such a number is conceivable for eukaryotic enhancers, but this analysis nonetheless emphasizes that equilibrium systems—even those with biologically salient numbers of binding sites—struggle to achieve realistic decision times in the presence of significant non-cognate factor interference.

How do non-equilibrium gene circuits fare? Already for the simple model shown in Figure 6.4B, we found a marked non-equilibrium improvement, decreasing the decision time from 24,000 burst cycles at equilibrium (Figure 6.4D) down to 1,500 (Figure 6.4E) for $w/c \approx 1,400$. We next used our parameter sweep algorithm to examine the impact of increasing the number of molecular conformations ($N_{LC} > 2$) while holding the number of binding sites fixed at one. This revealed further non-equilibrium improvements, particularly at high w/c: whereas the two conformation system could go no faster than 1,500 cycles when $w/c = 1,400$, we find that gene circuits with three locus conformations can drive decisions between c_0 and c_1 in as little as 104 bursts (Figure 6.4E); a full order of magnitude better than equilibrium genes with five binding sites (Figure 6.4D). Adding a fourth conformation improves this bound further to 83 burst cycles; just good enough to dip below the 86 burst limit for the Mouse system. Moreover, this $N_{LC} = 3$ system exhibits remarkable robustness to non-cognate factor interference, sustaining this same level of performance all the way up to $w/c \approx 10^4$ (Figure 6.4E).

These results suggest that, in biological contexts where the ratio of wrong-to-right activator concentrations exceeds the intrinsic binding affinity difference between them (α), energy dissipation effectively shifts from being a biochemical shortcut for transmitting information faster to, increasingly, becoming a necessary precondition for driving cellular decisions within biologically salient timescales. Yet Figure 6.4E also reveals one binding-site systems have a performance ceiling: to further improve, non-equilibrium gene circuits likely require multiple molecular steps ($N_{LC} \geq 2$) *and* multiple activator binding sites ($N_B \geq 2$).

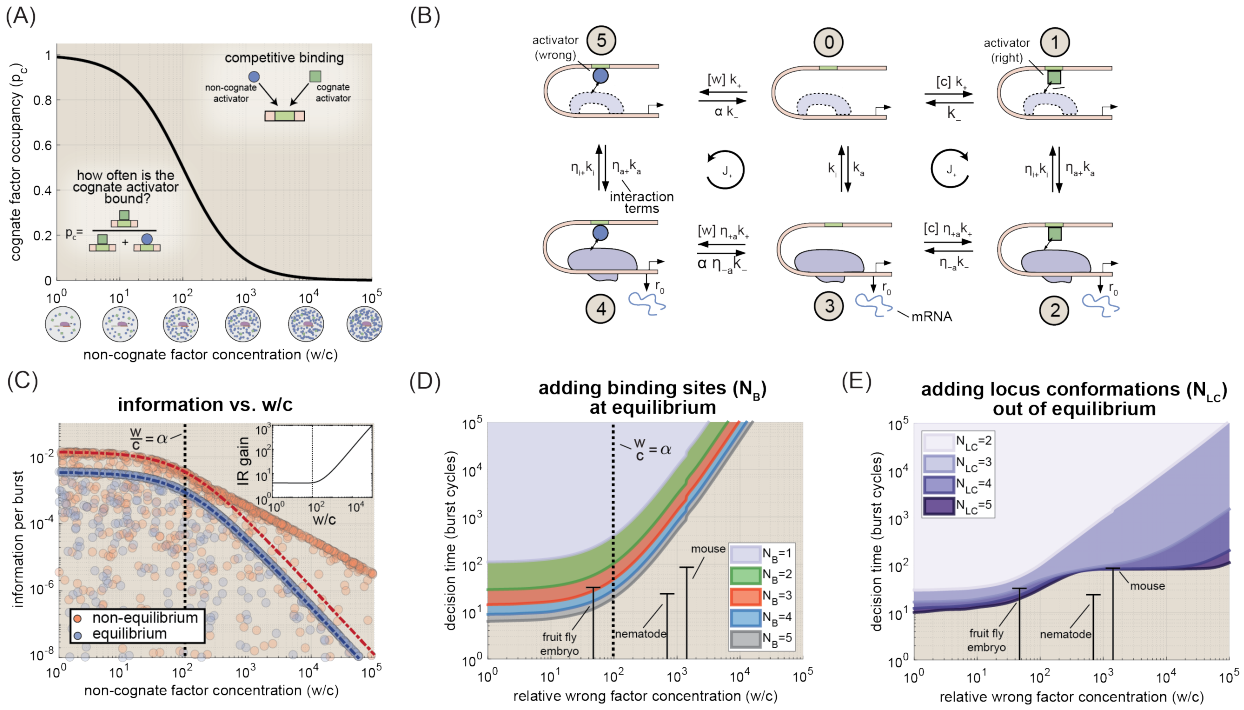

Figure 6.4: **Energy dissipation is key to driving cellular decisions in the presence of non-cognate factor interference.** **(A)** Plot of cognate factor occupancy at a single binding site as a function of relative non-cognate factor concentration. **(B)** Incorporating non-cognate activator binding leads to a six state model that features both a right and a wrong activation pathway. **(C)** Numerical results for the maximum achievable information rate for equilibrium (blue circles) and non-equilibrium (red circles) gene circuits with one activator binding site and two locus conformations (illustrated in (B)) as function of the relative concentration of non-cognate activators w/c. Blue dashed line indicates upper IR bound at equilibrium. Red lines indicate non-equilibrium IR bounds (see main text). Vertical dashed line indicates point where non-cognate factor concentration (w) equals the the cognate factor concentration multiplied by the affinity factor (αc). **(D)** Parameter sweep results showing decision times for equilibrium gene circuits with 1-5 activator binding sites as a function of w/c. **(E)** Decision times for non-equilibrium gene circuits with variable 2-5 locus conformations. (All results assume $\alpha = k_-^w/k_- = 100$. All decision time quantities assume $\epsilon = 0.32$.)

6.3.6 Information limit is defined by non-equilibrium tradeoffs between specificity and intrinsic sharpness

To uncover the mechanism driving the large non-equilibrium increases in IR uncovered in Figure 6.4, we broke the IR gain depicted in the panel inset of Figure 6.4C into its constituent components. Figure 6.5A plots the relative non-equilibrium gains in S and P (S/S^{eq} and P/P^{eq}) as a function of w/c for information-maximizing realizations of the six state gene circuit model shown in Figure 6.4B. The plot reveals that IR-maximizing gene circuits consistently utilize energy to drive sharpness above its equilibrium limit (S/S^{eq}¿1), while precision is maintained at or below its equilibrium limit ($P/P^{eq} \lesssim 1$). More to the point, we see that the *degree* to which non-equilibrium gene circuits amplify S increases dramatically as w/c grows larger, from a factor of 2 when $w/c \approx 1$ to a factor of 100 when $w/c \approx 10^4$ (Figure 6.5A). Thus, they key to understanding how energy increases IR at large w/c lies, once again, in understanding transcriptional sharpness.

We find that the upper non-equilibrium limit on S can be expressed as a function of the specificity (f), such that

$$S \leq \underbrace{\frac{f}{\frac{w}{c} + f}}_{\substack{\text{specificity} \\ \text{factor } (p_c)}} \times \underbrace{S_0}_{\substack{\text{intrinsic} \\ \text{sharpness}}}, \tag{6.6}$$

where we see that the S bound breaks naturally into two pieces, the specific bound fraction, p_c (defined in Equation 6.4), and a quantity that we will term the intrinsic sharpness (S_0), which is defined as a gene circuit's normalized sharpness absent non-cognate factor binding; that is, when $w = 0$.

To probe the interplay between intrinsic sharpness and specificity, we employed parameter sweeps for the six state system in Figure 6.4A at and away from equilibrium (Figure 6.5B). At equilibrium, this analysis indicated that $S_0 \leq 1$ (consistent with Figure 6.3A) and confirmed that $f^{eq} = \alpha$, just as for the case of the equilibrium binding scenario considered at the start of Section 6.3.5. Indeed, we find that this equality applies for *all* gene circuits operating at equilibrium irrespective of the number of binding sites , placing strict limits on information transmission at equilibrium when w/c is large.

Away from equilibrium, systems can overcome these constraints, achieving up to a two-fold increase in S_0 and increasing specificity by up to an additional factor of α to reach an upper limit of α^2 (Figure 6.5B). This hundredfold increase in f is consistent with the gain in S observed in Figure 6.5A, suggesting that the sharpness gain at high w/c arises from non-equilibrium increases in specificity. But why not spend energy to increase sharpness as well and achieve $S/S_0^{eq} = 2 \times \alpha = 200$? The simple answer is that non-equilibrium gains in intrinsic sharpness and specificity cannot be realized simultaneously. Instead, our analysis reveals a steep tradeoff between specificity and intrinsic sharpness away from equilibrium, with the maximum value of $S_0 = 2$ only realizable when specificity is at its equilibrium level ($f = \alpha$) and vice versa (Figure 6.5B). We find that the bound describing this tradeoff (red dashed line in Figure 6.5B) a simple analytic form, which allows us to express S as a function

of the specificity, f:

$$\mathrm{S} \leq \underbrace{\frac{f}{\frac{w}{c}+f}}_{\substack{\text{specificity}\\ \text{factor } (p_c)}} \times \underbrace{\left(\frac{\alpha^2+\alpha f-2f}{\alpha f-f}\right)}_{\substack{\text{intrinsic}\\ \text{sharpness } (\mathrm{S}_0)}}, \tag{6.7}$$

where we assume that $\alpha \leq f \leq \alpha^2$. See Appendix C.2.16 for a derivation of Equation 6.7. As with the non-equilibrium trade-offs between sharpness and precision, this incompatibility stems from the fact that sharpness and specificity require distinct and incompatible underlying molecular architectures. And although we've focused on the simple model shown Figure 6.4B, we find similar non-equilibrium tradeoffs between that f and S_0 for more complex molecular architectures (Figure C.3B). Thus, we conclude that these specificity gains come at the cost of diminished intrinsic sharpness.

The inexorable tradeoff between S_0 and f illustrated in Figure 6.5B means that gene loci must "choose" between allocating energy to maximize intrinsic sharpness and allocating energy to maximize specificity. To examine how this the concentration of non-cognate factors shapes this tradeoff, we took IR-maximizing non-equilibrium gene circuits spanning the relevant range of w/c values for systems with 2-5 locus conformations and calculated S_0 and f for each. Figure 6.5C shows the results of this analysis, illustrating the relative non-equilibrium gains in intrinsic sharpness and specificity, respectively, for IR-maximizing gene circuits as a function of w/c.

We find that the relative non-cognate factor concentration, w/c, defines a shifting optimality landscape. At low non-cognate factor concentrations, maximally informative gene circuits spend energy exclusively to maximize sharpness ($\mathrm{S}_0 > \mathrm{N_B}$ for all systems on the left-hand side of Figure 6.5C) at the cost of equilibrium levels of specificity ($f/\alpha = 1$). Thus our model predicts that at low levels of non-cognate factor interference—as would be experienced, for instance, in developing fruit fly embryos—non-equilibrium mechanisms are not required to buffer against non-cognate factor interference, and allocating energy to maximize intrinsic sharpness constitutes the optimal regulatory strategy. However, once w/c surpasses the affinity factor α, IR maximization starts to disfavor sharpness (see decreasing S_0 near $w/c = 10^2$ in Figure 6.5C) and, instead, increasingly depends on enhancing specificity to non-equilibrium levels (Figure 6.5C). Taken together, these results indicate that the optimal molecular strategy for transmitting information is not fixed, but will change according to a scale that is set by the relative magnitudes of level of non-cognate factor interference, w/c, and the kinetic binding differences between cognate and non-cognate factors, α.

6.3.7 Experimental signatures of non-equilibrium processes in transcriptional regulation

So far, we have demonstrated that energy dissipation can, in principle, increase the rate of information transmission in gene circuits. However, determining whether gene circuits actually leverage energy dissipation to do so constitutes a different challenge. In this section, we examine how simple experiments can serve to identify signatures of non-equilibrium

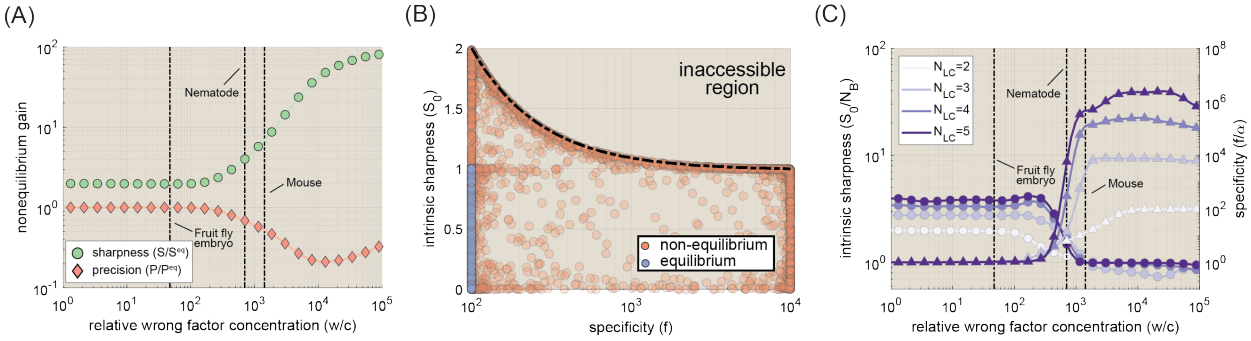

Figure 6.5: **A shifting optimality landscape for information transmission. (A)** Plot showing non-equilibrium gains in sharpness and precision as a function of w/c for six state ($N_B = 1, N_{LC} = 2$) gene circuits found to drive maximum information rates. IR−maximizing gene circuits are drawn from optimal systems uncovered in the parameter sweeps from Figure 6.4E. Values above 1 indicate that the system is dissipating energy to enhance performance. Black line indicates "break even" point where non-equilibrium value is equal to the equilibrium maximum. See Figure C.3A for results for systems with $N_{LC} > 2$. **(B)** Plot of numerical results for tradeoffs between intrinsic sharpness (S_0) and specificity (f) for equilibrium and non-equilibrium networks (blue and red circles, respectively). Note that equilibrium gene circuits have no horizontal dispersion because all are constrained to have $f = \alpha$. Black dashed line indicates bound predicted by Equation 6.7. **(C)** Plot of non-equilibrium gains in intrinsic sharpness and specificity for IR-maximizing gene circuits as a function of w/c. Values above 1 indicate that the system is dissipating energy to enhance sharpness or specificity. Note that left and right axes have different scales. (α was set to 100 for all plots shown.)

performance in real biological systems. For clarity of exposition, we focus on the six state gene circuit shown in Figure 6.4B. Yet while the quantitative details will vary for gene circuits with different molecular architectures, this simple case study nevertheless serves to illustrate a broadly applicable set of experimental and analytical approaches that can be used to test whether energy is being harnessed to enhance transcriptional performance in real biological systems.

Recent works have shown that strict equilibrium limits on transcriptional sharpness can be calculated if the number of activator binding sites is known, suggesting that sharpness might serve as an accessible signature of non-equilibrium regulatory mechanisms [82, 231]. Does this conclusion hold when we consider the impact of non-cognate factor binding? Equation 6.7 predicts that the upper S limit should decrease as w/c increases (dashed lines in Figure 6.6A) and numerical parameter sweeps of S vs. w/c confirm this fact, (Figure 6.6A). This w-dependence must be accounted for in order to correctly interpret experimental measurements.

For instance, consider the case when $w/c = 10^3$ (dashed vertical line in Figure 6.6A); a plausible value for, e.g., mammalian systems [94, 35, 37]. Our model predicts that the

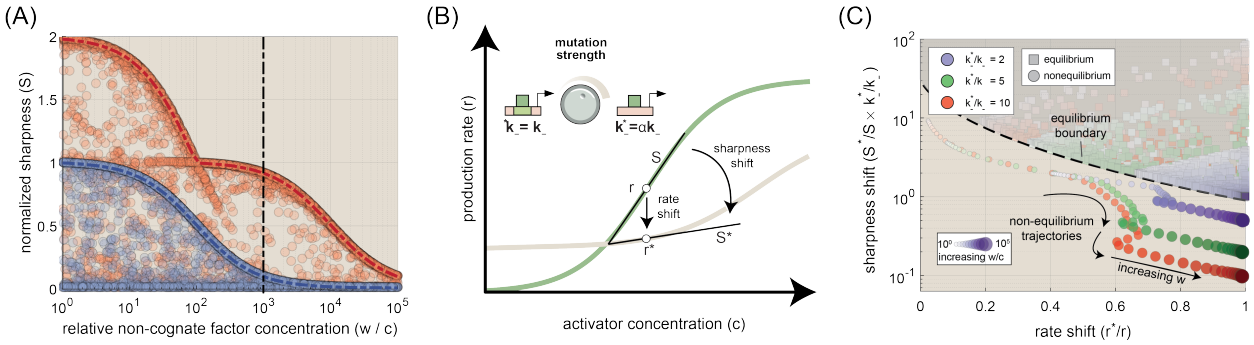

Figure 6.6: **Experimental signatures of non-equilibrium processes in transcriptional regulation. (A)** Observed sharpness as a function of w/c for equilibrium (blue circles) and non-equilibrium (red) gene circuits. Dashed line indicates point where $w/c = 10^3$. **(B)** Illustration of proposed binding site perturbation experiments. Reducing site specificity is predicted to reduce both observed sharpness, S, and the mRNA production rate, r. **(C)** Phase space plot of predicted sharpness shift versus rate shift for equilibrium (squares) and non-equilibrium (circles) gene circuits at three different perturbation strengths. Shading indicates w/c value (darker shades correspond to higher values). Additionally, circle size indicates w/c magnitude for non-equilibrium circuits. We see that, regardless of non-cognate concentration and perturbation strength, non-equilibrium systems do not cross the equilibrium boundary (dashed line).

maximum achievable S value for non-equilibrium gene circuits is 0.91, far exceeding the *true* equilibrium sharpness limit of 0.09 when $w/c = 10^3$ (blue dashed line in Figure 6.6A), but it falls *below* the the "naive" bound of S = 1 that one would calculate if not accounting for w (see also Figure C.4A). Thus, failing to account for non-cognate factor interference could mask strong non-equilibrium signatures in the data, highlighting the importance of incorporating regulatory cross-talk into models of transcription. However, accurately measuring w/c may be challenging (if not impossible) in many experimental settings, since w comprises the aggregate activity of all non-cognate activator species.

In light of this challenge, we propose a complementary experimental approach to search for signatures of non-equilibrium gene regulation that is more robust to the experimental uncertainty regarding w/c. As illustrated in Figure 6.6B, this method involves measuring changes to gene expression at $C = c^*$ that result from point mutations to the activator binding site (see, e.g., [248, 231]), and which thereby lead to a faster unbinding rate, k_-^*, for cognate activators ($k_-^*/k_- > 1$). Whereas w/c may be difficult to estimate in many biological contexts, robust algorithms exist to predict binding energies from DNA binding footprints [174], allowing for accurate predictions of how much a particular mutation will perturb the relative binding kinetics of the specific activator species. We employ two metrics to quantify the resulting change in gene expression: the fold change in the mRNA production rate (r^*/r) and in the normalized sharpness (S^*/S), with each defined as the quantity corresponding to the mutated binding site divided by its corresponding wild-type value (Figure 6.6B).

Overall, we find that IR-optimized non-equilibrium gene circuits are highly sensitive to changes in cognate activator specificity, and that this sensitivity can be used to probe for non-equilibrium behavior. At low w/c levels ($w/c \lesssim 10^3$), mutated non-equilibrium circuits exhibit larger shifts in their rate of transcription than can be achieved at equilibrium (Figure C.4B). Meanwhile, IR-optimized non-equilibrium systems experience a substantially larger drop in sharpness when $w/c > 10^3$ than even maximally sensitive equilibrium circuits (Figure C.4C). as a result, when combined, S^*/S and r^*/r define a perturbation response space in which non-equilibrium gene circuits that transmit information at optimal (or near optimal) levels are completely disjoint from equilibrium systems. This is illustrated in Figure 6.6C, which plots our model's predictions for the sharpness fold change ($k\_S^*/Sk\_^*$) vs. r^*/r under three different binding site perturbation strengths for equilibrium and non-equilibrium gene circuits (squares and circles, respectively). Results are shown for non-equilibrium systems subjected to w/c ranging from 1 to 10^5 (increasing circle size and shading indicate larger w/c). Despite the wide range of perturbation strengths and non-cognate factor concentrations examined, we see that non-equilibrium systems never cross the equilibrium boundary (dashed line). Thus, by measuring S^*/S and r^*/r, researchers can obtain clear-cut signatures on non-equilibrium regulation, even if the parameter w/c is unknown.

6.4 Discussion

Gene regulation is central to cellular function. Yet, despite decades of biochemical and genetic studies that have established a reasonably complete "parts list" of the molecular components driving eukaryotic transcription [161], and despite recent advances in our ability to track how the these pieces assemble in space [227] and in time [166, 59, 179], we nonetheless lack quantitative models that can predict how transcriptional control emerges from molecular interactions at the gene locus. Thermodynamic models of transcription, which assume that gene circuits operate at equilibrium, have previously been employed with considerable success in the context of bacterial systems [241]. However, the presence of ATP-dependent processes—such as chromatin remodeling [329], PIC assembly [289], and Pol II initiation [321]—within the eukaryotic transcriptional cycle suggests that equilibrium models may be insufficient to capture how eukaryotic gene circuits sense and respond to input transcription factor concentrations. Thus, there is an urgent need for theoretical frameworks that can probe how non-equilibrium mechanisms reshape the transcriptional input-output function and, ultimately, redefine the limits of transcriptional control.

Here, we have employed simple kinetic models of transcription to investigate how energy dissipation within the transcriptional cycle impacts the rate at which a gene circuit can drive cellular decisions. We used the rate of information transmission as a quantitative lens to examine how energy-dependent changes to measurable features of the transcriptional input-output function (Figure 6.1A) impact the timescale of cellular decision-making (Figure 6.1C and D). We found that biologically plausible rates of energy-dissipation can drive significant gains in IR, but that the regulatory mechanisms underlying these gains change

depending on the level of interference from non-cognate factor binding (w/c).

When w/c is below the intrinsic affinity difference between cognate and non-cognate factors (α), IR is maximized by harnessing energy to push transcriptional sharpness beyond its equilibrium limits (Figure 6.5C). Conversely, once $w/c > \alpha$, IR-maximization begins to instead demand that genes harness energy to increase transcriptional specificity in order to buffer against non-cognate factor interference (Figure 6.5C). And, whereas equilibrium gene circuits with sufficiently many binding sites (3 or more in our model) can drive biologically plausible decision times when $w/c < \alpha$ (Figure 6.4D), equilibrium gene regulatory mechanisms break down when w/c exceeds this limit, suggesting that energy dissipation may be indispensable in systems where non-cognate factor interference is sufficiently large (compare Figure 6.4D and E).

Performance tradeoffs dictate limits of information transmission away from equilibrium

A central result of this work is that, although energy dissipation can increase transcriptional sharpness, precision, and specificity *individually*, these gains cannot be realized simultaneously. Gene circuits cannot "have it all," even when operating away from equilibrium, and this places inherent limits on how rapidly genes can transmit information. When non-cognate factor binding is negligible, we showed that IR is dictated by a tradeoff between sharpness (S) and precision (P). And although previous works have established that energy expenditure can boost sharpness [82, 231] and, to a lesser extent, that it can suppress transcriptional noise [249], to our knowledge, we are the first to report a tradeoff between these two features of the transcriptional input-output function. As a result of this tradeoff, gene circuits must "choose" whether to spend energy to enhance sharpness or precision and, for all models considered, we discovered that the information rate was maximized by systems that boosted transcriptional sharpness (not precision) above its equilibrium limit (Figure 6.3A, Figure C.15A and B).

In a similar fashion, our analysis revealed that non-equilibrium gains in specificity and sharpness cannot be realized simultaneously (Figure 6.5B and Figure C.3B). Intuitively, this incompatibility arises from the fact that intrinsically sharp systems are tuned to amplify the concentration-dependent activator binding rates, whereas specific systems amplify differences in *unbinding* rates between cognate and non-cognate activator species. As noted above, our model predicts that w/c defines a shifting optimality landscape, wherein non-equilibrium gene circuits that maximize intrinsic sharpness drive the fastest decisions when $w/c \leq \alpha$, but where the optimal strategy begins to shift from increasing sharpness to, instead, engaging in activator proofreading when $w/c > \alpha$ (Figure 6.5C). The potential for this kind of context-dependent shift from sharp to specific gene circuits was recently noted in [113], although in this case sharpness was only investigated at its equilibrium limit. Here, we provide quantitative predictions for IR-maximizing how gene circuits navigate this sharpness-specificity tradeoff.

Conformational flexibility amplifies non-equilibrium performance gains

Another key finding of this work is that the presence of multiple locus conformations can amplify non-equilibrium gains in transcriptional sharpness (Figure 6.3C). This has the somewhat surprising implication that, away from equilibrium, the number of activator binding sites does not dictate an absolute upper limit on sharpness and, therefore, on the rate of information transmission. To achieve this sharpness gain, multi-conformation gene circuits mimic equilibrium genes with multiple binding sites, leveraging multiple successive regulatory interactions distributed across time to achieve the same effect as multiple binding sites distributed in space (Figure 6.3D). Our result is evocative of a recent study by Biddle and colleagues [16], in which the authors examined *equilibrium* gene circuits and demonstrated that systems with multiple conformational degrees of freedom could achieve sharper, more flexible, transcriptional input-output functions; although these systems still adhered to the fundamental equilibrium bound of $S \leq N_B$. Thus, our findings serve to further emphasize potential benefits of the conformational complexity of the eukaryotic gene cycle.

We also found that gene circuits with multiple locus conformations can realize dramatic increases in transcriptional specificity, such that $f \leq \alpha^{N_{LC}}$. This finding is in accordance with previous results in the kinetic proofreading literature showing that multiple molecular steps can function to enhance the specificity of molecular processes [224], and extends the findings of a recent work that examined transcriptional specificity in systems with up to three locus conformations [272]. Yet there exists an important asymmetry between sharpness and specificity: whereas the addition of activator binding sites can increase S at equilibrium, energy dissipation constitutes the *only* route to increasing f above the intrinsic affinity factor α. Thus, the decisive role that energy dissipation plays in facilitating timely decisions when w/c follows from the fact that functions to overcome a fundamental limitation of eukaryotic gene circuits—the lack of binding specificity—that no equilibrium mechanism can address.

Equilibrium regulatory schemes may be sufficient in many real biological systems

Yet, while activator proofreading may be critical when w/c is large, our analysis suggests that it is unlikely to constitute a universal constraint on gene regulatory architectures. Indeed, we find that even relatively simple equilibrium architectures with 3-5 binding sites should suffice to drive timely cellular decisions in "low interference" systems such as the fruit fly embryo (Figure 6.4D). Moreover, while simple estimates based on genomic transcription factor abundances suggest that many eukaryotic systems have the potential (at last in principle) to fall above the $w/c = \alpha$ interference limit, these estimates likely represent upper bounds on w/c, since it is well established that different cell types selectively express distinct subsets transcription factors [48, 176, 131]. In addition, we note that the relative size of the concentration difference between c_1 and c_0 ($\delta c/c$) plays a key role in dictating the rate of information transmission (Equation 6.2), and will vary across different biological contexts. It would thus be interesting to use the quantitative tools presented in this work to enumerate the space of viable equilibrium and non-equilibrium gene circuit architectures

for specific biological systems in which the relative magnitudes of w/c and $\delta c/c$ are well established.

On a more conceptual note, it is thought-provoking to regard of the $\delta c/c$ term in Equation 6.2 as measuring the size of an input chemical potential driven by some upstream signaling source: if the external environment does enough work to drive c_1 and c_0 far apart, then equilibrium regulatory architectures may suffice to distinguish them. Correspondingly, energy dissipation in gene circuits could compensate for weak input gradients. This framing points to the importance of modelling information transmission along relays of genes, rather than single gene loci (see, e.g. [292, 253]). For instance, the problem of non-cognate factor interference considered here becomes substantially more pressing if it impacts multiple steps of a signalling cascade.

Different frameworks for examining the impact of non-cognate factor binding

In considering the impact of non-cognate factor binding, we drew inspiration from a previous study by by Cepeda-Humerez and colleagues that examined competition between cognate and non-cognate transcription factors to bind and activate a single gene locus [35]. This formulation of the problem is distinct from the approach taken in two recent works, which addressed the problem of specificity from the perspective of a single activator species that interacts with two different gene loci: a cognate (with specific binding sites) and a non-cognate locus (without) [272, 113]. While both have proven fruitful, we favor the "single locus" approach pursued here, since it captures the effects of competitive binding between different species, which are an unavoidable reality of the crowded cellular environments.

Moreover, we find that this shift in perspectives has meaningful consequences for our understanding of how off-target binding impacts gene regulation. A previous study found that the equilibrium limit of $f = \alpha$ could only be achieved at the cost of high levels of transcriptional noise [272]. Yet, we find that this tradeoff evaporates once competitive binding between cognate and non-cognate factors is considered, since, in this case f is fixed at α (Figure 6.5B). We also find that the upper limits of transcriptional sharpness decrease as w/c increases (Equation 6.7 and Figure 6.6A). Previous studies have pointed transcriptional sharpness as a key potential indicator of non-equilibrium optimization [82, 231]. Our analysis reaffirms this idea, but, crucially, reveals that the relative concentration of non-cognate factors (w/c) must be accounted for in order to accurately assess whether a particular system is performing above the equilibrium limit (Figure 6.6A and B). For instance, an sharpness of 0.9 falls below the equilibrium limit for the six state gene circuit shown in Figure 6.4B when $w/c \approx 1$, but is an order of magnitude above the limit when $w/c \approx 10^3$ (Figure 6.6A).

Future directions

In this work, we have considered tradeoffs between sharpness, precision, and specificity. Yet, beyond these "real time" performance characteristics, our work also has implications for how gene regulatory architectures evolve. Specifically, the efficacy of the mutagenesis studies discussed in Section 6.3.7 depends on our somewhat unexpected finding that

IR-maximizing non-equilibrium gene circuits are significantly *more* sensitive to mutation-induced reductions to binding site affinity than their equilibrium counterparts (Figure 6.6C and Figure C.4B and C). This sensitivity points to an intriguing tension between the extant system's performance on the one hand (as quantified by IR), and its robustness random mutations in regulatory sequence on the other.

This work also represents a natural starting point for examining the regulatory impact of neutral DNA sequences. While we have considered gene loci with varying numbers of *specific* activator binding sites, real enhancers also contain significant stretches of "neutral" DNA that contains no binding sites, as well as weak activator sites that fall below typical thresholds used to identify specific sites [302, 267]. This focus on specific sites is widespread in theoretical studies of transcription [82, 231, 35, 166], despite the fact that experimental studies have established the importance of weak binding sites in the context of certain genes [267, 64, 86]. Moreover, recent efforts at synthetic enhancer reconstitution have pointed to the importance of supposedly neutral stretches or regulatory DNA [302], and it seems plausible from a theoretical perspective that these stretches, where cognate and non-cognate activator species bind with equal affinity, could have important effects on the input-output function in systems where $w/c > \alpha$. As a starting point, we propose that the kinetic models utilized here could readily be extended to feature some combination of specific and neutral sites. More ambitiously, the field would benefit from the introduction of continuous, rather than discrete, theoretical models that admit non-equilibrium phenomena while accounting for the reality that transcription factors interact with a continuum of sites along enhancer DNA.

Ultimately, however, the key to unraveling the molecular mechanisms by which gene loci sense and respond to transcription factor concentrations lies in the coupling of theoretical models with careful experimental measurements. To this end, we advocate for the expanded use of theoretically tractable synthetic enhancer systems where the number and identity of binding sites are well established, and where intervening DNA sequence is carefully engineered to minimize binding specificity (e.g. using SiteOut [83]). Several recent studies constitute promising initial steps in this direction [248, 231, 302, 158]. Additionally, synthetic transcription factor systems, which can be engineered to act orthogonally to endogenous regulatory networks, represent an intriguing experimental platform for investigating questions relating to transcriptional specificity [152, 63]. Lastly, statistical methods that infer how transcription factor concentrations impact the kinetics of the transcriptional cycle (e.g. [331, 168, 54, 28]) have shown promise as a way to connect macroscopic experimental measurements to theoretical models of the microscopic processes driving transcription. Looking ahead, holistic research efforts that integrate cutting-edge experiments, statistical methods, and theory will be key to bridging the as yet yawning gap between enhancer sequence and gene regulatory function.

Chapter 7

Fast, flexible inference of transcriptional dynamics with *burstMCMC*

Abstract

Transcriptional bursting is a ubiquitous feature of eukaryotic transcription. Yet, while the field is seeing increasingly widespread adoption of live imaging tools such as the MS2 and PP7 systems that permit real-time measurements of transcription in vivo, there continues to be a dearth of analytical tools that can link these measurement to quantitative models of gene regulation. Here we seek to provide such a tool, presenting *burstMCMC*, an efficient and flexible Bayesian framework for inferring burst parameters from live imaging measurements. We benchmark our model's performance using simulated data, and demonstrate its efficiency gains relative to our previously published method, cpHMM. We also show that *burstMCMC* can be used to accurately infer RNA Polymerase (RNAP) elongation times from single-color live imaging data, a key biological parameter for which no general inference method previously existed.

7.1 Introduction

The past decade has seen the widespread adoption of *in vivo* RNA fluorescence labeling technologies, such as the MS2/MCP [14] and PP7/PCP systems [36] as a means to directly visualize nascent mRNA molecules at individual gene loci. Yet, while these live imaging methods now represent a standard tool in the arsenal of many researchers in the study of gene regulatory processes, the field is only just beginning to harness the potential of these systems to serve as a quantitative foundation for investigating the molecular basis of transcriptional control. Here we seek to further this goal by presenting *burstMCMC*, a novel computational framework that uses Markov Chain Monte Carlo (MCMC) methods [107] to

uncover how transcription factor proteins modulate the dynamics of transcriptional bursting at target gene loci.

In a previous work, we developed the cpHMM inference method and applied it to dissect how transcription factors modulate transcriptional burst dynamics to give rise to the *even-skipped* stripe 2 pattern of gene expression in the developing fly embryo (Chapter 2 and [168]). A key innovation in this modeling approach was the introduction of a compound state formalism that could account for the presence of memory in MS2 and PP7 signals due to the finite amount of time (t_{elong}) required for Pol II molecules to transcribe the reporter gene (see Appendix A.3.4 for details). Yet, despite its utility, cpHMM has notable limitations, the most significant of which is the fact that the model's computational complexity (and, hence, the time required for inference) scales *exponentially* with the number of experimental time steps ($w = t_{elong}/\Delta\tau$) required for Pol II molecules to traverse the gene. In addition, cpHMM requires the user to input the elongation time w, a gene-specific parameter that is unknown *a priori* for most systems.

In what follows, we describe ongoing efforts to develop *burstMCMC*, with the aim both of addressing these limitations and other limitations of the original model and of developing novel inference capabilities. The crux of the exponential scaling issue noted above lies in the fact that, as a part of its inference, cpHMM must keep track of the likelihood of *all possible* promoter states at each time point along an MS2 trace, which scales as K^w (where K is the number of promoter activity states). Recently, however, Bowles and colleagues [28] showed that the number of models states that actually have a non-negligible likelihood at any given point in time can be several orders of magnitude smaller than K^w. As a result, only a few hundreds or thousands of states need to be tracked at any point in time to achieve accurate inference results. The authors use this insight to develop *burstInfer*, an inference approach that reduces burst inference times by truncating the space of possible model states at each experimental time point [28]. Yet, whereas cpHMM yields accurate inference results for systems with up to 3 distinct activity states and continuous promoter transitions—i.e. switching can occur at any point on or between experimental time steps—*burstInfer* only applies to systems with up to two activity states and, further, makes the assumption that promoter switching is discrete (i.e. happens only at experimental time points); an assumption that likely does not comport the true nature of transcriptional systems.

With *burstMCMC*, we take an alternative approach. Rather than algorithmically truncating state space to reduce model complexity, we use MCMC methods—originally developed to efficiently sample high-dimensional ensembles in Statistical Physics [210]—to instead sample the high-dimensional space of potential burst model states. We show that *burstMCMC* approaches cpHMM accuracy for three state transcriptional systems with both discrete and continuous promoter switching, while reducing computational complexity by orders of magnitude for biologically salient values of the elongation parameter, w. In addition, we demonstrate that *burstMCMC* can reliably infer the elongation time, w, thus removing a significant barrier to the application of burst inference to a wider array of genes and providing a new computational lens into a biological parameter that is of interest in its own right (see, e.g., [99, 190]).

Since this very much remains a work in progress, we close this chapter by sketching

out key next steps for model development. First, we lay out model improvements (many of which are already under way) that are aimed at increasing the speed, accuracy, and robustness of *burstMCMC* inference. Second, we discuss extensions to the model that probe new facets of the regulation of bursting dynamics, including hierarchical models for single cell burst inference and models with concentration-dependent parameters Finally, we draw upon concepts from Chapters 5 and 6 to speculate about how mean first passage times could serve as a computationally tractable link between phenomenological bursting models and bona fide molecular models that can shed direct light on the molecular mechanisms underlying the transcription factor control of burst dynamics.

7.2 Results

7.2.1 Introducing the model

In this section, we use a single fluorescent trace (Figure 7.1A) produced by a simple system with two promoter states (OFF and ON) to introduce the model's architecture, provide intuition for how *burstMCMC* infers hidden promoter states (Figure 7.1B), and to show how the resulting promoter state vector, $\hat{\boldsymbol{z}}$, is used to sample the posterior distributions of key model parameters: $p(\boldsymbol{\theta}|\boldsymbol{z}, \boldsymbol{y})$. See Appendix D for further details. We note that, although this discussion focuses on inference for a single fluorescence trace, all concepts and procedures presented along the way extend trivially to the case of multiple traces.

Consider the MS2 trace in Figure 7.1A, which depicts the total integrated intensity of the transcriptional locus over time. Our central aim is to use the observed fluorescent series, $\boldsymbol{y} = \{y_1, y_2, ..., y_T\}$, where T is the number of data points in the trace, to infer the unobserved time series of hidden (unobserved) promoter activity states, $\boldsymbol{z} = \{z_1, z_2, ..., z_T\}$. Figure 7.1C illustrates the basic architecture of our model. We assume a discrete model in which the promoter fluctuates between K different activity states, such that $z_t \in \{1, ...K\}$(K=2 for the illustration in Figure 7.1C). Transitions between the effective promoter states are assumed to be Markovian, meaning that the hidden promoter state z_t at time step t is conditionally dependent only on the state in the previous time step. This dependency is modeled through a $K \times K$ transition probability matrix $\boldsymbol{A} = p(z_t|z_{t-1})$, where A_{kl} is the probability of transitioning from the l^{th} state into the k^{th} state in the time interval $\Delta\tau$.

We assign a characteristic RNAP initiation rate, $r(k)$, with units of RNAP per minute, to each effective promoter state, $z(k)$. Thus, the number of RNAP molecules initiated between time steps $t-1$ and t will be $r(z_t)\Delta\tau$, $1 \leq k \leq K$. We define a fluorescence emission per time step for each state as $v(k) = F_{\text{RNAP}} r(k)\Delta\tau$, where F_{RNAP} is the fluorescence calibration factor that gives the fluorescence produced by each nascent RNAP molecule. F_{RNAP} can be experimentally determined using smFISH experiments (see Materials and Methods of Chapter 2), but this is not required for burst parameter inference. Since a new calibration must be determined separately on a microscope by microscope basis, it is often simplest to work with $v(k)$ in arbitrary fluorescence units.

Together, the $K \times K$ transition probability matrix $\boldsymbol{A}$ and K-element emission rate vector

$\boldsymbol{v}$ fully determine the dynamics of RNAP initiation at the gene promoter. For a particular activity trace, the sequence of hidden promoter states, $\boldsymbol{z}$, defines a corresponding sequence of emission rates, $\boldsymbol{e}$, where $e_t = \boldsymbol{v}(z_t)$ (second row in Figure 7.1C). If $\boldsymbol{e}$ was directly observable, then it would be possible to infer $\boldsymbol{A}$ and $\boldsymbol{v}$ directly. Instead, we are faced with the problem of interconverting between instantaneous emission states, $\boldsymbol{e}$, and the observed fluorescence, $\boldsymbol{y}$. In the original cpHMM model (as well as [28]), this is accomplished by switching from a simple model with K promoter states to compound model model with K^w compound states (see Appendix A.3.4 for further details). Here we take a different approach, by noting that we can then express the predicted fluorescence time series, $\hat{\boldsymbol{y}}$, as a convolution of $\boldsymbol{e}$ (bottom tier of Figure 7.1C), such that

$$\hat{\boldsymbol{y}} = \kappa * \boldsymbol{e}. \tag{7.1}$$

Here, κ is a kernel of length $\lceil w \rceil$ (ceiling of w), that encodes the fluorescence contribution of RNAP molecules at different positions along the gene body. We will take w as given for now, and will revisit this assumption in Section 7.2.4. Note that, unlike cpHMM, the *burstMCMC* formalism allows for the RNAP dwell time τ_{MS2} to be a fractional number of experimental time steps, such that w need not be an integer quantity.

Equation 7.1 provides an analytic mapping from hidden promoter states to experimental observable fluorescence values. However, the reverse problem of moving from fluorescent values to their underlying promoter states is nontrivial due to the presence of noise in the observed fluorescence signal:

$$\boldsymbol{y} = \kappa * \boldsymbol{e} + \boldsymbol{\epsilon}, \tag{7.2}$$

where $\boldsymbol{\epsilon}$ is a Gaussian noise vector with zero mean and a standard deviation given by the parameter σ, which is estimated as a part of *burstMCMC* inference. Estimating the sequence of hidden promoter activity states, $\boldsymbol{z}$, given the noisy signal $\boldsymbol{y}$ constitutes the central challenge of our inference problem. *burstMCMC* solves this problem efficiently by using Gibbs sampling to estimate hidden promoter states using the observed fluorescence trace $\boldsymbol{y}$ and parameter estimates $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{A}}, \hat{\boldsymbol{v}}, \hat{\sigma}\}$.

7.2.1.1 A Bayesian approach to burst parameter inference

In Bayesian parameter inference, our ultimate aim is to obtain an estimate of the joint posterior distribution of model parameters that reflects our prior modeling assumptions and our experimental data: $p(\boldsymbol{A}, \boldsymbol{v}, \sigma | \boldsymbol{y}, \boldsymbol{z})$. Estimating this joint distribution directly, however, is typically impractical in all but the simplest inference problems. The core insight behind Gibbs Sampling is that we can approximate this joint distribution by repeatedly sampling the *marginal* distributions of each parameter (e.g. $p(\boldsymbol{v}|\boldsymbol{A}, \sigma, \boldsymbol{y}, \boldsymbol{z})$), which tend to be much more tractable than the full joint distribution [34].

The process of sampling the marginal posterior distributions for each of our model parameters is greatly simplified by the judicious selection of special prior distributions—conjugate priors—that lead to well-defined posterior distributions with closed-form analytical expressions [244]. We thus adopt well established conjugate prior distributions for the transition probability matrix ($\boldsymbol{A}$), the emission rate vector ($\boldsymbol{v}$), and the fluorescent noise parameter (σ).

We emphasize, however, that in most use cases, these priors will be "weak", such that our prior guesses about the value of each parameter have little impact on the final estimates for each. The priors outlined below should thus be interpreted as statistical conveniences first and foremost, that provide an efficient, analytically tractable means of using the distribution of experimental observations, $p(\boldsymbol{y})$, to obtain estimates for the mean and uncertainty in our model parameters.

We assume that each column of $\boldsymbol{A}$ follows the Dirichlet distribution, which is a commonly used prior for discrete probability distributions, such that

$$p(\boldsymbol{a_i}) \sim \text{Dirichlet}(\boldsymbol{\alpha_0}), \tag{7.3}$$

where $\boldsymbol{\alpha_0}$ is a vector of positive real values of length K that, in the context of our model, can be thought of as pseudo-observations that encode prior guesses about the frequency of different state transitions.

As in the original formulation of cpHMM, estimation of the emission vector $\boldsymbol{v}$ given the vector of hidden states $\boldsymbol{z}$ can be formulated as a simple least squares linear regression problem. As such, we follow previous works that have employed Gibbs sampling in the context of linear regression models [203] and assume that $\boldsymbol{v}$ follows a K-dimensional multivariate normal distribution:

$$p(\boldsymbol{v}) \sim \mathcal{N}(\boldsymbol{v}_0, \boldsymbol{\Sigma}_0), \tag{7.4}$$

where the vector $\boldsymbol{v}_0$ encodes prior guesses for each emission rate and the covariance matrix $\boldsymbol{\Sigma}_0$ encodes our prior confidence in each guess.

Finally, we take σ to be inverse-gamma distributed (again, a standard assumption), such that

$$p(\sigma^2) \sim \text{Inverse-Gamma}(k, \theta), \tag{7.5}$$

where $1/\theta$ can be interpreted as encoding our prior guess about the scale of the Gaussian noise and k is a pseudocount variable that encodes our confidence in that guess.

At the outset of *burstMCMC* inference, we draw an initial set of parameter values, $\boldsymbol{\theta}_0$, from these priors that are then used to draw an initial sample of the latent state vector $\boldsymbol{z}$. We outline this process next, and then examine how the above priors shape our approach to sampling the marginal posterior distributions of $\boldsymbol{A}$, $\boldsymbol{v}$, and σ.

7.2.1.2 Sampling hidden promoter states

For a given set of model parameters, $\boldsymbol{\theta}_k$, we wish to obtain an estimate for the sequence of promoter states, $\boldsymbol{z}$, that corresponds to the observed MS2 trace, $\boldsymbol{y}$. Unlike the algorithm employed in Chapter 2, which estimated the probability of all possible promoter states at each time step, *burstMCMC* works by sampling discrete state values at each time step, such that the final outcome is a single sequence of promoter states, $\boldsymbol{z}_k$, drawn from the distribution $p(\boldsymbol{z}|\boldsymbol{\theta}_k, \boldsymbol{y})$. Intuitively, we can think of this sampling procedure as a stochastic simulation.

Figure 7.1C illustrates the process. We proceed by randomly selecting time points one by one from the set $t \in \{1, 2, ...T\}$ without replacement. For each time point t, we randomly draw a new promoter state z_t from $z_t \in \{1, 2, ...K\}$ based off of the marginal likelihood of

each state, conditioned on the system parameters ($\boldsymbol{\theta}_k$), the observed fluorescent values ($\boldsymbol{y}$), and the remainder of the hidden promoter state sequence ($\boldsymbol{z}_{\varnothing t}$): $p(z_t|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \boldsymbol{A}, \boldsymbol{y})$. We can express this as the product of two terms, such that

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \boldsymbol{A}, \boldsymbol{y}) = p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{A})p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \sigma, \boldsymbol{y}), \tag{7.6}$$

where the first term on the right-hand side captures the component of the probability derived from promoter state transitions and the second term captures the component relating to the fluorescent emission probabilities.

If $z_{t-1} = j$ and $z_{t+1} = l$, then we can express this up to a proportionality constant in terms of model parameters as

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{A}) \propto a_{ij}a_{li} \prod_{n=t}^{t+\lfloor w \rfloor} e^{-\frac{(\hat{y}_n - y_n)^2}{2\sigma^2}}, \tag{7.7}$$

where a_{mn} denotes the probability of switching from state n to state m, and where $\hat{y}_n$ reflects the predicted fluorescence at time n given that $z_t = i$. Note that the fluorescence component extends into the "future", reflecting the fact that the promoter state at time t impacts the predicted fluorescence, $\hat{\boldsymbol{y}}$, for time points up to and including $t + \lfloor w \rfloor$.

Although this approach leads to only a single estimate for $\boldsymbol{z}$ per MCMC iteration, samples taken across many model iterations provide an effective reflection of the space of possible promoter sequences that could correspond to the observed fluorescence trace $\boldsymbol{y}$. Next, we will briefly sketch how our promoter state vector, $\boldsymbol{z}_k$ can be used to sample from the posterior distributions of our key model parameters $\hat{\boldsymbol{\theta}} = \{\boldsymbol{A}, \boldsymbol{v}, \sigma\}$.

7.2.1.3 Using Gibbs sampling to estimate posterior parameter distributions

In this section, we outline how $\boldsymbol{z}$ can be used to obtain an updated sample of the transition probability matrix $\boldsymbol{A}$. See Appendix D.3 for a full discussion of the update procedures for $\boldsymbol{A}$, $\boldsymbol{v}$, and σ.

We draw updated samples separately for each column of $\boldsymbol{A}$. Conveniently, $\boldsymbol{A}$ is conditionally independent of $\boldsymbol{y}$, $\boldsymbol{v}$, and σ given the latent state vector $\boldsymbol{z}_k$. As a result, the update expression for each column of $\boldsymbol{A}$, $\boldsymbol{a}^i$, is Dirichlet-distributed, such that

$$p(\boldsymbol{a}^i_{k+1}|\boldsymbol{z}_k) \propto \text{Dirichlet}(\boldsymbol{n}^i_k + \boldsymbol{\alpha}^i_0), \tag{7.8}$$

where $\boldsymbol{\alpha}^i_0$ is a vector of length K where the jth element contains prior pseudo-counts that encode our prior expectation for the frequency of transitions *from* state i *to* state j, and where $\boldsymbol{n}^i_k$ is a vector of length K, where the jth element indicates the actual number of transitions "observed" in $\boldsymbol{z}_k$ from i to j.

From Equation 7.8 it is easy to see how our prior influences the posterior estimate: if the number of elements in $\boldsymbol{\alpha}_i$ is much less than the number of transitions out of state i, then our prior will have little impact, and the information gleaned from the data will dominate. Conversely, if the elements of $\boldsymbol{\alpha}^i_0$ are of order with (or larger than) elements of $\boldsymbol{n}_i$, then our

prior assumptions regarding the relative magnitudes of the transition probabilities out of state i will strongly influence our final posterior estimate of $\boldsymbol{a}_i$.

Figure 7.1D shows the results from repeated updates using Equation 7.8 over the course of 1,000 MCMC iterations. We see that, after an initial "burn-in" period, the model converges to the correct values for a_{21}, the probability switching into the ON state, and a_{12}, the probability of switching into the OFF state. Analogous update functions for $\boldsymbol{v}$ and σ are given in Appendix D.3. Figure 7.1E and F show the sampling results for these parameters, which, we see, also fluctuate about their true values. Figure 7.1F shows the univariate and bivariate distributions obtained for each parameter for the final 750 steps of the inference run. Overall, we see that the model obtains posterior distributions that accurately reflect the underlying burst parameters used to simulate the fluorescent trace for this simple scenario. Next, we turn to benchmarking *burstMCMC*'s performance in a more realistic (and challenging) inference scenario.

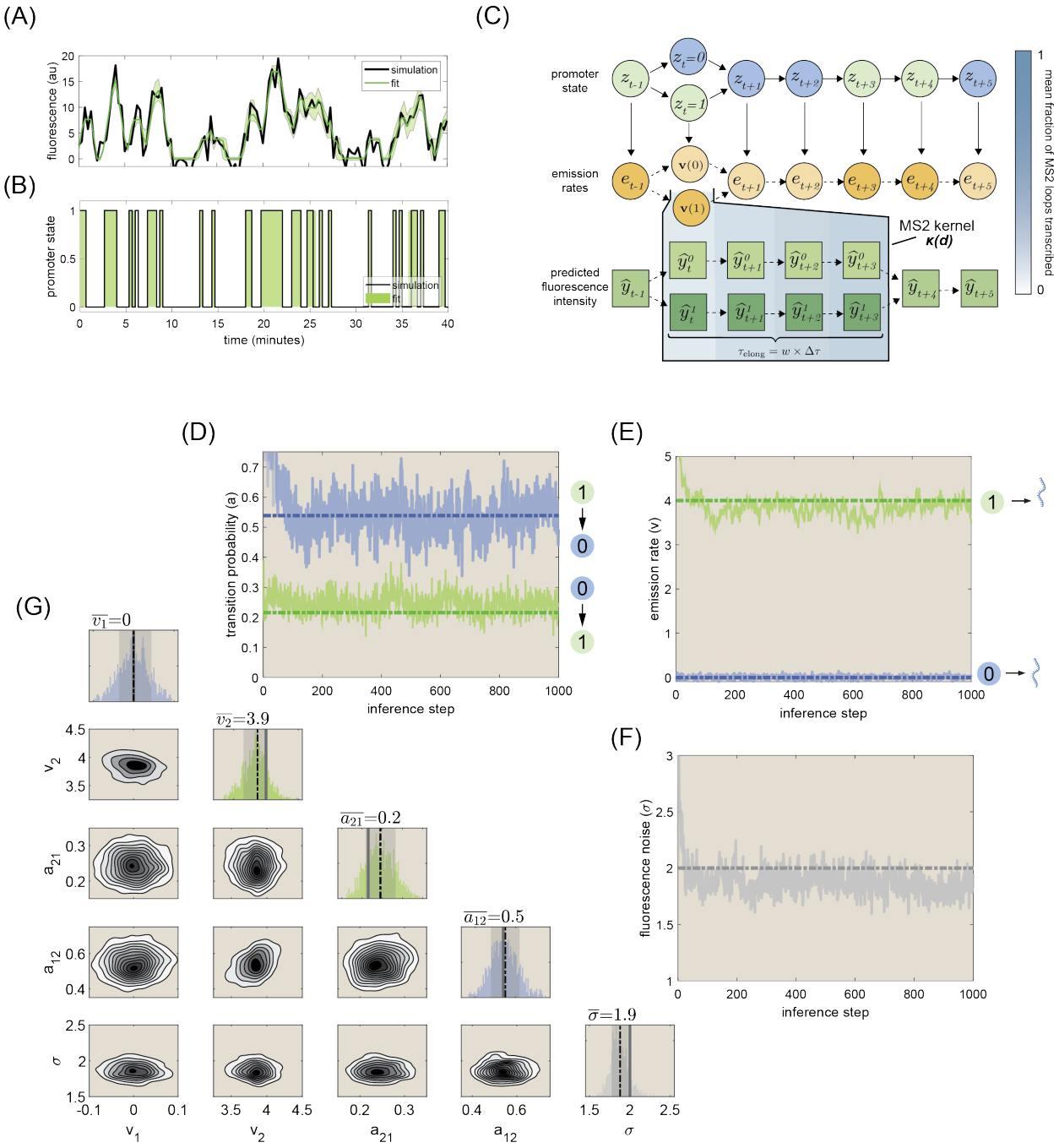
(A)
fluorescence (au)
simulation
fit
(B)
promoter state
time (minutes)
(C)
promoter state
emission rates
predicted fluorescence intensity
MS2 kernel
κ(d)
mean fraction of MS2 loops transcribed
(D)
transition probability (a)
inference step
(E)
emission rate (v)
inference step
(F)
fluorescence noise (σ)
inference step
(G)
v1
v2
a21
a12
σ

Figure 7.1: **A Bayesian framework for inferring burst parameters. (A)** Example of a fluorescent trace generated via stochastic simulation. Black trend indicates simulated experimental data, $\boldsymbol{y}$, (including noise). Green trend indicates model prediction, $\hat{\boldsymbol{y}}$, and shaded region indicates model uncertainty, as calculated across 750 MCMC samples. **(B)** Black trend gives true underlying sequence of promoter states, $\boldsymbol{z}$, and green shaded regions indicate model predictions. **(C)** Schematic illustrating *burstMCMC* model architecture, as well as the procedure for sampling the latent promoter state vector. The top row depicts a stochastic series of promoter states ($\boldsymbol{z}$), which, in turn, dictate a series of promoter emission rates, $\boldsymbol{e}$ (yellow circles). The key modeling difficulty comes from the fact that a single emission state impacts $\lceil w \rceil$ time steps of the fluorescence time series, $\hat{\boldsymbol{y}}$ (green squares). Thus, to sample the latent promoter state z_t, we must assess how different latent state values impact all relevant time steps of the fluorescence trace. **(D)** Time series of state transition probability values returned by Gibbs sampling in accordance with Equation 7.8. **(E)** and **(F)** Time series of Gibbs sampling results for the elements of the emission vector, $\boldsymbol{v}$, and the fluorescent noise parameter, σ. **(G)** Univariate and bivariate histograms depicting distribution of burst parameter values, as determined from the final 750 steps of *burstMCMC* inference. In univariate histograms, solid lines indicate true values and dashed lines indicate model predictions. We see good agreement in all cases, and also note minimal covariance in the bivariate parameter distributions.

7.2.2 Benchmarking *burstMCMC* performance using simulated data

In Appendix A.3.5, we use simulated fluorescent traces generated using a known, biologically relevant set of model parameters to validate the accuracy of the cpHMM inference. Here, we undertake a similar statistical validation for *burstMCMC* using comparable model parameters to those employed in Appendix A.3.5. As in Chapter 2, we focus on a 3 state bursting model (Figure 7.2A) that accounts for the presence of two distinct gene promoters within each diffraction-limited fluorescent spot (see Appendix A.3.6 for details). Before discussing the results, it is worth spilling a little ink to discuss an important distinction between discrete and continuous promoter switching dynamics.

7.2.2.1 Aside about continuous versus discrete promoter switching dynamics

Like cpHMM, *burstMCMC* is a fundamentally discrete model: it breaks promoter activity into discrete blocks of time equal to the experimental time resolution ($\Delta\tau$). The statistics of transitions between promoter states are given by the transition probability matrix $\boldsymbol{A}$, and may only occur at (not between) experimentally observed time points. In contrast, it seems reasonable to expect that real biological systems adhere to *continuous*, rather than discrete, dynamics described by a transition *rate* matrix, $\boldsymbol{R}$.

The presence of transitions between time points violates a key assumption of *burstMCMC*, and thus might be expected to lead to degraded performance. Thus, in assessing model performance, we will examine its accuracy both in the context of fluorescent traces

simulated using discrete switching dynamics and for fluorescent traces generated using the Gillespie Algorithm [110] that assume continuous switching dynamics. The former will indicate whether the model is performing as designed, and the latter will indicate how it fairs on data that more closely resemble what we expect from real experimental data.

7.2.2.2 Validation results

We generated ten independent data sets for the continuous and discrete switching cases, each containing 3,000 total data points (25 fluorescent traces 120 time steps long). We used identical model transition rates to those employed in the original validation of cpHMM. Emission rate and noise parameters were also selected to be comparable to those used for cpHMM validation. See Table 7.1 for a full description of system parameters. For each dataset, we ran 125 independent MCMC inference chains in parallel for 5,000 iterations, and report results only for the best fitting (most likely) run out of those 125. This procedure is analogous to our approach for cpHMM, which took the most likely of 25 independent runs in order to overcome the presence of multiple local minima and increase the odds of obtaining accurate parameter estimates. We note that both the number of chains and the number of MCMC steps used is conservative, and that accurate results can generally be achieved with significantly less computational input.

Figure 7.2B shows *burstMCMC* inference results for the emission rates of the three activity states in our model. We see that, for fluorescence data generated using discrete promoter switching (blue diamonds), *burstMCMC* recovers precisely the correct values of each of the three rates (indicated as yellow circles). The model also generally yields accurate results for traces generated assuming continuous promoter switching (red squares), although we note a slight underestimate for state 3 (10.8 ± 0.51 vs. 12 AU/min).

What about transition rates? Consistent with our approach in Chapters 2 and 3, we report transition rates, rather than transition probabilities. These were obtained by taking the matrix logarithm of the $\boldsymbol{A}$ matrices returned by *burstMCMC* inference:

$$\boldsymbol{R} = \frac{1}{\Delta\tau} \log(\boldsymbol{A}), \tag{7.9}$$

where $\Delta\tau$ is the experimental time resolution. Figure 7.2C shows our inference for the six transition rates connecting the 3 activity states in our model. For both continuous and discrete data, we find that *burstMCMC* accurately recovers all transition rate values, with the exception of the (very fast) transition rate from state 3 to state 2 (far right side of Figure 7.2C). In this case, the model infers values that are significantly too low (3.2 vs. 4.2 events per minute). Although cpHMM was found to successfully recover this rate, we found in the corresponding work that transition rates that imply more than one transition per time step generally present an inference challenge. Thus, it is not a surprise that this rate was inferred the least accurately. We note that *burstMCMC* successfully recovers this rate for discrete data with lower noise levels. Thus, a key next step will be to improve inference procedures to permit accurate inference of fast (but biologically salient) transition rates for data with realistic levels of fluorescence noise.

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, 6, 12 AU/min
Measurement noise (σ)	1.5 AU
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})	28 sec
Duration of each trace	40 min
Number of time points per dataset	3,000
Number of MCMC iterations	5,000
Number of MCMC chains	125
Number of traces per dataset	25
Number of independent datasets	10

Table 7.1: Parameter values used for generating synthetic datasets in the statistical validation of the *burstMCMC*.

How much does our underestimation of r_{12} impact our overall interpretation of the model? To answer this question, we reframed the inference results from Figure 7.2B and C in terms of the effective two state framework employed in Chapters 2 and 3, where the system is assumed to be on if either gene promoter is active. This leads to a simple effective two state model (right side of Figure 7.2A) with two transition rates and two emission rates. This model is simpler to interpret, and so represents the predominant means of reporting burst parameter results in publication. Happily, we find that *burstMCMC* accurately recovers all transition and emission rates for this effective two state system for both discrete and continuous (Figure 7.2D and E).

7.2.3 *burstMCMC* significantly reduces computational complexity

A key motivation for this work was the high computational cost imposed by cpHMM inference for systems where the memory parameter w is large. In the cpHMM model, the majority of this cost comes from the Expectation step, wherein the model updates the probability of each of the K^w states for each time point in the data set. Figure 7.3A shows the average amount of time required for a single expectation step as a function of w, demonstrating that the computation time for this procedure scales exponentially. From a practical standpoint, this means that, to be feasible, inference must be conducted on large computing clusters for $w \geq 7$, and cannot practicably be carried out once $w \gtrsim 9$. On the other hand, we find that the time required for *burstMCMC* to update its estimate of the hidden promoter state vector $\boldsymbol{z}$—the analogous step in its inference cycle—is essentially decoupled from w.

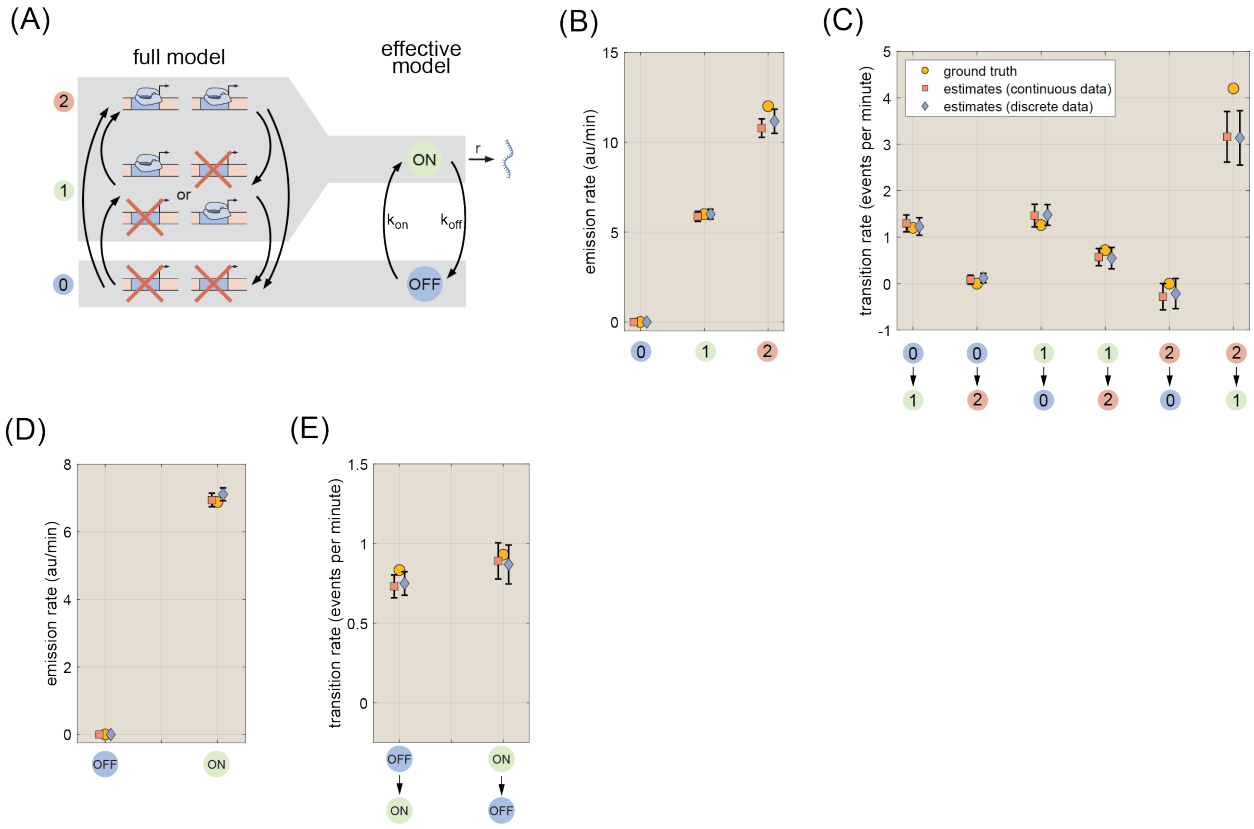

Figure 7.2: **Benchmarking *burstMCMC* performance using simulated data. (A)** We sought to validate model performance by simulating data for a gene locus with burst dynamics comparable to those inferred for the *even-skipped* stripe 2 gene in Chapter 2. We use a three state model there to capture the presence of two distinct gene loci—each of which can be ON or OFF—within each diffraction-limited spot. This three state model can be condensed into an effective two state model (right-hand side) where we consider the locus on if either promoter is active. **(B)** Inference results for emission rates of each promoter state. Circles indicate true parameter values used to generate the data, diamonds indicate inference results obtained using idealized experimental traces that assume discrete promoter switching, and squares indicate results for more realistic traces generated with continuous promoter switching. The model yields accurate results in all cases, save for a slight underestimate of the high initiation rate for continuous data. **(C)** Inference results for model transition rates. *burstMCMC* infers all rates accurately, save for the large transition rate from $2 \rightarrow 1$. **(D)** and **(E)** Inference results for emission and transition rates for effective two state model illustrated on right-hand side of (A). Model accurately recovers all effective parameters.

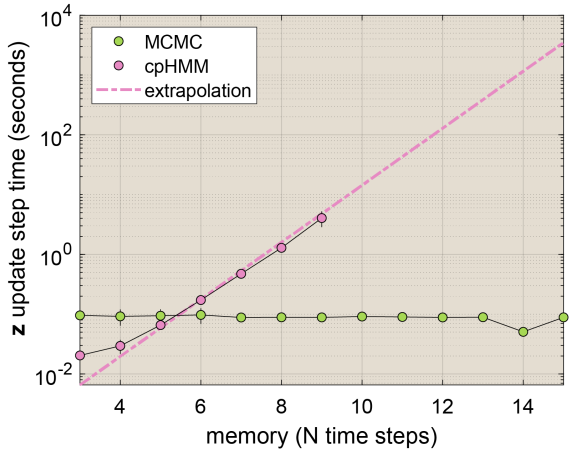

Figure 7.3: ***burstMCMC* significantly reduces computational cost of burst parameter inference. (A)** Green trend compares the average time required for *burstMCMC* to resample $\boldsymbol{z}$ (green circles) to the time needed for a single expectation step in cpHMM inference (pink circles) as a function of w. Dashed pink line extrapolates cpHMM performance to regimes that could not be calculated directly given computational constraints. Whereas cpHMM computational time scales exponentially with w, we see that the trend for *burstMCMC* is essentially flat.

Thus, we see that *burstMCMC* has the potential to solve a core limitation of the cpHMM method. We note, however, that more work is required to establish how other elements of *burstMCMC* inference scale with w. For instance, it is likely that the number of independent MCMC chains or the number of MCMC steps per chain needs to be increased for large w to account for the increased complexity of the inference problem. Future work will seek to establish how this scaling plays out in practice.

Beyond its improved computational efficiency, another great strength of *burstMCMC* is its malleability: the flexibility of MCMC methods means that it is straightforward to add parameters to the baseline model outlined in previous sections. In the next section, we demonstrate how this can be done for a crucial system parameter: the number of elongation steps w.

7.2.4 Inferring RNAP elongation times

Beyond its computational complexity, another factor limiting the application of the cpHMM method to a wider array of genes and biological systems is the need to specify w,

$$w = \frac{\tau_{\text{elong}}}{\Delta\tau}, \tag{7.10}$$

where $\Delta\tau$ is set by the temporal resolution of data acquisition and τ_{elong} is the elongation time, which is unknown *a priori*. In Chapter 2, we presented a method that uses derivatives

of the autocorrelation of fluorescence traces to estimate w (see Appendix A.3.10). However, this method is sensitive to noise due to its dependence on taking derivatives of experimental data, and can be difficult to interpret by non-experts since no methods exist for estimating w and its uncertainty aside from direct inspection of the autocorrelation derivatives themselves. Thus, there is a need for a robust, interpretable method for estimating w to complement the core functionalities of cpHMM and *burstMCMC*.

Here, we demonstrate how *burstMCMC* can be extended to permit the accurate inference of w. To do this, we add a standard Metropolis Hastings step to each MCMC iteration, in which a new length, w' (again, permitted to be a non-integer quantity), is proposed, which leads to a new MS2 kernel, κ' of length w'. We then use κ' to generate a new prediction for the fluorescent trace using

$$\hat{\boldsymbol{y}}' = \kappa' * \boldsymbol{z}. \tag{7.11}$$

Since the hidden promoter state vector is unchanged, the only difference in the total likelihood of the model lies in the error between predicted and actual fluorescence traces:

$$p(w'|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{y}) \propto \prod_{t=1}^{T} e^{-\frac{(\hat{y}_t - y_t)^2}{2\sigma^2}}. \tag{7.12}$$

Equation 7.12 is then used accept or reject w' using the standard Metropolis Hastings procedure.

Figure 7.4A shows the result of incorporating this procedure to standard *burstMCMC* inference for systems with a wide range of different w values:

$$w \in \{4.1, 5.7, 7.3, 8.9, 10.4\} \tag{7.13}$$

In general, we find that MH sampling is ineffective at estimating the true w in most cases. Instead, individual *burstMCMC* chains (solid colored lines in Figure 7.4A) tend to remain "stuck" near the initial guess that was used to initialize them, irrespective of the corresponding true w value (dashed lines of the same color). For instance, note how two of the 3 MCMC chains for $w = 8.9$ (blue) remain stuck near their initiation point at $w \approx 5$. This problem reflects a wider challenge with burst parameter inference, namely the presence of multiple local minima that can trap the system and lead to sub-optimal performance.

In this case, we find that it is possible to overcome this issue by using a simple implementation of simulated annealing, a method in which an artificial temperature parameter, T_a, is modulated over time to control how models explore complex, non-convex optimization landscapes [119]. In our case, we find that it is sufficient to apply T_a to σ, while leaving other model parameters unchanged. Specifically, take $\sigma_a = T_a \hat{\sigma}_k$, where $\hat{\sigma}_k$ is the initial estimate returned by Gibbs sampling. We take T_a to be a decreasing exponential function of the MCMC step, k (Figure 7.4B), such that

$$T_a = 2e^{-k/100} + 1. \tag{7.14}$$

This has the effect of artificially elevating σ early on in *burstMCMC* inference, which, in turn, "loosens" the likelihood expression given in Equation 7.12 and makes it easier for the

system to escape local minima. σ is then "cooled" back off to normal levels ($T_a = 1$) to obtain accurate parameter estimates.

Figure 7.4C shows the results of memory inference runs that incorporate simulated annealing. The impact of applying T_a is clear from the large fluctuations in w_k early on in the inference runs of each MCMC chain. This time, we find that the model robustly recovers the true w value across all elongation times tested, indicating that these early fluctuations help the model to overcome local minima for w and find the best-fitting value for each system.

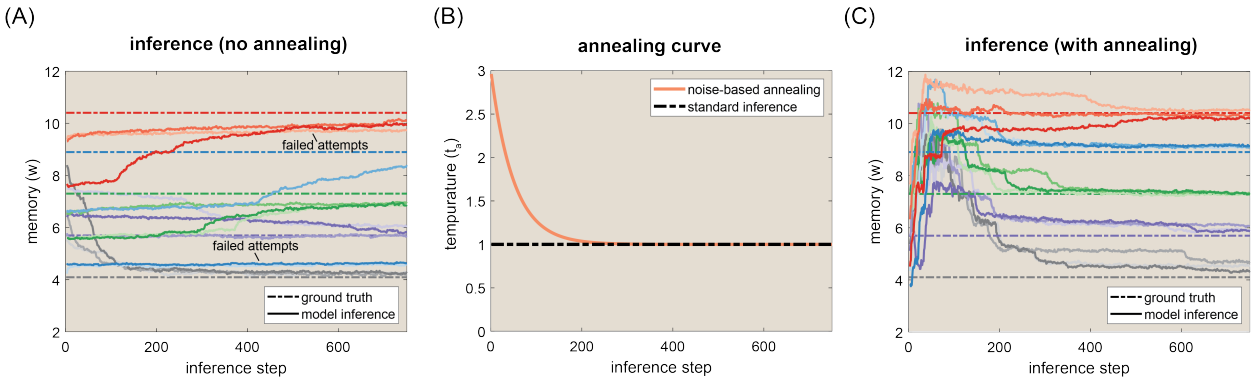

Figure 7.4: **Inferring RNAP elongation times with *burstMCMC*.** **(A)** Inference results for different w values. Solid lines indicate results for different MCMC chains. Dashed lines of same color indicate the correct w values. We find that the model frequently fails to infer the correct value for w. **(B)** Red curve illustrates effective temperature parameter that is used to elevate σ early on in *burstMCMC* inference. **(C)** Inference results for runs that incorporate simulated annealing. In this case, we find that the model robustly recovers the true elongation time for every system tested.

7.3 Future directions: a statistical bridge between live imaging data and quantitative models of transcription

7.3.1 Improving model accuracy

The first priority moving forward will be to augment the current inference framework to resolve the minor accuracy issues noted in Figure 7.2B and C. To realize this goal, we are actively pursuing multiple directions, all of which amount to efforts to improve the accuracy with which the model decodes the fluorescent vector $\boldsymbol{y}$ to achieve a representation of the hidden sequence of promoter states, $\boldsymbol{z}$. The first of these is adjusting the inference such that sampling is conducted using an ensemble of latent state chain replicates, rather than a single state vector, with the hope that multiple replicates will provide a more robust representation

of the distribution of possible promoter states at each time point. This may be particularly important for the accurate inference of rare states, such as state 2 in the model used for validation, which is visited less than 10% of the time. Most simply, this could be implemented by duplicating the current approach for sampling $\boldsymbol{z}$ to run simultaneously across multiple replicates. In such approach, each $\boldsymbol{z}$ representation would be updated independently, and then the replicate chains would be pooled in order to update $\boldsymbol{\theta}$. Alternatively, we could implement an approach more akin to Sequential Monte Carlo (particle filtering) [191], wherein multiple particles are tracked and collectively (rather than independently) resampled. Careful thought would need to be given to how to implement this approach in the context of a system with memory (i.e. κ of length greater than 1), like the one considered here.

A related approach that has shown some initial promise is parallel tempering [120]. There are many variations on this approach, but the basic idea is to have a standard MCMC chain estimating $\boldsymbol{z}$ connected to a "resevoir" of higher-temperature chains with $T_a > 1$. The higher temperature chains are situated in a "flatter" likelihood space, and thus can explore parameter space more efficiently than the $T_a = 1$ chain. The key is to allow for "leaks" between chains, most simply via a series of horizontal Metropolis Hastings transfer moves, that permit the sharing of information between high temperature chains and the $T_a = 1$ chain. In addition to increased accuracy, this approach can also increase the rate of model convergence [120]. More testing is needed to determine whether this approach improves the performance of *burstMCMC*. A third, closely related option is to extend the simulated annealing approach outlined in Section 7.2.4 to the broader problem of inferring $\boldsymbol{A}$, $\boldsymbol{v}$, and σ.

7.3.2 Model extensions: input-output functions and single-cell burst parameters

Thanks to recent innovations in live imaging techniques (see, e.g., [24]) it is increasingly feasible to measure input concentrations of transcription factor ("TF") proteins alongside the output transcriptional activity for a wide range of factors and target genes. With cpHMM, it is already possible to examine burst parameter trends as a function of transcription factor concentration by conducting inference on multiple sets of traces, each grouped by average TF concentration. There are two key challenges to this approach, however. First, each group must be sufficiently large (typically $N \gtrsim 20$ cells) to permit robust inference, which places limits on the granularity with which the transcriptional input-output function can be probed. Second, in contexts such as developmental systems, transcription factor concentrations are typically dynamic in time, a dynamism which current inference approaches fail to capture and, as a result, simply average over in one way or another. Yet, as we saw in Chapters 2 and 3, temporal control of transcription can play an important role in driving pattern formation.

To overcome these limitations, I propose extending burstInfer to permit one or more burst parameters to be explicit functions of input transcription factor concentration. For instance, for the two state model we could imagine a model in which the on rate k_{on} has the

form of a simple Hill function, such that

$$k_{\text{on}} = k_{\text{on}}^0 \frac{[TF]^H}{K_d^H + [TF]^H}, \tag{7.15}$$

where $[TF]$ is the concentration of the regulatory factor, k_{on}^0 is the maximum on rate value, and H and K_d control the function sharpness and sensitivity, respectively. Such an approach would remove the need to break transcription traces into groups for inference, allowing for inference to be conducted over full groups, increasing statistical power, as well as the granularity of the inferred input-output relationship. In addition, it would automatically account for changes in burst parameter values within individual traces over time.

From a technical perspective, incorporating these kinds of input dependencies into the *burstMCMC* framework is straight-forward; however it remains an open question whether accurate recovery of the additional parameters is possible. In addition, it is conceptually difficult to conceive of how we would apply this kind of framework to a 3+ state systems, with 6+ rate parameters and 3+ emission parameters. It would likely prove necessary to coarse-grain the system in some way in these cases, such as those sketched in the next section.

A separate but related direction involves extending *burstMCMC* to explicitly capture the hierarchical nature of transcriptional systems: while spatial-temporal distributions of regulatory inputs lead to similar mean behaviors between gene loci within neighboring cells, a wealth of literature has also the presence of extensive cell-to-cell variability (see, e.g., [81] and also Chapter 4). To capture this reality, the structure of *burstMCMC* could be updated to be explicitly hierarchical, assuming that transcriptional burst parameters in cells from the same spatial or temporal group are "drawn" from the same distributions, but allowing for each cell, i to have a unique set of parameter values, $\boldsymbol{\theta}_i$. For instance, we could stipulate that the emission rate for state 2, v_2 for each cell is drawn from a normal distribution that has a mean value, $\overline{v_2}$ that is equivalent the v_2 estimate returned by standard *burstMCMC*, but which also has a finite spread, σ_{v2}, that indicates how tightly constrained this emission value is within different populations of cells.

Two potential applications of this hierarchical approach would be (i) achieving superior single-trace estimates of $\boldsymbol{z}$ and (ii) allowing for practitioners to query not only which parameters are regulated *on average* but, also, the degree to which different aspects of transcriptional systems vary within cells that experience similar transcriptional inputs. In other words, how tightly controlled (or not) is each burst parameter? It would be intriguing, for instance, to ask this question of sister gene loci in two spot experiments that are derived from same cell (see Chapter 4). Indeed, the very concept of doing this problematicizes the concept of intrinsic noise as defined in Chapter 4 (and originally in [81]), which assumes that burst parameters themselves are fixed values within each population gene loci considered.

7.3.2.1 Using first passage times to connect burst parameters to molecular models

In principle, the same approaches that *burstMCMC* uses to estimate burst parameter values could be extended to estimate full molecular models like those examined in Chapters 5 and

6. However, in practice, it would likely be impractical to sample every molecular reaction explicitly. There may yet be a way to incorporate molecular models into burst parameter inference, however. In Chapters 5 and 6 we developed analytical methods to calculate effective bursting timescales from more detailed molecular models. Appendix C.2.4.4 in particular provides a generic method for calculating effective two state off and on rates (k^*_{off} and k^*_{on}) for arbitrarily complex molecular models. These effective parameters could be used to sample a set of molecular parameters, $\boldsymbol{m}$ conditional on current burst parameter estimates: $p(\boldsymbol{m}|\boldsymbol{A}, \boldsymbol{v}, \sigma, \boldsymbol{z})$. Likewise, updated estimates for for $\boldsymbol{z}$ could be made to reflect $\boldsymbol{m}$.

Of course, information is lost in looking only at the mean values of these effective parameters, since the distribution of waiting times to exit different activity states need not always be exponential, depending on the molecular model $\boldsymbol{m}$ (see, e.g. Section 5.4 in Chapter 5). One possible way around this would be to estimate passage times empirically from $\boldsymbol{z}$ and use this to constrain the molecular model parameters $\boldsymbol{m}$. A second would be to increase the number of activity states, K, until observed first passage times for each state are approximately exponential. This would indicate that the model size matches the number of rate-limiting steps present in the underlying transcriptional cycle.

Altogether, we believe that the improvements to *burstMCMC* sketched above will lead to a flexible and robust framework for inferring kinetic models from live imaging data. Ultimately, we hope that the method can serve as a statistical bridge between raw fluorescent traces, transcriptional bursting, and truly molecular models of transcriptional control.

Bibliography

[1] Simon Alamos et al. "Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics". In: *Nature Plants 2021 7:8* 7.8 (Aug. 2021), pp. 1037–1049. ISSN: 2055-0278. DOI: 10.1038/s41477-021-00976-0. URL: https://www.nature.com/articles/s41477-021-00976-0.

[2] et al. Alberts B, Johnson A, Lewis J. "Drosophila and the Molecular Genetics of Pattern Formation: Genesis of the Body Plan." In: *Molecular Biology of the Cell.* 2002. URL: https://www.ncbi.nlm.nih.gov/books/NBK26906/.

[3] Md Zulfikar Ali et al. "Probing Mechanisms of Transcription Elongation Through Cell-to-Cell Variability of RNA Polymerase". In: *Biophysical Journal* 118.7 (Apr. 2020), pp. 1769–1781. ISSN: 15420086. DOI: 10.1016/j.bpj.2020.02.002. URL: http://www.cell.com/article/S0006349520301193/fulltext%20http://www.cell.com/article/S0006349520301193/abstract%20https://www.cell.com/biophysj/abstract/S0006-3495(20)30119-3.

[4] C. David Allis et al. *Epigenetics*. Second edition. Cold Spring Harbor, New York: CSH Press, Cold Spring Harbor Laboratory Press, 2015, xiv, 984 pages. ISBN: 9781936113590 (hardcover).

[5] Ignacio Arganda-Carreras et al. "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification". In: *Bioinformatics* 33.15 (Aug. 2017), pp. 2424–2426. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTX180. URL: https://academic.oup.com/bioinformatics/article/33/15/2424/3092362.

[6] D. N. Arnosti et al. "The eve stripe 2 enhancer employs multiple modes of transcriptional synergy". In: *Development* 122.1 (1996), pp. 205–14.

[7] C. R. Bartman et al. "Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping". In: *Mol Cell* 62.2 (2016), pp. 237–247. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2016.03.007.

[8] Caroline R. Bartman et al. "Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation". In: *Molecular Cell* 73.3 (Feb. 2019), 519–532.e4. ISSN: 10974164. DOI: 10.1016/j.molcel.2018.11.004. URL: https://doi.org/10.1016/j.molcel.2018.11.004.

[9] Golan Bel, Brian Munsky, and Ilya Nemenman. "The simplicity of completion time distributions for common complex biochemical processes". In: *Physical Biology* 7.1 (2010). ISSN: 14783975. DOI: 10.1088/1478-3975/7/1/016003.

[10] Nezha S. Benabdallah et al. "Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation". In: *Molecular Cell* 76.3 (2019), 473–484.e7. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.07.038. URL: https://doi.org/10.1016/j.molcel.2019.07.038.

[11] Augusto Berrocal et al. "Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene". In: *bioRxiv* (June 2018), p. 335901. DOI: 10.1101/335901. URL: https://www.biorxiv.org/content/10.1101/335901v2.full.

[12] Augusto Berrocal et al. "Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene". In: *bioRxiv* (June 2018), p. 335901. DOI: 10.1101/335901. URL: https://www.biorxiv.org/content/10.1101/335901v2.full.

[13] Augusto Berrocal et al. "Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene". In: *eLife* 9 (Dec. 2020). DOI: 10.7554/elife.61635.

[14] E Bertrand et al. "Localization of ASH1 mRNA particles in living yeast". In: *Mol Cell* 2.4 (1998), pp. 437–445. ISSN: 1097-2765 (Print) 1097-2765 (Linking). DOI: S1097-2765(00)80143-4[pii].

[15] John W Biddle, Maximilian Nguyen, and Jeremy Gunawardena. "Negative reciprocity , not ordered assembly , underlies the interaction of Sox2 and Oct4 on DNA". In: (2019), pp. 1–30.

[16] John W. Biddle et al. *Allosteric conformational ensembles have unlimited capacity for integrating information*. Dec. 2020. DOI: 10.1101/2020.12.10.420117. URL: https://doi.org/10.1101/2020.12.10.420117.

[17] Lacramioara Bintu et al. "Dynamics of epigenetic regulation at the single-cell level". In: *Science* 351.6274 (2016), pp. 720–724. ISSN: 10959203. DOI: 10.1126/science.aab2956.

[18] Christopher M Bishop. *Pattern recognition and machine learning*. New York : Springer, [2006] ©2006. URL: https://search.library.wisc.edu/catalog/9910032530902121.

[19] W. J. Blake et al. "Phenotypic consequences of promoter-mediated transcriptional noise". In: *Mol Cell* 24.6 (2006), pp. 853–65. ISSN: 1097-2765 (Print) 1097-2765 (Linking). DOI: S1097-2765(06)00744-1[pii]10.1016/j.molcel.2006.11.003.

[20] J. T. Blankenship and E. Wieschaus. "Two new roles for the Drosophila AP patterning system in early morphogenesis". In: *Development* 128.24 (2001), pp. 5129–38. ISSN: 0950-1991 (Print) 0950-1991 (Linking).

[21] Hinrich Boeger et al. *From structural variation of gene molecules to chromatin dynamics and transcriptional bursting*. June 2015. DOI: 10.3390/genes6030469.

[22] A. N. Boettiger and M. Levine. "Rapid Transcription Fosters Coordinate snail Expression in the Drosophila Embryo". In: *Cell Rep* (2013). ISSN: 2211-1247 (Electronic). DOI: 10.1016/j.celrep.2012.12.015.

[23] J. P. Bothma et al. "Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo". In: *Elife* 4 (2015). ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.07956.

[24] J. P. Bothma et al. "LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos". In: *Cell* (2018). ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2018.03.069.

[25] Jacques Bothma and Michael Levine. "Development: Lights, Camera, Action - The Drosophila Embryo Goes Live!" In: *Curr Biol* 23 (2013), R965.

[26] Jacques P Bothma et al. "Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos". In: *Proceedings of the National Academy of Sciences* 111.29 (2014), pp. 10598–10603. ISSN: 0027-8424. DOI: 10.1073/pnas.1410022111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1410022111%7B%5C%%7D5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/24994903.

[27] Jacques P. Bothma et al. "Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.29 (July 2014), pp. 10598–10603. ISSN: 10916490. DOI: 10.1073/pnas.1410022111.

[28] Jonathan R Bowles et al. "Scalable inference of transcriptional kinetic parameters from MS2 time series data". In: *Bioinformatics* 38.4 (Jan. 2022), pp. 1030–1036. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTAB765. URL: https://academic.oup.com/bioinformatics/article/38/4/1030/6426074.

[29] Gregory D. Bowman and Michael G. Poirier. "Post-translational modifications of histones that influence nucleosome dynamics". In: *Chemical Reviews* 115.6 (2015), pp. 2274–2295. ISSN: 15206890. DOI: 10.1021/cr500350x.

[30] Simon M.G. Braun et al. "Rapid and reversible epigenome editing by endogenous chromatin regulators". In: *Nature Communications* 8.1 (2017). ISSN: 20411723. DOI: 10.1038/s41467-017-00644-y. URL: http://dx.doi.org/10.1038/s41467-017-00644-y.

[31] L. Bronstein, C. Zechner, and H. Koeppl. "Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population". In: *Methods* 85 (2015), pp. 22–35. ISSN: 1095-9130 (Electronic) 1046-2023 (Linking). DOI: 10.1016/j.ymeth.2015.05.012.

[32] Martin S. Buckley et al. "Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation". In: *Genes and Development* 28.1 (2014), pp. 14–19. ISSN: 08909369. DOI: 10.1101/gad.231886.113.

[33] Long Cai, Nir Friedman, and X. Sunney Xie. "Stochastic protein expression in individual cells at the single molecule level". In: *Nature* 440.7082 (Mar. 2006), pp. 358–362. ISSN: 14764687. DOI: 10.1038/nature04599. URL: https://www.nature.com/articles/nature04599.

[34] George Casella and Edward I. George. "Explaining the Gibbs Sampler". In: *The American Statistician* 46.3 (Aug. 1992), p. 167. ISSN: 00031305. DOI: 10.2307/2685208.

[35] Sarah A. Cepeda-Humerez, Georg Rieckh, and Gašper Tkačik. "Stochastic Proofreading Mechanism Alleviates Crosstalk in Transcriptional Regulation". In: *Physical Review Letters* 115.24 (2015), pp. 1–5. ISSN: 10797114. DOI: 10.1103/PhysRevLett.115.248101.

[36] Jeffrey A. Chao et al. "Structural basis for the coevolution of a viral RNA-protein complex". In: *Nature Structural and Molecular Biology* 15.1 (Jan. 2008), pp. 103–105. ISSN: 15459993. DOI: 10.1038/nsmb1327.

[37] Varodom Charoensawan, Derek Wilson, and Sarah A. Teichmann. "Genomic repertoires of DNA-binding transcription factors across the tree of life". In: *Nucleic Acids Research* 38.21 (Nov. 2010), p. 7364. DOI: 10.1093/NAR/GKQ617. URL: /pmc/articles/PMC2995046/%20/pmc/articles/PMC2995046/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995046/.

[38] Hongtao Chen et al. "Dynamic interplay between enhancer–promoter topology and gene activity". In: *Nature Genetics* 50.9 (Sept. 2018), pp. 1296–1303. ISSN: 1061-4036. DOI: 10.1038/s41588-018-0175-z. URL: http://www.nature.com/articles/s41588-018-0175-z.

[39] Jiji Chen et al. "Single-molecule dynamics of enhanceosome assembly in embryonic stem cells". In: *Cell* 156.6 (Mar. 2014), pp. 1274–1285. ISSN: 10974172. DOI: 10.1016/j.cell.2014.01.062.

[40] Xuanze Chen et al. "Study of RNA Polymerase II Clustering inside Live-Cell Nuclei Using Bayesian Nanoscopy". In: *ACS Nano* 10.2 (2016), pp. 2447–2454. ISSN: 1936086X. DOI: 10.1021/acsnano.5b07257.

[41] A. Chestier and M. Yaniv. "Rapid turnover of acetyl groups in the four core histones of simian virus 40 minichromosomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 76.1 (1979), pp. 46–50. ISSN: 00278424. DOI: 10.1073/pnas.76.1.46.

[42] Won Ki Cho et al. "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates". In: *Science* 361.6400 (2018), pp. 412–415. ISSN: 10959203. DOI: 10.1126/science.aar4199.

[43] Won-Ki Cho et al. "RNA Polymerase II cluster dynamics predict mRNA output in living cells." In: *eLife* 5 (2016). ISSN: 2050-084X. DOI: 10.7554/eLife.13617. URL: http://www.ncbi.nlm.nih.gov/pubmed/27138339%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4929003.

[44] Shasha Chong et al. "Mechanism of transcriptional bursting in bacteria". In: *Cell* 158.2 (July 2014), pp. 314–326. ISSN: 10974172. DOI: 10.1016/j.cell.2014.05.038.

[45] Sandeep Choubey. "Nascent RNA kinetics: Transient and steady state behavior of models of transcription". In: *Physical Review E* 97 (2018). DOI: 10.1103/PhysRevE.97.022402. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.97.022402.

[46] Sandeep Choubey, Jane Kondev, and Alvaro Sanchez. "Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules". In: *PLOS Computational Biology* 11.11 (Nov. 2015). Ed. by Mustafa Khammash, e1004345. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004345. URL: https://dx.plos.org/10.1371/journal.pcbi.1004345.

[47] Sandeep Choubey, Jane Kondev, and Alvaro Sanchez. "Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells". In: *Biophysj* 114.9 (2018), pp. 2072–2082. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.03.031. URL: https://doi.org/10.1016/j.bpj.2018.03.031.

[48] Mudra Choudhury and Stephen A. Ramsey. "Identifying Cell Type-Specific Transcription Factors by Integrating ChIP-seq and eQTL Data–Application to Monocyte Gene Regulation". In: *Gene Regulation and Systems Biology* 10 (Dec. 2016), p. 105. ISSN: 11776250. DOI: 10.4137/GRSB.S40768. URL: /pmc/articles/PMC5156548/%20/pmc/articles/PMC5156548/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5156548/.

[49] Robert Christian and Jean-Michell Marin. *Bayesian Essentials with R Second Edition*. Ed. by George Casella et al. Swcond. Springer, 2013. URL: http://www.springer.com/series/417.

[50] Jonathan R. Chubb et al. "Transcriptional Pulsing of a Developmental Gene". In: *Current Biology* 16.10 (2006), pp. 1018–1025. ISSN: 09609822. DOI: 10.1016/j.cub.2006.03.092.

[51] Ibrahim I. Cisse et al. "Real-time dynamics of RNA polymerase II clustering in live human cells". In: *Science* 341.6146 (2013), pp. 664–667. ISSN: 10959203. DOI: 10.1126/science.1239053.

[52] D. E. Clyde et al. "A self-organizing system of repressor gradients establishes segmental complexity in Drosophila". In: *Nature* 426.6968 (2003), pp. 849–53. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature02189.

[53] Leighton J. Core, Joshua J. Waterfall, and John T. Lis. "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters". In: *Science* (2008). ISSN: 00368075. DOI: 10.1126/science.1162228.

[54] Adam M Corrigan et al. "A continuum model of transcriptional bursting." In: *eLife* 5 (2016), e13051. ISSN: 2050-084X. DOI: 10.7554/eLife.13051. arXiv: arXiv:1011.1669v3. URL: http://elifesciences.org/content/5/e13051.

[55] Adam M. Corrigan et al. "A continuum model of transcriptional bursting". In: *eLife* 5.FEBRUARY2016 (Feb. 2016). ISSN: 2050084X. DOI: 10.7554/eLife.13051.

[56] A. Coulon and D. R. Larson. "Fluctuation Analysis Dissecting Transcriptional Kinetics with Signal Theory". In: *Methods in Enzymology*. Vol. 572. Academic Press Inc., 2016, pp. 159–191. DOI: 10.1016/bs.mie.2016.03.017.

[57] A. Coulon et al. "Kinetic competition during the transcription cycle results in stochastic RNA processing". In: *Elife* 3 (2014). ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.03939.

[58] Antoine Coulon et al. "Eukaryotic transcriptional dynamics: From single molecules to cell populations". In: *Nature Reviews Genetics* 14.8 (2013), pp. 572–584. ISSN: 14710056. DOI: 10.1038/nrg3484.

[59] Antoine Coulon et al. "Eukaryotic transcriptional dynamics: from single molecules to cell populations." In: *Nature reviews. Genetics* 14.8 (Aug. 2013), pp. 572–84. ISSN: 1471-0064. DOI: 10.1038/nrg3484. URL: http://www.ncbi.nlm.nih.gov/pubmed/23835438%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3807637.

[60] A. J. Courey and S. Jia. "Transcriptional repression: the long and the short of it". In: *Genes Dev* 15.21 (2001), pp. 2786–96.

[61] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. 2006, p. 776. ISBN: 0471241954. URL: https://www.bibsonomy.org/bibtex/22e9bfa879286689a14feb55b69d326c1/ywhuang%20http://www.amazon.com/Elements-Information-Edition-Telecommunications-Processing/dp/0471241954.

[62] J. Crocker, G. R. Ilsley, and D. L. Stern. "Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors". In: *Nat Genet* 48.3 (2016), pp. 292–8. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.3509.

[63] Justin Crocker and David L. Stern. "TALE-mediated modulation of transcriptional enhancers in vivo". In: *Nature methods* 10.8 (Aug. 2013), pp. 762–767. ISSN: 1548-7105. DOI: 10.1038/NMETH.2543. URL: https://pubmed.ncbi.nlm.nih.gov/23817068/.

[64] Justin Crocker et al. "Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness". In: *Cell* 160.0 (Jan. 2015), p. 191. ISSN: 10974172. DOI: 10.1016/J.CELL.2014.11.041. URL: /pmc/articles/PMC4449256/%20/pmc/articles/PMC4449256/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4449256/.

[65] J D'ans et al. *Application of The Theory of Diffusion-Controlled Reactions to Enzym Kinetics*. Tech. rep. 26. 1952, p. 532. URL: https://pubs.acs.org/sharingguidelines.

[66] Xavier Darzacq et al. "In vivo dynamics of RNA polymerase II transcription". In: *Nature Structural and Molecular Biology* 14.9 (2007), pp. 796–806. ISSN: 15459993. DOI: 10.1038/nsmb1280.

[67] Roger B Deal, Jorja G Henikoff, and Steven Henikoff. "Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones". In: *Science* 328.5982 (2010), pp. 1161–1165. ISSN: 1095-9203. DOI: 10.1126/science.1186777. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2879085%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[68] J. Desponds et al. "Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos". In: *PLoS Comput Biol* 12.12 (2016), e1005256. ISSN: 1553-7358 (Electronic) 1553-734X (Linking). DOI: 10.1371/journal.pcbi.1005256.

[69] Jonathan Desponds, Massimo Vergassola, and Aleksandra M Walczak. "A mechanism for hunchback promoters to readout morphogenetic positional information in less than a minute". In: *eLife* 9 (July 2020). ISSN: 2050084X. DOI: 10.7554/elife.49758.

[70] Jonathan Desponds et al. "Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos". In: *PLOS Computational Biology* 12.12 (2016), e1005256. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005256. URL: http://dx.plos.org/10.1371/journal.pcbi.1005256.

[71] Frank Devilbiss and Doraiswami Ramkrishna. "Addressing the Need for a Model Selection Framework in Systems Biology Using Information Theory". In: *Proceedings of the IEEE* 105.2 (2017), pp. 330–339. ISSN: 15582256. DOI: 10.1109/JPROC.2016.2560121.

[72] Wolfgang Driever and Christiane Nüsslein-Volhard. "The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner". In: *Cell* 54.1 (July 1988), pp. 95–104. URL: http://linkinghub.elsevier.com/retrieve/pii/0092867488901833%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/43/432F230F-D9C9-4076-B770-80CC2DD4E9C3%20papers3://publication/doi/10.1016/0092-8674(88)90183-3.

[73] Julien O Dubuis, Reba Samanta, and Thomas Gregor. "Accurate measurements of dynamics and reproducibility in small genetic networks". In: *Molecular systems biology* 9.639 (2013), pp. 1–15. ISSN: 1744-4292. DOI: 10.1038/msb.2012.72. URL: http://dx.doi.org/10.1038/msb.2012.72%7B%5C%%7D5Cnpapers2://publication/doi/10.1038/msb.2012.72.

[74] Julien O Dubuis et al. "Positional information, in bits". In: *Proceedings of the National Academy of Sciences* 110.41 (Oct. 2013), pp. 16301–16308. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1315642110%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/1D/1D2E4D81-89C8-4F72-8048-E2855B9ACFDE%20papers3://publication/doi/10.1073/pnas.1315642110.

[75] Jeremy Dufourt et al. "Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs". In: *Nature Communications* 9.1 (Dec. 2018), p. 5194. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07613-z. URL: http://www.nature.com/articles/s41467-018-07613-z.

[76] Elizabeth Eck et al. "Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility". In: *bioRxiv* (Mar. 2020), p. 2020.01.27.922054. DOI: 10.1101/2020.01.27.922054. URL: https://doi.org/10.1101/2020.01.27.922054.

[77] Elizabeth Eck et al. "Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility". In: *bioRxiv* (Mar. 2020), p. 2020.01.27.922054. DOI: 10.1101/2020.01.27.922054. URL: https://doi.org/10.1101/2020.01.27.922054.

[78] B. A. Edgar, G. M. Odell, and G. Schubiger. "Cytoarchitecture and the patterning of fushi tarazu expression in the Drosophila blastoderm". In: *Genes Dev* 1.10 (1987), pp. 1226–37. ISSN: 0890-9369 (Print) 0890-9369 (Linking).

[79] Bruce A. Edgar et al. "Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo". In: *Cell* 47.5 (Dec. 1986), pp. 747–754. ISSN: 00928674. DOI: 10.1016/0092-8674(86)90517-9.

[80] Avigdor Eldar and Michael B. Elowitz. *Functional roles for noise in genetic circuits*. Sept. 2010. DOI: 10.1038/nature09326. URL: /pmc/articles/PMC4100692/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4100692/.

[81] Michael B. Elowitz et al. "Stochastic gene expression in a single cell". In: *Science (New York, N.Y.)* 297.5584 (Aug. 2002), pp. 1183–1186. ISSN: 1095-9203. DOI: 10.1126/SCIENCE.1070919. URL: https://pubmed.ncbi.nlm.nih.gov/12183631/.

[82] Javier Estrada et al. "Information Integration and Energy Expenditure in Gene Regulation." In: *Cell* 166.1 (June 2016), pp. 234–44. ISSN: 1097-4172. DOI: 10.1016/j.cell.2016.06.012. URL: http://www.ncbi.nlm.nih.gov/pubmed/27368104%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4930556.

[83] Javier Estrada et al. "SiteOut: An Online Tool to Design Binding Site-Free DNA Sequences". In: *PLOS ONE* 11.3 (Mar. 2016), e0151740. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0151740. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0151740.

[84] W. D. Fakhouri et al. "Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo". In: *Mol Syst Biol* 6 (2010), p. 341. ISSN: 1744-4292 (Electronic) 1744-4292 (Linking). DOI: msb200997[pii]10.1038/msb.2009.97.

[85] Julia Falo-Sanjuan et al. "Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling Correspondence". In: *Developmental Cell* 50 (2019), pp. 411–425. DOI: 10.1016/j.devcel.2019.07.002. URL: https://doi.org/10.1016/j.devcel.2019.07.002.

[86] Emma K. Farley et al. "Suboptimization of developmental enhancers". In: *Science (New York, N.Y.)* 350.6258 (Oct. 2015), p. 325. ISSN: 10959203. DOI: 10.1126/SCIENCE.AAC6948. URL: /pmc/articles/PMC4970741/%20/pmc/articles/PMC4970741/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970741/.

[87] K. Featherstone et al. "Spatially coordinated dynamic gene transcription in living pituitary tissue". In: *Elife* 5 (2016), e08494. ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.08494.

[88] Andrea M Femino et al. *Visualization of Single RNA Transcripts in Situ*. Tech. rep. 5363. 1998, pp. 585–590.

[89] Matthew P. Fish et al. "Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid". In: *Nature protocols* 2.10 (2007), pp. 2325–2331. ISSN: 1750-2799. DOI: 10.1038/NPROT.2007.328. URL: https://pubmed.ncbi.nlm.nih.gov/17947973/.

[90] V. E. Foe and B. M. Alberts. "Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis". In: *Journal of cell science* 61 (1983), pp. 31–70. ISSN: 0021-9533. DOI: 10.1242/JCS.61.1.31. URL: https://pubmed.ncbi.nlm.nih.gov/6411748/.

[91] K. M. Forrest and E. R. Gavis. "Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila". In: *Curr Biol* 13.14 (2003), pp. 1159–68. ISSN: 0960-9822 (Print) 0960-9822 (Linking). DOI: S0960982203004512[pii].

[92] C. C. Fowlkes et al. "A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm". In: *Cell* 133.2 (2008), pp. 364–74. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2008.01.053.

[93] M. Frasch and M. Levine. "Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila". In: *Genes Dev* 1.9 (1987), pp. 981–95. ISSN: 0890-9369 (Print) 0890-9369 (Linking).

[94] Tamar Friedlander et al. "Intrinsic limits to gene regulation by global crosstalk". In: *Nature Communications 2016 7:1* 7.1 (Aug. 2016), pp. 1–12. ISSN: 2041-1723. DOI: 10.1038/ncomms12307. URL: https://www.nature.com/articles/ncomms12307.

[95] Miki Fujioka, James B. Jaynes, and Tadaatsu Goto. "Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression". In: *Development (Cambridge, England)* 121.12 (Dec. 1995), pp. 4371–4382. ISSN: 0950-1991. DOI: 10.1242/DEV.121.12.4371. URL: https://pubmed.ncbi.nlm.nih.gov/8575337/.

[96] Miki Fujioka et al. "Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients". In: *Development (Cambridge, England)* 126.11 (1999), pp. 2527–2538. ISSN: 0950-1991. DOI: 10.1242/DEV.126.11.2527. URL: https://pubmed.ncbi.nlm.nih.gov/10226011/.

[97] Takashi Fukaya, Bomyi Lim, and Michael Levine. "Enhancer Control of Transcriptional Bursting". In: *Cell* 166.2 (July 2016), pp. 358–368. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2016.05.025. URL: https://www.sciencedirect.com/science/article/pii/S0092867416305736.

[98] Takashi Fukaya, Bomyi Lim, and Michael Levine. "Enhancer Control of Transcriptional Bursting." In: *Cell* 166.2 (July 2016), pp. 358–368. ISSN: 1097-4172. DOI: 10.1016/j.cell.2016.05.025. URL: http://www.ncbi.nlm.nih.gov/pubmed/27293191%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4970759.

[99] Takashi Fukaya, Bomyi Lim, and Michael Levine. "Rapid Rates of Pol II Elongation in the Drosophila Embryo." In: *Current biology : CB* 27.9 (May 2017), pp. 1387–1391. ISSN: 1879-0445. DOI: 10.1016/j.cub.2017.03.069. URL: http://www.ncbi.nlm.nih.gov/pubmed/28457866%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5665007.

[100] J. C. Fung et al. "Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations". In: *J Cell Biol* 141.1 (1998), pp. 5–20. ISSN: 0021-9525 (Print) 0021-9525 (Linking).

[101] Eileen E.M. Furlong and Michael Levine. "Developmental enhancers and chromosome topology". In: *Science* 361.6409 (2018), pp. 1341–1345. ISSN: 10959203. DOI: 10.1126/science.aau0320.

[102] Bjoern Gaertner and Julia Zeitlinger. "RNA polymerase II pausing during development". In: *Development (Cambridge)* 141.6 (2014), pp. 1179–1183. ISSN: 09501991. DOI: 10.1242/dev.088492.

[103] Hernan G. Garcia and Thomas Gregor. "Live Imaging of mRNA Synthesis in Drosophila". In: *Methods in molecular biology (Clifton, N.J.)* 1649 (2018), pp. 349–357. ISSN: 1940-6029. DOI: 10.1007/978-1-4939-7213-5_23. URL: https://pubmed.ncbi.nlm.nih.gov/29130209/.

[104] Hernan G. Garcia et al. "Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning." In: *Current biology* 23.21 (2013), pp. 2140–2145. ISSN: 18790445. DOI: 10.1016/j.cub.2013.08.054. URL: http://dx.doi.org/10.1016/j.cub.2013.08.054.

[105] J. Christof M. Gebhardt et al. "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells". In: *Nature Methods* (2013). ISSN: 15487091. DOI: 10.1038/nmeth.2411.

[106] Charles J Geyer. *Introduction to Markov Chain Monte Carlo*. Tech. rep. 2011.

[107] Charles J. Geyer. "Practical markov chain monte carlo". In: *Statistical Science* 7.4 (Nov. 1992), pp. 473–483. ISSN: 08834237. DOI: 10.1214/ss/1177011137. URL: https://projecteuclid.org/journals/statistical-science/volume-7/issue-4/Practical-Markov-Chain-Monte-Carlo/10.1214/ss/1177011137.full%20https://projecteuclid.org/journals/statistical-science/volume-7/issue-4/Practical-Markov-Chain-Monte-Carlo/10.1214/ss/117.

[108] J M Ghuysen et al. "Dynamics of Replication-Independent Histone Turnover in Budding Yeast". In: *Science* 315.March (2007), pp. 1405–1409.

[109] D. T. Gillespie. "General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. ISSN: 0021-9991.

[110] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *Journal of Physical Chemistry*. Vol. 81. 25. American Chemical Society, 1977, pp. 2340–2361. DOI: 10.1021/j100540a008. URL: https://pubs.acs.org/sharingguidelines.

[111] I Golding et al. "Real-time kinetics of gene activity in individual bacteria". In: *Cell* 123.6 (2005), pp. 1025–1036. ISSN: 0092-8674 (Print).

[112] Tadaatsu Goto, Paul Macdonald, and Tom Maniatis. "Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues". In: *Cell* 57.3 (May 1989), pp. 413–422. ISSN: 00928674. DOI: 10.1016/0092-8674(89)90916-1.

[113] Rok Grah, Benjamin Zoller, and Gašper Tkačik. "Nonequilibrium models of optimal enhancer function". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.50 (Dec. 2020), pp. 31614–31622. ISSN: 10916490. DOI: 10.1073/PNAS.2006731117/-/DCSUPPLEMENTAL. URL: https://www.pnas.org/content/117/50/31614%20https://www.pnas.org/content/117/50/31614.abstract.

[114] Rok Grah, Benjamin Zoller, and Gašper Tkačik. "Normative models of enhancer function". In: *bioRxiv* (Apr. 2020), p. 2020.04.08.029405. DOI: 10.1101/2020.04.08.029405. URL: https://doi.org/10.1101/2020.04.08.029405.

[115] S. Gray, P. Szymanski, and M. Levine. "Short-Range Repression Permits Multiple Enhancers to Function Autonomously within a Complex Promoter". In: *Genes Development* 8.15 (1994), pp. 1829–1838. ISSN: 0890-9369.

[116] Thomas Gregor et al. "Probing the Limits to Positional Information". In: *Cell* 130.1 (July 2007), pp. 153–164. ISSN: 00928674. DOI: 10.1016/j.cell.2007.05.025. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867407006629.

[117] Thomas Gregor et al. "Probing the Limits to Positional Information". In: *Cell* 130.1 (July 2007), pp. 153–164. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2007.05.025. URL: https://www.sciencedirect.com/science/article/pii/S0092867407006629.

[118] Bo Gu et al. "Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements". In: *Science* 359.6379 (2018), pp. 1050–1055. ISSN: 10959203. DOI: 10.1126/science.aao3136.

[119] Thomas Guilmeau, Emilie Chouzenoux, and Victor Elvira. "Simulated Annealing: A Review and a New Scheme". In: *IEEE Workshop on Statistical Signal Processing Proceedings* 2021-July (2021), pp. 101–105. DOI: 10.1109/SSP49050.2021.9513782.

[120] Sanjana Gupta et al. "Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology". In: *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018* (2018), pp. 690–697. DOI: 10.1109/PDP2018.2018.00114. arXiv: 1801.09831.

[121] Keren Bahar Halpern. "Bursty Gene Expression in the Intact Mammalian Liver". In: *Molecular Cell* 58 (2015), pp. 147–156. DOI: 10.1016/j.molcel.2015.01.027. URL: http://dx.doi.org/10.1016/j.molcel.2015.01.027.

[122] J. M. Halstead et al. "Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals". In: *Science* 347.6228 (2015), pp. 1367–671. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aaa3380.

[123] Petter Hammar et al. "Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation". In: *Nature genetics* 46.4 (2014), p. 405. ISSN: 15461718. DOI: 10.1038/NG.2905. URL: /pmc/articles/PMC6193529/%20/pmc/articles/PMC6193529/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193529/.

[124] Anders S. Hansen et al. "CTCF and cohesin regulate chromatin loop stability with distinct dynamics". In: *eLife* 6 (2017), pp. 1–33. ISSN: 2050084X. DOI: 10.7554/eLife.25776.

[125] K. Harding et al. "Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila." In: *The EMBO Journal* 8.4 (1989), p. 1205. ISSN: 02614189. DOI: 10.1002/j.1460-2075.1989.tb03493.x. URL: /pmc/articles/PMC400935/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC400935/.

[126] Nathaniel A. Hathaway et al. "Dynamics and memory of heterochromatin in living cells". In: *Cell* 149.7 (2012), pp. 1447–1460. ISSN: 00928674. DOI: 10.1016/j.cell.2012.03.052. URL: http://dx.doi.org/10.1016/j.cell.2012.03.052.

[127] Tyler Heist, Takashi Fukaya, and Michael Levine. "Large distances separate coregulated genes in living Drosophila embryos". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.30 (2019), pp. 15062–15067. ISSN: 10916490. DOI: 10.1073/pnas.1908962116.

[128] O. Hendy et al. "Differential context-specific impact of individual core promoter elements on transcriptional dynamics". In: *Mol Biol Cell* 28.23 (2017), pp. 3360–3370. ISSN: 1939-4586 (Electronic) 1059-1524 (Linking). DOI: 10.1091/mbc.E17-06-0408.

[129] Telmo Henriques et al. "Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals". In: *Molecular Cell* 52.4 (2013), pp. 517–528. ISSN: 10972765. DOI: 10.1016/j.molcel.2013.10.001. URL: http://dx.doi.org/10.1016/j.molcel.2013.10.001.

[130] Telmo Henriques et al. "Widespread transcriptional pausing and elongation control at enhancers". In: *Genes and Development* 32.1 (2018), pp. 26–41. ISSN: 15495477. DOI: 10.1101/gad.309351.117.

[131] Gilbert L. Henry et al. "Cell type–specific genomics of Drosophila neurons". In: *Nucleic Acids Research* 40.19 (Oct. 2012), p. 9691. ISSN: 03051048. DOI: 10.1093/NAR/GKS671. URL: /pmc/articles/PMC3479168/%20/pmc/articles/PMC3479168/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3479168/.

[132] K. L. Hey et al. "A stochastic transcriptional switch model for single cell imaging data". In: *Biostatistics* 16.4 (2015), pp. 655–69. ISSN: 1468-4357 (Electronic) 1465-4644 (Linking). DOI: 10.1093/biostatistics/kxv010.

[133] Terrell L. Hill. *Free Energy Transduction and Biochemical Cycle Kinetics*. Springer New York, 1989, p. 119. ISBN: 9781461235583.

[134] Denes Hnisz et al. "A Phase Separation Model for Transcriptional Control". In: *Cell* 169.1 (Mar. 2017), pp. 13–23. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2017.02.007. URL: https://www.sciencedirect.com/science/article/pii/S009286741730185X?via%7B%5C%%7D3Dihub.

[135] David M. Holloway and Alexander V. Spirov. "Transcriptional bursting in Drosophila development: Stochastic dynamics of eve stripe 2 expression". In: *PLOS ONE* 12.4 (Apr. 2017). Ed. by Barbara Jennings, e0176228. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0176228. URL: http://dx.plos.org/10.1371/journal.pone.0176228.

[136] J J Hopfield. "Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity (protein synthesis/DNA replication/amino-acid recognition)". In: 71.10 (1974), pp. 4135–4139. URL: http://www.pnas.org/content/71/10/4135.full.pdf.

[137] A. Huang et al. "Decoding temporal interpretation of the morphogen Bicoid in the early Drosophila embryo". In: *Elife* 6 (2017). ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.26258.

[138] E. Jane Albert Hubbard. "The C. elegans germ line: a model for stem cell biology". In: *Developmental dynamics : an official publication of the American Association of Anatomists* 236.12 (Dec. 2007), p. 3343. ISSN: 10588388. DOI: 10.1002/DVDY.21335. URL: /pmc/articles/PMC2949268/%20/pmc/articles/PMC2949268/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2949268/.

[139] Garth R Ilsley et al. "Cellular resolution models for even skipped regulation in the entire Drosophila embryo." In: *eLife* 2 (Aug. 2013), e00522. ISSN: 2050-084X. DOI: 10.7554/eLife.00522. URL: http://www.ncbi.nlm.nih.gov/pubmed/23930223%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3736529.

[140] V. Jackson et al. "Studies on highly metabolically active acetylation and phosphorylation of histones". In: *Jour nal of Biological Chemistry* 250.13 (1975), pp. 4856–4863. ISSN: 00219258.

[141] Johannes Jaeger. "The gap gene network." In: *Cellular and molecular life sciences : CMLS* 68.2 (Jan. 2011), pp. 243–74. ISSN: 1420-9071. DOI: 10.1007/s00018-010-0536-y. URL: http://www.ncbi.nlm.nih.gov/pubmed/20927566%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3016493.

[142] Johannes Jaeger et al. "Dynamic control of positional information in the early Drosophila embryo." In: *Nature* 430.6997 (July 2004), pp. 368–371. URL: http://www.nature.com/doifinder/10.1038/nature02678%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/EA/EA0F5F41-D1DB-4A22-8EF5-E5B2C5743FDC%20papers3://publication/doi/10.1038/nature02678.

[143] Johannes Jaeger et al. "Dynamic control of positional information in the early Drosophila embryo." In: *Nature* 430.6997 (July 2004), pp. 368–371. URL: http://www.nature.com/doifinder/10.1038/nature02678%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/EA/EA0F5F41-D1DB-4A22-8EF5-E5B2C5743FDC%20papers3://publication/doi/10.1038/nature02678.

[144] S. M. Janicki et al. "From silencing to gene expression: real-time analysis in single cells". In: *Cell* 116.5 (2004), pp. 683–98. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

[145] H. Janssens et al. "Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even skipped* gene". In: *Nat Genet* 38.10 (2006), pp. 1159–65. ISSN: 1061-4036 (Print).

[146] J. Jiang, T. Hoey, and M. Levine. "Autoregulation of a segmentation gene in Drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element". In: *Genes Dev* 5.2 (1991), pp. 265–77. ISSN: 0890-9369 (Print) 0890-9369 (Linking).

[147] J. Jiang and M. Levine. "Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen". In: *Cell* 72.5 (1993), pp. 741–52. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

[148] P. Jiang et al. "Natural variation of the expression pattern of the segmentation gene even-skipped in melanogaster". In: *Dev Biol* 405.1 (2015), pp. 173–81. ISSN: 1095-564X (Electronic) 0012-1606 (Linking). DOI: 10.1016/j.ydbio.2015.06.019.

[149] Daniel L. Jones, Robert C. Brewster, and Rob Phillips. "Promoter architecture dictates cell-to-cell variability in gene expression". In: *Science* 346.6216 (Dec. 2014), pp. 1533–1536. ISSN: 10959203. DOI: 10.1126/science.1255301.

[150] Iris Jonkers, Hojoong Kwak, and John T. Lis. "Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons". In: *eLife* 2014.3 (2014), pp. 1–25. ISSN: 2050084X. DOI: 10.7554/eLife.02407.

[151] Iris Jonkers and John T. Lis. "Getting up to speed with transcription elongation by RNA polymerase II". In: *Nature Reviews Molecular Cell Biology* 16.3 (2015), pp. 167–177. ISSN: 14710080. DOI: 10.1038/nrm3953.

[152] Ami M. Kabadi and Charles A. Gersbach. "Engineering Synthetic TALE and CRISPR/Cas9 Transcription Factors for Regulating Gene Expression". In: *Methods (San Diego, Calif.)* 69.2 (2014), p. 188. ISSN: 10959130. DOI: 10.1016/J.YMETH.2014.06.014. URL: /pmc/articles/PMC4175060/%20/pmc/articles/PMC4175060/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4175060/.

[153] Michael H. Kagey et al. "Mediator and cohesin connect gene expression and chromatin architecture". In: *Nature* 467.7314 (Sept. 2010), pp. 430–435. ISSN: 0028-0836. DOI: 10.1038/nature09380. URL: http://www.nature.com/articles/nature09380.

[154] Stefan R. Kassabov et al. "SWI/SNF unwraps, slides, and rewraps the nucleosome". In: *Molecular Cell* 11.2 (2003), pp. 391–403. ISSN: 10972765. DOI: 10.1016/S1097-2765(03)00039-X.

[155] Yael Katan-Khaykovich and Kevin Struhl. "Dynamics of global histone acetylation and deacetylation in vivo: Rapid restoration of normal histone acetylation status upon removal of activators and repressors". In: *Genes and Development* 16.6 (2002), pp. 743–752. ISSN: 08909369. DOI: 10.1101/gad.967302.

[156] T.B. Kepler and T.C. Elston. "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations." In: *Biophys J.* 81 (2001), pp. 3116–36.

[157] Soile V.E. Keränen et al. "Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution II: dynamics". In: *Genome Biology* 7.12 (Dec. 2006), R124. ISSN: 14747596. DOI: 10.1186/GB-2006-7-12-R124. URL: /pmc/articles/PMC1794437/%20/pmc/articles/PMC1794437/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1794437/.

[158] Yang Joon Kim et al. "Predictive modeling reveals that higher-order cooperativity drives transcriptional repression in a synthetic developmental enhancer". In: *bioRxiv* (2021), pp. 1–55. URL: https://doi.org/10.1101/2021.07.28.454075%7B%5C%%7D0Ahttps://www.biorxiv.org/content/10.1101/2021.07.28.454075v1.

[159] Hiroshi Kimura and Peter R. Cook. "Kinetics of core histones in living human cells: Little exchange of H3 and H4 and some rapid exchange of H2B". In: *Journal of Cell Biology* 153.7 (2001), pp. 1341–1353. ISSN: 00219525. DOI: 10.1083/jcb.153.7.1341.

[160] Minoru S.H. Ko. "A stochastic model for gene induction". In: *Journal of theoretical biology* 153.2 (Nov. 1991), pp. 181–194. ISSN: 0022-5193. DOI: 10.1016/S0022-5193(05)80421-7. URL: https://pubmed.ncbi.nlm.nih.gov/1787735/.

[161] Roger D. Kornberg. "The molecular basis of eukaryotic transcription". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.32 (Aug. 2007), pp. 12955–12961. ISSN: 00278424. DOI: 10.1073/PNAS.0704138104/ASSET/573DB0B8-41E5-4EFE-A261-EFFE5771389F/ASSETS/GRAPHIC/ZPQ0260768180018.JPEG. URL: https://www.pnas.org.

[162] Tony Kouzarides. "Chromatin Modifications and Their Function". In: *Cell* 128.4 (2007), pp. 693–705. ISSN: 00928674. DOI: 10.1016/j.cell.2007.02.005.

[163] Arnaud R. Krebs et al. "Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters". In: *Molecular Cell* 67.3 (Aug. 2017), 411–422.e4. ISSN: 10974164. DOI: 10.1016/j.molcel.2017.06.027.

[164] Timo Kuhn et al. "Single molecule tracking and analysis framework including theory-predicted parameter settings". In: *Scientific Reports* 11.1 (May 2021), pp. 1–12. ISSN: 2045-2322. DOI: 10.1038/s41598-021-88802-7. URL: https://www.nature.com/articles/s41598-021-88802-7.

[165] Nicholas C Lammers et al. "Multimodal transcriptional control of pattern formation in embryonic development". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.2 (2020), pp. 836–847. ISSN: 10916490. DOI: 10.1073/pnas.1912500117.

[166] Nicholas C. Lammers et al. *A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting*. Dec. 2020. DOI: 10.1016/j.ceb.2020.08.001. arXiv: 2008.09225.

[167] Nicholas C. Lammers et al. *Code from "A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting"*. 2020. URL: https://github.com/GarciaLab/TranscriptionalTimescalesReview.git.

[168] Nicholas C. Lammers et al. "Multimodal transcriptional control of pattern formation in embryonic development". In: *Proceedings of the National Academy of Sciences* 117.2 (Jan. 2020), pp. 836–847. ISSN: 0027-8424. DOI: 10.1073/PNAS.1912500117. URL: https://www.pnas.org/content/117/2/836.

[169] Alex H Lang et al. "Thermodynamics of statistical inference by cells". In: *Physical Review Letters* 113.14 (2014). ISSN: 10797114. DOI: 10.1103/PhysRevLett.113.148103. arXiv: 1405.4001.

[170] Daniel R Larson et al. *Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene Downloaded from*. Tech. rep. 2011. URL: http://science.sciencemag.org/.

[171] Moyra Lawrence, Sylvain Daujat, and Robert Schneider. "Lateral Thinking: How Histone Modifications Regulate Gene Expression". In: *Trends in Genetics* 32.1 (2016), pp. 42–56. ISSN: 13624555. DOI: 10.1016/j.tig.2015.10.007. URL: http://dx.doi.org/10.1016/j.tig.2015.10.007.

[172] P. A. Lawrence et al. "Borders of parasegments in Drosophila embryos are delimited by the fushi tarazu and even-skipped genes". In: *Nature* 328.6129 (1987), pp. 440–2. ISSN: 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/328440a0.

[173] Peter A. (Peter Anthony) Lawrence. "The making of a fly : the genetics of animal design". In: (1995), p. 228.

[174] Daniel D. Le et al. "Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding". In: *Proceedings of the National Academy of Sciences* 115.16 (Apr. 2018), E3702–E3711. ISSN: 0027-8424. DOI: 10.1073/PNAS.1715888115. URL: https://www.pnas.org/content/115/16/E3702%20https://www.pnas.org/content/115/16/E3702.abstract.

[175] Joel L. Lebowitz and Herbert Spohn. "A gallavotti-cohen-type symmetry in the large deviation functional for stochastic dynamics". In: *Journal of Statistical Physics* 95.1-2 (1999), pp. 333–365. ISSN: 00224715. DOI: 10.1023/a:1004589714161. arXiv: 9811220 [cond-mat]. URL: https://link.springer.com/article/10.1023/A:1004589714161.

[176] Bum Kyu Lee et al. "Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells". In: *Genome Research* 22.1 (Jan. 2012), p. 9. ISSN: 10889051. DOI: 10.1101/GR.127597.111. URL: /pmc/articles/PMC3246210/%20/pmc/articles/PMC3246210/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3246210/.

[177] Chang Hwan Lee, Heaji Shin, and Judith Kimble. "Dynamics of Notch-Dependent Transcriptional Bursting in Its Native Context". In: *Developmental Cell* 50.4 (Aug. 2019), 426–435.e4. ISSN: 18781551. DOI: 10.1016/j.devcel.2019.07.001.

[178] Chang Hwan Lee, Heaji Shin, and Judith Kimble. "Dynamics of Notch-Dependent Transcriptional Bursting in Its Native Context". In: *Developmental Cell* 50.4 (Aug. 2019), 426–435.e4. ISSN: 1534-5807. DOI: 10.1016/J.DEVCEL.2019.07.001.

[179] Tineke L Lenstra et al. "Transcription Dynamics in Living Cells". In: *Annual Review of Biophysics* 45 (2016), pp. 25–47. ISSN: 19361238. DOI: 10.1146/annurev-biophys-062215-010838. URL: www.annualreviews.org.

[180] Tineke L. Lenstra et al. "Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription". In: *Molecular Cell* 60.4 (Nov. 2015), pp. 597–610. ISSN: 10974164. DOI: 10.1016/j.molcel.2015.09.028.

[181] Tineke L. Lenstra et al. "Transcription Dynamics in Living Cells". In: *Annual Review of Biophysics* 45.1 (2016), pp. 25–47. ISSN: 1936-122X. DOI: 10.1146/annurev-biophys-062215-010838.

[182] Congxin Li et al. "Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation". In: *Cell Systems* (2018), pp. 1–15. ISSN: 24054712. DOI: 10.1016/j.cels.2018.01.012. URL: http://linkinghub.elsevier.com/retrieve/pii/S2405471218300127.

[183] Gu Li et al. "Rapid spontaneous accessibility of nucleosomal DNA". In: *Nature Structural and Molecular Biology* 12.1 (2005), pp. 46–53. ISSN: 15459993. DOI: 10.1038/nsmb869.

[184] Bomyi Lim et al. "Visualization of Transvection in Living Drosophila Embryos". In: *Molecular Cell* 70.2 (Apr. 2018), 287–296.e6. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2018.02.029. URL: https://www.sciencedirect.com/science/article/pii/S1097276518301746?via%7B%5C%%7D3Dihub.

[185] Timothée Lionnet and Carl Wu. "Single-molecule tracking of transcription protein dynamics in living cells: seeing is believing, but what are we seeing?" In: *Current Opinion in Genetics & Development* 67 (Apr. 2021), pp. 94–102. ISSN: 0959-437X. DOI: 10.1016/J.GDE.2020.12.001.

[186] Timothée Lionnet et al. "A transgenic mouse for in vivo detection of endogenous labeled mRNA". In: *Nature Methods* 8.2 (2011), pp. 165–170. ISSN: 15487091. DOI: 10.1038/nmeth.1551.

[187] Shawn C Little, Mikhail Tikhonov, and Thomas Gregor. "Precise developmental gene expression arises from globally stochastic transcriptional activity." In: *Cell* 154.4 (Aug. 2013), pp. 789–800. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.07.025. URL: http://www.ncbi.nlm.nih.gov/pubmed/23953111%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3778922.

[188] Shawn C Little, Mikhail Tikhonov, and Thomas Gregor. "Precise developmental gene expression arises from globally stochastic transcriptional activity." In: *Cell* 154.4 (Aug. 2013), pp. 789–800. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.07.025. URL: http://www.ncbi.nlm.nih.gov/pubmed/23953111%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3778922.

[189] Shawn C Little et al. "The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA". In: *PLoS Biology* 9.3 (Mar. 2011). Ed. by Claude Desplan, e1000596. URL: http://dx.plos.org/10.1371/journal.pbio.1000596%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/3F/3F4DA034-CBC4-48CD-9BC9-9F1B10A925D9%20papers3://publication/doi/10.1371/journal.pbio.1000596.

[190] Jonathan Liu et al. "Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage". In: *PLOS Computational Biology* 17.5 (May 2021), e1008999. ISSN: 1553-7358. DOI: 10.1371/JOURNAL.PCBI.1008999. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008999.

[191] Jun S Liu et al. "Sequential Monte Carlo Methods for Dynamic Systems". In: *Journal of the American Statistical Association* 93 (1998), pp. 1032–1044. ISSN: 0162-1459. DOI: 10.1080/01621459.1998.10473765. URL: https://www.tandfonline.com/action/journalInformation?journalCode=uasa20.

[192] Tanguy Lucas et al. "3 minutes to precisely measure morphogen concentration". In: *PLOS Genetics* 14.10 (Oct. 2018), e1007676. ISSN: 1553-7404. DOI: 10.1371/JOURNAL.PGEN.1007676. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007676.

[193] Tanguy Lucas et al. "Live imaging of bicoid-dependent transcription in Drosophila embryos." In: *Current biology : CB* 23.21 (2013), pp. 2135–2139. ISSN: 18790445. DOI: 10.1016/j.cub.2013.08.053.

[194] Connor H. Ludwig and Lacramioara Bintu. "Mapping chromatin modifications at the single cell level". In: *Development* 146.12 (2019). ISSN: 14779129. DOI: 10.1242/dev.170217.

[195] Michael Z. Ludwig et al. "Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness". In: *PLoS Genetics* 7.11 (Nov. 2011). Ed. by Harmit S. Malik, e1002364. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002364. URL: http://dx.plos.org/10.1371/journal.pgen.1002364.

[196] C. L. Luengo Hendriks et al. "Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline". In: *Genome Biology* 7.12 (2006), R123. ISSN: 1465-6914 (Electronic) 1465-6906 (Linking). DOI: 10.1186/gb-2006-7-12-r123.

[197] S. E. Luria and M. Delbrück. "MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE". In: *Genetics* 28.6 (1943).

[198] Paul M. Macdonald, Phil Ingham, and Gary Struhl. "Isolation, structure, and expression of even-skipped: a second pair-rule gene of Drosophila containing a homeo box". In: *Cell* 47.5 (Dec. 1986), pp. 721–734. ISSN: 0092-8674. DOI: 10.1016/0092-8674(86)90515-5. URL: https://pubmed.ncbi.nlm.nih.gov/2877745/.

[199] Sebastian J. Maerkl and Stephen R. Quake. "A systems approach to measuring the binding energy landscapes of transcription factors". In: *Science* 315.5809 (Jan. 2007), pp. 233–237. ISSN: 00368075. DOI: 10.1126/science.1131007. URL: https://science.sciencemag.org/content/315/5809/233%20https://science.sciencemag.org/content/315/5809/233.abstract.

[200] Sohail Malik and Robert G. Roeder. "Mediator: A Drawbridge across the Enhancer-Promoter Divide". In: *Molecular Cell* 64.3 (Nov. 2016), pp. 433–434. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2016.10.024. URL: https://www.sciencedirect.com/science/article/pii/S1097276516306694.

[201] Mattias Mannervik. "Control of Drosophila embryo patterning by transcriptional co-regulators". In: *Experimental cell research* 321.1 (Feb. 2014), pp. 47–57. ISSN: 1090-2422. DOI: 10.1016/J.YEXCR.2013.10.010. URL: https://pubmed.ncbi.nlm.nih.gov/24157250/.

[202] Manu et al. "Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors." In: *PLoS computational biology* 5.3 (Mar. 2009). Ed. by Stanislav Shvartsman, e1000303. URL: http://dx.plos.org/10.1371/journal.pcbi.1000303%20file:///Users/paulinamustafa/Dropbox/Library.papers3/Files/4E/4EAE43E6-0527-4C4E-9BBC-02FEBC0F551C.pdf%20papers3://publication/doi/10.1371/journal.pcbi.1000303.

[203] Jean-Michel Marin and Christian P. Robert. *Bayesian Essentials with R*. Springer Texts in Statistics. New York, NY: Springer New York, 2014. ISBN: 978-1-4614-8686-2. DOI: 10.1007/978-1-4614-8687-9. URL: http://link.springer.com/10.1007/978-1-4614-8687-9.

[204] Sarah Marzen, Hernan G. Garcia, and Rob Phillips. *Statistical mechanics of Monod-Wyman-Changeux (MWC) models*. May 2013. DOI: 10.1016/j.jmb.2013.03.013. URL: /pmc/articles/PMC3786005/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3786005/.

[205] Frank J Massey. "The Kolmogorov-Smirnov Test for Goodness of Fit". In: *Source: Journal of the American Statistical Association* 46.253 (1951), pp. 68–78.

[206] Davide Mazza et al. "A benchmark for chromatin binding measurements in live cells". In: *Nucleic Acids Research* (2012). ISSN: 03051048. DOI: 10.1093/nar/gks701.

[207] S. L. McKnight and Jr. Miller O. L. "Electron microscopic analysis of chromatin replication in the cellular blastoderm Drosophila melanogaster embryo". In: *Cell* 12.3 (1977), pp. 795–804. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

[208] Steven L. McKnight and Oscar L. Miller. "Post-replicative nonribosomal transcription units in D. melanogaster embryos". In: *Cell* 17.3 (1979), pp. 551–563. ISSN: 00928674. DOI: 10.1016/0092-8674(79)90263-0.

[209] Pankaj Mehta et al. *A high-bias, low-variance introduction to Machine Learning for physicists*. Tech. rep. 2019. arXiv: 1803.08823v3.

[210] Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (Dec. 1953), pp. 1087–1092. ISSN: 00219606. DOI: 10.1063/1.1699114. URL: https://aip.scitation.org/doi/abs/10.1063/1.1699114.

[211] Ron Milo and Rob Phillips. *Cell Biology by the Numbers*. Garland Science, Dec. 2015. DOI: 10.1201/9780429258770. URL: https://www.taylorfrancis.com/https://www.taylorfrancis.com/books/mono/10.1201/9780429258770/cell-biology-numbers-ron-milo-rob-phillips.

[212] Mustafa Mir et al. "Dense bicoid hubs accentuate binding along the morphogen gradient". In: *Genes and Development* 31.17 (Sept. 2017), pp. 1784–1794. ISSN: 15495477. DOI: 10.1101/gad.305078.117.

[213] Mustafa Mir et al. "Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos". In: (2018). DOI: 10.1101/377812. URL: http://dx.doi.org/10.1101/377812.

[214] Mustafa Mir et al. "Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos". In: *eLife* 7.Dv (2018), pp. 1–27. ISSN: 2050084X. DOI: 10.7554/eLife.40497.

[215] Leonid A. Mirny. "Nucleosome-mediated cooperativity between transcription factors". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.52 (Dec. 2010), pp. 22534–22539. ISSN: 00278424. DOI: 10.1073/pnas.0913805107.

[216] Tom Misteli et al. "Dynamic binding of histone H1 to chromatin in living cells". In: *Nature* 408.6814 (2000), pp. 877–881. ISSN: 00280836. DOI: 10.1038/35048610.

[217] Nacho Molina et al. "Stimulus-induced modulation of transcriptional bursting in a single mammalian gene". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.51 (Dec. 2013), pp. 20563–20568. ISSN: 00278424. DOI: 10.1073/pnas.1312310110.

[218] Maël Montévil et al. *Theoretical principles for biology: Variation*. Oct. 2016. DOI: 10.1016/j.pbiomolbio.2016.08.005.

[219] T. Morisaki et al. "Real-time quantification of single RNA translation dynamics in living cells". In: *Science* 352.6292 (2016), pp. 1425–9. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aaf0899.

[220] Tatsuya Morisaki et al. "Single-molecule analysis of transcription factor binding at transcription sites in live cells". In: *Nature Communications* 5 (2014). ISSN: 20411723. DOI: 10.1038/ncomms5456.

[221] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. "Reconciling Kinetic and Equilibrium Models of Bacterial Transcription". In: *bioRxiv* (June 2020), p. 2020.06.13.150292. DOI: 10.1101/2020.06.13.150292. URL: https://doi.org/10.1101/2020.06.13.150292.

[222] Brian Munsky, Zachary Fox, and Gregor Neuert. *Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics*. Sept. 2015. DOI: 10.1016/j.ymeth.2015.06.009.

[223] T. Muramoto et al. "Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation." In: *Proc Natl Acad Sci U S A* (2012). ISSN: 1091-6490. DOI: 1117603109[pii]10.1073/pnas.1117603109.

[224] Arvind Murugan, David A. Huse, and Stanislas Leibler. "Speed, dissipation, and error in kinetic proofreading". In: *Proceedings of the National Academy of Sciences* 109.30 (July 2012), pp. 12034–12039. ISSN: 0027-8424. DOI: 10.1073/PNAS.1119911109. URL: https://www.pnas.org/content/109/30/12034%20https://www.pnas.org/content/109/30/12034.abstract.

[225] Damien Nicolas et al. "What shapes eukaryotic transcriptional bursting?" In: *Mol. BioSyst.* 13.7 (June 2017), pp. 1280–1290. ISSN: 1742-206X. DOI: 10.1039/C7MB00154A. URL: http://xlink.rsc.org/?DOI=C7MB00154A.

[226] J Ninio. "Kinetic amplification of enzyme discrimination." In: *Biochimie* 57.5 (1975), pp. 587–95. ISSN: 0300-9084. URL: http://www.ncbi.nlm.nih.gov/pubmed/1182215.

[227] Eva Nogales, Robert K. Louder, and Yuan He. "Structural Insights into the Eukaryotic Transcription Initiation Machinery". In: *Annual Review of Biophysics* 46.1 (2017), pp. 59–83. ISSN: 1936-122X. DOI: 10.1146/annurev-biophys-070816-033751.

[228] Reinhard Obst. "The Timing of T Cell Priming and Cycling". In: *Frontiers in Immunology* 6.NOV (2015), p. 1. ISSN: 16643224. DOI: 10.3389/FIMMU.2015.00563. URL: /pmc/articles/PMC4633513/%20/pmc/articles/PMC4633513/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633513/.

[229] O. Padovan-Merhar et al. "Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms". In: *Mol Cell* 58.2 (2015), pp. 339–52. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2015.03.005.

[230] A. Pare et al. "Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting". In: *Curr Biol* 19.23 (2009), pp. 2037–42. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: S0960-9822(09)01848-X[pii]10.1016/j.cub.2009.10.028.

[231] Jeehae Park et al. "Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity". In: (2019), pp. 1–25.

[232] D. S. Parker et al. "The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity". In: *Sci Signal* 4.176 (2011), ra38. ISSN: 1937-9145 (Electronic). DOI: 10.1126/scisignal.2002077.

[233] Jean Peccoud and Bernard Ycart. *Markovian modeling of gene-product synthesis*. 1995. DOI: 10.1006/tpbi.1995.1027.

[234] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (Jan. 2012), pp. 2825–2830. ISSN: 15324435. DOI: 10.48550/arxiv.1201.0490. arXiv: 1201.0490. URL: https://arxiv.org/abs/1201.0490v4.

[235] José E Pérez-Ortín et al. "Eukaryotic mRNA Decay: Methodologies, Pathways, and Links to Other Stages of Gene Expression". In: (2013). DOI: 10.1016/j.jmb.2013.02.029. URL: http://dx.doi.org/10.1016/j.jmb.2013.02.029.

[236] Isabelle S. Peter and Eric H. Davidson. *Genomic control process : development and evolution*. London, UK ; San Diego, CA, USA: Academic Press is an imprint of Elsevier, 2015, xii, 448 pages. ISBN: 9780124047297 0124047297.

[237] Mariela D Petkova et al. "Optimal Decoding of Cellular Identities in a Genetic Network". In: *Cell* 176 (2019), 844–855.e15. DOI: 10.1016/j.cell.2019.01.007. URL: https://doi.org/10.1016/j.cell.2019.01.007.

[238] Mariela D. Petkova et al. "Optimal Decoding of Cellular Identities in a Genetic Network". In: *Cell* 176.4 (Feb. 2019), 844–855.e15. ISSN: 10974172. DOI: 10.1016/j.cell.2019.01.007.

[239] Rob Phillips. "Napoleon Is in Equilibrium". In: *Annual Review of Condensed Matter Physics* 6.1 (2015), pp. 85–111. ISSN: 1947-5454. DOI: 10.1146/annurev-conmatphys-031214-014558. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev-conmatphys-031214-014558.

[240] Rob Phillips and Nigel Orme. *The Molecular Switch: Signaling and Allostery*. Princeton University Press, 2020. ISBN: 9780691200248. DOI: 10.2307/j.ctvx5w8pf. URL: http://www.jstor.org/stable/j.ctvx5w8pf.

[241] Rob Phillips et al. "Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression". In: *Annu. Rev. Biophys* 48 (2019), pp. 121–163. DOI: 10.1146/annurev-biophys-052118. URL: https://doi.org/10.1146/annurev-biophys-052118-.

[242] Rob Phillips et al. "Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression". In: *Annu. Rev. Biophys* 48 (2019), pp. 121–163. DOI: 10.1146/annurev-biophys-052118. URL: https://doi.org/10.1146/annurev-biophys-052118-.

[243] M. Rabinowitz. "Studies on the cytology and early embryology of the egg of Drosophila melanogaster". In: *Journal of Morphology* 69.1 (1941), pp. 1–49. ISSN: 0362-2525. DOI: DOI10.1002/jmor.1050690102.

[244] Howard Raiffa and Robert. Schlaifer. *Applied statistical decision theory*. Wiley, 2000, p. 356. ISBN: 978-0-471-38349-9.

[245] Arjun Raj and Alexander van Oudenaarden. "Single-Molecule Approaches to Stochastic Gene Expression". In: *Annual Review of Biophysics* 38.1 (June 2009), pp. 255–270. ISSN: 1936-122X. DOI: 10.1146/annurev.biophys.37.032807.125928.

[246] Arjun Raj et al. "Imaging individual mRNA molecules using multiple singly labeled probes". In: *Nature Methods* 5.10 (2008), pp. 877–879. ISSN: 15487091. DOI: 10.1038/nmeth.1253.

[247] Arjun Raj et al. "Stochastic mRNA synthesis in mammalian cells". In: *PLoS Biology* (2006). ISSN: 15457885. DOI: 10.1371/journal.pbio.0040309.

[248] Armando Reimer et al. "Minimal Synthetic Enhancers Reveal Control of the Probability of Transcriptional Engagement and its Timing by a Morphogen Gradient". In: *SSRN Electronic Journal* (2021). DOI: 10.2139/ssrn.3887798. URL: https://doi.org/10.1101/2021.07.10.451524.

[249] Georg Rieckh and Gašper Tkačik. "Noise and information transmission in promoters with multiple internal states". In: *Biophysical Journal* 106.5 (Mar. 2014), pp. 1194–1204. ISSN: 00063495. DOI: 10.1016/j.bpj.2014.01.014. arXiv: 1307.8075. URL: http://dx.doi.org/10.1016/j.bpj.2014.01.014.

[250] Joseph Rodriguez, Daniel R Larson, and Larson ARjatscls. "Transcription in Living Cells: Molecular Mechanisms of Bursting". In: (2020), pp. 1–24. DOI: 10.1146/annurev-biochem-011520. URL: https://doi.org/10.1146/annurev-biochem-011520-.

[251] Joseph Rodriguez et al. "Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity". In: *Cell* 176.1-2 (Jan. 2019), 213–226.e18. ISSN: 10974172. DOI: 10.1016/j.cell.2018.11.026.

[252] Nitzan Rosenfeld et al. "Gene regulation at the single-cell level". In: *Science* 307.5717 (Mar. 2005), pp. 1962–1965. ISSN: 00368075. DOI: 10.1126/science.1106914. URL: https://pubmed.ncbi.nlm.nih.gov/15790856/.

[253] Tuhin Subhra Roy et al. "Information transmission in a two-step cascade: interplay of activation and repression". In: *Theory in Biosciences* 140.3 (2021), pp. 295–306. ISSN: 16117530. DOI: 10.1007/s12064-021-00357-3. arXiv: 2109.12901.

[254] Katja N. Rybakova et al. "Multiplex Eukaryotic Transcription (In)activation: Timing, Bursting and Cycling of a Ratchet Clock Mechanism". In: *PLOS Computational Biology* 11.4 (Apr. 2015). Ed. by Ilya Ioshikhes, e1004236. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004236. URL: https://dx.plos.org/10.1371/journal.pcbi.1004236.

[255] A. Sanchez and J. Kondev. "Transcriptional control of noise in gene expression". In: *Proc Natl Acad Sci U S A* 105.13 (2008), pp. 5081–6.

[256] Alvaro Sanchez and Ido Golding. *Genetic determinants and cellular constraints in noisy gene expression*. Dec. 2013. DOI: 10.1126/science.1242975. URL: https://science.sciencemag.org/content/342/6163/1188%20https://science.sciencemag.org/content/342/6163/1188.abstract.

[257] Alvaro Sanchez et al. "Effect of promoter architecture on the cell-to-cell variability in gene expression". In: *PLoS Computational Biology* 7.3 (Mar. 2011). ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1001100.

[258] M. Sasai and P. G. Wolynes. "Stochastic gene expression as a many-body problem". In: *Proc Natl Acad Sci U S A* 100.5 (2003), pp. 2374–9. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.2627987100.

[259] R. Sayal et al. "Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early Drosophila embryo". In: *Elife* 5 (2016). ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.08445.

[260] M J Schnitzer and S M Block. "Statistical kinetics of processive enzymes". In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 60. 1995, pp. 793–802. DOI: 10.1101/SQB.1995.060.01.085.

[261] Stefan Schoenfelder and Peter Fraser. "Long-range enhancer–promoter contacts in gene expression control". In: *Nature Reviews Genetics* 20.8 (2019), pp. 437–455. ISSN: 14710064. DOI: 10.1038/s41576-019-0128-0. URL: http://dx.doi.org/10.1038/s41576-019-0128-0.

[262] Clarissa Scholes, Angela H Depace, and A ´ Lvaro Sá Nchez. "Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle". In: *Cell Systems* 4 (2017), pp. 97–108. DOI: 10.1016/j.cels.2016.11.012. URL: http://dx.doi.org/10.1016/j.cels.2016.11.012http://dx.doi.org/10.1016/j.cels.2016.11.012.

[263] E. Segal et al. "Predicting expression patterns from regulatory sequence in Drosophila segmentation". In: *Nature* 451.7178 (2008), pp. 535–40. ISSN: 1476-4687 (Electronic). DOI: nature06496[pii]10.1038/nature06496.

[264] T. N. Senaratne et al. "Investigating the Interplay between Sister Chromatid Cohesion and Homolog Pairing in Drosophila Nuclei". In: *PLoS Genet* 12.8 (2016), e1006169. ISSN: 1553-7404 (Electronic) 1553-7390 (Linking). DOI: 10.1371/journal.pgen.1006169.

[265] Adrien Senecal et al. "Transcription factors modulate c-Fos transcriptional bursts". In: *Cell Reports* 8.1 (2014), pp. 75–83. ISSN: 22111247. DOI: 10.1016/j.celrep.2014.05.053.

[266] Alexander S. Serov, Alexander J. Levine, and Madhav Mani. "Abortive Initiation as a Bottleneck for Transcription in the Early Drosophila Embryo". In: (Jan. 2017). arXiv: 1701.06079. URL: http://arxiv.org/abs/1701.06079.

[267] Amir Shahein et al. "Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance". In: *bioRxiv* (2021).

[268] V. Shahrezaei and P. S. Swain. "Analytical distributions for stochastic gene expression". In: *Proc Natl Acad Sci U S A* 105.45 (2008), pp. 17256–61.

[269] Jayasha Shandilya and Stefan G.E. Roberts. "The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling". In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819.5 (May 2012), pp. 391–400. ISSN: 1874-9399. DOI: 10.1016/J.BBAGRM.2012.01.010. URL: https://www.sciencedirect.com/science/article/pii/S1874939912000284.

[270] Wanqing Shao and Julia Zeitlinger. "Paused RNA polymerase II inhibits new transcriptional initiation". In: *Nature Genetics* 49.7 (2017), pp. 1045–1051. ISSN: 15461718. DOI: 10.1038/ng.3867.

[271] Robert Shelansky and Hinrich Boeger. "Nucleosomal proofreading of activator–promoter interactions". In: *Proceedings of the National Academy of Sciences* 117.5 (Feb. 2020), pp. 2456–2461. ISSN: 0027-8424. DOI: 10.1073/PNAS.1911188117. URL: https://www.pnas.org/content/117/5/2456.

[272] Robert Shelansky and Hinrich Boeger. "Nucleosomal proofreading of activator–promoter interactions". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.5 (2020), pp. 2456–2461. ISSN: 10916490. DOI: 10.1073/pnas.1911188117.

[273] A. W. Shermoen, M. L. McCleland, and P. H. O'Farrell. "Developmental control of late replication and S phase length". In: *Curr Biol* 20.23 (2010), pp. 2067–77. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: 10.1016/j.cub.2010.10.021.

[274] E. D. Siggia and M. Vergassola. "Decisions on the fly in cellular sensory systems". In: *Proceedings of the National Academy of Sciences* 110.39 (Sept. 2013), E3704–E3712. ISSN: 0027-8424. DOI: 10.1073/pnas.1314081110. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1314081110.

[275] Daniel Silk et al. "Model Selection in Systems Biology Depends on Experimental Design". In: *PLoS Computational Biology* 10.6 (2014), p. 1003650. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1003650. URL: www.bbsrc.ac.uk.

[276] D S Sivia and J Skilling. *Data Analysis - A Bayesian Tutorial*. 2nd. Oxford Science Publications. Oxford University Press, 2006.

[277] Samuel O. Skinner et al. "Single-cell analysis of transcription kinetics across the cell cycle". In: *eLife* 5.JANUARY2016 (Jan. 2016). ISSN: 2050084X. DOI: 10.7554/eLife.12175.001.

[278] S. Small, A. Blair, and M. Levine. "Regulation of even-skipped stripe 2 in the Drosophila embryo". In: *EMBO J* 11.11 (1992), pp. 4047–57.

[279] S. Small and M. Levine. "The initiation of pair-rule stripes in the Drosophila blastoderm". In: *Curr Opin Genet Dev* 1.2 (1991), pp. 255–60. ISSN: 0959-437X (Print) 0959-437X (Linking).

[280] S. Small et al. "Transcriptional regulation of a pair-rule stripe in Drosophila". In: *Genes Dev* 5.5 (1991), pp. 827–39. ISSN: 0890-9369 (Print) 0890-9369 (Linking).

[281] L. H. So et al. "General properties of transcriptional time series in Escherichia coli". In: *Nat Genet* 43.6 (2011), pp. 554–60. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.821.

[282] Heehwa G. Son et al. "RNA surveillance via nonsense-mediated mRNA decay is crucial for longevity in daf-2/insulin/IGF-1 mutant C. elegans". In: *Nature Communications 2017 8:1* 8.1 (Mar. 2017), pp. 1–11. ISSN: 2041-1723. DOI: 10.1038/ncomms14749. URL: https://www.nature.com/articles/ncomms14749.

[283] Jasmin Speil et al. "Activated STAT1 transcription factors conduct distinct saltatory movements in the cell nucleus". In: *Biophysical Journal* 101.11 (2011), pp. 2592–2600. ISSN: 00063495. DOI: 10.1016/j.bpj.2011.10.006.

[284] Barbara Steurer et al. "Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.19 (2018), E4368–E4376. ISSN: 10916490. DOI: 10.1073/pnas.1717920115.

[285] G. Struhl, P. Johnston, and P. A. Lawrence. "Control of Drosophila Body Pattern by the Hunchback Morphogen Gradient". In: *Cell* 69.2 (1992), pp. 237–249. ISSN: 0092-8674.

[286] Svetlana Surkova et al. "Characterization of the Drosophila segment determination morphome." In: *Developmental biology* 313.2 (Jan. 2008), pp. 844–62. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2007.10.037. URL: http://www.ncbi.nlm.nih.gov/pubmed/18067886%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2254320.

[287] David M. Suter et al. "Mammalian genes are transcribed with widely different bursting kinetics". In: *Science* 332.6028 (2011), pp. 472–474. ISSN: 00368075. DOI: 10.1126/science.1198817.

[288] David M. Suter et al. *Origins and consequences of transcriptional discontinuity*. Dec. 2011. DOI: 10.1016/j.ceb.2011.09.004.

[289] Dylan J. Taatjes. "The Continuing SAGA of TFIID and RNA Polymerase II Transcription". In: *Molecular Cell* 68.1 (Oct. 2017), pp. 1–2. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2017.09.028.

[290] Katjana Tantale et al. "A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting". In: *Nature Communications* 7.1 (Nov. 2016), p. 12248. ISSN: 20411723. DOI: 10.1038/ncomms12248. URL: http://www.nature.com/articles/ncomms12248.

[291] Sheila S. Teves et al. "A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes". In: *eLife* 7 (2018), pp. 1–22. ISSN: 2050084X. DOI: 10.7554/eLife.35621.

[292] Mikhail Tikhonov et al. "Only accessible information is useful: insights from gradient-mediated patterning". In: (). DOI: 10.1098/rsos.150486. URL: http://dx.doi.org/10.1098/rsos.150486%20http://rsos.royalsocietypublishing.org..

[293] Gasper Tkacik, Curtis G Callan, and William Bialek. "Information flow and optimization in transcriptional regulation." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.34 (Aug. 2008), pp. 12265–70. ISSN: 1091-6490. DOI: 10.1073/pnas.0806077105. URL: http://www.ncbi.nlm.nih.gov/pubmed/18719112%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2527900.

[294] Gašper Tkačik and Aleksandra M. Walczak. *Information transmission in genetic regulatory networks: A review*. Apr. 2011. DOI: 10.1088/0953-8984/23/15/153102. arXiv: 1101.4240. URL: https://iopscience.iop.org/article/10.1088/0953-8984/23/15/153102%20https://iopscience.iop.org/article/10.1088/0953-8984/23/15/153102/meta.

[295] Miroslav Tomschik et al. "Fast long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.9 (2005), pp. 3278–3283. ISSN: 00278424. DOI: 10.1073/pnas.0500189102.

[296] Huy Tran et al. "Precision in a rush: Trade-offs between reproducibility and steepness of the hunchback expression pattern". In: *PLoS Computational Biology* 14.10 (2018). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1006513. URL: https://doi.org/10.1371/journal.pcbi.1006513.

[297] Edward Tunnacliffe, Adam M. Corrigan, and Jonathan R. Chubb. "Promoter-mediated diversification of transcriptional bursting dynamics following gene duplication". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.33 (Aug. 2018), pp. 8364–8369. ISSN: 10916490. DOI: 10.1073/pnas.1800943115.

[298] Akif Uzman. "Genes and signals: Ptashne, M., Gann, A." In: *Biochemistry and Molecular Biology Education* 30.5 (Sept. 2002), pp. 340–341. ISSN: 14708175. DOI: 10.1002/bmb.2002.494030059994. URL: http://doi.wiley.com/10.1002/bmb.2002.494030059994.

[299] K. J. Venken et al. "P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster". In: *Science* 314.5806 (2006), pp. 1747–51. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1134426.

[300] K. J. Venken et al. "Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster". In: *Nat Methods* 6.6 (2009), pp. 431–4. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: nmeth.1331[pii]10.1038/nmeth.1331.

[301] B. J. Vincent, J. Estrada, and A. H. DePace. "The appeasement of Doug: a synthetic approach to enhancer biology". In: *Integr Biol (Camb)* 8.4 (2016), pp. 475–84. ISSN: 1757-9708 (Electronic) 1757-9694 (Linking). DOI: 10.1039/c5ib00321k.

[302] Ben J. Vincent, Javier Estrada, and Angela H. DePace. "The appeasement of Doug: a synthetic approach to enhancer biology". In: *Integrative Biology* 8.4 (Apr. 2016), pp. 475–484. ISSN: 1757-9694. DOI: 10.1039/C5IB00321K. URL: http://xlink.rsc.org/?DOI=C5IB00321K.

[303] Ben J. Vincent et al. "Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos". In: *PLOS Genetics* 14.9 (Sept. 2018). Ed. by Michael Levine, e1007644. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1007644. URL: https://dx.plos.org/10.1371/journal.pgen.1007644.

[304] A Wald. "Sequential Tests of Statistical Hypotheses". In: *The Annals of Mathematical Statistics* 16.2 (May 1945), pp. 117–186. ISSN: 00034851. URL: http://www.jstor.org/stable/2235829.

[305] A. Wald and J. Wolfowitz. "Optimum Character of the Sequential Probability Ratio Test". In: *The Annals of Mathematical Statistics* 19.3 (Sept. 1948), pp. 326–339. ISSN: 0003-4851. DOI: 10.1214/aoms/1177730197. URL: https://projecteuclid.org/euclid.aoms/1177730197.

[306] C. Wang et al. "Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells". In: *Cell* 165.4 (2016), pp. 990–1001. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2016.04.040.

[307] Yaolai Wang et al. "Gene transcription in bursting: a unified mode for realizing accuracy and stochasticity". In: *Biological Reviews* 94.1 (Feb. 2019), pp. 248–258. ISSN: 1469185X. DOI: 10.1111/brv.12452.

[308] Søren Warming et al. "Simple and highly efficient BAC recombineering using galK selection". In: *Nucleic acids research* 33.4 (2005), pp. 1–12. ISSN: 1362-4962. DOI: 10.1093/NAR/GNI035. URL: https://pubmed.ncbi.nlm.nih.gov/15731329/.

[309] R. Warrior and M. Levine. "Dose-dependent regulation of pair-rule stripes by gap proteins and the initiation of segment polarity". In: *Development* 110.3 (1990), pp. 759–67. ISSN: 0950-1991 (Print) 0950-1991 (Linking).

[310] J. H. Waterborg. "Histone synthesis and turnover in alfalfa. Fast loss of highly acetylated replacement histone variant H3.2". In: *Journal of Biological Chemistry* 268.7 (1993), pp. 4912–4917. ISSN: 00219258.

[311] Ward Whitt. "Asymptotic Formulas for Markov Processes with Applications to Simulation". In: *Operations Research* 40.2 (Apr. 1992), pp. 279–291. ISSN: 0030-364X. DOI: 10.1287/opre.40.2.279. URL: https://pubsonline.informs.org/doi/abs/10.1287/opre.40.2.279.

[312] G. S. Wilkie et al. "Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei". In: *Curr Biol* 9.21 (1999), pp. 1263–6. ISSN: 0960-9822 (Print) 0960-9822 (Linking). DOI: S0960-9822(99)80509-0[pii].

[313] Felix Wong and Jeremy Gunawardena. "BB49CH10_Gunawardena ARjats.cls Gene Regulation in and out of Equilibrium". In: (2020). DOI: 10.1146/annurev-biophys-121219. URL: https://doi.org/10.1146/annurev-biophys-121219-.

[314] Felix Wong and Jeremy Gunawardena. *Gene Regulation in and out of Equilibrium*. May 2020. DOI: 10.1146/annurev-biophys-121219-081542. URL: https://pubmed.ncbi.nlm.nih.gov/32375018/%20https://pubmed.ncbi.nlm.nih.gov/32375018/?dopt=Abstract.

[315] B. Wu et al. "Translation dynamics of single mRNAs in live cells and neurons". In: *Science* 352.6292 (2016), pp. 1430–5. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aaf1084.

[316] Heng Xu et al. "Combining protein and mRNA quantification to decipher transcriptional regulation". In: *Nature Methods* 12.8 (July 2015), pp. 739–742. ISSN: 15487105. DOI: 10.1038/nmeth.3446.

[317] Heng Xu et al. "Combining protein and mRNA quantification to decipher transcriptional regulation". In: *Nature Methods* 12.8 (Aug. 2015), pp. 739–742. ISSN: 1548-7091. DOI: 10.1038/nmeth.3446. URL: http://www.nature.com/articles/nmeth.3446.

[318] Heng Xu et al. "Combining protein and mRNA quantification to decipher transcriptional regulation". In: *Nature Methods* 12.8 (Aug. 2015), pp. 739–742. ISSN: 1548-7091. DOI: 10.1038/nmeth.3446. URL: http://www.nature.com/articles/nmeth.3446.

[319] Heng Xu et al. "Stochastic Kinetics of Nascent RNA". In: *Physical Review Letters* 117.12 (Sept. 2016), p. 128101. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.117.128101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.117.128101.

[320] Shigehiro Yamada et al. "The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output". In: *Current Biology* 29.8 (Apr. 2019), 1387–1393.e5. ISSN: 0960-9822. DOI: 10.1016/J.CUB.2019.03.019. URL: https://www.sciencedirect.com/science/article/pii/S0960982219303173?via%7B%5C%%7D3Dihub.

[321] Ming Yan and Jay D. Gralla. "Multiple ATP-dependent steps in RNA polymerase II promoter melting and initiation." In: *The EMBO Journal* 16.24 (Dec. 1997), p. 7457. ISSN: 02614189. DOI: 10.1093/EMBOJ/16.24.7457. URL: /pmc/articles/PMC1170345/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1170345/.

[322] X. Yan et al. "Dynamics of Translation of Single mRNA Molecules In Vivo". In: *Cell* 165.4 (2016), pp. 976–89. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2016.04.034.

[323] Ahmet Yildiz et al. "Kinesin Walks Hand-Over-Hand". In: *Science* 303.5658 (Jan. 2004), pp. 676–678. ISSN: 00368075. DOI: 10.1126/science.1093753. URL: www.sciencemag.orghttp://science.sciencemag.org/.

[324] S. Yunger et al. "Single-allele analysis of transcription kinetics in living mammalian cells". In: *Nat Methods* 7.8 (2010), pp. 631–3. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: nmeth.1482[pii]10.1038/nmeth.1482.

[325] C. Zechner et al. "Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings". In: *Nat Methods* 11.2 (2014), pp. 197–202. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.2794.

[326] Barry M. Zee et al. "In vivo residue-specific histone methylation dynamics". In: *Journal of Biological Chemistry* 285.5 (2010), pp. 3341–3350. ISSN: 00219258. DOI: 10.1074/jbc.M109.063784.

[327] D. Zenklusen, D. R. Larson, and R. H. Singer. "Single-RNA counting reveals alternative modes of gene expression in yeast". In: *Nat Struct Mol Biol* 15.12 (2008), pp. 1263–71. ISSN: 1545-9985 (Electronic) 1545-9985 (Linking). DOI: nsmb.1514[pii]10.1038/nsmb.1514.

[328] Zhengjian Zhang et al. "Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis." In: *Genes & development* 30.18 (Sept. 2016), pp. 2106–2118. ISSN: 1549-5477. DOI: 10.1101/gad.285395.116. URL: http://www.ncbi.nlm.nih.gov/pubmed/27798851%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5066616.

[329] Coral Y. Zhou et al. "Mechanisms of ATP-Dependent Chromatin Remodeling Motors". In: *Annual Review of Biophysics* 45 (July 2016), pp. 153–181. ISSN: 19361238. DOI: 10.1146/annurev-biophys-051013-022819. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-051013-022819.

[330] B Zoller, S C Little, and T Gregor. "Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting". In: *Cell* 175.3 (2018), 835–847 e25. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2018.09.056.

[331] Benjamin Zoller, Shawn C Little, and Thomas Gregor Correspondence. "Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting In Brief Multiple gene expression boundaries in a developing embryo arise from the same strategy of modulated transcriptional bursting rates". In: *Cell* 175 (2018), pp. 835–847. DOI: 10.1016/j.cell.2018.09.056. URL: https://doi.org/10.1016/j.cell.2018.09.056.

[332] Benjamin Zoller et al. "Structure of silent transcription intervals and noise characteristics of mammalian genes". In: *Mol Syst Biol* 11 (2015). DOI: 10.15252/msb.

Appendix A

SI for Multimodal transcriptional control of pattern formation in embryonic development

A.1 Supplementary Figures

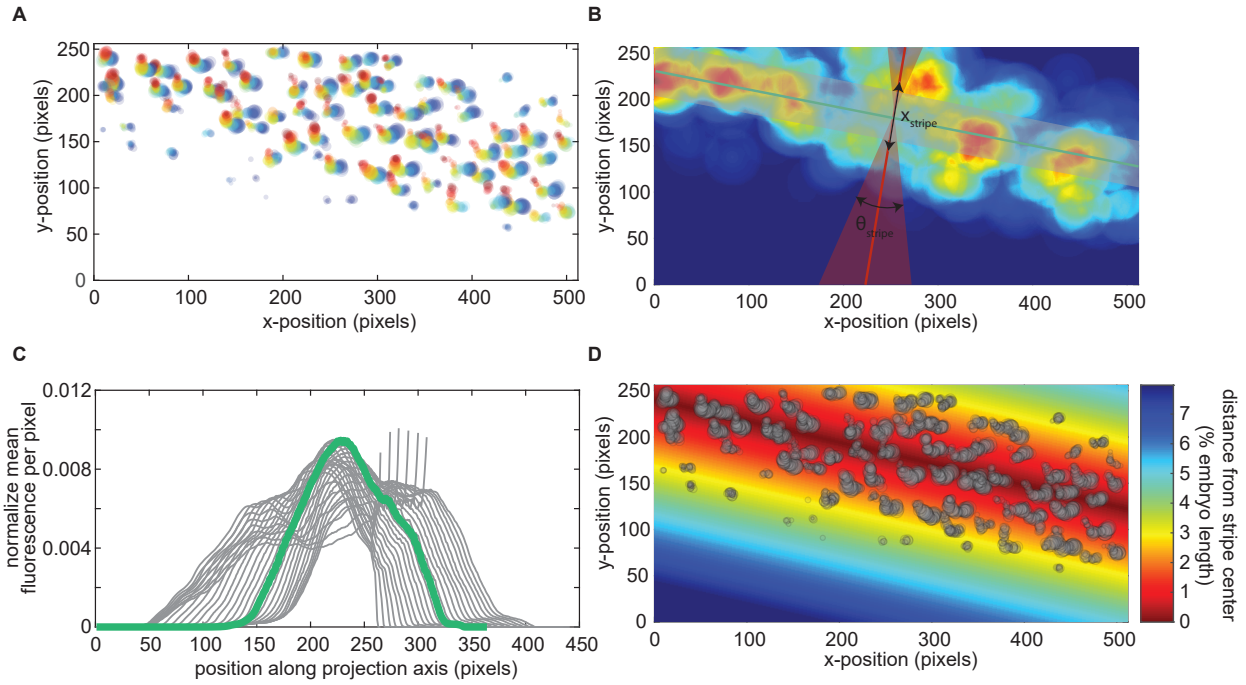

Figure A.1: **Aligning stripes from multiple embryos.** In order to minimize alignment errors when combining data from across multiple *Drosophila* embryos, an automated routine was employed to define a new experimental axis for each data set based upon the spatial distribution of transcriptional activity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

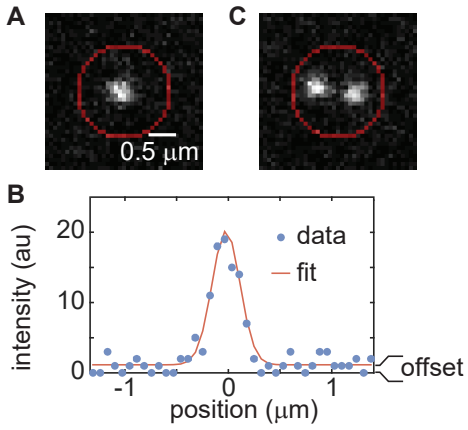

Figure A.2: **Integrating MS2 Spots.** **(A)** Sites of nascent transcription are identified in 3D using the Weka segmentation plugin for FIJI. Once identified, as described in [104], the Z-plane corresponding to the maximum fluorescence intensity is determined. On this Z-plane, fluorescence of a site is measured by integrating raw pixel intensities in a circular region around the fluorescent MS2 spot of a predefined area (indicated by the red circle) and then subtracting off the background intensity obtained by fitting a 2D Gaussian profile as outlined in (B). **(B)** X-Z projection of 2D Gaussian function fitted to MS2 spot shown in (A). Background intensity is estimated using the offset value fo this Gaussian fit. The per-pixel offset is then multiplied by the area of the integration region. This background value is then subtracted from the fluorescence integrated across the area shown in (A). **(C)** The radius was chosen to be large enough to integrate the intensities from both sister chromatids, even when they are spatially separated and distinguishable .

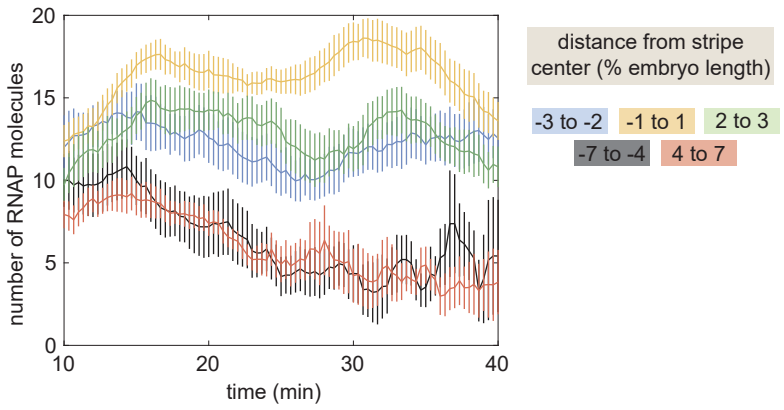

Figure A.3: **Mean transcriptional activity.** Mean transcriptional activity as a function of time for different positions along the stripe. (Average over 11 embryos, error bars indicate bootstrap estimate of the standard error of the mean. See Materials and Methods).

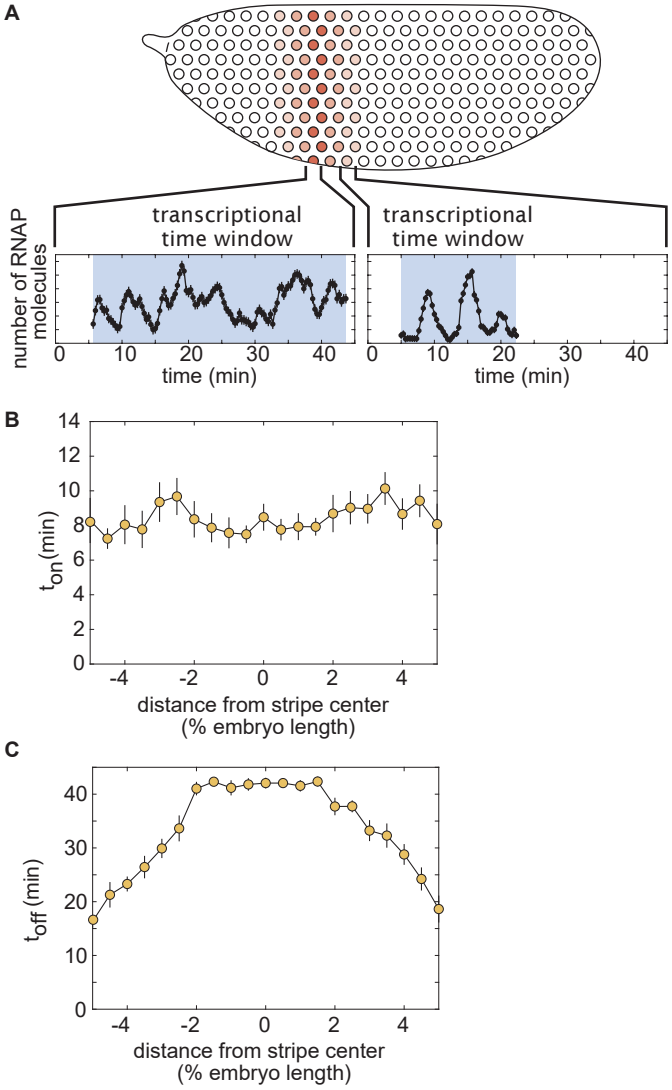

Figure A.4: **Regulation of the transcriptional time window. (A)** Single-nucleus measurements reveal that the duration of transcription is modulated along the stripe and that nuclei transcribe in a burst-like fashion. **(B)** Time for nuclei to activate transcription after mitosis, t_{on}, as a function 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).

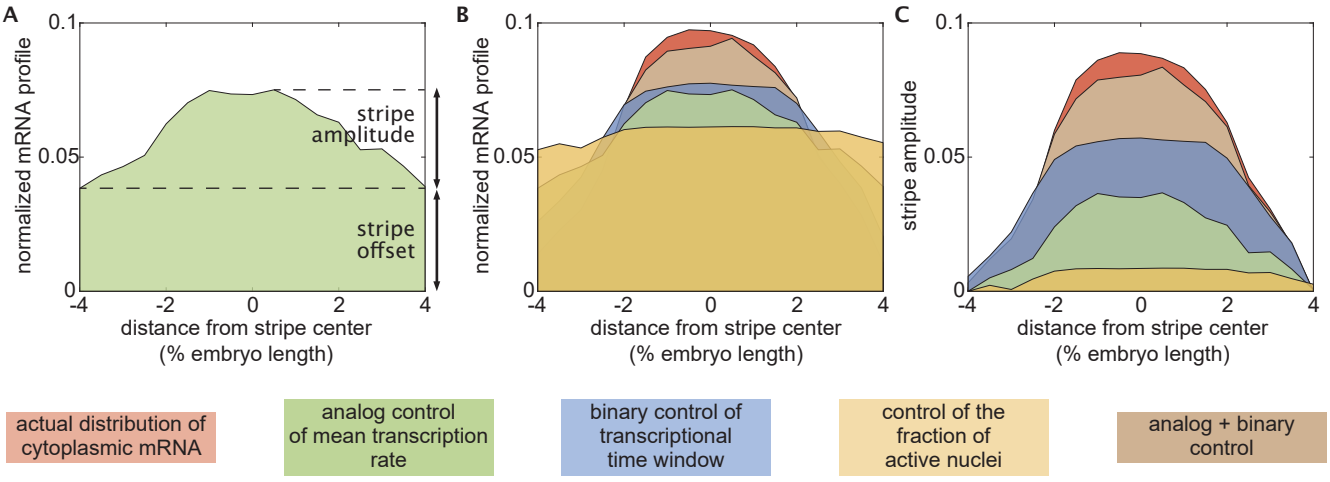

Figure A.5: **Definition of stripe amplitude.** **(A)** The normalized mRNA profile for the stripe can be separated into an offset and an amplitude. **(B)** Normalized mRNA profiles and **(C)** stripe amplitude for the cytoplasmic pattern of mRNA as well as for the contributions from the various regulatory strategies.

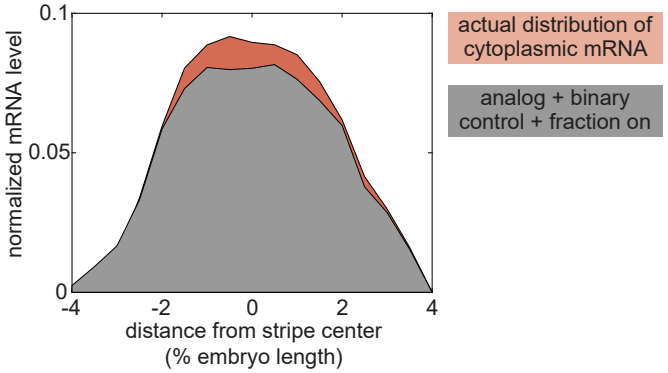

Figure A.6: **Joint effect of mean rate, binary control, and fraction of active nuclei.** Including of the predicted effect of anterior-posterior-dependent modulation of the fraction of active nuclei has little effect on the predicted cytoplasmic mRNA profile (compare brown profile in Figure 2.1G, gray 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.

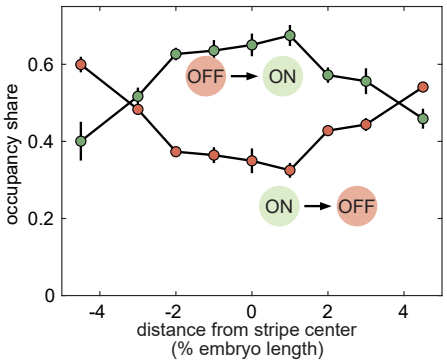

Figure A.7: **Fraction of time spent in each transcriptional state**. Fraction of time spent in the ON and OFF states as a function of the position along the stripe. (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

A.2 Extended Materials and Methods

A.2.1 Data processing

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

This routine, described graphically in Figure A.1, used observed spatial patterns of fluorescence measured from 30 minutes into nuclear cycle 14—the approximate time at which the mature stripe is first established [27]— to the time of last observation ($\geq$40 min) to find the natural position and orientation of the mature stripe. Generally, the *eve* stripes run roughly perpendicular to the anterior-posterior (AP) axis of the embryo; however, the approach allowed for the possibility that the true orientation of the *eve* 2 stripe deviated from the orientation implied by manual estimates of the anterior posterior axis. Thus, a variety of orientations for the natural stripe axis were considered, ranging between $\pm$ 15 degrees from the line perpendicular to the stripe with the manually specified anterior posterior axis. For each orientation, a sliding window of 4% embryo length in width was used to find the position along the proposed orientation that captured the largest fraction of the total fluorescence emitted by the mature stripe. The orientation and position that maximized the amount of fluorescence captured within this window defined a line through the field of view that was taken as the stripe center. All anterior-posterior positions used for subsequent analyses were defined relative to this center line.

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

A.2.2 cpHMM inference

To account for finite RNA polymerase (RNAP)elongation times, a compound state Markov formalism was developed in which the underlying two-promoter system—assumed to have three states (see Figure 2.4E,F)—was transformed into a system with 3^w compound gene states, where w indicates the number of time steps needed for an RNAP molecule to traverse the full transcript (see Figure A.36). These compound gene states played the role of the "hidden" states within the traditional HMM formalism. See Appendix A.3.4 for details regarding the model's architecture. Following this transformation from promoter states to compound gene states, it was possible to employ a standard version of the expectation-maximization (EM) algorithm, implemented using custom-written scripts in Matlab, to estimate bursting parameters from subsets of experimental traces (Appendix A.3.4). The scripts are available at the GarciaLab/cpHMM GitHub repository. Bootstrap sampling was used to estimate the standard error in our parameter estimates. Subsets of 8,000 data points were used to generate time-averaged parameter estimates. In order to accurately capture the time-averaged dynamics across the entirety of nuclear cycle 14, the full length of each experimental trace was used for time averaged inference. Sample sizes for windowed inference varied due to data set limitations. When possible, samples of 4,000 points were used. Only data points falling within a 15 minute window centered about the time point of interest were included in windowed inference runs. Inference was not conducted for spatiotemporal regions for which fewer than 1,250 time points were available. A minimum of 10 bootstrap samples were used to estimate each parameter value reported in this work. Reported values represent the median taken across bootstrap samples.

A.2.3 Input-output logistic regressions

The input-output analysis presented in Figure 2.7 utilized input transcription factor data from immunostaining experiments presented in [74], as well as live measurements of a Bicoid-GFP fusion courtesy of Jonathan Liu and Elizabeth Eck. Logistic regression parameters were estimated in Matlab using the *fmincon* function. See Appendix A.3.8 for further details.

A.2.4 Bootstrap error calculation

Bootstrap resampling was used frequently throughout this work to estimate the standard error in a variety of reported quantities, from trends estimated directly from raw experimental data in Figure 2.1 to cpHMM inference results presented in Figure 2.5 and Figure 2.6. In this

procedure, multiple bootstrap replicates, y^i_{boot} are generated by sampling with replacement from the pool of available experimental data, Y (see, e.g. [**Efron2016**]). The parameter of interest (say, $t_{\text{on}}(x)$) is then calculated for each replicate and the mean of these estimates is taken as the bootstrap estimate of the parameter value, $\hat{t}_{\text{on}}(x)$, while the standard deviation across the pool of bootstrap parameter estimates is used to approximate the standard error in our estimate of $t_{\text{on}}(x)$. In our case, simply performing this procedure across the available pool of nuclei failed to account for biological variability that exists from embryo to embryo. To account for this, we introduced a hierarchical bootstrapping procedure. The first step in this procedure was to draw bootstrap samples from across the 11 embryos used in this study. Because these samples were taken with replacement, most bootstrap samples excluded some embryos out of the original set of 11 and included duplicates (or triplicates, etc.) of others. Each embryo-level bootstrap defined a subset of nuclei. The final set of nuclei used for parameter estimation was generated by performing another round of bootstrap sampling on this pool. Bootstrap averages and standard errors were then calculated as described above. This two-step procedure thus accounts for both embryo-to-embryo and nucleus-to-nucleus variability.

We note that the limited number of data points available for many spatiotemporal regions prevented us from performing this two-tiered bootstrap procedure in the case of our time-dependent cpHMM inference (Figure 2.6D-F and Appendix A.3.7–Figure A.30D-E). In these cases, we used all available sets (essentially skipping the first bootstrap resampling step) and took bootstrap samples from amongst available nuclei as in step two of the procedure described above.

A.2.5 Absolute calibration of MS2 signal

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

Thus, in order to frame our results with respect to units with a clear physical interpretation, we calibrated our fluorescence measurements in terms of absolute numbers of mRNA molecules. This calibration was also used to inform our Poisson loading sensitivities (Appendix A.3.4). To calculate this calibration for our *eve* stripe 2 data, we relied on measurements reported by a previous study that utilized MS2 in conjunction with single molecule FISH to establish a calibration factor, α, between the integrated MS2 signal, F_{MS2}, and the number of mRNA molecules produced at a single transcriptional locus, N_{FISH}, [104] given by

$$\alpha = \frac{N_{\text{FISH}}}{F_{\text{MS2}}}. \tag{A.1}$$

This calibration factor can be used to estimate the average contribution of a single mRNA molecule to the observed (instantaneous) fluorescent signal. While the values for the parame-

ters in A.1 reported here pertain to the transcriptional output driven by the Bicoid activated P2 enhancer and promoter during nuclear cycle 13, the calibration should generalize to all measurements taken using the same microscope.

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

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

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

$$F_1 = \frac{1}{\alpha} = f_{\max}\frac{\frac{1}{2}L_I + L_{II}}{v_{\text{elong}}}, \tag{A.3}$$

such that

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

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

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

resulting in

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

Measurements performed in [104] give N_{FISH} to be 220 ($\pm$ 30) mRNA per nucleus and v_{elong} to be 1.5 ($\pm$ 0.14) kb/min. Experimental measurements on the P2 enhancer (courtesy of Elizabeth Eck, Maryam Kazemzadeh-Atoufi and Jonathan Liu) indicate that the total fluorescence per nucleus, F_{MS2}, is 9,600 ($\pm$320) AU minutes. For the reporter gene used to take these measurements, L_I and L_{II} are 1.275 kb and 4.021 kb, respectively. Thus, we obtain

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

$$= 13 \pm 1.7 \text{ AU/RNAP}. \tag{A.7}$$

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

A.3 Appendices

A.3.1 Theoretical model to predict cytoplasmic mRNA levels given from *in vivo* measurements of transcriptional activity

Derivation details

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

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = \underbrace{R(x,t)}_{\text{transcription rate}} - \underbrace{\gamma\,\text{mRNA}(x,t)}_{\text{degradation rate}}, \tag{A.8}$$

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

$$\frac{\text{dmRNA}}{\text{d}t}(x,t) = -\gamma\,\text{mRNA}(x,t), \tag{A.9}$$

such that degradation is now the only contribution to the change of mRNA concentration in time. Note that, in these two equations, we have ignored the contribution of mRNA diffusion. Previous measurements have estimated a diffusion coefficient of mRNA of 0.09 $\mu\text{m}^2/\text{s}$ [122] and a typical mRNA degradation rate of 0.14 min^{-1} [78]. Given these numbers, we expect an *eve* mRNA molecule to diffuse approximately 6 μm, 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 2.3G), we expect diffusion to play a minimal role in stripe formation. Finally, note that we are also ignoring the delay between transcriptional initiation and the delivery of an mRNA molecule to the cytoplasm as a result of nuclear export. This delay would affect the *timing* of pattern formation, but would leave our conclusions about the relative role of transcriptional bursting and the regulation of the duration of the transcriptional time window unaffected.

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

$$R(x) \approx \langle R(x,t) \rangle_t. \tag{A.10}$$

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

lated mRNA is given by

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

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

$$\text{mRNA}_{\text{active}}(x,t) = \frac{R(x)}{\gamma}, \tag{A.12}$$

where mRNA production and degradation are balanced.

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

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

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

$$\text{mRNA}_{\text{full}}(x,t) = \underbrace{\frac{R(x)}{\gamma}}_{\text{analog control}} \times \underbrace{e^{-\gamma(t-\min(t_{\text{off}}(x),t))}\left(1 - e^{-\gamma(\min(t_{\text{off}}(x),t)-t_{\text{on}}(x))}\right)}_{\text{binary control}}, \tag{A.14}$$

which can be simplified slightly to yield

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

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

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

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

Accounting for multiple transcriptional states

In the main text, Equation 2.3 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 A.16 to obtain an expression for the predicted amount of cytoplasmic mRNA that includes the burst parameters inferred by our cpHMM

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

While we present our results in terms of an effective two-state model in the main text, the presence of two transcriptional loci within each observed fluorescent spot suggests that the system is more naturally described using a three-state kinetic model. Here, we extend the framework presented in Equation A.17 to a scenario in which there are three distinct system states: 0 promoters on (0), 1 promoter on (1), and both promoters on (2) (see Figure 2.4). We begin with a general expression for this scenario that takes the contribution from the analog control term shown in Equation A.16 to be a sum over the output of each of the 3 activity states, namely,

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

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

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

$$0 = \boldsymbol{R}(x)\boldsymbol{\pi}(x), \quad \text{(A.19)}$$

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

$$\boldsymbol{R}(x) = \begin{bmatrix} -k_{01}(x) & k_{10}(x) & 0 \\ k_{01}(x) & -k_{10}(x)-k_{12}(x) & k_{21}(x) \\ 0 & k_{12}(x) & -k_{21}(x) \end{bmatrix}. \quad \text{(A.20)}$$

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

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

$$0 = -\pi_0 k_{01} + \pi_1 k_{10} \tag{A.21}$$

$$0 = \pi_0 k_{01} - \pi_1\big(k_{10} + k_{12}\big) + \pi_2 k_{21} \tag{A.22}$$

$$0 = \pi_1 k_{12} - \pi_2 k_{21}. \tag{A.23}$$

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

$$1 = \pi_0 + \pi_1 + \pi_2. \tag{A.24}$$

With this in mind, we can solve Equation A.21 for π_1 to find

$$\pi_1 k_{10} = \pi_0 k_{01} \tag{A.25}$$

$$\pi_1 = \pi_0 \frac{k_{01}}{k_{10}}. \tag{A.26}$$

Next, we use the normalization condition to eliminate π_2 from Equation A.23:

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

By combining this result with Equation A.26, we obtain

$$\pi_0 \frac{k_{01}}{k_{10}} k_{12} = (1 - \pi_0 - \pi_0 \frac{k_{01}}{k_{10}}) k_{21} \tag{A.28}$$

$$\pi_0 \frac{k_{01} k_{12}}{k_{10} k_{21}} = 1 - \pi_0 \frac{k_{10} + k_{01}}{k_{10}} \tag{A.29}$$

$$\pi_0 = \frac{k_{10} k_{21}}{k_{10} k_{21} + k_{01} k_{21} + k_{01} k_{12}}. \tag{A.30}$$

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

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

And finally π_2:

$$\pi_2 = 1 - \pi_0 - \pi_1$$

$$= \frac{k_{01}k_{12}}{k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}}. \tag{A.32}$$

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

$$\text{mRNA}(x,t) = p_{\text{active}}(x) \underbrace{\frac{1}{\gamma}\Big(r_1(x)\frac{k_{01}(x)k_{21}(x)}{\kappa(x)} + r_2(x)\frac{k_{01}(x)k_{12}(x)}{\kappa(x)}\Big)}_{\text{analog control}} \times \underbrace{\big(e^{-\gamma(t-\min(t_{\text{off}}(x),t))} - e^{-\gamma(t-t_{\text{on}}(x))}\big)}_{\text{binary control}}, \tag{A.33}$$

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

$$\kappa = k_{10}k_{21} + k_{01}k_{21} + k_{01}k_{12}. \tag{A.34}$$

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

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

Here we provide expressions relating the effective two-state parameters presented in the main text to parameters from the full three-state model. As we have done throughout this work, we take the transition rates between states (0) and (2) of the 3-state model to be negligible (consistent with inference results, see Appendix A.3.7). First, the on rate, $k_{\text{on}}^{\text{eff}}$ is directly equivalent to the transition rate between states (0) and (1), that is,

$$k_{\text{on}}^{\text{eff}} = k_{01}. \tag{A.35}$$

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

$$k_{\text{off}}^{\text{eff}} = \frac{\pi_1 k_{10}}{\pi_1 + \pi_2} \tag{A.36}$$

$$= \frac{k_{01}k_{21}k_{10}}{k_{01}k_{21} + k_{01}k_{12}} \tag{A.37}$$

$$= \frac{k_{21}k_{10}}{k_{21} + k_{12}}. \tag{A.38}$$

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

$$r^{\text{eff}} = \frac{\pi_1 r_1 + \pi_2 r_2}{\pi_1 + \pi_2} \tag{A.39}$$

$$= \frac{r_1 k_{01} k_{21} + r_2 k_{01} k_{12}}{k_{01} k_{21} + k_{01} k_{12}} \tag{A.40}$$

$$= \frac{r_1 k_{21} + r_2 k_{12}}{k_{21} + k_{12}}. \tag{A.41}$$

A.3.2 Measuring the amount of produced mRNA

Here, we outline our methodology for estimating rates of mRNA production depicted in Figure 2.3A and B, as well as the total cytoplasmic mRNA levels per nucleus shown in Figure 2.3G.

Calculating rates of mRNA production

The observed fluorescence signal at transcriptional loci as a function of time, $F(t)$, is linearly related to the number of actively transcribing RNAP molecules. Thus, after a period equal to the amount of time needed for an RNAP molecule to transcribe the gene, τ_{elong}, the number of new mRNAs added to the cytoplasm will be proportional to $F(t)$ [27], that is,

$$F(t) \propto M(t + \tau_{\text{elong}}) - M(t), \tag{A.42}$$

where $M(t)$ indicates the total number of mRNA molecules that have been produced up to time t. We can relate this fluorescence signal to absolute numbers of RNAP molecules using the calibration procedure described in the Extended Materials and Methods. Drawing from the derivation provided in the SI Methods of [27], we take the rate of mRNA production at time t to be approximately equal to the observed fluorescence at time $t - \frac{\tau_{\text{elong}}}{2}$,

$$F\Big(t - \frac{\tau_{\text{elong}}}{2}\Big) \propto \frac{\mathrm{d}M(t)}{\mathrm{d}t}. \tag{A.43}$$

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

Figure 2.3A depicts the time-averaged rate of mRNA production for each nucleus within the experimental field of view for one of our 11 live imaging movies. For each nucleus, this quantity was obtained by averaging the fluorescence across all observed time points, from when the nucleus first turned on (t_{on}) through to the final time point where expression was detected (t_{off}), which is taken to be *either* the time at which the nucleus transitioned into a quiescent state *or* the time depicted in the figure (40 minutes into nuclear cycle 14)—whichever came first. Thus the average rate of mRNA production for nucleus i is obtained from

$$R_i = \frac{\alpha \sum_{t=t_{\text{on}}}^{t_{\text{off}}} F_i(t)}{\sum_{t=t_{\text{on}}}^{t_{\text{off}}} 1}, \tag{A.44}$$

where the denominator is the total number of time points over which the averaging is performed and α indicates a conversion factor with units of RNAP molecules per AU per unit time that accounts for two factors: (1) the conversion from fluorescence to absolute numbers of RNAP molecules and (2) the dwell time of RNAP molecules on the gene. From Equation A.7, we know that we have one RNAP molecule for every 13 AU. The second factor is analogous to the $v_{\text{elong}}/(L_I + L_{II})$ component of Equation A.5, but with the appropriate lengths and elongation rates for the *eve* stripe 2 reporter. See *SI Appendix*, section A.3.10 for details about how we estimate the elongation time for our experimental system. Here, we will simply quote the results that $\alpha = 0.037$ RNAP molecules per AU per minute. Thus, the mean rate of mRNA production of a single nucleus is given by

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

A similar procedure was performed to estimate the average rate of mRNA production for each region along the anterior-posterior that is depicted in Figure 2.3B. This time, however, we summed over observed fluorescence values for *all* nuclei within the relevant region and time period and divided by the total number of time points such that

$$R_x = \frac{0.037 \sum_{i=1}^{N} \sum_{t=t_{\text{on}}}^{t_{\text{off}}} F_i(t)}{\sum_{i=1}^{N} \sum_{t=t_{\text{on}}}^{t_{\text{off}}} 1}, \tag{A.46}$$

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

Calculating full mRNA profiles

In contrast to the production rates calculated above, determining the relative contributions to stripe formation from each regulatory strategy depicted in Figure 2.3G did not require an AU to RNAP calibration. Thus, we capture the calibration factor, along with all other proportionality constants, with a generic term β, with the expectation that β will drop out from all consequential stripe contribution calculations. For a given region along the axis of the embryo, the average observed fluorescence across all N nuclei (active, quiescent, and those that never engaged in transcription) within the region of interest was used as a proxy for the instantaneous rate of mRNA production per nucleus, given by

$$\begin{aligned} \frac{\mathrm{d}M(x,t)}{\mathrm{d}t} &= \frac{\beta}{N} \sum_{i=1}^{N} F_i(x,t) \\ &= \beta \langle F(t) \rangle_x. \end{aligned} \tag{A.47}$$

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

Having obtained an expression for the rate of mRNA production as a function of space and time, we next sought to account for the degradation of mRNA over time. As indicated

in the main text, we assumed a constant rate of mRNA decay, γ, over space and time. The next section in this appendix provides evidence for the validity of this assumption. For a constant mRNA decay rate, calculating the average concentration of mRNA amounts to taking a weighted sum over all preceding production rates for a position of interest, where the weight terms account for the effects of mRNA decay and are of the form $e^{-\gamma t}$. Thus, we summed over all time points for each region of interest to estimate the total amount of cytoplasmic mRNA present on average, yielding the quantity on the left-hand side of Equation 2.2, namely,

$$\mathrm{mRNA}(x,t) = \beta \sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \big\langle F(t-n\Delta\tau) \big\rangle. \tag{A.48}$$

Here $\Delta\tau$ is the experimental time resolution, and $T = \frac{t}{\Delta\tau}$ denotes the number of measurements taken through time t. The exponential term within the summand on the right-hand side captures the effects of mRNA decay (*see* Appendix A.3.1). Finally, to calculate the normalized mRNA profile shown in Figure 2.3G (red), the estimates for the total accumulated mRNA per nucleus found using Equation A.48 must be divided by the sum across all spatial regions considered, namely

$$\mathrm{mRNA}_{\mathrm{norm}}(x_j,t) = \frac{\sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \big\langle F(t-n\Delta\tau) \big\rangle_{x_j}}{\sum_{x_i \in X} \sum_{n=1}^{T} e^{-\gamma(t-n\Delta\tau)} \big\langle F(t-n\Delta\tau) \big\rangle_{x_i}}, \tag{A.49}$$

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

$$\mathrm{mRNA}_{\mathrm{full}}(x_j,t) = \mathrm{mRNA}_{\mathrm{norm}}(x_i,t) - \min_{x} \Big(\mathrm{mRNA}_{\mathrm{norm}}(x_j,t) \Big). \tag{A.50}$$

Validating the fluorescence model

We employed a stochastic simulation to test the validity of the relation proposed in Equation A.47 and, more generally, of the approximate equality between time-lagged fluorescence and mRNA production asserted in Equation A.42. Simulated traces were generated using the Gillespie Algorithm [109], adjusted to allow system parameters to vary in time. We assumed an effective off rate of 0.667 transitions per minute and an initiation rate of 16 RNAP molecules per minute. To generate a temporal trend, we varied the effective on rate from an initial value of 1.6 transitions per minute (0 to 12 minutes of our simulation) to a basal value of 0.5 transitions per minute (27.5 to 35 minutes of our simulation). Figure A.8 compares the rate of mRNA production predicted using the time-lagged average of simulated fluorescence traces, to the true rate of mRNA production.

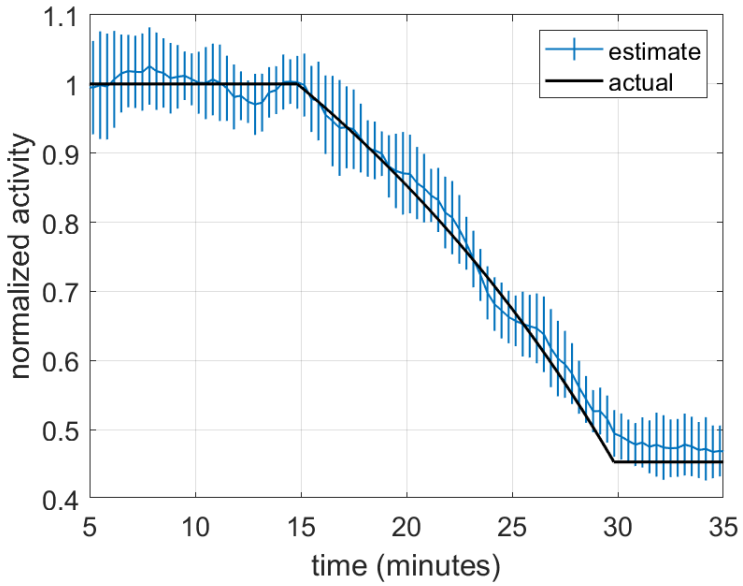

Figure A.8: **Comparison of actual and estimated mRNA production rates.** The black trend indicates the true rate of mRNA production as a function of time used for the simulation of transcription. The blue trend indicates the rate estimated from these simulated fluorescence traces. Error bars indicate standard deviation across 10 replicates containing 100 simulated traces each.

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

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

The predicted profile due to binary control of the transcriptional time window alone (Figure 2.3G, blue) was calculated following the same procedure as for the full mRNA profile described above, save for the fact that, in this case, instantaneous fluorescence values for individual nuclei were converted to binary indicator variables ($f_i(t)$) that were set equal to 1 if $t < t^i_{\text{off}}$ and 0 otherwise. Additionally, only nuclei that were active at *some point* during nuclear cycle 14 were included in order to distinguish the effects of the transcriptional time window (Figure 2.1C) from the control of the fraction of active nuclei (Figure 2.1D). 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

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

$$= \frac{N_c(x,t)}{N(x)}, \tag{A.51}$$

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

$$\mathrm{mRNA}_{\mathrm{binary}}(x,t) = \sum_{n=1}^{T} e^{-\gamma(t-n\Delta t)} \frac{N_c(x,n\Delta t)}{N(x)}. \tag{A.52}$$

The steps for calculating the the normalized binary mRNA levels comprising the blue profile in Figure 2.3G from Equation A.52 are identical to those shown for the full mRNA profile in Equation A.49 and Equation A.50 and are therefore not repeated here.

Comparison between predicted and measured cytoplasmic mRNA profiles

As a check for the validity of our approach to predicting levels of cytoplasmic mRNA from live imaging data (Equation A.49 and Equation A.50), we sought to compare our model's predictions to existing mRNA FISH data for the endogenous *eve* stripe 2 [92]. 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 [11] for details). This was done to minimize potential differences with the activity of the endogenous gene. Most notably, unlike the BAC reporter construct, the minimal reporter construct used for the majority of this work does not contain an enhancer sequence that is responsible for driving *eve* expression late in nuclear cycle 14 [146].

The researchers who generated the mRNA FISH data used the percent invagination of cellular membranes through cellularization as a means to break individual fixed embryos into rough temporal cohorts [196]. We cross-referenced the invagination ranges for each temporal group in the FISH data from [196] with data provided by precise measurements of invagination for different time points in [74] to obtain estimates for the range of times encapsulated by each of these cohorts.

We elected to use the cohort comprised of embryos with ages ranging between 38 to 48 minutes into nuclear cycle 14 because this range was much narrower than the preceding cohort and because the stripe appeared to be relatively stable during this time period. We note that the authors of [196] measured invagination on the ventral surface of the embryo, while the authors in [74] used the dorsal surface. This difference could lead to inconsistencies, since invagination is known to proceed more rapidly on the ventral side of the embryo [196]. However, the authors in [196] reported that this discrepancy is minimal up to the point where cell membrane extension has progressed to approximately 40% of its eventual full extent. The lower and upper bounds on the percent membrane invagination for the chosen cohort are 26% and 50% respectively. Thus, we expect the time estimate derived for the beginning of the period to be reasonably accurate, since dorsal and ventral membrane

progression was reported to be comparable during this period. Moreover, to the degree that ventral invagination outpaces dorsal invagination at the end of our period of interest, this would result in an over-estimation of ending time. Thus, if anything, the true temporal window encompassed by the selected cohort may actually be tighter than 10 minutes, since the ending time might in fact be earlier than 48 minutes into nuclear cycle 14. Given the relative stability of the stripe profile during this period of development, we do not expect this potential discrepancy to have a material impact on our conclusions.

Appendix A.3.2 Figure A.9 summarizes the results of this comparison. To account for uncertainty regarding the precise dorsal-ventral (DV) orientation of embryos within our live-imaging set, we compared our model's predictions to mRNA measurements for a range of DV positions, encompassed by the green-shaded profile. We found a high degree of agreement between model predictions and reported levels of cytoplasmic mRNA. This conclusion is relatively insensitive to our assumptions regarding the average lifetime of *eve* mRNA as shown by the blue and red lines in the figure (predictions assuming mRNA lifetimes of 7 and 15 minutes, respectively). We thus concluded that the assumptions underlying our model for predicting cytoplasmic mRNA levels from *in vivo* single-cell transcriptional activity measurements are valid.

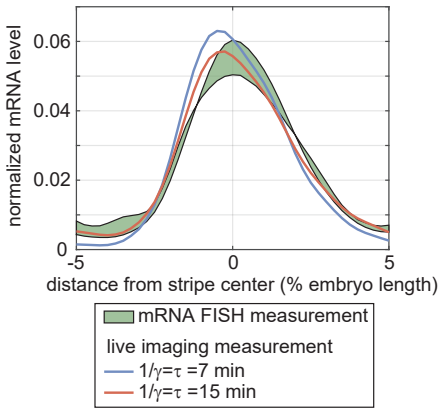

Figure A.9: **Comparison of predicted cytoplasmic mRNA by live-imaging measurements to direct measurements by FISH.** In an effort to check the validity of our modelling assumptions, we compared the predicted cytoplasmic mRNA profile stemming from live-imaging measurements of stripe 2 of an *eve* reporter from a BAC containing the full *eve* locus [11] to direct measurements of *eve* cytoplasmic mRNA levels using FISH [196]. Here, the blue and red lines indicate our model's predictions under two different assumptions for the rate of mRNA degradation, and the shaded green profile indicates the range of reported mRNA levels for different ranges of dorsal-ventral positions. Comparisons indicate a high degree of agreement between prediction and measurement, indicating that our modelling assumptions are justified.

Sensitivity of results to mRNA lifetime assumption

In the main text we assume a degradation rate for *eve* of 0.14 min^{-1} (corresponding to a lifetime of roughly $\tau = 7$ min). Since, to our knowledge, the decay rate of *eve* mRNA has not been measured directly, we follow [27] and base this estimate off of measurements for another of the pair rule genes, *fushi tazu* (*ftz*) [79]. In this section, we examine the degree to which the apparent contributions of each regulatory strategy (Figure 2.1) change under different assumptions for *eve* mRNA lifetime. Rather than conducting an exhaustive survey, we instead focus primarily on two limiting cases: rapid mRNA decay ($\tau = 1$ min) and *no* mRNA decay ($\tau = \infty$).

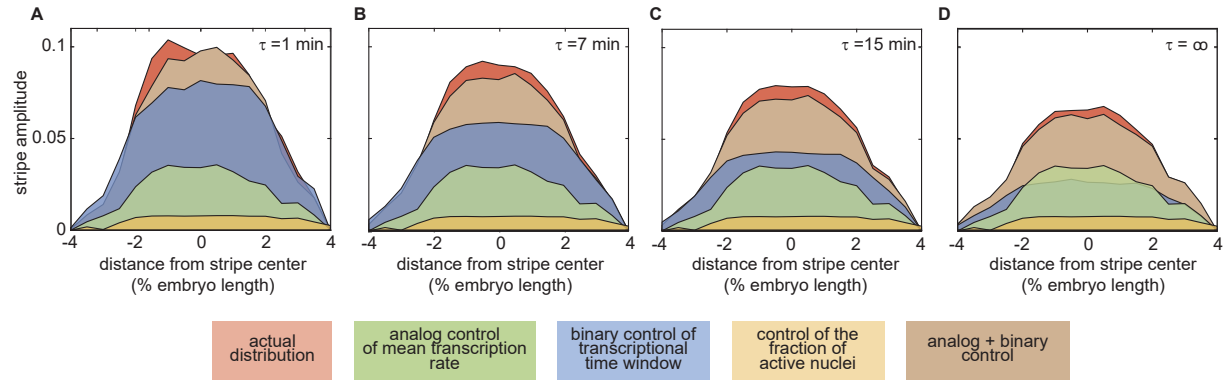

Figure A.10: **Sensitivity of regulatory strategy contribution to assumed mRNA lifetime.** The average lifetime of *eve* mRNA is a significant assumed parameter in our model. This figure compares the predicted contributions of each regulatory strategy for the mRNA lifetime assumed in the main text ($\tau = 7$ min) to limiting cases in which mRNA is assumed to decay almost instantaneously ($\tau = 1$ min) on the one hand, and infinitely slowly on the other ($\tau = \infty$). Even at these extremes, the central conclusion that the stripe is formed via the join action of mean rate modulation (green profile) and the time window (blue profile) remains intact. As expected, the relative contribution of the time window is sensitive to the assumed τ, yet even in the limit of no significant mRNA decay, its impact is still of order with the effect of mean rate modulation.

Figure A.10 summarizes the results of our analysis. We find that, regardless of the assumed mRNA lifetime, our model predicts that *eve* stripe 2 is formed almost entirely via the interplay between the binary control of the transcriptional time window and the analog modulation of the mean rate of transcription (compare brown and red profiles in Figure A.10). However, we find that the relative importance of each factor depends, somewhat, on the assumed decay rate. In the case of rapid mRNA decay, as well as for the decay rate assumed in the main text, the time window (blue profile) is clearly the dominant factor in driving pattern formation (Figure A.10A and B). If we assume the true mRNA lifetime is 15 minutes, slightly more than double our best guess of 7 minutes, we find that the time window is still

predicted to contribute slightly more to stripe formation, but that the two contributions are now of order with one another (Figure A.10C). Finally, in the limit where there is effectively no mRNA decay, the effects of the mean rate and time window are roughly equivalent (Figure A.10D). 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 [78]. 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.

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 2.1 and Figure 2.3 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 [11]. Since this BAC contains the full *eve* regulatory locus, it likely provides a better proxy for stripe formation in the endogenous context than the isolated *eve* 2 reporter. Figure A.11 shows the results of this analysis. As with the reporter construct used in the main text (Figure A.11A), we find that in the endogenous context (Figure A.11B) 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 2.1G). Interestingly, unlike the reporter construct, the full predicted profile (red profile) that accounts for the interplay between mRNA decay temporal fluctuations in the mean rate of mRNA production differs substantially from the simpler model (brown profile) that approximates mRNA production as constant over time. We speculate that this difference is attributable to the influence of the "late enhancer"—which is present in the *eve* BAC but *not* in the reporter—that takes over control of *eve* activity late in nuclear cycle 14. Further work will be necessary to fully elucidate the regulatory impact of this late element on the formation of the mature *eve* stripe pattern.

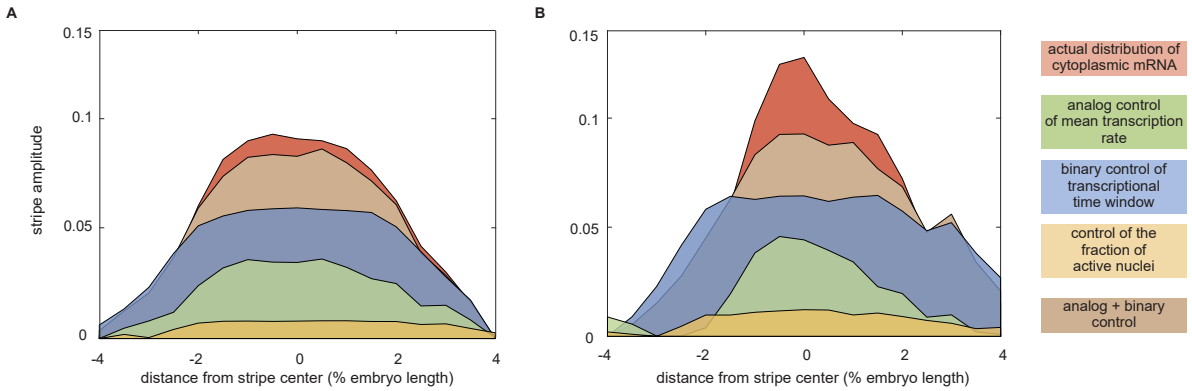

Figure A.11: **Regulatory strategy contributions to *eve* stripe 2 formation in endogenous context. (A)** Control strategy contributions for the reporter *eve* stripe 2 reporter construct (reproduced from Figure 2.3G), **(B)** Predicted contributions for *eve* BAC. As with the reporter construct, the formation of *eve* stripe 2 in the context of the full *eve* regulatory locus is dominated by the interplay between mean rate modulation (green) and control of the time window of transcriptional activity (blue).

A.3.3 Accounting for effects of experimental detection threshold

A number of analyses in this work rely (directly or indirectly) on the estimation of when gene loci first enter into a transcriptionally engaged state (t_{on}) as well as when they stop transcribing (t_{off}). These quantities are estimated using our live imaging data. Because live imaging experiments are subject to a detection limit (below which dim loci will not register as being active), there is a potential for bias in our estimates of t_{on} and t_{off}, as well as other metrics that derive in part from these quantities (duration of transcriptional activity, for instance). Similarly, there is a potential for bias in the measurements of the fraction of active nuclei as well as the transcriptional time window (Figure 2.3C-F) and the contributions of these regulatory strategies to pattern formation depicted in Figure 2.3G. These quantities were calculated assuming that undetected loci produced *no* mRNA when, in fact, it is possible that they produce at low levels periodically or even throughout the whole nuclear cycle. To estimate the nature and severity of these potential biases, we first estimated the detection limit for our live imaging experiments. Where appropriate, we used this limit to determine its potential effect on our conclusions.

Estimating the detection limit

We followed a methodology that was laid out in a previous work that employed the MS2 system in the fly to estimate the detection limit of our live imaging experiments [104]. Specifically, we calculated the minimum observed fluorescence value for each gene locus in our dataset. We then fit a Gaussian distribution to this set of minimum values to estimate the detection limit for our dataset. As shown in Figure A.12, this procedure returned an

estimated detection limit of 54 AU $\pm$ 1 which, according to our estimate of the absolute calibration, corresponds to approximately 4 RNAP molecules.

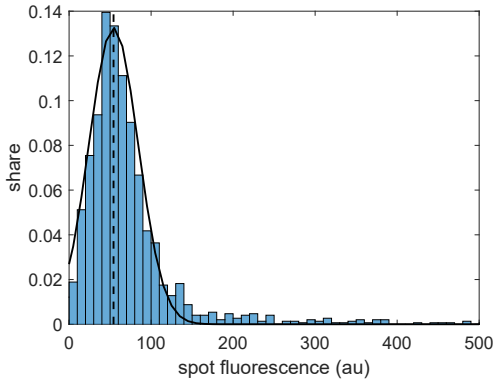

Figure A.12: **Estimating the fluorescence detection limit.** Distribution of minimum observed fluorescence values for each fluorescence trace in our data set (N=1484). The solid black line indicates the Gaussian probability density function that best fits the empirical data. The vertical dashed line indicates the inferred detection limit of 54 AU $\pm$ 1.

Estimating detection threshold artifacts for $t_{\mathbf{on}}$ and $t_{\mathbf{off}}$

As illustrated in Figure A.13A, the presence of a detection threshold will generally lead to estimated on-times that are too late and estimated off-times that are too early. To gain quantitative estimates for these biases, we used the results of our burst parameter inference shown in Figure 2.5 to simulate 50 activity trajectories for each region along the anterior-posterior axis that accurately recapitulated observed position-specific burst dynamics. For each simulated trace, we imposed a detection threshold of 54 AU and examined how far the resulting threshold-impacted estimates for t_{on} and t_{off} diverged from the true on and off times. Averaging these effects across 50 traces for each anterior-posterior region indicated that the detection limit did lead to biases. Specifically, we found that it would result in a 30-60 second *overestimate* (too late) of t_{on} (Figure A.13B) and a 20-50 second *underestimation* (too early) of t_{off} (Figure A.13C). These errors compound for our estimate of the duration of the transcriptional time window, leading to underestimates of between 50 and 100 seconds (Figure A.13D). As shown in the figure, while statistically significant, comparisons between raw and threshold-adjusted estimates for these three quantities reveal that these biases are small compared to the quantities of interest, and thus have a minimal effect on the observed trends.

In addition to the estimation of average on and off times as a function of AP position, analyses presented in Figure 2.6C (fraction of quiescent nuclei over time) and Figure 2.7A,D, and E (input/output inference for the transcriptional time window time window) of the main text, as well as Figure A.34 in this SI, required us to distinguish between transcriptionally engaged and quiescent nuclei. In these cases we employed a simpler definition of t_{off}, taking

it to be equal to the last time at which fluorescence was observed for a given locus. According to the results for our t_{off} analysis in Figure A.13B, we would expect the impact of this simpler (but more empirical) approach to be minimal, leading to a slight overestimation of t_{off} (1-2 minutes).

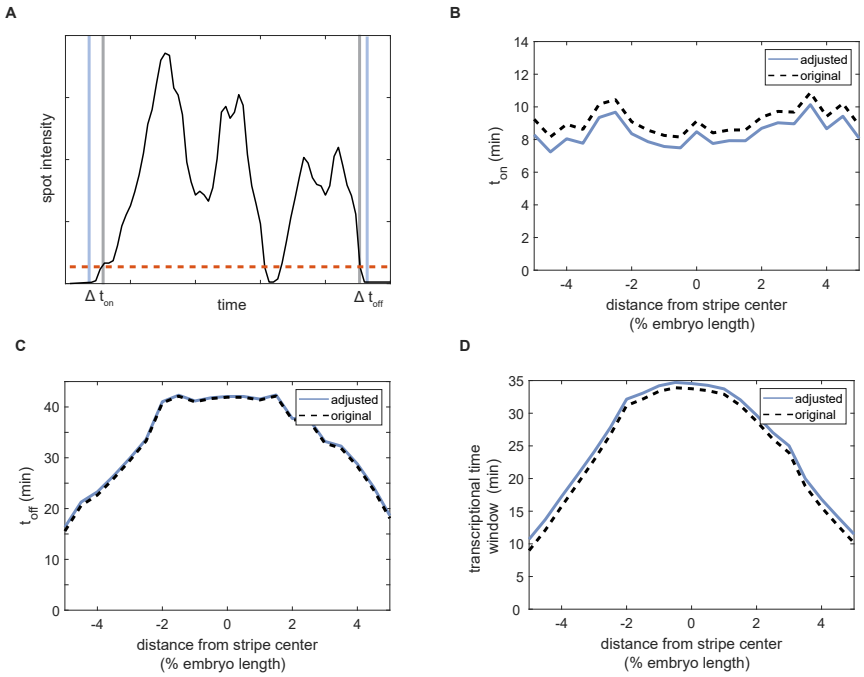

Figure A.13: **Accounting for effects of detection limit on off and on time calculations. (A)** Example of a simulated fluorescence trace. Red dashed line indicates the estimated detection limit of 54 AU. Blue lines indicate the true on and off times and gray lines indicate the *apparent* on and off times due to the effects of the detection threshold. **(B-D)** Plots of raw and adjusted anterior-posterior trends for (B) t_{on}, (C) t_{off}, and (D) the duration of the time window ($t_{\text{off}} - t_{\text{on}}$).

Possibility of basal expression does not impact conclusions about regulatory strategy contributions

This work invokes two regulatory strategies, the control of the transcriptional time window and of the fraction of active nuclei (Figure 2.1C,D), that assume that undetected loci are transcriptionally silent; that is, that they do not produce any mRNA. Yet, due to the detection limit of our experiment, we cannot completely rule out the possibility that some undetected loci are actually expressing at a basal level. We thus examined what effects, if any, the existence of such basal expression would have on the conclusions presented in this work.

First, we examined how the predicted contributions of each regulatory control strategy in Figure 2.3G would change in the presence of basal expression. We first examined what would

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

We next examined an even more extreme case wherein *all* undetected loci—both those that were never detected as transcriptionally active and those that shut off early as a part of the control of the transcriptional time window—expressed at the basal rate throughout the nuclear cycle (Figure A.14C). As expected, incorporating basal activity at all undetected loci has the effect of decreasing the overall prominence of the stripe pattern, since the expression floor is effectively raised from 0 to 54 AU. However, despite this change, the core conclusion that the analog control of the mean rate and binary control of the transcriptional time window jointly drive pattern formation (Figure 2.3) remains valid. Thus, we conclude that these results are robust to the possible existence of basal activity amongst quiescent nuclei.

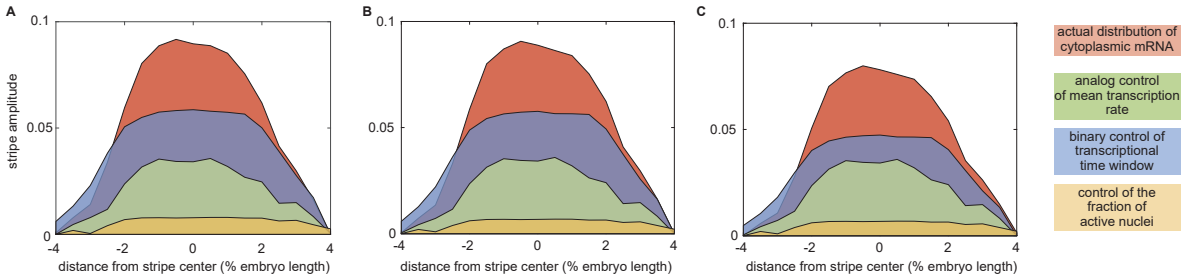

Figure A.14: **Impact of possible basal expression on control strategy contributions to stripe formation.** Predicted contributions of the regulation of the mean rate (green), time window (blue), and fraction of active nuclei (yellow) to pattern formation. **(A)** Predicted contributions assuming no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

Any undetected basal expression would have radically different burst dynamics

To remain undetected a reporter gene must never have more than four nascent RNAP molecules actively transcribing along its length. In the main text, we invoke the two-state model of bursting to describe the highly stochastic expression patterns exhibited by *eve* gene loci. A fundamental feature of this model is that gene loci transition between an active and an inactive state (see Figure 2.4). If basally expressing loci also adhered to this model, the burst parameters (k_{on}, k_{off}, and r) would need to be such that the bursts were either small enough or rare enough such that basal loci never contain more than four actively transcribing RNAP molecules at any time point. With this in mind, let us consider the consequences of tuning each burst parameter (see Figure 2.6A) to satisfy this requirement.

Figure A.16A features a simulated fluorescence trace generated using the bursting parameters inferred at the anterior flank of the stripe (anterior-posterior position -4 in Figure 2.5D and E). How would these bursting dynamics have to change in order to seemingly silent nuclei to be transcribing at a low, basal rate?

First, if the difference between basally expressing and detectable loci stems from a reduction in k_{on}, it would still be the case that bursts, when they occurred, would last about 90 s ($1/k_{\text{off}}$ in Figure 2.5E) and load RNAP molecules at a rate of about 1/3 s^{-1} (Figure 2.5D), resulting in the loading of approximately 32 RNAP molecules over the duration of the burst. This number of RNAP molecules on the gene could not escape detection, as it is significantly over our detection limit of four RNAP molecules. Thus, as illustrated in Figure A.16B, the only way for loci to avoid detection in this scenario is for k_{on} to be so low that it is unlikely for *a single* burst to occur within the approximately 40 minute time window of our observation. If we require that the odds of observing a single burst during the 40 minute window of observation be at most 90%, then we have

$$p_{\text{burst}} \leq 0.10 \tag{A.53}$$

$$1 - e^{-40k_{\text{on}}} \leq 0.10 \tag{A.54}$$

$$-k_{\text{on}} \geq \frac{\ln 0.9}{40} \tag{A.55}$$

$$k_{\text{on}} \leq 0.003 \text{ min}^{-1}, \tag{A.56}$$

Where p_{burst} is the probability of a burst occuring. We see that this requirement results in a k_{on} value that is *two orders of magnitude* lower than what is observed in detected loci (Figure A.15, low k_{on}). Such a locus would be active a mere 0.4% of the time (compared to

around 50% among observed loci). As a result, even if these loci did transcribe with bursting dynamics, these dynamics would be so different from those considered in this work that basally expressing loci would constitute a qualitatively distinct population from detected, high-expressing loci, and not a mere population of active loci that happened to present slightly lower fluorescence values.

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

$$p_{\text{long}} \leq 0.10 \tag{A.57}$$
$$e^{-0.2k_{\text{off}}} \leq 0.10 \tag{A.58}$$
$$k_{\text{off}} \geq -\frac{\ln 0.1}{0.2} \tag{A.59}$$
$$k_{\text{off}} \geq 11.5 \text{ min}^{-1}. \tag{A.60}$$

Here, p_{long} denotes the likelihood of a burst longer than 0.2 minutes. The k_{off} value consistent with these constraints is already an order of magnitude larger the k_{off} inferred for transcriptionally engaged loci. However we have not yet accounted for the fact that are typically multiple bursts over the course of the nuclear cycle and we require that it be unlikely for *any* single burst to crest the detection limit. This means that the basal k_{off} would need to be larger still. If we assume (conservatively) that there will be approximately 10 bursts per 40 minute time period, then we have

$$(1 - p_{\text{long}})^{10} \geq 0.9 \tag{A.61}$$
$$e^{-0.2k_{\text{off}}} \leq 1 - 0.9^{0.1} \tag{A.62}$$
$$k_{\text{off}} \geq -\frac{\ln(0.011)}{0.2} \tag{A.63}$$
$$k_{\text{off}} \geq 23 \text{ min}^{-1}. \tag{A.64}$$

Thus we see that this additional requirement implies a k_{off} of at least 23min^{-1}, $\sim$ 30-times higher than the values observed for this magnitude throughout this work. As illustrated in Figure A.16C, this would lead to a dramatically different kind of activity (compare to Figure A.16A). In such a scenario, bursts would only last 3 seconds on average, with only a single RNAP loaded per burst. Thus, as with k_{on}, we conclude that the dramatic modulation of k_{off} necessary to explain the presence of basally expressing loci would demand that these loci behave qualitatively different from active ones.

Finally, if we assume that basal expression is realized by decreasing the rate of RNAP loading during transcriptional bursts, similar arguments to those considered above for k_{off} indicate that an undetected gene locus must load RNAP molecules at a rate no faster than 0.5/min to remain undetected (Figure A.15); a rate that is 40 times higher than that of active loci.

All of the above scenarios strain the bursting model to the breaking point. If k_{on} is modulated in basally expressing loci, then this would imply that basal gene loci are active a mere 0.4% of the time (Figure A.16A). If k_{off} or r are modulated, then "bursts" would consist, on average, of a single polymerase loading event (Figure A.16C and D). Further studies employing single-molecule techniques will be needed to establish the presence or absence of low-level expression. In the meantime, we conclude that while we cannot rule out the existence of basal expression, we *can* confidently state that, if it exists, such activity must be radically different in character from the burst dynamics observed amongst the "transcriptionally engaged" loci we identify in the main text.

Figure A.15: table
Comparing burst characteristics of three possible basal expression schemes to "normal" loci expressing above the detection limit.

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

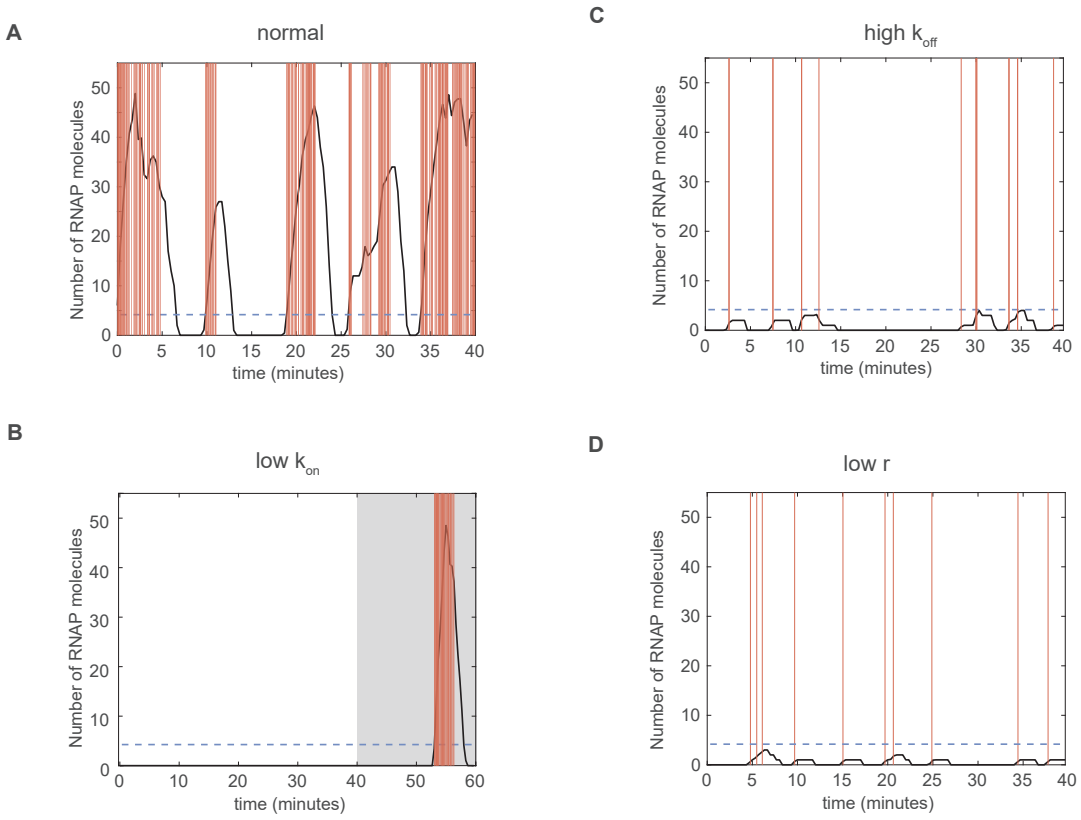

Figure A.16: **Possible basal burst dynamics scenarios.** Simulated activity traces illustrating bursting dynamics for basally expressing gene loci. Vertical red lines indicate RNAP initiation events. Dashed blue line indicates detection limit. **(A)** illustrative trace generated using burst parameters from anterior stripe flank. **(B-D)** hypothetical basal traces with **(B)** reduced k_{on} (0.003 min^{-1}), **(C)** elevated k_{off} (23.7 min^{-1}), and **(D)** reduced RNAP loading rate (0.5 RNAP min^{-1}). Gray region in (B) indicates the time period coming after 40 minutes of observation.

A.3.4 The compound-state hidden Markov model

Model introduction

To model the dynamics of an observed fluorescence series, $\boldsymbol{y} = \{y_1, y_2, ..., y_T\}$, where T is the number of data points in a trace, we assume that, at each time step, the sister promoters can be in one of K effective states. In the analysis of *eve* stripe 2 data, we use a simple model with the number of effective states equal to three ($K = 3$). The method, however, allows for more complex transcription architectures with higher numbers of states. Transitions between the effective promoter states are assumed to be Markovian, meaning that the hidden promoter state z_t at time step t is conditionally dependent only on the state in the previous time step.

This dependency is modeled through a $K \times K$ transition probability matrix $\boldsymbol{A} = p(z_t|z_{t-1})$, where A_{kl} is the probability of transitioning from the l^{th} state into the k^{th} state in the time interval $\Delta\tau$, where $\Delta\tau$ is the data sampling resolution. We assign a characteristic RNAP initiation rate, $r(k)$, with units of RNAP per minute, to each effective promoter state, $z(k)$, $1 \leq k \leq K$. Thus, the number of polymerases initiated between time steps $t-1$ and t will be $r(z_t)\Delta\tau$. Because the fluorescence intensity contributed by each polymerase depends on the number of transcribed MS2 stem loops, the contribution will vary with the position of the polymerase on the gene. In our transcription model we assume that polymerase elongation takes place at a constant rate. Therefore, if τ_{MS2} is the time it takes to transcribe the MS2 loops, the fluorescence contribution of an RNAP molecule will initially grow linearly ($\tau \leq \tau_{\text{MS2}}$) and will then stay constant for the remainder of transcription ($\tau_{\text{MS2}} \leq \tau \leq \tau_{\text{elong}}$). Given this time dependence, we define a maximum fluorescence emission per time step for each state as $v(k) = F_{\text{RNAP}} r(k)\Delta\tau$, $1 \leq k \leq K$, where F_{RNAP} is the fluorescence calibration factor determined using smFISH experiments (see Materials and Methods).

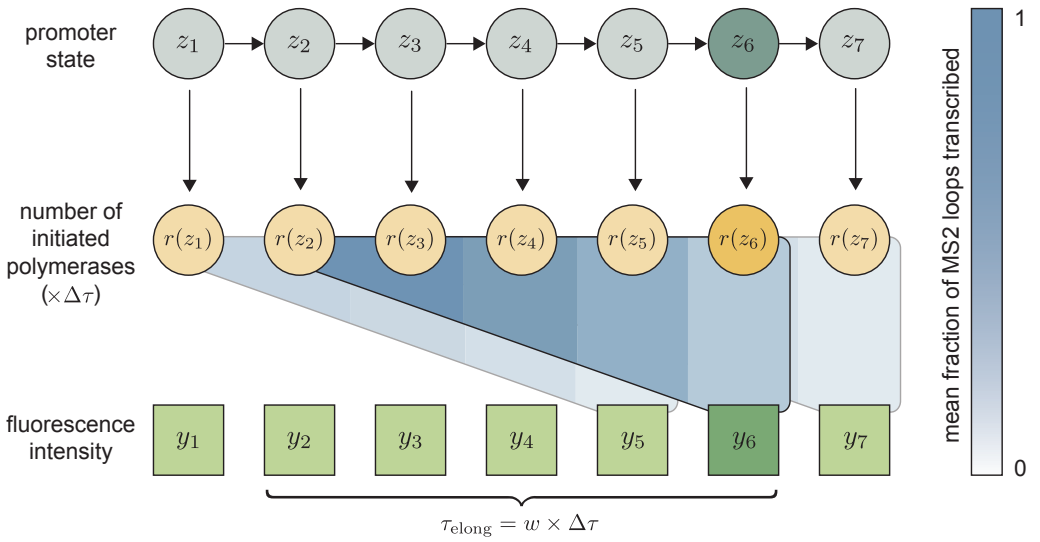

Figure A.17: **Schematic overview of the cpHMM architecture.** The sister promoters are modeled as undergoing a series of Markovian transitions between effective transcriptional states (z_t). Each promoter state uniquely determines the number of polymerases initiated in a single time step ($r(z_t)\Delta\tau$). Fluorescence emissions from polymerases initiated in the most recent w steps combine to produce the observed fluorescence intensity (y_t). The color bar indicates the mean fraction of MS2 loops that have been transcribed and contribute fluorescence at the moment of observation. The color corresponding to the more recently initiated polymerases is therefore lighter (fewer loops transcribed) than that corresponding to polymerases initiated at earlier times (more loops transcribed).

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

memory. Here $\Delta\tau$ indicates the observational time resolution, a quantity set by experimental conditions. The time required for an RNAP molecule to transcribe our reporter gene (τ_{elong}) is *a priori* unknown. We developed an autocorrelation-based method to estimate τ_{elong} directly from our experimental data (see Appendix A.3.10 and [56]). The observation y_t at time step t conditionally depends not only on the hidden promoter state z_t, but also on the hidden states in the previous w time steps, $\{z_t, z_{t-1}, ..., z_{t-w+1}\}$. To be able to describe the observed system dynamics through a hidden Markov model, the observation at time step t needs to be conditionally independent from the states at earlier time steps. We therefore introduce the concept of a compound state, $s_t = \{z_t, z_{t-1}, ..., z_{t-w+1}\}$, which, together with the set of model parameters, $\boldsymbol{\theta}$, is sufficient to define the probability distribution of the observation y_t, thereby satisfying the Markov condition. Since $z_t \in \{1, ..., K\}$, each compound state can take one of K^w different values, $s_t \in \{1, ..., K^w\}$. While the number of possible compound states is K^w, only K different transitions are allowed between them, since the most recent $w-1$ promoter states are deterministically passed from one compound state to the next, i.e. the last $w-1$ elements in $s_{t+1} = \{z_{t+1}, z_t, ..., z_{t-w+2}\}$ are present in s_t as well. The schematic overview of the cpHMM architecture is shown in Figure A.17.

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

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

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

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

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{A.66}$$

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

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{\boldsymbol{s}=\{s_1, s_2, ..., s_T\}} p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}). \tag{A.67}$$

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

We note that the notion of a compound state was also introduced in an earlier work [55] to account for the memory effect in hidden Markov modeling of *actin* transcription and then an EM methodology was applied to learn the kinetic parameters from MS2-based transcription data. Unlike their approach, however, we do not explicitly model the recruitment of individual RNAP molecules, but instead, assign a continuous RNAP initiation rate to each promoter state. Additionally, our model estimates the magnitude of the background noise present in the experimentally measured fluorescence signal, whereas the model presented in [55] 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 A.3.5 we demonstrate that relaxing the continuous RNAP loading assumption when generating synthetic data does not significantly affect the accuracy of the cpHMM inference.

Expectation-maximization (EM) algorithm

Consistent with standard EM approaches (cf. Bishop [18], Chapter 13), at each iteration we maximize the lower bound of the logarithm of the likelihood using the current estimate of the model parameters, namely,

$$\hat{\boldsymbol{\theta}}_{k+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{A.68}$$

$$\mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{\boldsymbol{s}=\{s_1,s_2,...,s_T\}} p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k)\, \log p(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) \leq \log\, p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{A.69}$$

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

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

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

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

$$\log p(y_t|s_t, \boldsymbol{v}, \sigma) = \frac{1}{2}\sum_{i=1}^{K^w} s_t^i \left(\log\lambda - \log(2\pi) - \lambda(y_t - V_i(\boldsymbol{v}))^2\right), \tag{A.72}$$

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

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

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

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

where $\lceil n_{\text{MS2}} \rceil$ and $\lfloor n_{\text{MS2}} \rfloor$ are the ceiling and the floor of n_{MS2}, respectively. The dependence of the weighting function $\kappa(d)$ on the position for a specific choice of parameters is illustrated in Figure A.18.

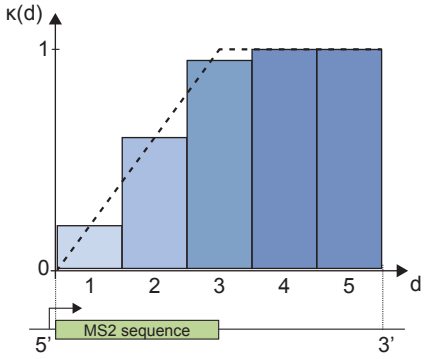

Figure A.18: **The weighting function $\kappa(d)$ evaluated at different positions along the genome.** The dashed line represents the fraction of the MS2 loops transcribed at a given position. Parameters used for plotting: $\tau_{\text{elong}} = 100$ sec, $\tau_{\text{MS2}} = 50$ sec, $\Delta\tau = 20$ sec, $w = \tau_{\text{elong}}/\Delta\tau = 5$, $n_{\text{MS2}} = \tau_{\text{MS2}}/\Delta\tau = 2.5$.

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

$$V_i(\boldsymbol{v}) = F_{i,:}\boldsymbol{v}, \tag{A.74}$$

where the i^{th} row of the $K^w \times K$ matrix $\boldsymbol{F}$ 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 $\boldsymbol{F}$ corresponding to the compound state $s = \{1, 1, 3, 2, 3\}$ will be $[\kappa(1)+\kappa(2), \kappa(4), \kappa(3)+\kappa(5)]$.

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

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

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

$$\langle s_t^i\rangle = \sum_{\boldsymbol{s}=\{s_1,s_2,\ldots,s_T\}} s_t^i\, p(\boldsymbol{s}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{A.76}$$

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

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

$$\langle s_t^i\rangle = \sum_{s_1,s_2,\ldots,s_T} s_t^i\, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t} s_t^i\, p(s_t|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k), \tag{A.78}$$

$$\langle s_t^i s_{t-1}^j\rangle = \sum_{s_1,s_2,\ldots,s_T} s_t^i s_{t-1}^j\, p(s_1, s_2, \ldots, s_T|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \sum_{s_t,s_{t-1}} s_t^i s_{t-1}^j\, p(s_t, s_{t-1}|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k). \tag{A.79}$$

Following the conventional implementation of the forward-backward algorithm (cf.[18], Chapter 13), we use the Markov property of the promoter state dynamics, together with the sum and products rules of probability, to write

$$p(s_t|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \frac{\alpha_t(s_t)\beta_t(s_t)}{p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k)}, \tag{A.80}$$

$$p(s_{t-1}, s_t|\boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) = \frac{\alpha_{t-1}(s_{t-1})\, p(y_t|s_t, \hat{\boldsymbol{\theta}}_k)\, p(s_t|s_{t-1}, \hat{\boldsymbol{\theta}}_k)\beta_t(s_t)}{p(\boldsymbol{y}|\,\hat{\boldsymbol{\theta}}_k)}, \tag{A.81}$$

$$\alpha_t(i) = p(y_1, \ldots, y_t, s_t = i|\,\hat{\boldsymbol{\theta}}_k), \tag{A.82}$$

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

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

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

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

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

$$\alpha_1(i) = p(y_1 | s_1 = i, \hat{\boldsymbol{\theta}}_k)\, p(s_1 = i | \hat{\boldsymbol{\theta}}_k), \tag{A.86}$$

$$\beta_T(i) = 1, \tag{A.87}$$

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

$$\left(\sum_{s_T=1}^{K^w} p(s_T | \boldsymbol{y}, \hat{\boldsymbol{\theta}}_k) \right) p(\boldsymbol{y} | \hat{\boldsymbol{\theta}}_k) \equiv p(\boldsymbol{y} | \hat{\boldsymbol{\theta}}_k) = \sum_{s_T=1}^{K^w} \alpha_T(s_T). \tag{A.88}$$

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

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

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

$$\text{fluorescence emission rates: } \hat{\boldsymbol{v}} = \boldsymbol{M}^{-1} \boldsymbol{b}, \text{ where} \tag{A.90}$$

$$M_{mn} = \sum_{t=1}^{T} \sum_{i=1}^{K^w} \langle s_t^i \rangle F_{in} F_{im}, \tag{A.91}$$

$$b_m = \sum_{t=1}^{T}\sum_{i=1}^{K^w}\langle s_t^i\rangle y_t F_{im}, \tag{A.92}$$

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

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

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

Pooled inference on multiple traces

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

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{s}_{1:N}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\boldsymbol{y}_n, \boldsymbol{s}_n|\boldsymbol{\theta}), \tag{A.95}$$

$$p(\boldsymbol{s}_n|\boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) = p(\boldsymbol{s}_n|\boldsymbol{y}_n, \hat{\boldsymbol{\theta}}_k), \quad 1 \le n \le N. \tag{A.96}$$

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

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

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

$$\mathcal{L}(\boldsymbol{\theta}\,|\boldsymbol{y}_{1:N}, \hat{\boldsymbol{\theta}}_k) = \sum_{n=1}^{N}\sum_{i=1}^{K^w}\sum_{k=1}^{K}\langle s_1^i(n)\rangle C_{ki}\log\pi_k$$

$$+\frac{1}{2}\sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{i=1}^{K^w}\langle s_t^i(n)\rangle\left(\log\lambda-\log(2\pi)-\lambda(y_t(n)-F_{i,:}\boldsymbol{v})^2\right)$$
$$+\sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{i,j=1}^{K^w}\sum_{k,l=1}^{K}B_{ij}\langle s_t^i(n)s_{t-1}^j(n)\rangle C_{ki}C_{lj}\log A_{kl}, \tag{A.98}$$

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

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

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

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

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

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

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

Execution of the cpHMM method

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

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

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

To initialize the fluorescence emission rates, $\boldsymbol{r}$, and the Gaussian precision parameter, $\lambda=1/\sigma^2$, we first treat the fluorescence data $\boldsymbol{y}_{1:N}$ as identical and independently distributed (i.i.d.) and use a simplified time-independent EM algorithm to find their optimal values (cf. Bishop [18], Chapter 13). We initialize the highest emission rate by randomly choosing a

value between 70% and 130% of the highest emission rate inferred by the i.i.d. approach. The lowest emission rate is initialized to 0 because of the apparent silent periods in the activity traces. The remaining $(K-2)$ emission rates are initialized by choosing random values between 0 and the highest emission rate. Finally, we initialize the Gaussian noise σ by randomly choosing a value between 50% and 200% of the noise inferred by the i.i.d. approach.

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

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

Windowed cpHMM

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

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

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

$$\log p(s_{T_1}|\boldsymbol{\pi}^{(T_1-w+1)}, \boldsymbol{A}) = \log \left(p(z_{T_1-w+1}|\boldsymbol{\pi}^{(T_1-w+1)}) \prod_{t=T_1-w+2}^{T_1} p(z_t|z_{t-1}, \boldsymbol{A}) \right)$$

$$= \sum_{i=1}^{K^w} \sum_{n=1}^{K} s^i_{T_1} D^w_{ni} \log \pi_n^{(T_1-w+1)} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \sum_{k,l=1}^{K} s^i_{T_1} D^{d-1}_{ki} D^d_{li} \log A_{kl}. \tag{A.107}$$

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

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

$$\hat{A}_{mn} = \frac{\sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{mi} C_{nj} + \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{mi} D^d_{ni} \log A_{mn}}{\sum_{k=1}^{K} \sum_{t=T_1}^{T_2} \sum_{i,j=1}^{K^w} B_{ij} \langle s^i_t s^j_{t-1} \rangle C_{ki} C_{nj} + \sum_{k=1}^{K} \sum_{i=1}^{K^w} \sum_{d=2}^{w} \langle s^i_{T_1} \rangle D^{d-1}_{ki} D^d_{ni} \log A_{kn}}. \tag{A.108}$$

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

A.3.5 Statistical validation of cpHMM

We validated cpHMM for the three-state ($K = 3$) architecture schematically illustrated in Figure A.19A by generating synthetic trajectories of effective promoter states using the Gillespie algorithm [109] and adding Gaussian noise to the resulting activity traces. Parameters in Figure A.20 were used for data generation. Pooled inferences were conducted on 20 independent datasets, each containing 9,000 data points, representative of the number of experimental data points in a central stripe region. The top panel of Figure A.19B shows the kinetic architecture used to simulate the promoter trajectory in Figure A.19C (yellow) as it switches through the multiple possible states. This promoter trajectory leads to the simulated trace of Figure A.19D (red). Using cpHMM, we found the best fitted path for our observable (Figure A.19D, black) and the corresponding most likely promoter state trajectory (Figure A.19C, blue).

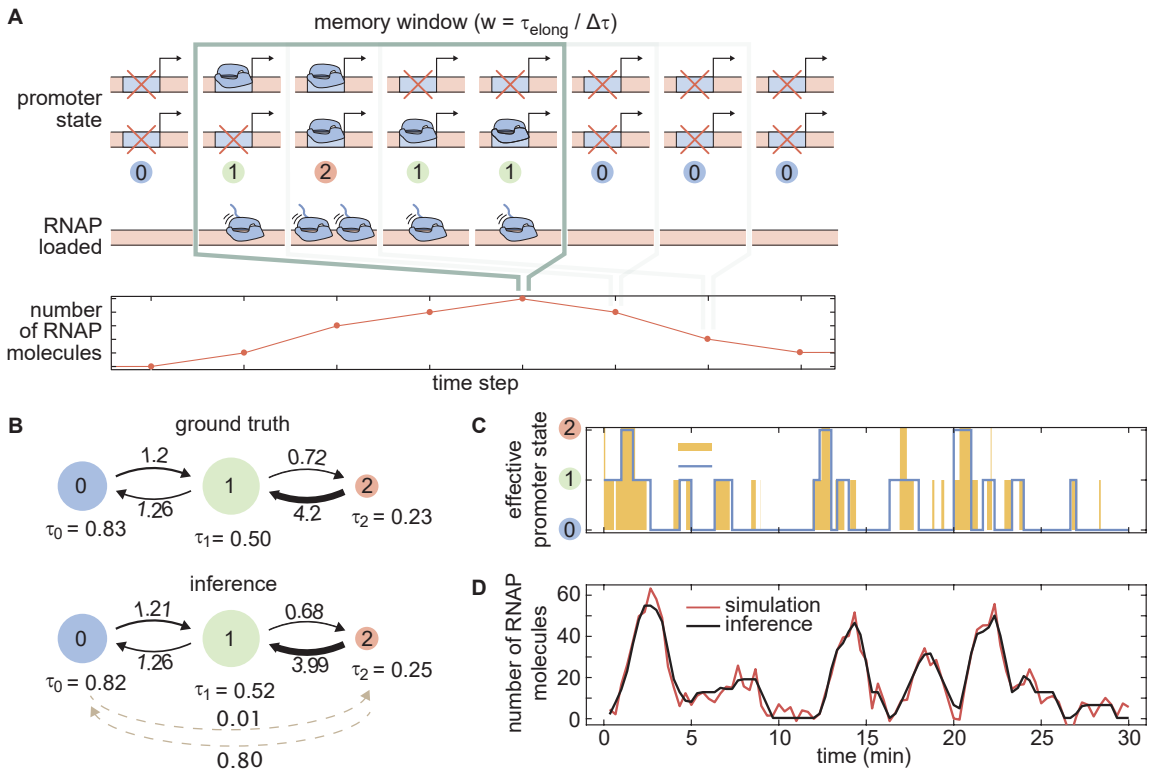

Figure A.19: **Statistical validation of cpHMM. (A)** Three-state cpHMM architecture where ON and OFF promoter states on each sister chromatid result in an effective three-state model. The trajectory of effective promoter states over the memory time window given by the elongation time dictates the number of RNAP molecules loaded onto the gene. **(B)** Flow diagrams of promoter states and transition rates for the true parameters used to simulate trajectories (top) and corresponding average inference results obtained from 20 independent datasets (bottom). The area of each state circle is proportional to the relative state occupancy, and the thickness of the arrows is proportional to the transition rates. Dashed lines correspond to inferred transitions with very slow rates that were absent in the simulation. Rates are in min^{-1} and dwell times are in min. Error bars for the mean inferred parameters are shown in Figure A.21. **(C)** Sample simulated promoter activity trace (yellow) generated using the parameters in (B), overlaid with the best fitted trace (blue) obtained using the Viterbi algorithm [**Viterbi1967**]. **(D)** Simulated and best fitted observable number of RNAP molecules corresponding to the promoter trajectory shown in (C).

Figure A.20: table

Parameter values used for generating synthetic datasets in the statistical validation of the model. In order to perform this validation, we chose parameters that approximated those obtained through the cpHMM inference on experimental data shown in Figure 2.5.

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

As shown in Figure A.19B and Figure A.21, comparison of the simulated and inferred parameters indicates that we reliably recovered the parameters used to generate our simulated data with high precision. We accurately inferred transition rates, dwell times, fraction of time spent in each state, and the rates of RNAP loading over 20 independent datasets of simulated traces.

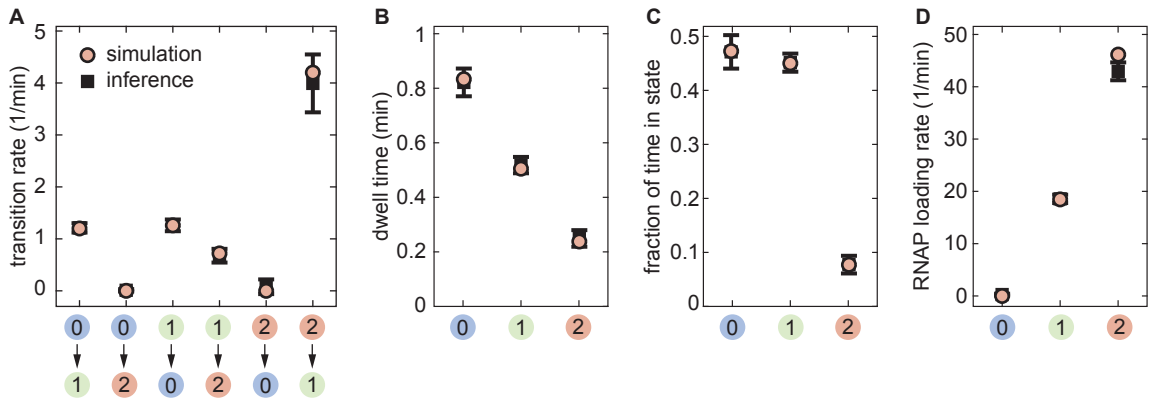

Figure A.21: **Inference statistics for the cpHMM validation.** The true and inferred values of **(A)** transition rates, **(B)** dwell times in states, **(C)** state occupancies, and **(D)** RNAP loading rates are compared. Statistics on the inferred values are obtained from 20 independently generated datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates).

Validation details

We used the relation between the transition rate matrix, **R**, and the inferred transition probability matrix, **A**, defined in Appendix A.3.4 to obtain estimates of the transition rates,

namely,

$$\mathbf{A} = e^{\mathbf{R}\Delta\tau}, \tag{A.109}$$

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

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

Continuous vs. Poisson promoter loading

To demonstrate the validity of our choice to use continuous RNAP initiation rates in the transcription model (Appendix A.3.4), we repurposed our simulation to, instead of considering a constant rate of RNAP loading, explicitly account for individual RNAP loading events when generating the traces. We assumed that individual polymerase molecules traverse at a constant elongation rate ($v_{\text{elong}} = 46$ bp/sec, Appendix A.3.10) and that their arrival to the promoter region has a Poisson waiting time distribution, provided that the promoter is cleared from the previous polymerase molecule which has a finite footprint size of $l_{\text{RNAP}} = 50$ bp [**Rice1992**]. This led to a two-step model for the process of RNAP initiation, with Poisson-distributed wait times for the recruitment of RNAP to the promoter followed by a finite wait period as the RNAP cleared the promoter—a process taken to be approximately deterministic. With this information in hand, we expressed the mean loading time of RNAP at a single promoter (r_1^{-1}) as the sum of the mean time of polymerase arrival at an empty promoter, $\langle\tau_{\text{arrival}}\rangle$, and the time required to clear it after arrival, $\frac{l_{\text{RNAP}}}{v_{\text{elong}}}$, that is,

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

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

We performed inference on these simulated traces using cpHMM with the objective of determining whether a Poisson loading rate had an effect on the obtained parameters. As shown in Figure A.22, when the data is generated using Poisson RNAP loading, cpHMM slightly overestimates the high transition rate, but otherwise manages to accurately recover the model parameters. This therefore justifies our modeling approach of assigning continuous RNAP initiation rates to each promoter state, instead of explicitly modeling the recruitment of individual polymerases.

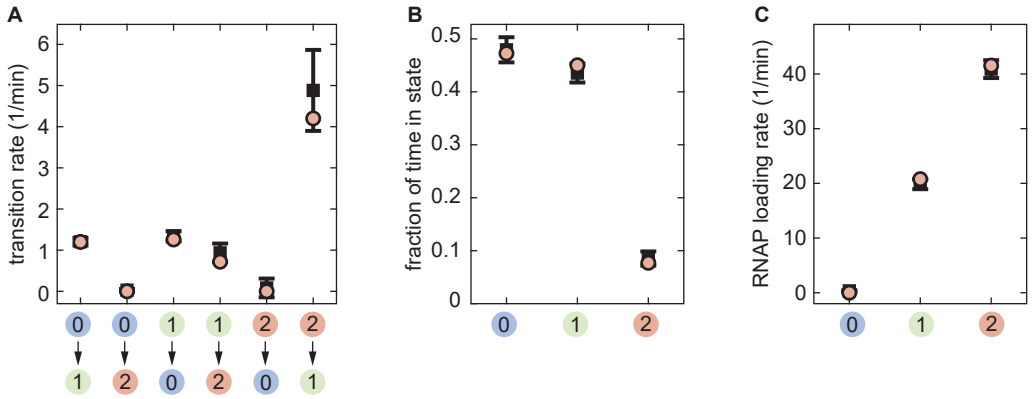

Figure A.22: **Validation of cpHMM on Poisson RNAP loading data. (A)** Transition rates, **(B)** state occupancies and **(C)** RNAP loading rates inferred from 15 independently generated datasets assuming Poisson loading of RNAP. (Error bars represent one standard deviation calculated across these 15 independent replicates.)

Sensitivity of cpHMM to data sampling resolution

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

Figure A.23 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 $= 0.05\text{s}^{-1}$) becomes comparable to the fastest transition rate (0.07s^{-1}). Except for the fastest transition rate, all other rates are inferred accurately for the whole spectrum of sampling resolutions (Figure A.23A). The accuracy of inferred state occupancies is also remarkably high, making it robust to variations in the data sampling rate (Figure A.23B). The high RNAP loading rate tends to be underestimated for slower sampling resolutions, which is reasonable since the chances of promoter leaving state (2) during a single time step become greater, effectively reducing the net rate of loaded RNAP molecules per time step (Figure A.23C). 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.

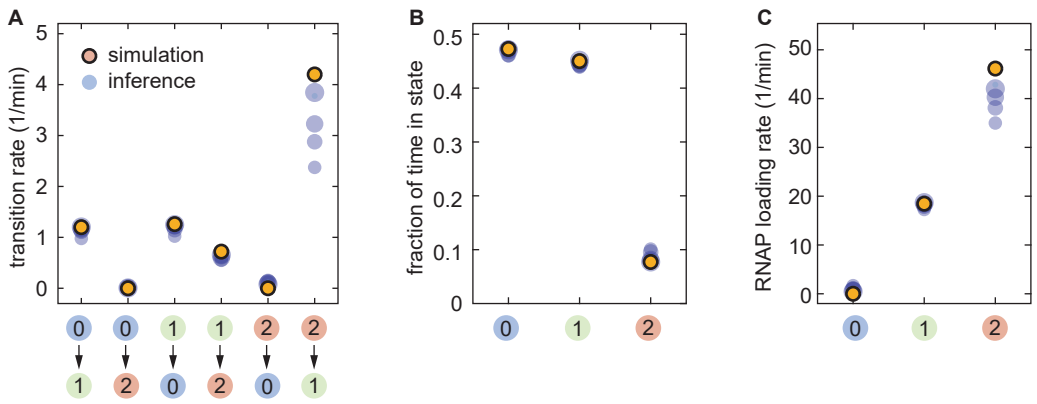

Figure A.23: **Sensitivity of cpHMM to data sampling resolution. (A)** Transition rates, **(B)** state occupancies and **(C)** RNAP loading rates inferred from datasets generated with varying time resolutions. Transparent circles represent averages over 20 independently generated samples. The increasing size of the blue circles corresponds to higher data sampling resolutions (largest: 20s, smallest: 46s).

Performance of cpHMM in different kinetic regimes

Thus far, the validation of cpHMM was performed on datasets that were generated using parameters similar to those inferred for the *eve* promoter. These parameters have characteristic low ON rates (k_{01}, k_{12}) and a high OFF rate (k_{21}), where "low" and "high" are relative to the data sampling frequency, which for our experimental setup is 3/min. To assess the utility of our inference method for a generic choice of model parameters, we performed additional inference studies in three different parameter regimes: low ON rates and low OFF rates (A.3.5–Figure A.24A-C), high ON rates and low OFF rates (Figure A.24D-F), and high ON rates and high OFF rates (Figure A.24G-I).

As expected, the inference is the most accurate when the data sampling frequency is greater than the transition rates (Figure A.24A-C), in which case multiple transitions within a single time frame occur only rarely, making our discrete Markovian representation of the state dynamics a valid approximation. The largest deviations of the inferred model parameters from their ground truth values occur when the ON rates are high and the OFF rates are low (Figure A.24D-F). Since the promoter rarely remains in the lower initiation states (0 or 1) for the entire duration of a frame and tends to rapidly transfer into a higher initiation state (1 or 2, respectively), the rates of RNAP loading for states 0 and 1 are significantly overestimated (Figure A.24F). Despite the inaccuracies in estimating the RNAP loading rates, all transition rates, with the exception of k_{10}, are inferred with a high accuracy (Figure A.24D). Remarkably, the deviations caused by the high ON rates get substantially suppressed when the OFF rates are also made comparably high (Figure A.24G-I). This can be thought of as a consequence of an effective counterbalancing between unwanted ON and OFF transitions within a single time frame.

Overall, these additional studies, together with the statistical validation studies discussed

earlier (Figure A.21), elucidate the domain of applicability of cpHMM: the method performs accurate inference when the ON/OFF transition rates are respectively slow/slow, slow/high, or high/high; and is not successful in accurately inferring some of the model parameters when the ON rates are high, but the OFF rates are low. We hope that these characteristics of the method will be useful in informing the design of promoter architectures and new experiments.

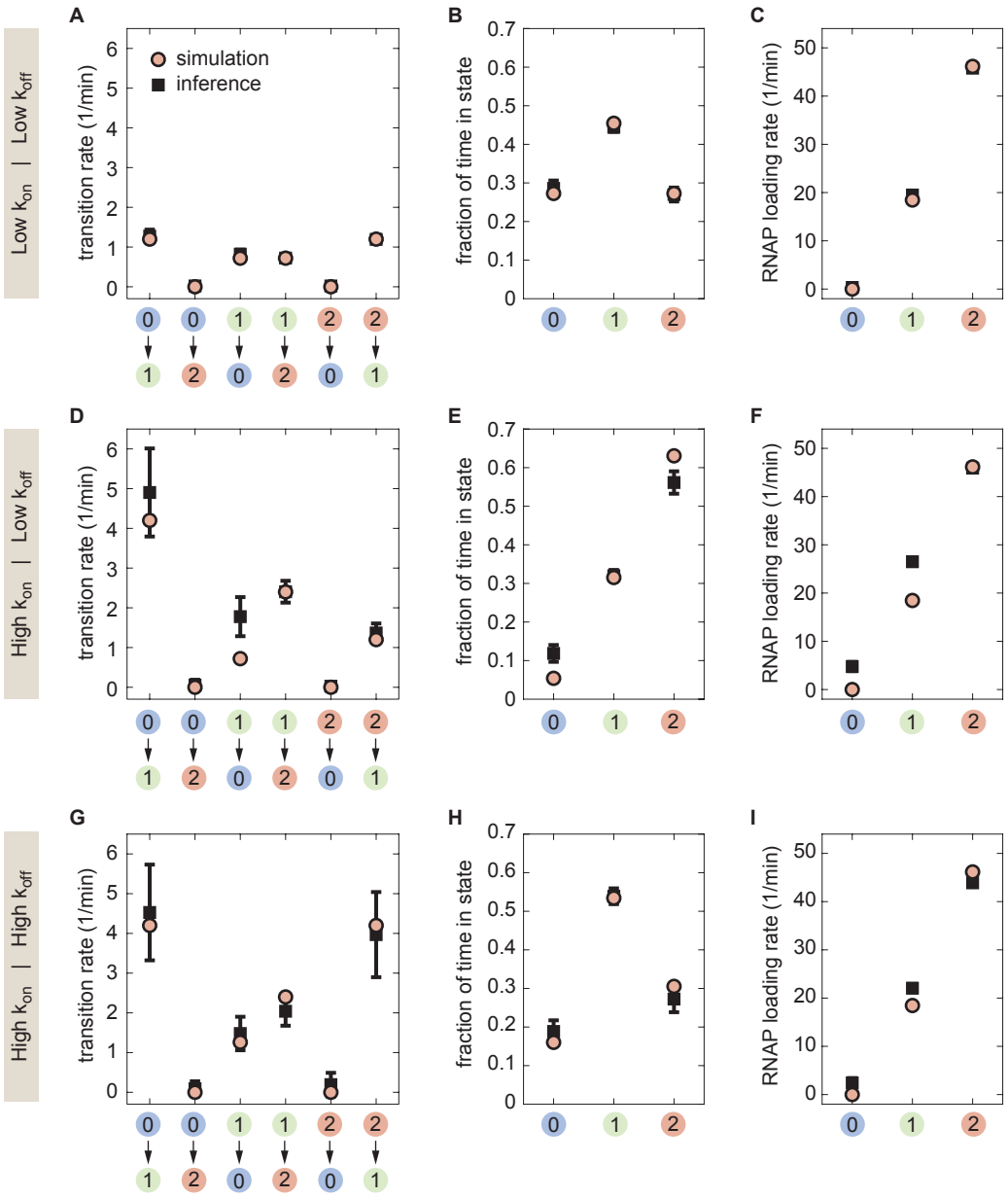

Figure A.24: **Study of cpHMM performance for different choices of the ON/OFF transition rates** Comparison of inference performance for different ON/OFF rates using a data sampling frequency of 3/min. **(A-C)** low/low, (**D-F**) high/low, (**G-I**) high/high. The statistics of inferred model parameter values is obtained from 20 independent datasets. (Error bars indicate one standard deviation calculated across these 20 independent replicates.)

Windowed cpHMM

To check that our windowed cpHMM was capable of fitting time-varying data, we conducted statistical validation using simulated traces exhibiting various time-dependent trends in the bursting parameters. We studied three scenarios that mimicked ways in which bursting parameters could, in principle, be modulated to drive the onset of transcriptional quiescence: a decrease in k_{on} over time, an increase in k_{off} and a decrease in r. We also studied the case of increasing k_{on}, as this was the strongest temporal trend observed in our experimental data. Figure A.25 summarizes the results for these validation tests.

For each test, 100 simulated traces, 40 minutes in length, were generated ($\Delta\tau$ =20 s) that exhibited the desired parameter trends. Consistent with our approach to the experimental data, a sliding window of 15 minutes was used for inference, meaning that for each inference time, τ_{inf}, all data points within 7.5 minutes of τ_{inf} were included in the inference. This led to inference groups consisting of 4500 data points, with the exception of the first and last time points, each of which had 3700 data points (first and last $w+1$ points were excluded from inference). Transition and initiation rates shown in Figure A.25 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 Figure A.27A), as these were found to provide the most faithful indication of underlying system trends.

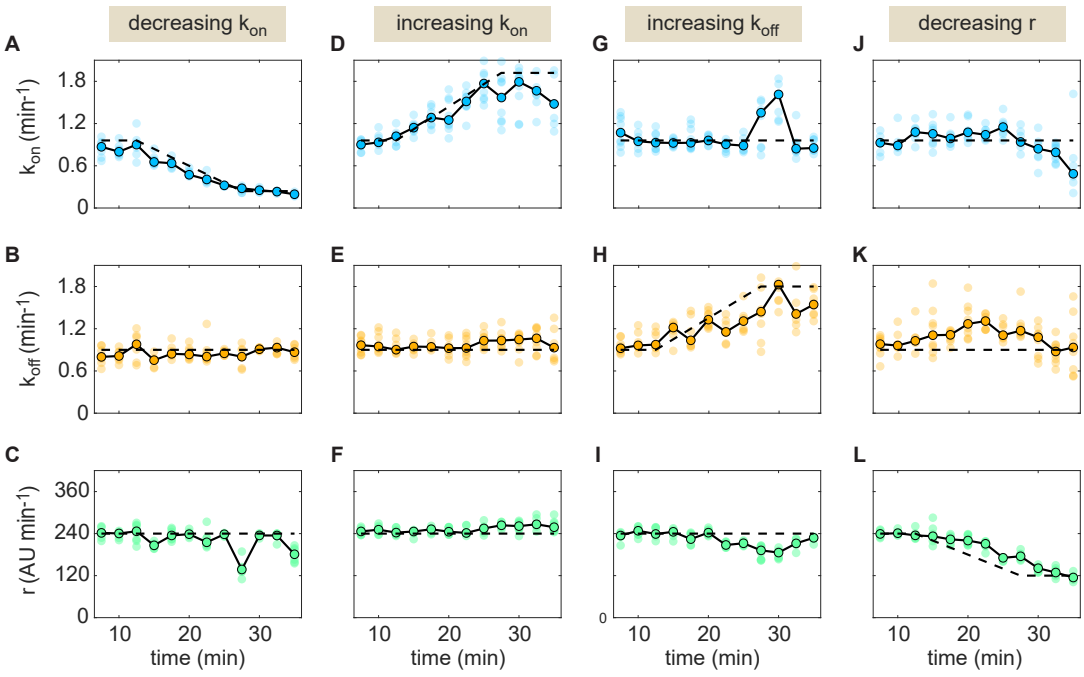

Figure A.25: **Validation of windowed cpHMM inference.** The method's accuracy was tested for four distinct sets of parameter time trends. Results for each scenario are organized by column. In each plot, the black dashed line indicates the true parameter value as a function of time. Connected points (outlined in black) indicate the median inferred parameter value at each time point across 10 distinct replicates. Translucent points indicate inference values from individual replicates. Thus, the dispersion of these replicates at a given time point indicates the precision of the inference.

For each scenario, we assessed whether and to what degree the windowed cpHMM method could accurately recover the temporal profiles. In general, the method was found to perform quite well within the parameter regimes that were tested. For both the increasing and decreasing k_{on} scenarios (Figure A.25A-C,D-E), windowed cpHMM inference accurately captured the modulation in k_{on} with no significant variation evident in the r and k_{off} trends. In the case of increasing k_{off} (Figure A.25G-I), we observed deviations in k_{on} and r from their true values at the inflection point of the k_{off} curve (around 30 min). However, the deviation in r is relatively mild and the "blip" in k_{on}, while of larger magnitude, is comprised of only two time points and so would likely not be mistaken for a legitimate indication of underlying system behavior. In the case of a decrease in the initiation rate (Figure A.25J-L) we observe a $\sim$ 5 min delay in the model response. We attribute this delay to the finite dwell time of RNAP molecules on the gene (in this case τ_{elong} =140 sec, although further studies will be needed to determine why the observed lag appears larger than the elongation time). In addition, we note a degradation in the precision of the inference of k_{on} and k_{off} at low r (RHS of Figure A.25J, K).

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

A.3.6 Sister chromatids

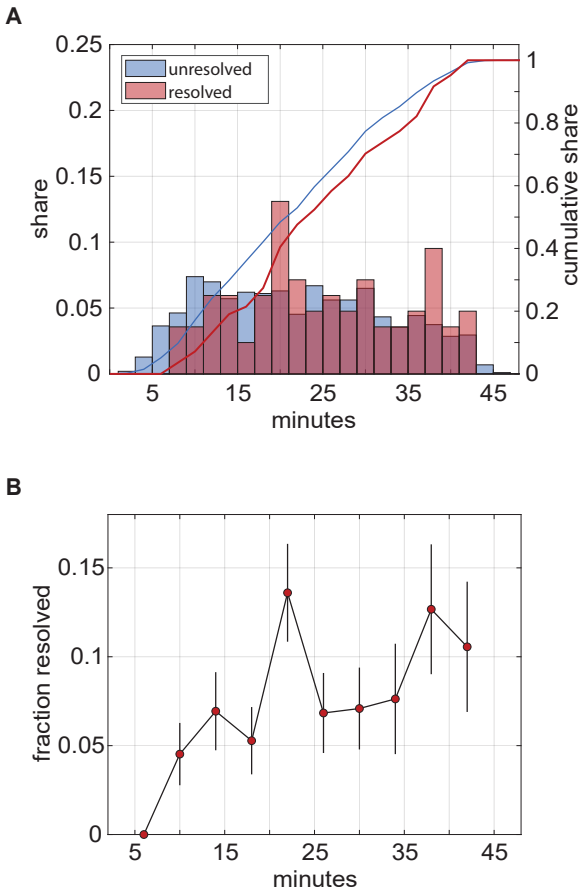

Figure A.26: **Live imaging data indicate timing of sister chromatid appearance. (A)** Distribution of observation times for frames in which chromatids were resolveable (red) and diffraction-limited (blue). Bars indicate empirical probability distribution function. Lines indicate cumulative density function. Data indicate the presence of chromatids by no later than 7-8 minutes into nuclear cycle 14. **(B)** Fraction of frames featuring resolved chromatids as a function of time. Trend suggests replication of relevant portion of genome across all observed nuclei is completed by approximately 10 minutes into nuclear cycle 14. Initial lag is likely attributable—at least in part—to stochastic turn-on times between sister *eve* loci and lower fluorescence levels early on in the nuclear cycle.

Detection of sister chromatid appearance

Previous studies have indicated that the *D. melanogaster* genome is quickly replicated at the beginning of each nuclear cycle in early development [243, 273] , suggesting that each diffraction-limited spot in our imaging data likely contains two distinct *eve* loci. We sought

use our live imaging data to verify whether genome replication occurred early enough in the nuclear cycle such that the presence of the replicated promoters would have to be taken into account. While the two *eve* loci are located within a diffraction-limited spot for the majority of frames in our data, there are a subset of frames in which two distinct puncta can be clearly observed due to fluctuations in the separation between chromatids (see Figure 2.4D). 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 -Figure A.26A, 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 A.4B). Our results indicate that, at the very least, the genomic region containing our *eve* stripe 2 reporter is replicated within *some* nuclei by 6-8 minutes into nuclear cycle 14. Figure A.26B tracks the share of total observations for which we detected resolved puncta as a function of time. A systematic delay in DNA replication would be expected to result in a progressive increase in this metric over time. However, such a trend is not evident. While we see *no* resolved sister loci between 4 and 8 minutes (first point in the plot in Figure A.26B), this absence could be attributed to other factors at play early on in nuclear cycle 14. For example, part of this apparent lag could be attributable to the fact that loci are, on average, dimmer early on in the nuclear cycle, which could mask the presence of two *eve* loci by reducing the probability of both producing observable amounts of fluorescence at the same time. It is also possible that the precise timing of locus turn-on varies for each sister locus, as it does for loci in different nuclei. Regardless, even if the initial rise between 6 and 10 minutes in Figure A.26B *is* reflective of the replication of the locus during this period of time, the relative stability of the frequency of resolved loci from 10 minutes onward indicates that this process is restricted to the first few minutes of transcription. Additional experiments are needed to further elucidate the interplay between DNA replication and the onset of transcription. Regardless, the examination of our live imaging data supports the conclusion that the majority of our data consist of diffraction limited spots containing two distinct *eve* loci.

Probing for interactions between sister chromatids

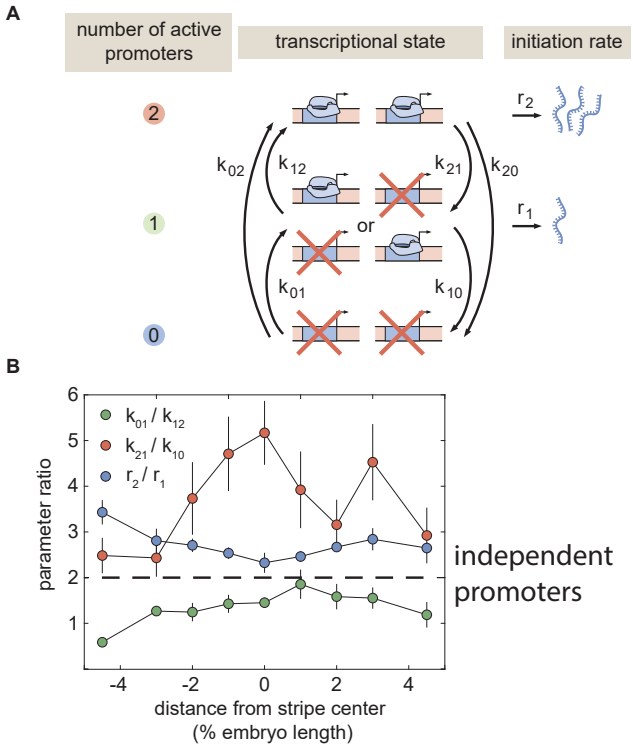

Figure A.27: **Probing combined transcription of sister chromatids. (A)** Revised three-state model of promoter switching within a fluorescent punctum that accounts for the combined action of both sister chromatids. **(B)** Summary of bursting parameter ratios. All three bursting parameter ratios deviate from their expected values under the independence assumption given by the horizontal dashed line. (Error bars indicate magnitude of difference between first and third quartiles of cpHMM inference results for bootstrap samples of experimental data over multiple embryos. See Materials and Methods for details)

If each fluorescent punctum contains two promoters (Figure 2.4D), then it is necessary to revisit the widely used two-state model of transcriptional bursting. In this revised scenario, each promoter on one of the sister chromatids undergoes fast ON/OFF switching. Therefore, each spot (encompassing two identical loci) can be in one of three states: (0) both promoters OFF, (1) one promoter ON and the other OFF, and (2) both promoters ON (Figure A.27B). States (1) and (2) are expected to exhibit different rates of RNAP loading, r_1 and r_2, respectively. See Appendix A.3.1 and Appendix A.3.4 and for details regarding the implementation of this three-state model.

The presence of two transcriptional loci within each fluorescent punctum suggests three constraints on the relationship between bursting parameters in the model shown in Figure A.27A. 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 pro-

moters transitioning simultaneously should be negligible; we expect no transitions between states (0) and (2) such that $k_{02} = k_{20} = 0$. Finally, if the promoters switch between their states in an independent manner, then there will be an extra constraint on their transitions rates. For example, there are two paths to transition from (0) to (1) as either promoter can turn on in this case. However, there is only one possible trajectory from (1) to (2) because only one promoter has to turn 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 [189, 330], classic electron microscopy work suggests a scenario in which sister chromatids are tightly correlated in their transcriptional activity [207, 208]. Given this uncertainty regarding chromatid independence, we elected to employ a general three-state model that makes no assumptions about the nature and strength of sister chromatid interactions. In addition to permitting greater flexibility, this agnostic approach also meant that the structure of the kinetic model returned by cpHMM inference provided clues regarding the nature of the coupling between sister loci. Specifically, we examined the ratios between the high and low on rates (k_{01} and k_{12}), off rates (k_{21} and k_{10}), and initiation rates (r_2 and r_1). A deviation from these expectations would indicate either that the two sister loci do not initiate RNAP independently (first constraint), or that they do not transition between activity states independently (second and third constraint).

Overall, our results suggest that the two loci are coupled to a nontrivial degree. We observe that the rate of initiation for the high state, $r_2(x)$, (corresponding to two active promoters) is consistently greater than twice the middle state, $r_1(x)$ (Figure A.27B, bliue). This trend suggests some sort of synergy in the RNAP initiation dynamics of the sister promoters. Even more strikingly, we observe that the rate of switching from (2) to (1), k_{21}, is *much* higher than twice the rate of switching from (1) to (0), k_{10}, (Figure A.27C, red). This indicates that each promoter is more likely to switch off when its sister locus is also active. This anti-correlation is consistent with some form of competition between the loci, a scenario that could arise, for instance, if local concentrations of activating TFs are limiting. In addition, we observe substantial variation in the relationship between the high and low on rates (k_{01} and k_{12}, respectively), ranging from one of near equality in the anterior flank to nearly the 2-to-1 ratio that would be expected of independent loci in the stripe center and posterior (Figure A.27C, green). Finally, as shown in Figure A.28, 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.

A.3.7 cpHMM inference sensitivities

Full three-state inference results

For the sake of simplicity, we presented our inference results in the main text using an effective two-state model in which two distinct active transcriptional states were combined into a single effective ON state (see Figure 2.4E and F). Here, for completeness, we include time-averaged and time-resolved inference results for the full three-state model where, as shown in Figure A.27, (0) corresponds to the state where both promoters are in the OFF state, (1) indicates the state where either promoter is in the ON state, and (2) represents the states where both promoters are in the ON state.

As indicated in the main text, the full three-state results (Figure A.28) exhibited the same trends as were evident in the effective two-state plots (Figure 2.5). In agreement with the effective two-state model, the rate of transcript initiation is not modulated to a significant degree across the stripe (Figure A.28D). Moreover, we once again see that activation rates, and specifically the rate of switching from OFF to the middle ON rate (states 0 and 1 in Figure A.28E) are strongly elevated in the stripe center.

Like the time-averaged results, time-resolved inference trends for the full three-state model agree closely with effective two-state results shown in main text (compare Figure A.29 to Figure 2.6D-F). Due to a lack of statistics for state (2), we show only transition rates into and out of the first active state (middle state in Figure 2.4E).

Two-state inference results

Although the presence of sister chromatids indicated that the three-state model was most appropriate for the *eve* stripe 2 system, we wanted to check that our conclusions were robust to this assumption. To do this, we conducted time-averaged and windowed inference assuming a simpler, two-state model (*see, e.g.* Figure 2.4B). Note that this approach is distinct from the *effective* two-state results presented in the main text. There, as outlined in Figure 2.4D-F, a three-state model was specified for inference and the results for the two active (ON) states were aggregated after the fact to simplify the presentation of the results. Conversely, here, we explicitly conducted inference using a two-state model.

Most of our findings remained unchanged in the context of the two-state model. Consistent with the three-state case, the two-state time-averaged cpHMM inference indicated that the fraction of time spent in an active state, rather than the rate of RNAP initiation, drives the difference in mRNA production rates across the stripe (Figure A.30A-C). Moreover, as with the three-state case, two-state results indicated that the bulk of this variation stem from modulation in k_{on} (Figure A.30C, green). Interestingly, whereas we did see a degree of spatial dependence in k_{off} for 3-states, we observed no such trend for 2-states (Figure A.30C, red). In general, this is not surprising, as our use of a simpler model likely means that multiple switching rates are being projected onto the k_{off} parameter. Specifically, if the *eve* stripe 2 system is indeed a true three-state system, then we would expect the two-state k_{off} estimate to reflect the joint action of the k_{10}, k_{21}, and k_{12} rates from the three-state model.

As a result, the spatial dependence of each one of these rates would get averaged out when combined onto k_{off}.

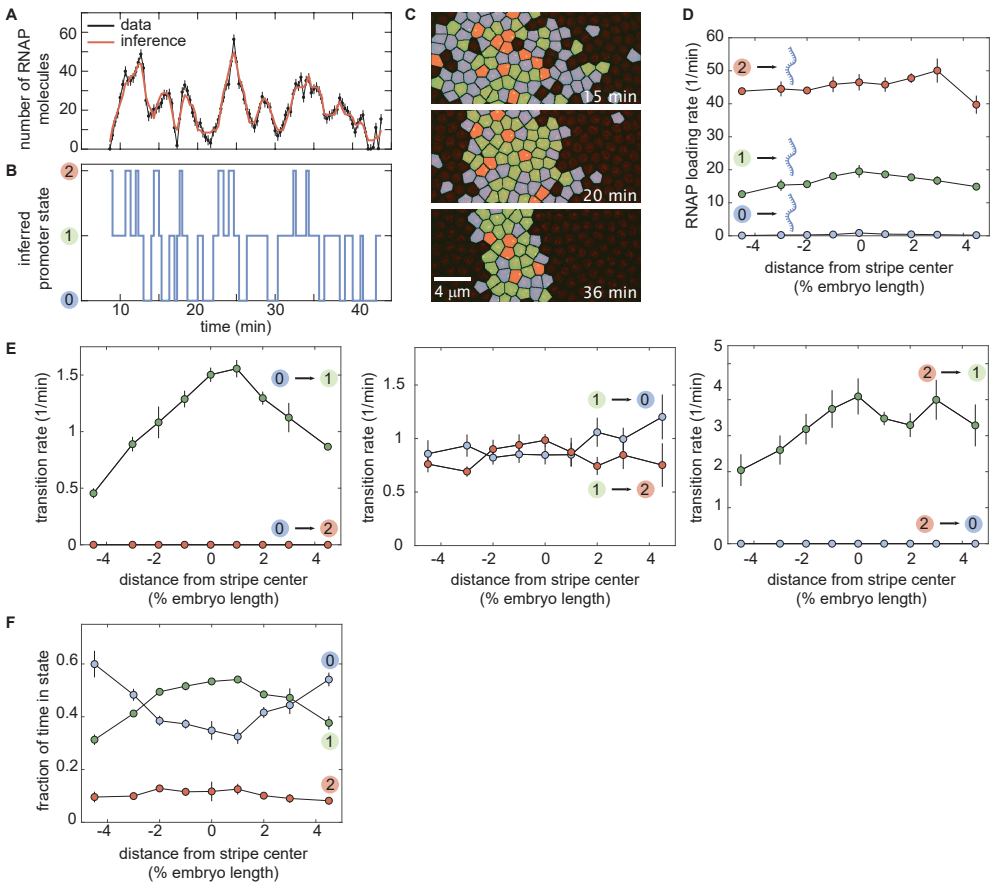

Figure A.28: **Full three-state results for time-averaged cpHMM inference.** **(A)** Representative experimental trace along with its best fit and **(B)** its most likely corresponding promoter state trajectory. **(C)** Instantaneous visualization of promoter state in individual cells throughout development through the false coloring of nuclei by promoter state (colors as in B). **(D)** The rate of initiation for each transcriptional state is not significantly modulated along the embryo. **(E)** Our cpHMM revealed that the transition rate between the OFF (0) and middle ON state (1) is up-regulated in the stripe center. In contrast, the rates of switching out of the middle and high ON states show little to no significant AP-dependent modulation. (**F**) The modulation of the rate of switching from 0 to 1 acts to increase the fraction of time the promoter spends in the active states in the stripe center. (A, error bars obtained from estimation of background fluorescent fluctuations, as described in Materials and Methods and [104]; D, E, and F, error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

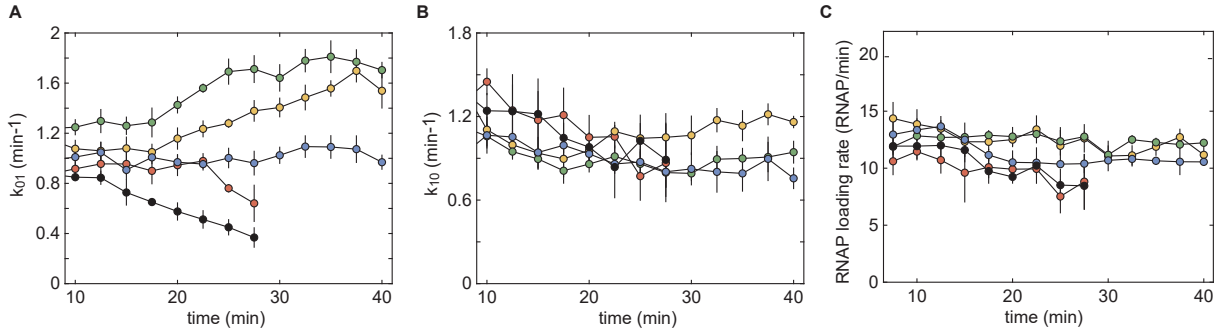

Figure A.29: **Full three-state results for time-dependent cpHMM inference. (A)** Transition rate from transcriptionally inactive state (0) to the first active state (1). Same trends evident as for effective 2 state model. **(B)** Transition rate from first on state (1) to OFF state (0). **(C)** Rate of transcript initiation in first on state (1) as a function of time. (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data taken across 11 embryos. See Materials and Methods for details.)

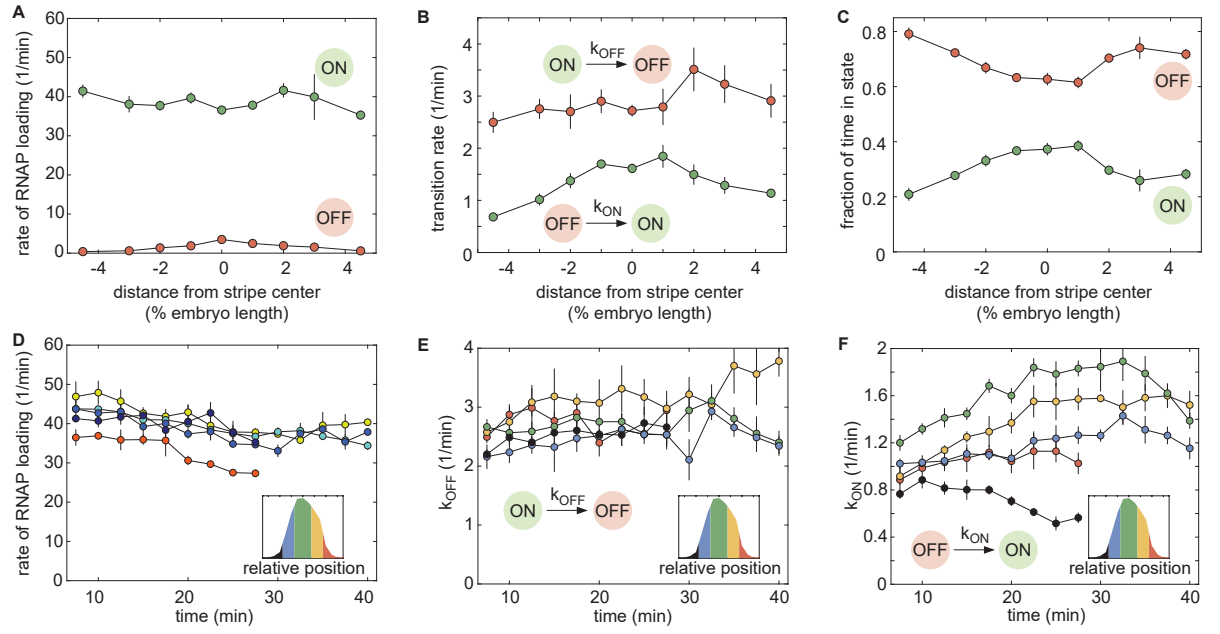

Figure A.30: **Two-state cpHMM inference. (A-C)** Time-averaged 2-state inference results. (A) Consistent with three-state inference results, we observed no significant modulation in the rate of initiation along the axis of the embryo. Moreover, we found that k_{on} (green plot in (B)) was modulated along the anterior-posterior axis to vary the amount of time the promoter spent in the ON state (green curve (C)). In a departure from the three-state case, we observed no significant spatial trend in k_{off}, though we noted a spike in k_{off} at 3% of the stripe center. **(D-F)** Time-resolved (windowed) two-state cpHMM results. (D) Consistent with the 3-state inference, we saw little to no modulation in the rate of RNAP loading r over time, although we noted a mild downward trend across all AP bins that was most pronounced in the posterior flank (red curve). (E) Two-state inference indicated no significant temporal trends in k_{off}. (F) k_{on} time trends largely agreed with the three-state case, although we noted that the decrease in k_{on} in the posterior flank that was apparent in the three-state results was not observable in this two-state context (Figure 2.6E, red). (Error bars indicate the magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrapped samples of experimental data. See Materials and Methods for details.)

As with the time-averaged case, we found that results for two-state windowed cpHMM were generally consistent with three-state trends. A notable exception to this rule was the absence of any significant decrease in k_{on} in the posterior stripe flank (Figure A.30F, red). This is not entirely surprising, as the trend returned by the three-state inference was relatively mild (Figure 2.6E, 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.

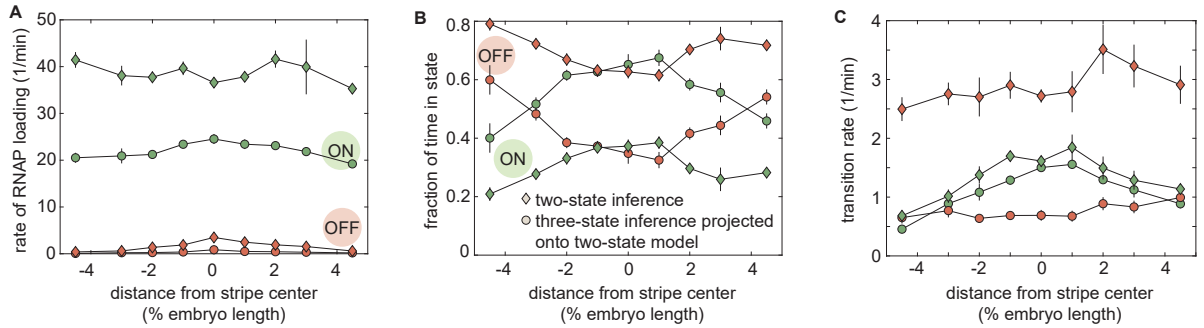

Figure A.31: **Comparing two- and three-state cpHMM inference results.** Three-state inference results can be presented in terms of a two-state model in which states (1) and (2) are aggregated into a single ON state (see Figure 2.4E and F). Here, color schemes are consistent with those employed in Figure A.30A-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 2.5. **(A)** Anterior-posterior-dependent trends in the rate of RNAP initiation are nearly identical between the true and effective initiation rates, however the initiation rate returned by two-state cpHMM inference (green squares) is roughly twice as large as that implied by the three state results (green circles). **(B)** As with the initiation rates, we observe similar trends between the true and effective cases, but substantial differences in magnitude. The effective two-state model recovers an ON state occupancy that is roughly double that returned by two state cpHMM inference. **(C)** While the ON rate trends and magnitudes are nearly identical, the OFF rate returned by two-state cpHMM inference is roughly triple that implied by three-state inference. Thus it is clear that this difference in OFF rate underlies the observed departures in both state occupancies (B) and state initiation rates (A). (Error bars indicate magnitude of the difference between the first and third quartiles of cpHMM inference results for bootstrap samples of experimental data. See Materials and Methods for details.)

As Figure A.31 makes clear, while anterior-posterior-dependent parameter trends are by and large consistent between the true and effective two state models, we do observe substantial differences in the absolute magnitudes of parameter values. These differences originate (directly or indirectly) from the three-fold difference in the value of k_{off} between the true and effective models (Figure A.31C, red squares and circles, respectively). The k_{off} value for the effective two-state model is defined as

$$k_{\text{off}} = \frac{k_{10}k_{21}}{k_{21} + k_{12}}. \tag{A.112}$$

See Appendix A.3.1 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 cpHMM inference implies significantly lower shares falling between 20 to and 40%. Since both models must reproduce the same mean production rate—this is an inherent feature of the experimental traces—we see that the two-state cpHMM inference returns an estimated initiation rate that is consistently twice as large as the initiation rate implied by the effective two-state model.

Thus, while most of the conclusions featured in this paper are robust to our choice of model architecture, this decision does, nonetheless hold important implications for how we

understand the underlying system. Further work is needed elucidate the root cause of this discrepancy and move towards a more concrete understanding of the correspondence between the structure of the model and that of the physical system.

A.3.8 Input-Output analysis details

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

Data sources

The input-output analysis presented in the main text made use of previously published data sets for the spatiotemporal concentration profiles of the gap genes Hunchback, Krüppel, and Giant (Figure A.32A, C and D). These data derive from elegant experiments in which individual embryos were co-immunostained for transcription factors of interest and precisely staged by measuring progressive cellularization over the course of nuclear cycle 14 to generate a time series of protein concentration profiles spanning the course of this period of development [74]. The Bicoid concentration data used for this analysis derives from live imaging experiments using a Bicoid-GFP fusion established by [116]. These data come courtesy of Jonathan Liu and Elizabeth Eck (Figure A.32B).

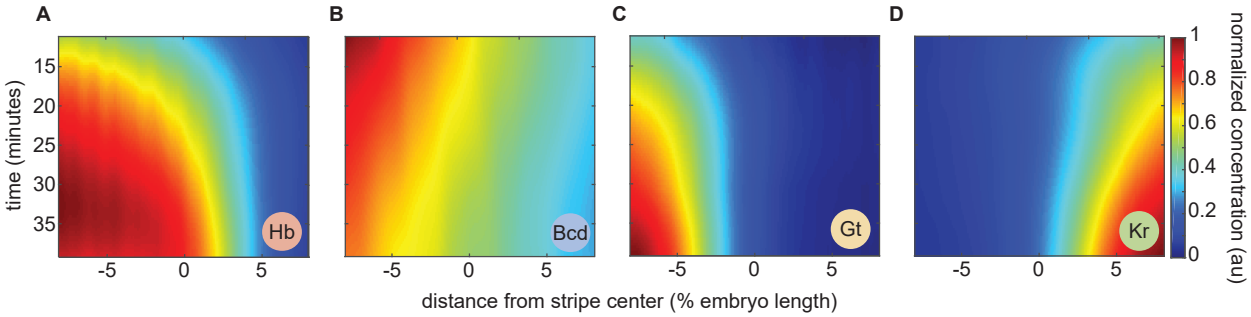

Figure A.32: **Spatiotemporal transcription factor concentration maps.** Heatmaps indicate normalized concentration profiles for the *eve* stripe 2 regulators **(A)** Hunchback, **(B)** Bicoid, **(C)** Giant, and **(D)** Krüppel as a function of space and time. In each case, levels were normalized relative to the maximum concentration observed within the spatiotemporal window of interest.

Data processing

To prepare the Krüppel, Giant, and Hunchback profiles for use in our logistic regression analysis, we adopted an approach similar to that described in [74]. Dorso-ventral orientation of embryos was found to have negligible effect on calculated intensity profiles and was ignored (i.e. all embryos were included, regardless of orientation). For each time point in nuclear

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

$$I_{\text{norm}} = \frac{I_{\text{raw}}}{\max(I_{\text{raw}}) - \min(I_{\text{raw}})} \tag{A.113}$$

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

Finally, we discovered that the anterior-posterior axes in our live imaging data (both for *eve* stripe 2 and Bicoid-GFP) were inconsistent with the axes employed by the fixed data reported by the authors in [74]. We addressed this issue by using *eve* stripe 2 as a fiduciary mark to register the positions of the fixed and live data sets. Specifically, we aligned the mRNA peak predicted by our model at 40 minutes into nuclear cycle 14 with the peak in second stripe of the *eve* protein profile at 40 minutes, as reported in [238].

Logistic regression framework

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

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

Inference details: transcriptional time window

For the time window input-output analysis, we considered only loci that were transcriptionally active for one or more time steps in nuclear cycle 14. Loci were classified as transcriptionally active for all time points between the first and last time points for which they

exhibited detectable levels of transcriptional activity and silent for all time points following their final shut-off for which their nuclei were still present in the experimental field of view. Time points preceding the onset of activity were discarded. Figure A.33A illustrates how this quantity varies over space and time in our experimental data. We considered a class of logistic regression models in which each transcription factor was permitted to appear at most once, thus requiring that each factor act on *eve2* in a uniform manner through space and time; i.e., the same protein could not activate expression on one stripe flank and repress on the other.

Inference details: transcriptional bursting

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

Results of unconstrained inference: time window

For the input-output inference results presented in the main text (Figure 2.7), we used prior knowledge about the regulatory function of each input transcription factor to constrain its range of permissible values in our inference. Specifically, we constrained the activators Bicoid and Hunchback to play activating roles in our model and, likewise, required that the repressors Krüppel and Giant played repressing roles. In several cases, this constrained inference led to models in which one or more transcription factors played no significant regulatory role (Bicoid and Hunchback for the time window and Bicoid for transcriptional bursting). In this section, we tested the sensitivity of the conclusions presented in the main text to our use of functional constraints by conducting unconstrained input-output inference runs.

The results of our unconstrained input-output inference for the transcriptional time window are identical to those presented in the main text. Despite the fact that no limitations were imposed on the regulatory function of each factor, we nonetheless recovered a model in which the two repressors, Giant and Krüppel, are necessary and sufficient to explain the onset of transcriptional quiescence in the stripe flanks. In agreement with the constrained case, we found that the addition of Hunchback and Bicoid to this two-repressor model had no qualitative effect on the output profile predicted by the model (Figure A.33B). A quantitative comparison of model fit scores confirmed that the addition of Hunchback and Bicoid did nothing to improve model fit (Figure A.33C). 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.

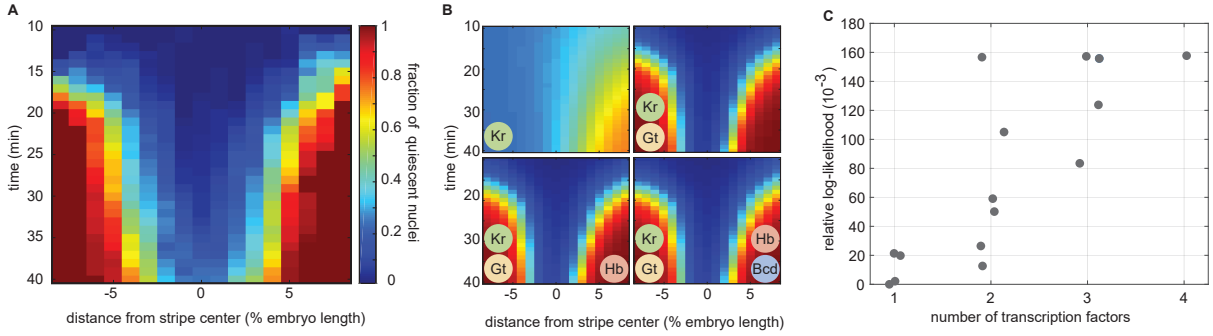

Figure A.33: **Unconstrained inference results for the transcriptional time window. (A)** Observed fraction of quiescent nuclei as a function of space and time. Identical data to that presented in Figure 2.7A. **(B)** Relaxing constraints on the functional nature of each transcription factor had no appreciable effect on the inference results. Profiles shown here are indistinguishable from those shown in Figure 2.7D. Once again, we find that the joint action of the repressors Giant and Krüppel is sufficient to explain the progressive onset of transcriptional quiescence in the stripe flanks. **(C)** A quantitative comparison of model fits reinforces the qualitative conclusions drawn from (B). Models including 3 and 4 transcription factors cannot improve on the fit achieved by the simpler double repressor model. Here blue dots indicate models for which only Giant and Krüppel make significant contributions to the model fit. This indicates that, while the 3 and 4 transcription factor models include additional parameters, these do not contribute appreciably to overall model fit, emphasizing the fact that these models behave, effectively, as double repressor models.

Results of unconstrained inference: transcriptional bursting

In the context of the transcriptional bursting input-output analysis, the removal of functional constraints led to a significantly more complex landscape of inferred regulatory models. While the functional roles of Krüppel, Giant, and Hunchback were consistent with the constrained case (repressing, repressing, and activating, respectively), Bicoid was consistently inferred to play a repressing role. Despite this complication, the three-factor Krüppel-Giant-Hunchback model favored by the constrained inference remained the best-fitting three-factor model (Figure A.34C, red circle). While the addition of Bicoid as a repressor to create a model dependent on all four input transcription factors led to a small improvement in model fit (Figure A.34C), comparison of this four-factor model's predicted activity profile with that of the Krüppel-Giant-Hunchback model revealed no material improvement in the model's agreement with the experimental data (Figure A.34B, 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 [278]. We thus conclude the Krüppel-Giant-Hunchback model remains the most plausible option in the unconstrained case.

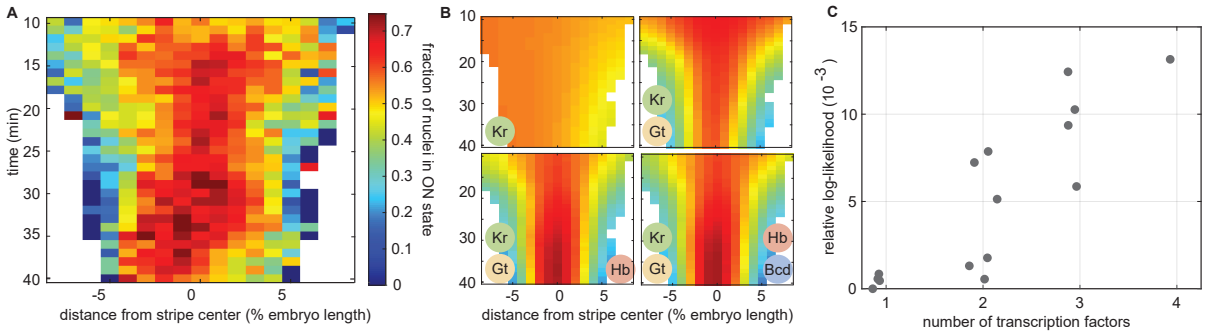

Figure A.34: **Unconstrained inference results for transcriptional bursting. (A)** Observed fraction of transcriptionally active nuclei in the ON (bursting) state. Identical data to that presented in Figure 2.7B. **(B)** As with time window, relaxing the constraints on the functional nature of each transcription factor did little to alter the inference results presented in the main text (compare to Figure 2.7E). As with the constrained results, the joint action of Giant, Krüppel, and Hunchback appears sufficient to explain the spatiotemporal activity pattern revealed by cpHMM inference. **(C)** A quantitative comparison of model fits.

A.3.9 Inherent limits of burst parameter inference

By definition, the onset of transcriptional quiescence coincides with the cessation of observable bursting activity. If this cessation is driven by changes in the bursting parameters as in scenario (ii) in Figure 2.6A, there is an inherent limit to the timescale of such changes that could be detected: changes that unfold over time scales of the same order or faster than the characteristic timescale of the process of transcriptional bursting itself (1-3 min) cannot be detected. Notably, this is not a limit of the cpHMM method, but, rather is inherent to the system—in order to infer bursting parameters, we must observe bursts and, in order to infer a change in parameters, we must have access to bursting activity that reflects this change. Thus, the characteristic frequency of bursts sets an insurmountable resolution limit for any kind of bursting parameter inference. To illustrate this limitation, we simulated three scenarios in which k_{on} decreases to 0 over periods 15, 5, and 1 min in length. We then sought to recover the trend in k_{on}. To emphasize that the limitations are not specific to cpHMM, but rather, are an inevitable consequence of the structure of the system, we used the *true* promoter trajectories (those used to generate the simulated data) to estimate k_{on}. These estimates thus represent the absolute best-case scenario for parameter inference, in which we recover the underlying behavior of the system exactly.

Figure A.35G-H indicate that, even with perfect knowledge of the bursting state at each gene locus, it is not possible to recover a change in the on rate that happens within the span of one minute. These results show that—even under ideal circumstances—there exists a time scale below which time-dependent burst parameter inference will fail to detect shifts in burst parameter values. The absence of bursts following the transition means that, not only are we unable to accurately recover the true trend, but we are also unable even to determine

whether any decrease in k_{on} occurred (on any time scale). Thus, in this scenario, it would be impossible to determine that a modulation in the bursting parameters—as opposed to a transition into some alternative, silent state—drives the onset of quiescence.

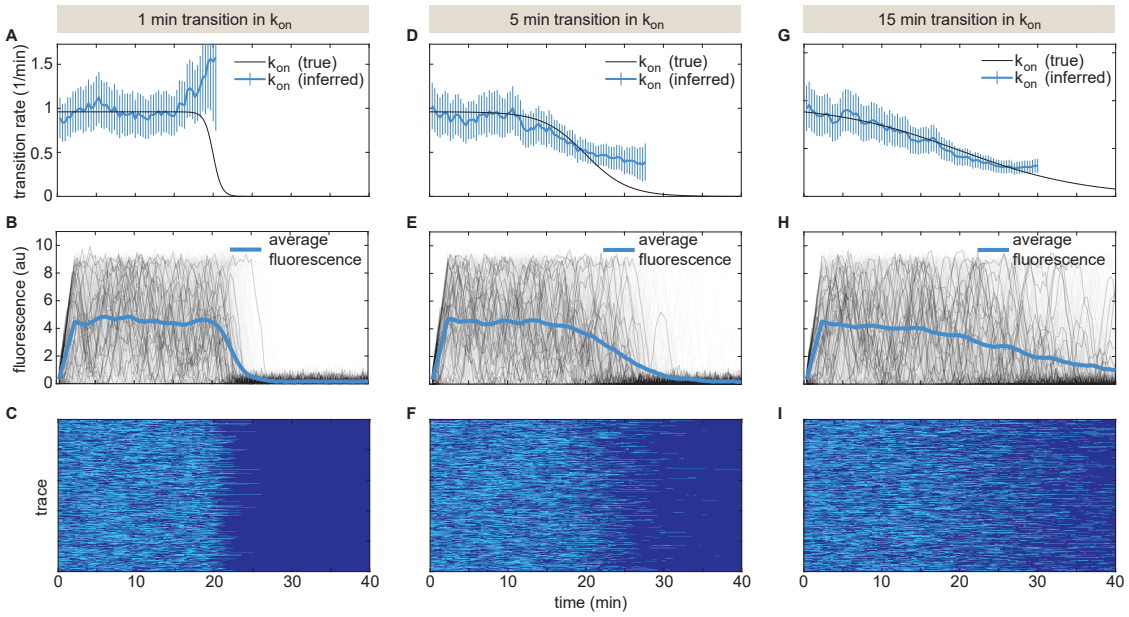

Figure A.35: **Limitations of burst parameter inference.** (**A**-**C**) 15-min transition. (**A**) Black curve indicates true k_{on} value as a function of time and blue curve indicates inferred value. Because the change unfolds on a time scale that is much slower than the bursting timescale, it is possible to accurately recover the underlying k_{on} trend from the fluorescent traces. (**B**) The temporal trend in the average fluorescence across simulated traces (blue curve) reflects this gradual decrease in k_{on}. Note that variation in simulated traces (gray) unfolds on a significantly faster timescale than the change in the mean. (**C**) Visualization of promoter switching. Light blue indicates ON periods and dark blue indicates OFF periods. The observation that bursts of activity are interspersed throughout the k_{on} transition makes it possible to recover the temporal trend. (**D**-**F**) 5 min transition. (**D**) We are able to recover first half of k_{on} trend, but due to the speed of transition, insufficient active traces remain to permit the accurate recovery of the full profile. (**E**, **F**) The onset of quiescence is much starker than in the 15 min case. Because the transition happens faster than in (**A**-**C**), there are fewer bursts that unfold during the transition and, hence, we have fewer reference points with which to infer the underlying trend. (**G**-**I**) 1 min transition. Here the k_{on} transition occurs on the timescale of a single burst. As a result, we are unable to recover the temporal trend. (**H**-**I**) The period of observation is divided in a nearly binary fashion. (A,D,G, error bars indicate 95 % confidence interval of exponential fits used to estimate k_{on}).

A.3.10 Determining the RNAP dwell time using autocorrelation

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

$$w = \frac{\tau_{\text{elong}}}{\Delta\tau}, \tag{A.114}$$

where $\Delta\tau$ is set by the temporal resolution of our data acquisition and τ_{elong} is the elongation time which is unknown *a priori*. Past studies have estimated elongation rates for other systems involved in early patterning in the *Drosophila* embryo, but there is substantial disparity between the reported values. A live imaging study of transcriptional activity driven by the *hunchback* P2 enhancer reported an elongation rate of $1.4 - 1.7$ $kb.min^{-1}$ [104]. However, a recent study of the same regulatory element reported elongation rates of $2.4 - 3.0$ $kb.min^{-1}$—nearly twice as fast [99]. These results suggested that RNAP elongation rates measured for other systems might not apply to our *eve* stripe 2 reporter. Thus, in order to ensure the validity of our inference, we developed an approach that uses the mean autocorrelation function of experimental fluorescence traces to estimate the elongation time directly from our data.

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

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

where μ_f is the average observed fluorescence, σ_f is the standard deviation of the fluorescence and E denotes the expectation value operator. As illustrated in Figure A.36A, the fact that it takes RNAP molecules some finite amount of time to traverse the gene implies that the observed fluorescence at a transcriptional locus at some time t, $F(t)$, will be correlated with preceding fluorescence values $F(t-\tau)$ so long as $\tau < \tau_{\text{elong}}$ because the two time points will share a subset of the same elongating RNAP molecules. As τ increases, the correlation between $F(t)$ and $F(t-\tau)$ due to these shared RNAP molecules will decay in a linear fashion until it reaches zero when $\tau = \tau_{\text{elong}}$ (Figure A.36B, blue curve).

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

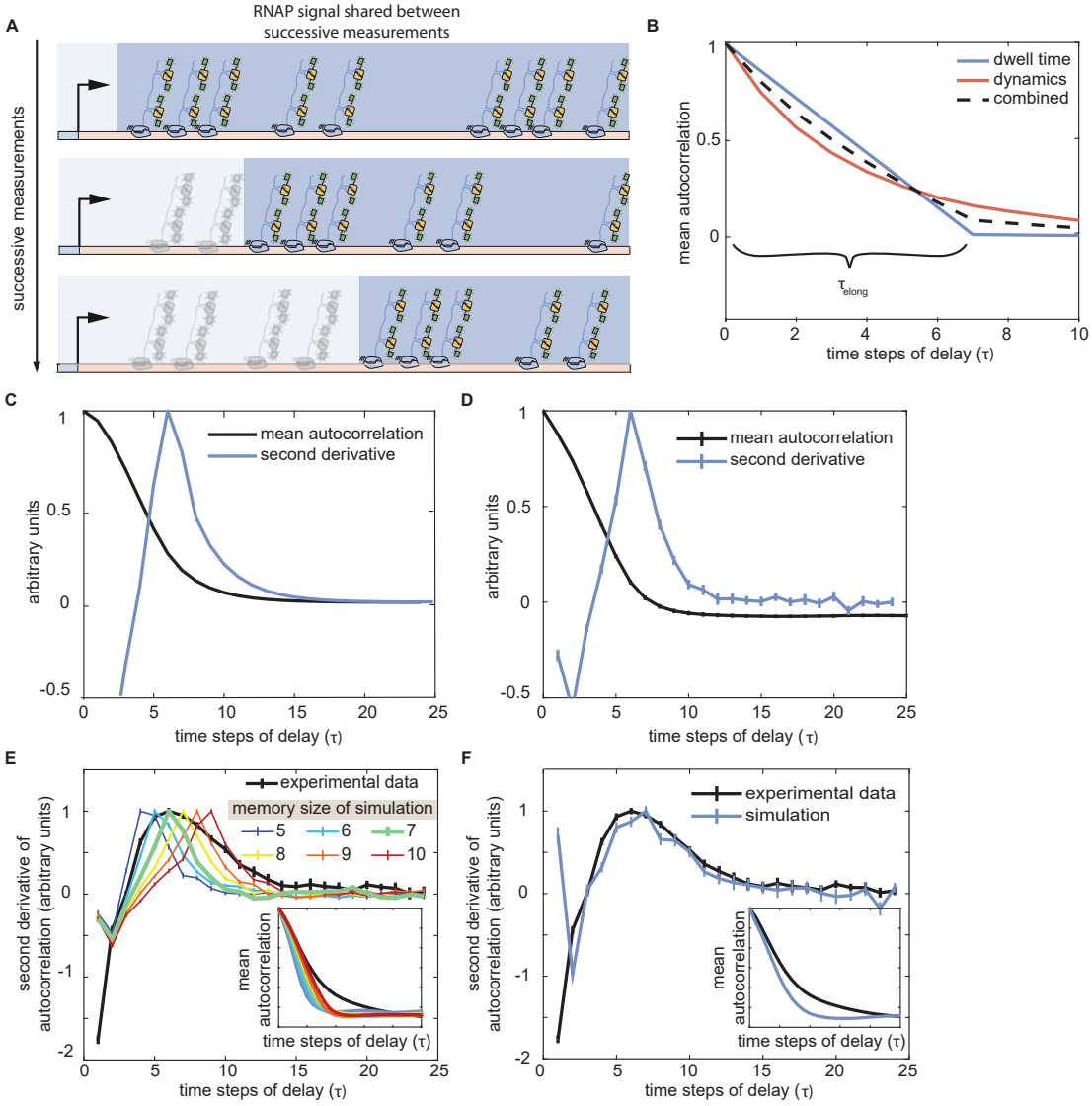

Figure A.36: **Using the autocorrelation of the fluorescence signal to estimate RNAP dwell time. (A)** It takes a finite amount of time for RNAP molecules to transcribe the full length of the reporter gene. As a result, successive fluorescence measurements will contain some of the same GFP-tagged RNAP molecules. Dark blue-shaded regions indicate the subset of RNAP molecules that are present on the gene for successive measurements. **(B)** This overlap causes successive measurements to be correlated, and the degree of correlation due to the overlap decays linearly, reaching zero when the separation between measurements is equal to the elongation time, τ_{elong} (blue curve). However, the trace autocorrelation function contains other signatures that can obscure the inflection induced by RNAP elongation dynamics. For instance, successive time points also exhibit correlation due to the promoter switching dynamics (red curve). **(C)** Theoretical analysis of the autocorrelation function and **(D)** stochastic simulations indicate that the second derivative of the mean autocorrelation function (dark blue curves) can be used to find the structural break in the function (black curves) that corresponds to τ_{elong}. Here, a peak at 6 time steps of delay indicates an elongation time of 7 times steps (140 s). **(E)** Simulated traces with elongation time of 7 time steps (green curve) exhibit a peak in the second derivative that co-

approximately, by the second largest eigenvalue of the Markov chain's transition rate matrix (Figure A.36B, red curve)

$$R_P(\tau) \sim e^{-\lambda_2 \tau}. \tag{A.116}$$

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

A theoretical analysis of $R_F(\tau)$ indicated that the second derivative of the mean autocorrelation function reliably exhibits a peak that can be use to read out the value of τ_{elong}. Figure A.36C shows the analytic prediction for the autocorrelation and second derivative when τ_{elong} is equal to 7 time steps ($w = 7$). We confirmed that the same second derivative approach works in the context of stochastic simulations using realistic parameters for the *eve* stripe 2 system (Figure A.36D). Having confirmed the efficacy of the autocorrelation method for simulated data, we next applied the same technique to uncover τ_{elong} for our experimental traces.

The black profile in Figure A.36E indicates the form of the autocorrelation second derivative for the set of traces used for cpHMM inference. We observed that, while there is a definite inflection point, the peak for the experimental data is much broader than for simulated traces. The most likely cause of this feature is the existence of variability in τ_{elong} (see below). From comparisons of the position of the second derivative peak for experimental traces with simulated profiles, we concluded that an elongation time of $w = 7$ ($\tau_{\text{elong}} = 140\,\text{s}$) best characterized our data (Figure A.36E, green curve). This implies that

$$\begin{aligned} v_{\text{elong}} &= \frac{6444\,\text{bp}}{140\,\text{s}} \\ &= 46\,\text{bp}\cdot\text{s}^{-1} \\ &= 2.8\,\text{kb}\cdot\text{min}^{-1}, \end{aligned} \tag{A.117}$$

where the length used represents the distance from the start of the MS2 step loop sequence to the end of the 3' end of the construct. Interestingly, this elongation rate falls within the $2.4 - 3.0\ \text{kb}\cdot\text{min}^{-1}$ range reported in [99].

Figure A.36F shows how a simple adjustment to our simulation approach, wherein the elongation time steps w for individual RNAP molecules were drawn from a Gaussian distribution with mean $\mu_w = 7$ and standard deviation $\sigma_w = 2.5$ time steps can qualitatively reproduce the wider profile observed in experimental data, indicating that our observations are indeed consistent with the presence of variability in RNAP elongation times. Additional experimental and theoretical work will be necessary to uncover the biological source of this variability.

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

cpHMM inference is insensitive to small changes in RNAP dwell time

Due to the uncertainty in our estimate of τ_{elong}, we conducted sensitivity estimates to ensure that our inference results were robust to our input assumption for w. As shown in Figure A.37, we conducted cpHMM inference on our experimental data assuming different values of w. Based upon our autocorrelation analysis, w values of 6, 7 and 8 seemed the most plausible candidates for the average system elongation time (see Figure A.36E). While small quantitative difference are apparent across these three cases (with a median coefficient of variation of 11%), the results for different values of w showed a constant offset throughout the embryo , such that qualitative trends were largely robust to the assumed w value.

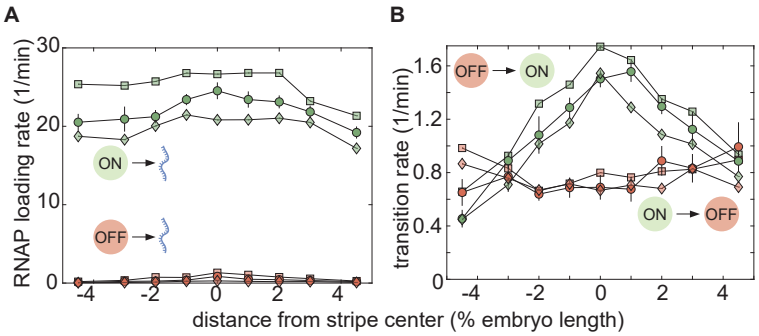

Figure A.37: **Elongation time sensitivities.** Square, circle, and diamond symbols denote inference results for memory time window values w of 6, 7, and 8, respectively. $w = 7$ plots are bolded. Bootstrap errors are shown for $w = 7$ case. **(A)** Initiation rates are robust to w assumption. **(B)** Transition rates also exhibit high degree of robustness to the w used for inference. (Error bars indicate magnitude of difference between first and third quartiles of cpHMM inference results for bootstrap samples of experimental data.

A.4 Movies

Video 1. Transcriptional activity of *eve* stripe 2 reported by MS2. Raw MS2 signal where fluorescent puncta report on the number of actively transcribing RNAP molecules.
Video 2. Mean rate of transcription of *eve* stripe 2 reported by MS2. Nuclei false colored by their time-averaged transcriptional activity (up to the depicted time point).
Video 3. Transcriptional time window. Nuclei along the stripe false colored after the duration of their transcriptional time window.
Video 4. Fraction of active nuclei. Nuclei along the stripe false colored according to whether they engaged in transcription at any time point during the nuclear cycle.
Video 5. Fluorescent puncta contain sister chromatids. Fluorescent puncta transiently separate to reveal the presence of sister chromatids as shown by the white circles throughout the movie.

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

Video 7. Average embryo containing all inputs and the output. Average concentrations of Bicoid (blue), Hunchback (red), Kŕuppel (green) and Giant (yellow) combined with the average transcriptional activity of the *eve* reporter (purple). (Hunchback, Kŕuppel and Giant data obtained from [74]).

Appendix B

SI for A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting

B.1 Literature summary of timescales of transcriptional bursting and associated molecular processes

In this section, we present a survey of timescales observed for transcriptional bursting across a broad swath of organisms (Appendix Table B.1). Further, we review *in vivo* and *in vitro* measurements that have revealed the timescales of the molecular transactions underlying transcription and its control.

Recent technological advances such as single-molecule tracking, live-cell imaging, and a variety of high-throughput sequencing methods, have revealed how eukaryotic transcription is driven by a dizzying array of molecular processes that span a wide range of timescales. The overview of these timescales presented in Figure 5.1E show how many of these processes are significantly faster than transcriptional bursting.

Chromatin accessibility is a central control point for regulating transcription in eukaryotes [76, 29]. DNA wrapped around nucleosome restricts transcription factor access [162, 29]. Multiple studies have determined the timescales of spontaneous DNA unwrapping and rewrapping to be around 0.01-5 s [154, 183, 295]. While unwrapping and rewrapping are probably too fast to directly lead to long transcriptional bursts, DNA unwrapping might represent a "foothold" by which factors transiently bind DNA and enact larger-scale, sustained chromatin modifications [29].

Interestingly, nucleosome turnover occurs over a longer timescale compatible with bursting, with multiple studies suggesting timescales of several minutes to hours [67, 108, 310, 216, 159]. Recent genome-wide studies have measured average nucleosome turnover time to be approximately 1 hour in the fly and in yeast [67, 108]. Further, histone modifications may modulate nucleosomal occupancy [29, 194, 171], and the half-life as well as addition of these modifications can also span a broad range of timescales compatible with bursting, from several minutes to days [41, 140, 155, 326, 17, 30, 126].

Once the chromatin is open, enhancers, DNA stretches containing transcription factor binding sites and capable of contacting promoters to control gene expression, become accessible. Transcription factor binding recruits co-factors and general transcription factors to the promoter, triggering the biochemical cascade that ultimately initiates transcription [58]. While the resulting bursts of RNAP initiation last from a few minutes to hours (Figure 5.1A-D), single-molecule live imaging has shown that transcription factor binding is a highly transient process, with residence times of 0.5-15 s [214, 212, 39, 220, 105, 206, 283, 328]. The vast majority of transcription factors bind DNA for seconds, but it is worth noting that some transcription factors and chromatin proteins can bind DNA for minutes [291].

However, the binding of transcription factors, the general transcriptional machinery, and RNAP to the DNA might be more complex than the simple cartoon picture of individual molecules engaging and disengaging from the DNA. For example, recent experiments have revealed that both mediator and RNAP form transient clusters with relatively short lifetimes in mammalian nuclei of 5-13 s 10 s, respectively [43, 42, 51, 40]. In addition, it is demonstrated that transcription factors can also form clusters *in vivo* [212, 214]. However, how these cluster dynamics relate to transcriptional activity remains unclear.

Further, enhancers and promoters are often separated by kbp to even Mbp. The mechanism by which enhancers find their target loci from such a large distance, and how this contact triggers transcription, remain uncertain and are reviewed in [101]. *In vivo* measurements of enhancer-promoter separation in the *Drosophila* embryo have shown that this distance fluctuates with a timescale of tens of seconds to several minutes [38, 127, 184] —timescales strikingly similar to those of bursting. However, recent work has cast doubt on the simple "lock and key" model of enhancer association (stable, direct contact between enhancers and promoters triggers transcription), suggesting instead that enhancers may activate cognate loci from afar and, in some cases, may activate multiple target loci simultaneously [38, 98, 184, 101, 118, 10, 261]. Many important questions remain about the nature of enhancer-driven activation and it remains to be seen whether enhancer association dynamics are generic aspect of eukaryotic transcriptional regulation, or whether they only pertain to a subset of organisms and genes.

A single transcriptional burst generally consists of multiple RNAP initiation events ($\sim$10-100 at a rate of 1/6-1/3 s when the promoter is ON in *Drosophila*, for instance) [27, 165]. The transcriptional bursting cycle thus encompasses a smaller, faster biochemical cycle in which RNAP molecules are repeatedly loaded and released by the general transcription machinery. One interesting hypothesis for the molecular origin of transcriptional bursting is that the OFF state between bursts is enacted by an RNAP molecule that becomes paused at the promoter, effectively creating a traffic jam [151]. Live imaging and genome-wide studies have shown that RNAP pausing before initiation is common in eukaryotes [151, 53, 102, 8] and that its half-life of up to 20 min can be consistent with transcriptional bursting [284, 130, 270, 163, 32, 150, 129, 66].

Although the dynamics of some of the molecular processes outlined above are compatible with the long timescales of transcriptional bursting, we still lack a holistic picture of how these kinetics are integrated to realize transcriptional bursts and, ultimately, to facilitate the regulation of gene expression by transcription factors.

We must also acknowledge that we still lie at the very beginning of a reckoning with the dynamics of transcriptional processes as measurements for some molecular processes results in a range of timescales that are difficult to reconcile. In particular, we still lack solid dynamic measurements regarding the assembly of the transcription preinitiation

complex. Yet, perhaps more egregious than the lack of any individual dynamical measurement is the lack of a comprehensive, quantitative, and predictive understanding of how these molecular processes interact with one another in time and space to give rise to transcriptional bursting.

Table B.1: **Literature summary of transcriptional bursting.** We attempted to summarize the duration of a single transcriptional burst from various organisms and genes. In the cases where the single-cell data is not available, such as in data stemming from smFISH experiments, we used population averaged T_{ON} and/or T_{OFF} values instead to give a sense on the timescales.

System	Method	Bursting Timescale	Reference
Bacteria			
in vitro	single-molecule assay	5-8 minutes	[44]
Tet system	MS2	$T_{ON} \approx 6$ minutes, $T_{OFF} \approx 37$ minutes	[111]
Fruit fly embryo			
even-skipped stripe 2	MS2	few minutes	[165, 27]
even-skipped	MS2	few minutes	[12]
Notch signaling	MS2	5-20 minutes	[85]
snail, Krüpple	MS2	5 minutes	[98]
gap genes: *hunchback, giant, Krüpple, knirps*	smFISH	$T_{ON} \approx 3$ minutes, $T_{OFF} \approx 6$ minutes	[330]
hunchback	MS2	few minutes	[70]
even-skipped stripe 2	MS2	few minutes	[27]
C. elegans			
Notch signaling	MS2	10-70 minutes	[177]
Human, Mouse			
TGF-β signaling	luciferase assay	few hours	[217]
TFF-1 signaling	MS2	few hours	[251]
HIV-1 viral gene	MS2	few minutes	[290]
liver genes	smFISH	$T_{ON} \approx 30$ minutes - 2 hours	[121]

mammalian genes	luciferase assay	few hours	[287]
Amoeba			
actin gene family	RNA-seq	few hours	[297]
actin gene family	MS2	10-15 minutes	[55]

Table B.2: **Summary of measured timescales of underlying molecular processes associated with transcription.** While the vast majority of transcription factors bind DNA for seconds, it is worth noting that some transcription factors (e.g. TATA-binding protein) and chromatin proteins (e.g. CTCF, Cohesin) can bind DNA for minutes. These outliers are not included in Figure 5.1

System	Organism	Experimental method	Timescale	Reference
Nucleosomal DNA Wrapping/Unwrapping				
Mononucleosomes	*In vitro* reconstitution	FRET	0.1-5 s	[295]
Mononucleosomes	*In vitro* reconstitution	FRET	10-250 ms	[183]
Mononucleosomes	*In vitro* reconstitution	Photochemical crosslinking	<1 s	[154]
Nucleosome Turnover				
Histone H3.3	*Drosophila* cell	Genome-wide profiling	1-1.5 h	[67]
Histone H3	Yeast	Genomic tiling arrays	∼1 h	[108]
Histone H2B, H3, and H4 tagged with GFP	Human cell	FRAP	several minutes	[159]
Histone H1 tagged with GFP	Human cell	FRAP	several minutes	[216]
Histone H3	Plant cell (Alfalfa)	Isotope labeling	several hours	[310]
Histone Modification				
dCas9 inducible recruitment	Mammalian cell	Single-cell imaging	several hours to days	[30]
rTetR inducible recruitment	Mammalian cell	Single-cell imaging	several hours to days	[17]
Chemical-mediated recruitment	Mammalian cell	Chromatin *in vivo* assay	several days	[126]
Histone H3	Human cell	Liquid chromatography, mass spectrometer and heavy methyl-SILAC labeling	several hours to days (half-maximal time of methylation)	[326]
Targeted recruitment	Yeast	ChIP	5-8min (reversal of targeted deacetylation) 1.5 min(reversal of targeted acetylation)	[155]
Histone H2a, H2b, H3, and H4	Mammalian cell	Isotope labeling	<15 min (acetylation half-life)	[41]

Histone H2, H2a and H2b	Mammalian cell	Isotope labeling	∼3 min (acetylation half-life)	[140]

Transcription Factor Binding				
Bicoid	*Drosophila* embryo	SMT	∼2 s	[214]
Bicoid	*Drosophila* embryo	SMT	∼1 s	[212]
Zelda	*Drosophila* embryo	SMT	∼5 s	[214]
Sox2	Mammalian cell	SMT	∼12-15 s	[39]
p53	Mammalian cell	SMT	∼3.5 s	[220]
p53	Mammalian cell	SMT, FRAP, FCS	∼1.8 s	[206]
Glucocorticoid receptor	Mammalian cell	SMT	∼8.1 s	[220]
Glucocorticoid receptor	Mammalian cell	SMT	∼1.45 s	[105]
STAT1	Mammalian cell	SMT	∼0.5 s	[283]
TFIIB	*In vitro* reconstitution	SMT	∼1.5 s	[328]
TATA-binding protein	Mammalian cell	SMT	1.5-2 min	[291]
Chromatin Protein Binding				
CTCF	Mammalian cell	SMT	∼1-2 min	[124]
Cohesin	Mammalian cell	SMT	∼22 min	[124]
Enhancer-Promoter Interaction				
snail shadow enhancer	*Drosophila* embryo	MS2, PP7 labeling	∼10-40 s (fluctuation cycle interval)	[127]
snail enhancer	*Drosophila* embryo	MS2, PP7 labeling	several minutes	[184]
endogenous *even-skipped* locus with homie insulator	*Drosophila* embryo	MS2, PP7 labeling	several minutes	[38]
Transcription Initiation				
even-skipped stripe 2 enhancer	*Drosophila* embryo	MS2 labeling	∼3 s (promoter ON)	[165]
HIV-1 promoter	Mammalian cell	MS2 labeling	∼4.1 s (promoter ON)	[290]
hb P2 enhancer	*Drosophila* embryo	MS2 labeling	∼6 s	[104]
RNAP Cluster Dynamics				
RNAP tagged with Dendra2	Mammalian cell	tcPALM	∼12.9 s (with small fraction of stable clusters)	[42]
RNAP tagged with Dendra2	Mammalian cell	tcPALM	∼8.1 s	[43]

RNAP tagged with Dendra2	Human cell	Bayesian nanoscopy	several seconds	[40]
RNAP tagged with Dendra2	Human cell	tcPALM	∼5.1 s	[51]
Mediator Cluster Dynamics				
Mediator tagged with Dendra2	Mammalian cell	tcPALM	∼11.1 s	[42]
Promoter-Proximal Pausing				
RNAP tagged with GFP	Human cell	FRAP	∼40 s	[284]
RNAP (genome-wide)	*Drosophila* cell	RNA sequencing	∼2-20 min	[130]
RNAP (genome-wide)	*Drosophila* cell	ChIP-nexus	∼5-20 min	[270]
RNAP (genome-wide)	*Drosophila* cell	Genome-wide footprinting	∼2.5-20 min	[163]
RNAP tagged with GFP	Intact *Drosophila* salivary glands		∼5 min	[32]
RNAP (genome-wide)	Mammalian cell	GRO-seq	∼6.9 min (average)	[150]
RNAP (genome-wide)	*Drosophila* cell	scRNA-seq	15-20 min (at genes with low activity)	[129]
LacO-tagged minimal CMV promoter	Human cell	MS2 labeling, FRAP	∼4 min	[66]

SMT: Single-molecule tracking

FRAP: Fluorescence recovery after photobleaching

FCS: Fluorescence correlation spectroscopy

FRET: Fluorescence resonance energy transfer

ChIP: Chromatin immunoprecipitation

PALM: Photo-activated localization microscopy

B.2 Two-state model calculations

As noted in the main text, the average initiation rate is equal to r times the fraction of time the promoter spends in this ON state p_{on},

$$\left\langle \text{initiation rate} \right\rangle = r\, p_{on}. \tag{B.1}$$

To predict the effect of bursting on transcription initiation, it is necessary to determine how p_{on} depends on the bursting parameters. In the mathematical realization of the two-state model shown in Figure 5.2A, the temporal evolution of p_{off}, the probability of being in the OFF state, and p_{on} is given by

$$\frac{dp_{off}}{dt} = -k_{on}\, p_{off} + k_{off}\, p_{on}, \tag{B.2}$$

and

$$\frac{dp_{on}}{dt} = k_{on}\, p_{off} - k_{off}\, p_{on}. \tag{B.3}$$

To solve these equations, we make the simplifying assumption that our system is in steady state such that p_{on} and p_{off} are constant in time. In this scenario, we can set the rates dp_{off}/dt and dp_{on}/dt to zero. We can then solve for k_{off} in terms of k_{on} resulting in

$$k_{off} = \frac{k_{on}\, p_{off}}{p_{on}}. \tag{B.4}$$

Plugging in $p_{on} + p_{off} = 1$ gives us

$$k_{off} = \frac{k_{on}\,(1 - p_{on})}{p_{on}}, \tag{B.5}$$

which can be solved in terms of k_{on}, k_{off}

$$p_{on} = \frac{k_{on}}{k_{on} + k_{off}}. \tag{B.6}$$

B.3 Molecular model calculations

Here we provide a brief overview of the calculations relating to the three theoretical models of transcription presented in Section 5.3: the independent binding model (Figure 5.3C), the cooperative binding model (Figure 5.3E) and the rate-limiting step model (Figure 5.3G). We also provide resources relating to the calculation of first-passage time distributions discussed in section 5.4.

B.3.1 Stochastic simulations

We make heavy use of stochastic simulations throughout this work. A custom-written implementation of the Gillespie Algorithm [110] was used to simulate trajectories for the various models discussed in the main text. These simulated trajectories were used to generate the activity trace plots in Figure 5.3D, F, and G, as well as the first-passage time distributions in Figure 5.4B-D. All code related to this project (including the Gillespie Algorithm implementation for stochastic activity trace generation) can be accessed on GitHub.

B.3.2 Independent binding model

All calculations in this section pertain to the independent binding model presented in Figure 5.3C.

B.3.2.1 Calculating state probabilities

Calculating the probability of each activity state is central to determining a system's overall transcriptional behavior. Because our mathematical model is a linear chain with no cycles (see Figure 5.3B), we can make progress towards calculating the state probabilities, p_i, by imposing detailed balance, which gives

$$p_n k_+(n) = p_{n+1} k_-(n+1), \tag{B.7}$$

where k_+ and k_- are the effective rates of adding and subtracting a single activator molecule that we define in Figure 5.3B. Plugging in Equation 5.5 and Equation 5.6 from the main text results in

$$p_n(N-n)k_{n,n+1} = p_{n+1}(n+1)k_{n+1,n}, \tag{B.8}$$

where, the rates $k_{n,n+1}$ and $k_{n+1,n}$ are the microscopic binding and unbinding rates defined in Figure 5.3A, respectively. Now we make use of the fact that there are only two unique microscopic rates in independent binding system: activator molecules bind at a rate $k_{n,n+1} = k^b = k_0^b[A]$, with $[A]$ being the activator concentration and k_0^b the binding rate constant, and unbind at a rate $k_{n,n-1} = k^u$. Plugging these values into Equation B.8 and rearranging leads to

$$p_{n+1} = \left(\frac{N-n}{n+1}\right)\left(\frac{k^b}{k^u}\right)p_n. \tag{B.9}$$

To further simplify the expression in Equation B.9, we write $\frac{k^u}{k^b}$ as a dissociation constant (K_d), resulting in

$$p_{n+1} = \left(\frac{N-n}{n+1}\right)\frac{p_n}{K_d}, \tag{B.10}$$

which has the form of a recursive formula for calculating state probabilities from their predecessors. For instance, for the case where $n = 0$ we have

$$p_1 = N\frac{p_0}{K_d}. \tag{B.11}$$

We can extend this logic to calculate the probability of any state, n, as a function of p_0, leading to

$$p_n = \frac{N!}{(N-n)!n!}\frac{p_0}{K_d^n} = W(n)\frac{p_0}{K_d^n}, \tag{B.12}$$

where the factorial terms captured by $W(n)$ on the far right-hand side can be thought of as accounting for the fact that a given number of activators bound, n, may correspond to multiple microscopic binding configurations (compare Figure 5.3A and B). Note that $W(0) = 1$, which means that Equation B.12 is valid even when $n = 0$. Finally, we impose the normalization condition that the sum of the state probabilities should be equal to 1, which leads to

$$p_n = \frac{p_0 W(n) K_d^{-n}}{p_0 \sum_{i=0}^{N} W(i) K_d^{-i}}. \tag{B.13}$$

Canceling out the factors of p_0 gives us our final expression for p_n, namely

$$p_n = \frac{W(n)K_d^{-n}}{\sum_{i=0}^{N} W(i)K_d^{-i}} = \frac{W(n)K_d^{-n}}{Z}, \tag{B.14}$$

where Z on the far righ-hand side indicates the sum of all state weights. Thus, given values of the rates k^b and k^u, which define K_d, we can calculate the probability of the system being in each binding state n. This probability is shown diagrammatically in the shading of the different states in Figure 5.3C.

B.3.2.2 Independent binding cannot produce bimodal transcriptional output

A basic requirement for bimodal transcriptionl behavior is that $p_0 > p_1$ and $p_N > p_{N-1}$, where N is the total number of binding sites. Couching this in terms of Equation B.14 leads to

$$\frac{p_0}{p_1} = \frac{1}{N}K_d > 1, \tag{B.15}$$

which simplifies to

$$K_d > N \tag{B.16}$$

for the low activity regime and

$$\frac{p_N}{p_{N-1}} = \frac{1}{N}\frac{1}{K_d} > 1, \tag{B.17}$$

leading to

$$K_d < \frac{1}{N} \tag{B.18}$$

for the high activity regime. Since K_d is set by the ratio $\frac{k^u}{k^b}$, which is constant for all states in the independent binding model, it is not possible for it to be simultaneously larger (Equation B.16) and smaller (Equation B.18) than the number of binding sites N. We thus conclude that independent binding is incompatible with bimodal transcription, regardless of the number of binding sites N.

B.3.2.3 Diffusion-limited binding

In the main text we state that we set $k^b = [A]k_0^b$ to 0.5 s^{-1} for the simulations shown in Figure 5.3C. This is convenient because it leads to a model where half the available sites are bound, on average. This choice is also physically reasonable. Bicoid concentrations in the embryo in the region of *hunchback* expression are on the order of 10 nM ($[A] \approx 10$ nM) [116], so that $k^b \approx 0.5$ s^{-1} thus implies a k_0^b of approximately 0.05 $nM^{-1}s^{-1}$. This is comfortably below the upper bound for k_0^b set by diffusion (0.1–10 $nM^{-1}s^{-1}$), which we take to be the speed limit for independent binding [65].

B.3.3 Cooperative binding

All calculations in this section pertain to the independent binding model presented in Figure 5.3E.

B.3.3.1 Deriving cooperativity weights

In Equation 5.7 we incorporated cooperative binding by adding multiplicative weights, ω, giving

$$k_{i,i+1} = k^b \omega^i. \tag{B.19}$$

This functional form follows from the assumption that each bound activator increases k^b by a constant factor $\omega \geq 1$. This leads the expression for $k_+(n)$

$$k_+^{coop}(n) = (N-n)\omega^n k_{n,n+1}, \tag{B.20}$$

which is a nonlinear function of n. Now, in analogy to the calculations presented in B.3.2.1, let's re-derive our expressions for p_n. To start, we have

$$p_{n+1} = \Big(\frac{N-n}{n+1}\Big)\Big(\frac{k^b}{k^u}\Big)\omega^n p_n. \tag{B.21}$$

Again expressing $\frac{k^u}{k^b}$ as a dissociation constant (K_d), we obtain

$$p_{n+1} = \Big(\frac{N-n}{n+1}\Big)\frac{\omega^n p_n}{K_d}. \tag{B.22}$$

We can also extend this logic to calculate the probability of any state, n, as a function of p_0, leading to

$$p_n = \frac{N!}{(N-n)!n!}\frac{\omega^{\frac{n(n-1)}{2}}p_0}{K_d^n} = W(n)\frac{\omega^{\frac{n(n-1)}{2}}p_0}{K_d^n}. \tag{B.23}$$

Finally, by requiring that all state probabilities sum to one, we obtain

$$p_n = \frac{W(n)\omega^{\frac{n(n-1)}{2}}K_d^{-n}}{Z}, \tag{B.24}$$

where Z again denotes the sum of all state weights as in Equation B.14. We have used these expressions to calculate the probability of each state shown using the shading in Figure 5.3E.

B.3.3.2 Cooperativity permits bimodal expression

Now, let's use Equation B.23 to examine how the addition of the cooperativity factor ω makes bimodal bursting possible. Recall that bimodal expression requires that $p_0 > p_1$ and $p_N > p_{N-1}$. For the low activity regime, cooperativity is not relevant and so the form of the requirement remains the same, namely

$$\frac{p_0}{p_1} = \frac{1}{N} K_d > 1. \tag{B.25}$$

However, things change in the high activity regime. Here, we have

$$\frac{p_N}{p_{N-1}} = \frac{1}{N} \frac{\omega^{N-1}}{K_d} > 1. \tag{B.26}$$

In stark contrast to the independent binding case, we see that the addition of ω makes it possible to realize both conditions simultaneously, opening the door to bimodal burst behaviors. Specifically, bimodality demands

$$K_d > N, \tag{B.27}$$

and

$$\omega > K_d^{\frac{2}{N-1}} \tag{B.28}$$

to be true. Not only do these requirements demonstrate that cooperativity is required to achieve bimodal bursting, they also indicate that K_d must be greater than the number of binding sites in the model, which corresponds to a system where, absent cooperative effects, activator binding is highly disfavored.

B.3.3.3 Cooperativity is necessary to simultaneously achieve kinetic trapping at both ends of the chain

The reasoning here closely mirrors the discussion from the previous section. To achieve kinetic trapping at both the high and low ends of the binding chain model simultaneously, we require (at a minimum) that $k_-(1) > k_+(1)$ and $k_-(N-1) < k_+(N-1)$. We can use Equation B.20 and Equation 5.6 to express these requirements in terms of system parameters. For the low activity regime, we have

$$\frac{k_-(1)}{k_+(1)} = \frac{K_d}{(N-1)\omega} > 1, \tag{B.29}$$

and for the higher regime we obtain

$$\frac{k_+(N-1)}{k_-(N-1)} = \frac{\omega^{N-1}}{(N-1)K^d} > 1. \tag{B.30}$$

We can simplify these requirements to obtain upper and lower bounds on ω, namely

$$\Big[K_d(N-1)\Big]^{\frac{1}{N-1}} \; < \; \omega \; < \; \frac{K_d}{N-1}. \tag{B.31}$$

We see that Equation B.31 implies restrictions on the relationship between K_d and N. Specifically, there must be a gap between the upper and lower bounds in Equation B.31 such that there exist viable ω values. This means that

$$\Big[K_d(N-1)\Big]^{\frac{1}{N-1}} < \frac{K_d}{N-1}, \tag{B.32}$$

must hold. Upon simplification, this gives

$$K_d > (N-1)^2. \tag{B.33}$$

Equation B.33 tells us that the dissociation constant must be larger than one (indeed, it must be larger than 25 for a 6 binding site system). This implies that the expression for the lower ω bound on the left-hand side of Equation B.31 is guaranteed to be greater than one as well, which indicates that cooperative interactions are necessary to realize kinetic traps on both ends of the chain.

B.3.4 First-passage time calculations

In this Review we used stochastic simulations (briefly outlined in B.3.1) to arrive at expectations for the form of first-passage time distributions for the cooperative binding and rate-limiting step models. All relevant scripts are available at GitHub. We also note that the functional forms for waiting time distributions can be using analytical methods such as Laplace Transforms. We do not provide the details for this approach here, but point the reader to [260, 9], as well as the sources cited therein, for more information.

Appendix C

SI for Competing constraints shape the limits of gene regulation out of thermodynamic equilibrium

C.1 Supplementary Figures

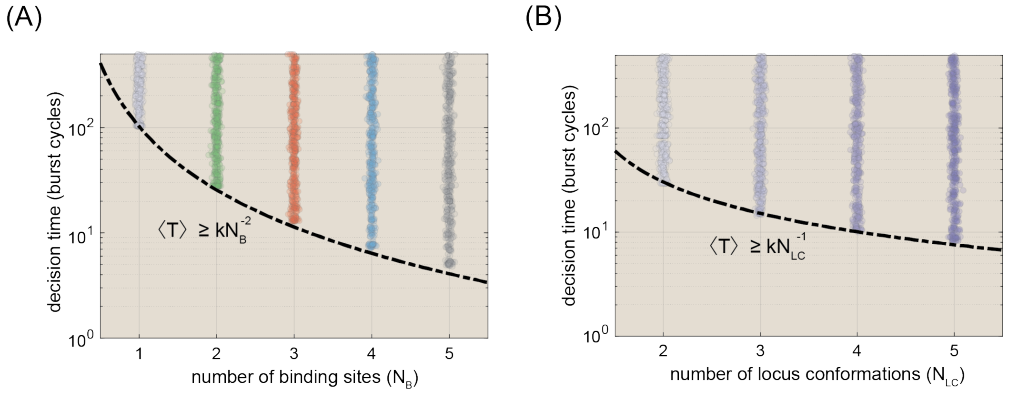

Figure C.1: **Decision times for different gene circuit architectures. (A)** Parameter sweep results for equilibrium gene circuits with different numbers of activator binding sites. Black dashed line indicates lower limit of the decision time and is a function of the form $\langle \mathrm{T} \rangle = k\mathrm{N}_\mathrm{B}^{-2}$, where k is a proportionality constant. **(B)** Plot of range of achievable decision times for non-equilibrium gene circuits with a single activator binding site ($\mathrm{N}_\mathrm{B} = 1$) as a function of the number of activation steps, N_LC. The dashed line indicates the lower decision time bound, and is a function of the form $\langle \mathrm{T} \rangle = k\mathrm{N}_\mathrm{LC}^{-1}$. All results shown assume an error probability of 32%.

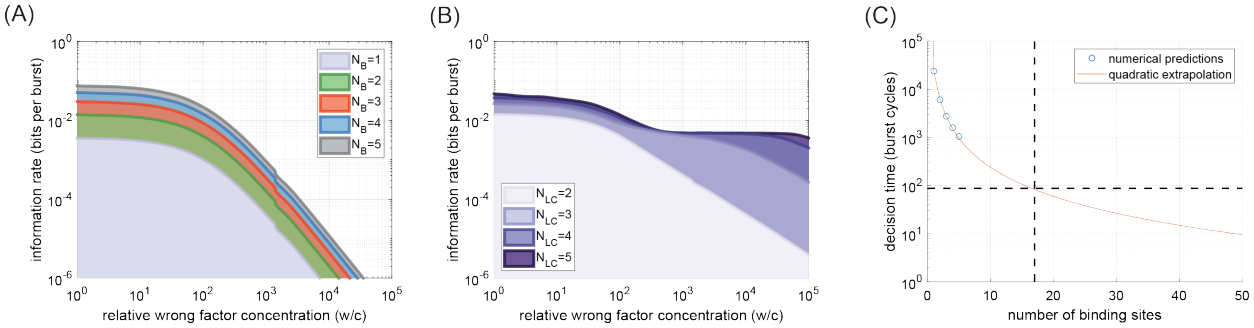

Figure C.2: **Supplemental data for IR vs. w/c analyses. (A)** Parameter sweep results showing the range of achievable information rates as a function of w/c for equilibrium gene circuits with 1-5 activator binding sites and one molecular activation step. **(B)** Sweep results for non-equilibrium gene circuits with 1-4 molecular steps and a single activator binding site. **(C)** Plot showing results of extrapolation analysis that uses numerical results for the minimum decision times for equilibrium gene circuits with 1-5 binding sites to estimate performance of larger models with may binding sites. Analysis indicates that at least 17 sites would be required to achieve plausible decision ties in the context of the mouse system.

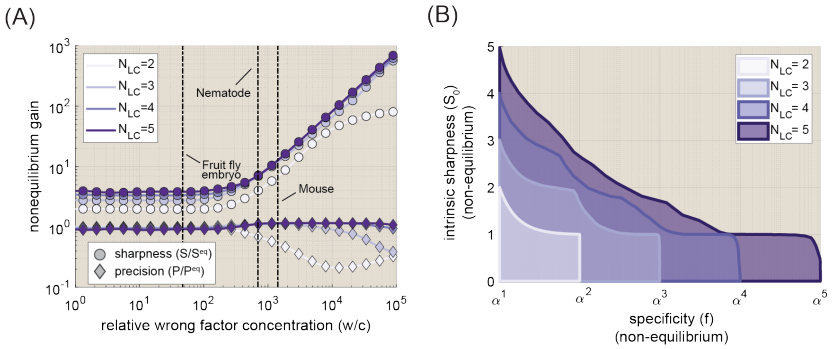

Figure C.3: **Supplemental results for main text Figure 6.5 (A)** Non-equilibrium sharpness and precision gains for IR-maximizing gene circuits with 2-5 locus conformations. **(B** Plot showing range of achievable specificity and intrinsic sharpness values for non-equilibrium gene circuits with with 2-5 locus conformations.

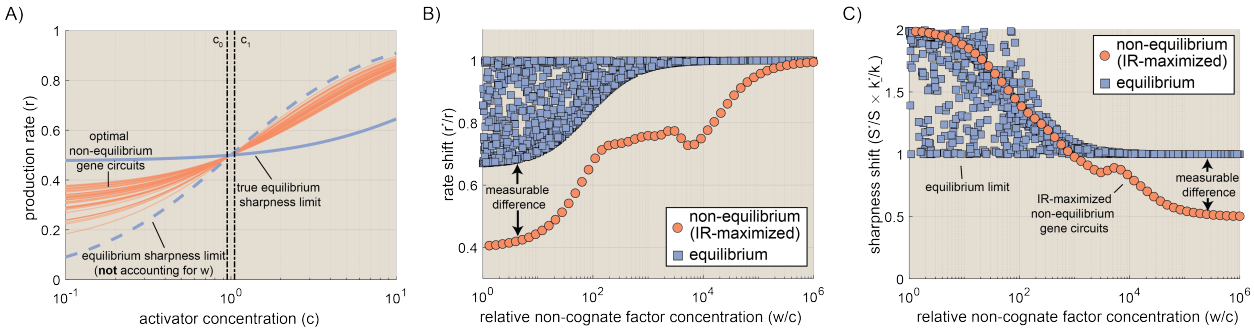

Figure C.4: **Additional experimental signature figures. (A)** Predicted induction curves for 50 near-optimal non-equilibrium gene circuits when $w = 10^3 c^*$, as well as the *true* induction curves for the sharpest achievable equilibrium curve (solid blue line) and the *naive* sharpness limit when not accounting for w (dashed line). Note that red curves fall above the true equilibrium limit but below the naive limit. **(B)** Predicted sharpness shift resulting from a binding site perturbation for equilibrium gene circuits (blue squares) and IR-maximizing non-equilibrium circuits (red circles). Non-equilibrium shift becomes markedly larger than equilibrium limit when $w/c > 10^3$. **(C)** Predicted rate shift upon perturbing the activator binding site. Non-equilibrium circuits are far more sensitive than equilibrium circuits when $w/c < 10^4$. The shift becomes negligible at higher values.

C.2 Appendices

C.2.1 Gaussian noise approximation

Throughout this work, we make the simplifying assumption that the intrinsic noise due in accumulated mRNA levels due to transcriptional bursting is approximately Gaussian. In this section, we use stochastic simulations to put this assumption to the quantitative test. The Markov chain central limit theorem states that the distribution of a quantity that is a function of a Markov chain (such as the transcription rate, r), will become approximately Gaussian as the number of iterations becomes large [106].

The question, then, is we can expect the accumulated transcriptional output to approach this limiting Gaussian distribution within timescales that are relevant to the decision times discussed in this work. To determine this, we used stochastic simulations [110] to track the distribution of the accumulated output 1,000 random realizations of the four state system shown in Figure 6.1B for 10,000 burst cycles. Each realization had a unique set of transition rates and, correspondingly, a unique average rate of transcription, $\overline{r} = \overline{a}r_0$. For each model realization, we ran 100 stochastic simulations and tracked the distribution of the apparent average transcription rate for each, as function of accumulation time. Figure C.5A shows the apparent mean rate across 100 simulations for a single illustrative gene circuit realization. Inset histograms indicate distribution of apparent transcription rates at different time points. As expected, we see that the apparent rates are initially highly dispersed; however, even after 50 burst cycles, we see that $p(\overline{r})$ has become a much narrower, roughly symmetrical distribution that appears approximately Gaussian.

To systematically assess the rate of convergence to normality, we utilized the simple One-sample Kolmogrov-Smirnov test ("kstest", [205]), which tests the null hypothesis that a vector of transcription outputs from realization i at time t, $\boldsymbol{r}_i(t)$, is drawn from a normal distribution. The test returns a p value corresponding to the probability of observing $\boldsymbol{r}_i(t)$ if the transcriptional output were truly Gaussian. In standard implementations $p \lesssim 0.05$ is taken to constitute strong evidence that the output is *not* Gaussian. Thus, to assess convergence to normality, we tracked this p value over time for each of the 1,000 gene circuit realizations.

Figure C.5B shows the average kstest p-values across 10 different sets of gene circuits, grouped by their average rate of transcription. In all cases, we see that noise profiles rapidly converge towards normality, such that all systems cross the (relatively conservative) threshold of $p = 0.1$ within 10 burst cycles (dashed line in Figure C.5B). Gene circuits near the tail ends of the induction curve ($\overline{r} \leq 0.1$ and $\overline{r} \geq 0.9$) take the longest to converge, which is likely because it takes longer for distributions near the boundaries to become symmetric about their mean; yet even these converge rapidly.

The fastest decisions discussed in the main text (Figure 6.4D and E), and most decision times considered are significantly longer. Thus, we conclude that the Gaussian noise approximation invoked throughout this work is justified.

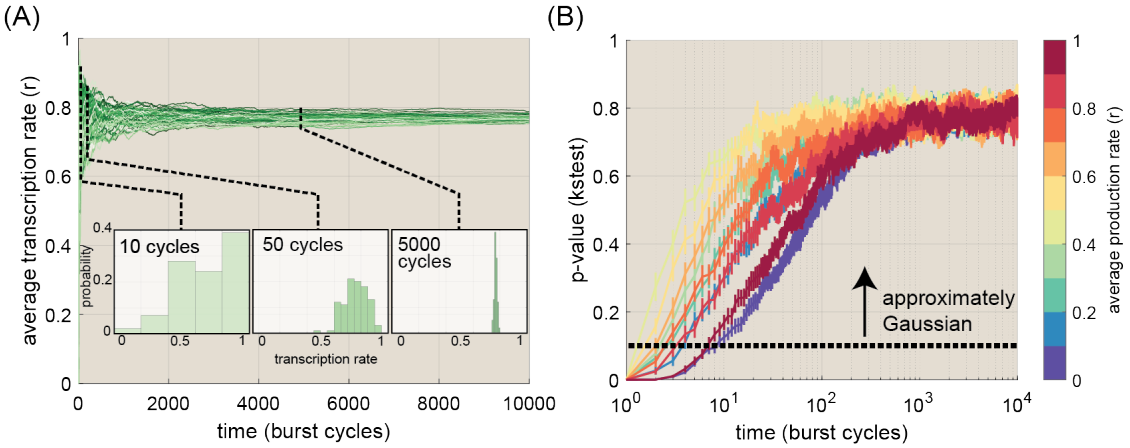

Figure C.5: **Testing validity of the Gaussian noise approximation. (A)** Illustrative plot showing average transcription rate as a function of the averaging time across 100 stochastic simulations of one realization of the four state model gene circuit. Inset histograms show distribution of apparent rates at three different time pints. We see that, as the accumulation time increases, the distributions get tighter and appear more Gaussian in shape. **(B)** Plot showing p-values of one-sample Kolmograv-Smirnov test. Different colors indicate average trends for systems with different average transcription rates. We see that systems near the low and high ends of the induction curve converge to Guassian form most slowly, but even these cross the $p = 0.1$ line within about 10 burst cycle. Error bars indicate bootstrap estimates of standard error calculated for each group.

C.2.2 Deriving the rate of information transmission for a gene locus

Motivated by [274], we define the rate of information transmission as the time derivative of the expected Kullback-Leibler (KL) Divergence between the two hypotheses ($C = c_0$ and $C = c_1$), given some accumulated mRNA level m, such that

$$\mathrm{IR} = \frac{d}{dt}\Big\langle \mathrm{D_{KL}}\big[p(c_1|m)||p(c_0|m)\big]\Big\rangle, \tag{C.1}$$

where $\mathrm{P}(c_0|m)$ and $\mathrm{P}(c_1|m)$ indicate (respectively) the conditional likelihood that the true value of C is c_0 and c_1 given the observed output m, and where the angled brackets indicate that we are dealing with the expected value of $\mathrm{D_{KL}}$ across many replicates. We refer readers information theory reference materials for a formal definition of $\mathrm{D_{KL}}$ (see, e.g., [61]); however, at an intuitive level it can be regarded as measuring how different two probability distributions are from one another. Thus, with Equation C.1, we define the rate of information production as the rate at which the two possibilities (c_1 or c_0?) become distinguishable from one another given the observed "evidence" (m).

We can write out the expected KL divergence from Equation C.1 more explicitly as the

weighted sum of log probability ratios:

$$\text{IR} = \frac{d}{dt}\Big(p_0\Big\langle \ln\frac{p(c_0|m)}{p(c_1|m)}\Big\rangle_0 + p_1\Big\langle \ln\frac{p(c_1|m)}{p(c_0|m)}\Big\rangle_1\Big), \tag{C.2}$$

where $\langle ...\rangle_i$ indicates the expectation taken assuming the true value of C to be c_i and where P_0 and P_1 indicate the priors on the true value of C, taken to be equal moving forward ($\text{P}_1 = P_0 = 1/2$). This formulation provides a more concrete for the sense in which IR is the information rate: as the conditional probabilities of the observed output given the true (numerators) and false (denominators) hypotheses about C diverge in favor of the true hypothesis, the log ratio terms will become large and positive. Thus a positive derivative corresponds to positive information production.

However, here we must recall that our focus here is to understand how the molecular architecture of gene loci impacts the transcriptional response and, ultimately, IR. Thus we wish to work in terms of $p(m|c)$—the conditional distribution of observed mRNA outputs given some input—rather than $p(c|m)$. To do this, we make use of Bayes' Theorem. We have:

$$\frac{p(c_0|m)}{p(c_1|m)}\frac{P(m)}{P(m)} = \frac{p(m|c_0)}{p(m|c_1)}\frac{p(c_0)}{p(c_1)}. \tag{C.3}$$

This expression becomes an equality if we assumed equal prior probabilities for our two hypotheses ($p(c_0) = p(c_1)$):

$$\frac{p(c_0|m)}{p(c_1|m)} = \frac{p(m|c_0)}{p(m|c_1)}. \tag{C.4}$$

Thus, we can use Equation C.4 to rewrite Equation C.2 as:

$$IR = \frac{d}{dt}\frac{1}{2}\Big(\Big\langle \ln\frac{p(m|c_0)}{p(m|c_1)}\Big\rangle_0 + \Big\langle \ln\frac{p(m|c_1)}{p(m|c_0)}\Big\rangle_1\Big). \tag{C.5}$$

We can think of the conditional probabilities, $p(m|c_i)$, in Equation C.5 as representing the full *stochastic* transcriptional response to some input activator concentration c_i. When these are approximately Gaussian (a condition discussed above in Appendix C.2.1), it becomes a straightforward exercise to solve for the expected log ratios in Equation C.5. We will solve for the case when $C = c_1$ in full. The c_0 case proceeds in precisely the same fashion. To start, we have

$$\Big\langle \ln\frac{p(m|c_1)}{p(m|c_0)}\Big\rangle_1 = \int_0^\infty p(m|c_1)\ln p(m|c_1)dg - \int_0^\infty p(m|c_1)\ln p(m|c_0)dg. \tag{C.6}$$

Recall that $g = rt$ is Gaussian with probability density function:

$$p(m|c_i) = \frac{e^{-\left(\frac{m-\overline{m}(c_i)}{2\sigma^2(c_i)}\right)^2}}{\sqrt{2\pi\sigma^2(c_i)}}. \tag{C.7}$$

Plugging Equation C.7 in for $\ln p(m|c_1)$ yields

$$\Big\langle \ln \frac{p(m|c_1)}{p(m|c_0)} \Big\rangle_1 = -\frac{1}{2}\ln\big(2\pi\sigma_m^2(c_1)\big) - \frac{1}{2} - \int_0^\infty p(m|c_1)\Big[-\frac{1}{2}\ln\big(2\pi\sigma_m^2(c_0)\big) - \frac{1}{2}\Big(\frac{\overline{m}(c_0)-g}{\sigma_m(c_0)}\Big)^2\Big]dm, \tag{C.8}$$

where we've recognized that the first integral will simply yield the standard expression for the entropy of a Gaussian random variable. Pulling constant factors out of the second integral leads to

$$\Big\langle \ln \frac{p(m|c_1)}{p(m|c_0)} \Big\rangle_1 = -\frac{1}{2}\ln\big(2\pi\sigma_m^2(c_1)\big) - \frac{1}{2} + \frac{1}{2}\ln\big(2\pi\sigma_m^2(c_0)\big) + \frac{1}{2\sigma_m^2(c_0)}\int_0^\infty p(m|c_1)\big[\overline{m}^2(c_0) - 2m\overline{m}(c_0) + m^2\big]dm. \tag{C.9}$$

Simplifying and recognizing that $\langle m^2\rangle_1 = \overline{m}^2(c_1) + \sigma_m^2(c_1)$ leads to:

$$\Big\langle \ln \frac{p(m|c_1)}{p(m|c_0)} \Big\rangle_1 = \frac{1}{2}\ln\frac{\sigma_m^2(c_1)}{\sigma_m^2(c_0)} - \frac{1}{2} + \frac{1}{2\sigma_m^2(c_0)}\Big[\overline{m}^2(c_0) - 2\overline{m}(c_0)\overline{m}(c_1) + \sigma_m^2(c_1)^2 + \overline{m}^2(c_1)^2\Big]. \tag{C.10}$$

Finally, we recall that $m = rt$ and $\sigma_m^2 = \sigma^2 t$, obtaining

$$\Big\langle \ln \frac{p(m|c_1)}{p(m|c_0)} \Big\rangle_1 = \frac{1}{2}\Big[\ln\frac{\sigma_r^2(c_0)}{\sigma_r^2(c_1)} + t\frac{\big(\overline{r}(c_1)-\overline{r}(c_0)\big)^2}{\sigma_r^2(c_0)} + \frac{\sigma_r^2(c_1)}{\sigma_r^2(c_0)} - 1\Big]. \tag{C.11}$$

Performing the same procedure for the case where $c = c_0$ yields:

$$\Big\langle \ln \frac{p(m|c_0)}{p(m|c_1)} \Big\rangle_0 = \frac{1}{2}\Big[\ln\frac{\sigma_r^2(c_1)}{\sigma_r^2(c_0)} + t\frac{\big(\overline{r}(c_0)-\overline{r}(c_1)\big)^2}{\sigma_r^2(c_1)} + \frac{\sigma_r^2(c_0)}{\sigma_r^2(c_1)} - 1\Big]. \tag{C.12}$$

Plugging Equation C.11 and Equation C.12 into Equation C.5 and taking the derivative with respect to time yields

$$\mathrm{IR} = \frac{1}{4}\frac{\big(\overline{r}(c_1)-\overline{r}(c_0)\big)^2\big(\sigma(c_1)^2 + \sigma(c_0)^2\big)}{\sigma(c_0)^2\sigma(c_1)^2}. \tag{C.13}$$

Next, if we assume that the difference between c_0 and c_1 is small (as stipulated in the main text), then $\sigma(c_0) \approx \sigma(c_1) \approx \sigma^2(c^*)$ and $\overline{r}(c_1) - \overline{r}(c_0) \approx \delta c dr/dc$, leading to

$$\mathrm{IR} = \frac{1}{2}\Big(\delta c\frac{dr}{dc}\Big)^2\frac{1}{\sigma(c^*)^2}. \tag{C.14}$$

Finally, we invoke the definitions of sharpness and precision given in Figure 6.1A, which leads to Equation 6.2 from the main text:

$$\mathrm{IR} = \frac{1}{2}\Big(\frac{\delta c}{c^*}\Big)^2 s^2 p^2. \tag{C.15}$$

C.2.3 Poisson noise from mRNA synthesis is negligible relative to noise from bursting

In this section, we provide support for the claim, made in the Section 6.3.2 of the main text, that Poisson noise due to mRNA synthesis is negligible relative to noise from transcriptional bursting. We take as our starting point Equation C.55 from Appendix C.2.10,

$$\mathrm{P} = \frac{\overline{a}(1-\overline{a})}{\sigma}, \tag{C.16}$$

which relates the normalized precision, P, to the bursting noise, σ, and the fraction of time a gene circuit spends in transcriptionally active states, $\overline{a}$. From Figure 6.3A, we see that $P \leq 1$ for the four state gene circuit shown in Figure 6.1B when the system is out of equilibrium, which, from Equation C.16, implies that

$$\sigma^2 \geq \overline{a}^2(1-\overline{a})^2 \tag{C.17}$$

for the 4 state system.

Thus, Equation C.18 gives a lower bound for the intrinsic variance in gene expression that arises due to transcriptional burst fluctuations at the gene locus. To see how to relate this to noise from mNRA synthesis, we need to take two more steps. First, we must recall that we are working in units of the burst cycle time, τ_b. Second, we must further recall that we set the actual rate of mRNA synthesis, r_0, equal to 1 throughout the main text. We must do away with these simplifications in order to relate σ^2 to synthesis noise. Accounting for these simplifications, the full expression for the noise floor, in "real" time units and accounting for the true rate of mRNA synthesis is

$$\sigma^2_{\mathrm{burst}} \geq \tau_b r_0^2 \overline{a}^2(1-\overline{a})^2. \tag{C.18}$$

Now, if we assume mRNA synthesis to be a Poisson process (following, e.g., [272]), we have that this component of the variance is simply equal to

$$\sigma^2_{\mathrm{mRNA}} = r_0 \overline{a}. \tag{C.19}$$

The key thing to notice about Equation C.19 is that mRNA synthesis noise is *independent* of the bursting timescale τ_b. Thus, as τ_b increases, $\sigma^2_{\mathrm{burst}}$ will increase in magnitude relative to σ^2_{mRNA}. Figure C.6A and B illustrate this fact, showing predicted bursting and mRNA synthesis variance components, respectively, as a function of the bursting time scale τ_b and the activity level ($\overline{a}$). All calculations assume an mRNA synthesis rate of 20 mRNA per minute, a rate based off of estimates from the fruit fly [168] and of consistent with measurements from other systems [290]. From Figure C.6A, we see that $\sigma^2_{\mathrm{burst}}$ peaks at $\overline{a} = 0.5$ and increases dramatically as we move rightward along the x axis. We emphasize that this represents a *lower* bound for maximally precise non-equilibrium gene circuits; most systems (including IR-optimized systems) will lie above this bound. In contrast Figure C.6B shows that noise from mRNA synthesis scales linearly with $\overline{a}$, and is constant in τ_b.

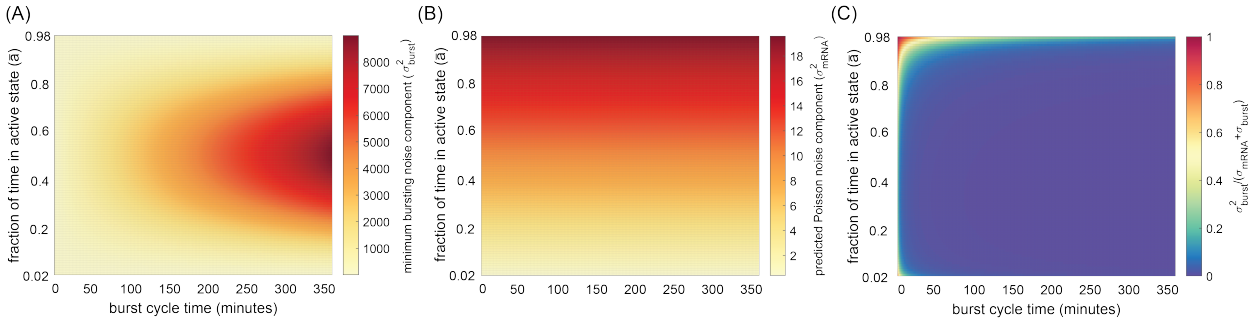

Figure C.6: **Testing contribution from mRNA synthesis noise. (A)** Heatmap showing lower bound of bursting component of variance for the non-equilibrium four state model shown in Figure 6.1B as a function of the fraction of time spent in the active state ($\overline{a}$) and the burst cycle time (τ_b). **(B)** Heatmap showing predicted variance component arising from mRNA synthesis. **(C)** Predicted relative contribution of mRNA synthesis noise to total intrinsic noise levels in gene expression. Note that contribution is only significant for rapidly bursting systems near the saturation point. (All calculations assume an mRNA synthesis rate of 20 per minute, in keeping with estimates from [166])

The total gene expression noise level is given by

$$\sigma^2_{\text{tot}} = \sigma^2_{\text{mRNA}} + \sigma^2_{\text{burst}}. \tag{C.20}$$

We can use this expression to calculate a lower bound on the relative contribution of mRNA synthesis noise to the overall intrinsic variance in gene expression. Figure C.6C shows the results of this calculation. We see that, with the exception of rapidly bursting systems near the saturation ($\overline{a} \approx 1$). Thus, we conclude that noise from transcriptional bursting constitutes the dominant source of gene expression noise for the vast majority of the parameter regimes relevant for the investigations in this paper, and that our decision to neglect Poisson noise from mRNA synthesis is reasonable.

C.2.4 Analytic expressions for key gene circuit characteristics

This section lays out analytic expression for key quantities that play a central role in the investigations undertaken over the course of the main text. We do not repeat derivations for expressions that are treated separately elsewhere in these Appendices, and avoid re-deriving expressions from scratch, unless they are novel to this work.

C.2.4.1 The transition rate matrix and activity vector

Consider a gene circuit g that has K different microscopic states. We assume that microscopic transitions between the molecular states that make up g are Markovian, such that our system can be modeled as a continuous time Markov chain (CTMC). It follows that the

behavior of g is fully determined by two quantities: the transition rate matrix, $\boldsymbol{Q}$ and the state activity vector, $\boldsymbol{a}$.

$\boldsymbol{Q}$ is a $K \times K$ matrix with off-diagonal elements that encode the rates with which the system switches between microscopic rates. For instance, q_{mn}—the element in the mth row and nth column of $\boldsymbol{Q}$—gives the transition rate going from state n to state m. The diagonal elements of $\boldsymbol{Q}$ are negative, and are scaled such that each column of $\boldsymbol{Q}$ sums to 0. The activity vector $\boldsymbol{a}$ is a binary vector of length K that contains a "1" for each state that is transcriptionally active, and a "0" for inactive states. We assume that both $\boldsymbol{Q}$ and $\boldsymbol{a}$ are fixed in time.

C.2.4.2 State probabilities, transcription rate, and transcriptional noise

A first step to calculating virtually all other quantities of interest is to obtain the steady-state vector, $\boldsymbol{\pi}$, which is a vector of length K that gives the steady state probability of finding the gene circuit of any one of the K microscopic states. Intuitively, we can obtain $\boldsymbol{\pi}$ by finding the right eigenvalue of $\boldsymbol{Q}$ with an eigenvalue of 0,

$$\boldsymbol{Q}\boldsymbol{v} = 0, \tag{C.21}$$

and imposing the additional constraint that the elements of $\boldsymbol{\pi}$ sum to 1, such that

$$\boldsymbol{\pi} = \frac{\boldsymbol{v}}{\sum_{i=1}^{K} v_i}. \tag{C.22}$$

This the steady state probability vector in hand, we can calculate the average transcription rate by taking the dot product of $\boldsymbol{a}$ and $\boldsymbol{\pi}$:

$$\overline{r} = r_0 \underbrace{\sum_{i=0}^{K} a_i \pi_i}_{\substack{\text{fraction of} \\ \text{time active } (\overline{a})}}, \tag{C.23}$$

where we define the quantity indicated by the underbrace as the average fraction of time, $\overline{a}$, that the system spends in the active state. Throughout the course of this work, we assume that r_0 is held fixed, such that the transcriptional activator may only impact transcription by altering microscopic transition rates in $\boldsymbol{Q}$ to alter $\boldsymbol{\pi}$. Further, since we take Poisson noise from mRNA synthesis to be negligible (see Appendix C.2.3), the absolute magnitude of r_0 is unimportant, and we set it to 1 for simplicity.

Next, we turn to obtaining an expression for the variance (noise) in gene expression. From Whitt 1992 [311], we have that

$$\sigma^2 = 2 \sum_{i=1}^{K} \sum_{j=1}^{K} a_i \pi_i z_{ij} a_j, \tag{C.24}$$

where z_{ij} is the element from ith row and jth column of what is known as the fundamental matrix, $\boldsymbol{Z}$ of our transition rate matrix, $\boldsymbol{Q}$. $\boldsymbol{Z}$ is a $K \times K$ matrix that plays an integral

role in the calculation of many key behaviors of a Markov chain. Once again drawing from Whitt, we can calculate $\boldsymbol{Z}$ using the formula

$$\boldsymbol{Z} = (\boldsymbol{\Pi} - \boldsymbol{Q})^{-1} - \boldsymbol{\Pi}, \tag{C.25}$$

where $\boldsymbol{\Pi}$ is a $K \times K$ matrix with each row equal to π.

C.2.4.3 Using the fundamental matrix to calculate first passage times

First passage times provide a useful conceptual tool for connecting microscopic fluctuations, which often are unobservable, with emergent dynamical behaviors, such as transcriptional bursting. The fundamental matrix provides an invaluable tool for doing this in the context of arbitrarily complex transcriptional systems. Once again, we start with an expression from Whitt 1992 [311] that relates off-diagonal elements of $\boldsymbol{Z}$ to first passage times between microscopic states:

$$z_{ji} = \pi_i[ET_{ei} - ET_{ij}], i \neq j. \tag{C.26}$$

Here, ET_{ij} is the mean expected first passage time to from state j to state i and ET_{ei} the first passage time to state i at equilibrium, defined as

$$ET_{ei} = \pi_i \sum_{j=1}^{K} \pi_j ET_{ij}. \tag{C.27}$$

Now, from [311] we also have that the diagonal elements of $\boldsymbol{Z}$ can be expressed as

$$z_{ii} = \pi_i ET_{ei}. \tag{C.28}$$

We can now combine Equations C.26 and C.28 to solve for the first passage time from state i to state j:

$$ET_{ij} = \frac{z_{ii} - z_{ji}}{\pi_i}. \tag{C.29}$$

C.2.4.4 Calculating the burst cycle time

First passage times are intimately related to a quantity of central importance throughout the text: the burst cycle time, τ_b, defined as the average time required for a system to complete one ON→OFF→ON cycle (Figure 6.1C). This is trivial in the case of a simple two state system with a single OFF and ON state and rates k_{on} and k_{off}. In this case, the burst cycle time is simply

$$\tau_b = \frac{k_{\text{on}} + k_{\text{off}}}{k_{\text{on}} k_{\text{off}}}. \tag{C.30}$$

The calculation becomes less trivial for systems with larger numbers of states, however. Fortunately, the concepts outlined above provide us with the tools necessary to derive a generic expression for τ_b that applies to systems of arbitrary complexity.

The essence of the procedure lies in calculating effective off and on rates (k^*_{off} and k^*_{on}) from $\boldsymbol{Q}$ using first passage times. We go through this procedure in detail for k^*_{on} and note that the same approach applies for k^*_{off}. The activity vector $\boldsymbol{a}$ partitions our system into M OFF states and N ON states. To calculate k^*_{on}, the first step is to estimate the expected amount of time it will take for the system to reach an ON state (any ON state) from each OFF state. We can do this defining a new transition rate matrix, $\boldsymbol{Q}^{\text{OFF}}$, that has dimensions $M+1 \times M+1$. The off-diagonal elements of the first M rows an M columns of $\boldsymbol{Q}^{\text{OFF}}$ are simply equal to the microscopic rates from $\boldsymbol{Q}$ that connect the M OFF states in our gene circuit.

The final row and column, however, are different and contain total fluxes into and out-of all ON states from each OFF state. An element in the final row of $\boldsymbol{Q}^{\text{OFF}}$ is given by

$$q^{\text{OFF}}_{m+1,i} = \sum_{j=1}^{K} a_j q_{ji}, \tag{C.31}$$

where a_j is the jth element of the activity vector, q_{ij} is a microscopic rate from the original transition rate matrix, and we assume the state i is in the set of OFF states. Thus, we see that each element of the last row of $\boldsymbol{Q}^{\text{OFF}}$ gives the total flux from each OFF state to *all* ON states in the gene circuit. The elements of the final column have a complementary definition:

$$q^{\text{OFF}}_{i,m+1} = \sum_{j=1}^{K} a_j q_{ij}. \tag{C.32}$$

With our condensed transition rate matrix thus defined, we can use Equations C.21 and C.22 to calculate $\boldsymbol{\pi}^{\text{OFF}}$ and Equation C.25 to calculate $\boldsymbol{Z}^{\text{OFF}}$. Then, we can use Equation C.29 to obtain a vector $\boldsymbol{et}^{ON}$ of length M, where each element i is defined as the expected first passage time from OFF state i back into *any* of the ON states. Specifically, we have that each element, i, is given by

$$et_i = \frac{z^{\text{OFF}}_{m+1,m+1} - z^{\text{OFF}}_{i,m+1}}{\pi_{m+1}}. \tag{C.33}$$

Thus, we have obtained a vector,$\boldsymbol{et}^{ON}$, of expected mean first passage times out of each OFF state into the set of N active transcriptional states. But how do we weight the different passage times in this vector to arrive at an overall average expectation for the amount of time required for the system to turn back ON following a transition into an OFF state? It's tempting here to use the stead-state probabilities of each OFF state given by $\boldsymbol{\pi}$, but this is actually not correct.

Instead, the key is to recognize that each OFF state should be weighted by the rate at which ON states switch into it. In other words, we weight OFF states by the probability that they are the initial state the system reaches upon turning OFF; the gateway into the OFF states. Mathematically, we encode these weights using the flux vector $\boldsymbol{f}^{\text{OFF}}$, which has M elements, each defined as

$$f^{\text{OFF}}_i = \sum_{j=1}^{K} a_j q_{ij} \pi_j. \tag{C.34}$$

Finally, we combine this expression with Equation C.33 to obtain an expression for the average reactivation time as a flux-weighted average of the first passage times out of each OFF state:

$$ET_{\text{OFF}\rightarrow\text{ON}} = \frac{1}{k^*_{\text{on}}} = \frac{\sum_{i=1}^{M} f_i et_i}{\sum_{i=1}^{M} f_i}. \tag{C.35}$$

As noted above, the calculations for k^*_{off} follow precisely the same logic, with the roles of the OFF and ON states switched. After this is done, the total burst cycle time, τ_b, is simply

$$\tau_b = ET_{\text{OFF}\rightarrow\text{ON}} + ET_{\text{ON}\rightarrow\text{OFF}}. \tag{C.36}$$

Equation C.35 is useful because it allows us to relate the (potentially quite complex) microscopic dynamics of a transcriptional system to emergent bursting timescales observed in live imaging experiments [166]. To our knowledge, we are the first to take this flux-weighted first passage time approach to modeling burst dynamics. We hope that the expressions provided here will prove useful to others seeking to pursue similar projects in the future.

A final useful feature of Equation C.35 is that the absolute size of τ_b scales inversely with the microscopic rates in $\boldsymbol{Q}$, such that we can decrease τ_b by a factor λ by simply multiplying $\boldsymbol{Q}$ by λ. We use this trick to renormalize all time-dependent metrics calculated over the course of our parameter sweeps to have units of burst cycle time. This is done by calculating τ_b for each new model realization we generate, and then multiplying its transition rate matrix by this quantity to generate a normalized rate matrix, namely

$$\boldsymbol{Q}^* = \tau_b \boldsymbol{Q}. \tag{C.37}$$

The adjusted matrix, $\boldsymbol{Q}^*$, is then used to calculate all relevant gene circuit characteristics.

C.2.4.5 A generic expression for the rate of energy dissipation

Equation 6.1 gives an expression for the rate of energy dissipation (also termed entropy production), Φ, in the context of the four state model shown in Figure 6.1B. This is a special case of a more general formula for Φ that applies to arbitrary molecular architectures. From [169, 175], we have

$$\Phi = \sum_{i=i}^{K} \sum_{j\neq i}^{K} \pi_i q_{ji} \ln\left(\frac{q_{ji}}{q_{ij}}\right). \tag{C.38}$$

We use Equation C.38 to calculate all energy dissipation rates given throughout the main text.

C.2.5 The Sequential Probability Ratio Test

Over half a century ago, Wald conceived of the Sequential Probability Ratio Test (SPRT) as a solution to the problem of making accurate decisions between two hypotheses, H_1 and H_0 in "real time" as relevant data is accruing [304]. Shortly thereafter, it was established that SPRT represents the optimal approach to sequential decision problems involving binary

decisions [305], meaning that it requires the fewest observations to achieve a desired level of accuracy. In this framework, a downstream receiver (in our case, downstream genes or other cellular processes) tracks the accrual of some signal (mRNA, and eventually protein) over time and compares how likely this accrued signal is under the two hypotheses to be distinguished (e.g. high or low activator concentration). In this work, we use the optimal nature of SPRT to set lower bounds on decision times that could be achieved given the transcriptional output of model gene loci. The essence of the test lies in tracking the relative likelihoods of our two hypotheses ($C = c_1$ and $C = c_0$) over time as more and more transcriptional output, m, accrues:

$$\frac{P_0}{P_1} = \frac{P(c_0|m)}{P(c_1|m)} \tag{C.39}$$

Figure Figure C.7A shows a stochastic simulation of how this ratio evolves over time for the output of a single model gene circuit. Although the true concentration in this case is c_0, we see that the two hypotheses are essentially indistinguishable early on. This is because the range of possible outputs given high and low activator concentrations overlap significantly early on (leftmost panel of Figure C.7B). However, as more and more time passes, the expected outputs (m) given the two possible inputs (c_1 and c_0) start to separate. We see that the ratio in their likelihoods diverges more and more in favor of c_0 ($P_0/P_1 >> 1$), corresponding to a higher and higher degree of certainty that c_0 is the correct choice.

This divergence, however, is non-monotonic and noisy, which reflects the stochastic nature of protein production at a single gene locus. It has been shown that the noisy divergence of the log of the probability ratio (which we will call L) can modeled as a 1-D diffusive process with average drift IR [274] given by

$$\text{IR} = \frac{d}{dt}\langle L \rangle. \tag{C.40}$$

In this framework, a "decision" is made when L crosses a so-called "decision boundary" (horizontal dashed lines in Figure C.7A). Siggia et al showed that the Gaussian diffusion approximation could be used to obtain an analytic expression for the expected time needed to make a decision. From Equation 15 in the supplement of [274], we have that:

$$\langle T \rangle = \frac{K}{2V \sinh \frac{VK}{D}} \left[e^{\frac{VK}{D}} + e^{-\frac{VK}{D}} - 2 \right], \tag{C.41}$$

where V is the same as IR from above (and in the main text) D encodes the diffusivity of decision process (essentially, how large the fluctuations are about its mean drift trajectory), and K is related to the log of the error tolerance parameter ϵ, such that

$$K = \log\left(\frac{1-\epsilon}{\epsilon}\right). \tag{C.42}$$

We note that Equation C.41 assumes equal priors regarding the likelihood of c_1 and c_0, and also assumes equal error tolerances for choosing incorrectly in either case [69].

If we take the accumulated transcriptional output of our gene circuit, $m = rt$, to be approximately Gaussian (see Appendix C.2.1), then it can be shown that D has the form:

$$D = \frac{(m_0 - m_1)^2(\sigma_0^6 + \sigma_1^6)}{4\sigma_0^4\sigma_1^4}, \tag{C.43}$$

where m_i and σ_i give the mean and variance in the accumulated transcriptional output, given that $C = c_i$. From Equation C.13 in Appendix C.2.2, we also have that

$$V = \mathrm{IR} = \frac{(m_0 - m_1)^2(\sigma_0^2 + \sigma_1^2)}{4\sigma_0^2\sigma_1^2}. \tag{C.44}$$

In a different context (exponential distributions, rather than Gaussian), Desponds and colleagues [69] demonstrated that $D \approx V$ when the difference between hypothesis—$\delta c/c^*$ in our case—is small. From Equations C.43 and C.44, we see that this also holds for the Gaussian case: when c_1 and c_0 are sufficiently close, σ_1 and σ_0 will be approximately equal, such that:

$$D \approx V \approx \frac{(m_0 - m_1)^2}{2\sigma^2}. \tag{C.45}$$

As demonstrated by [69], when $D \approx V$, Equation C.41 simplifies dramatically, yielding

$$\langle T \rangle = \log\left(\frac{1-\epsilon}{\epsilon}\right)\frac{1-2\epsilon}{\mathrm{IR}}, \tag{C.46}$$

which is Equation 6.3 from the main text. For correctness, we use the full expression (Equation C.41) to calculate all decision time quantities shown in the main text. However, since Equation C.46 holds quite well for the 10% concentration difference considered here, we give the simpler expression in the main text to aid the reader's intuition.

C.2.6 Implementation of parameter sweep algorithm

In this section, we describe the parameter sweep algorithm employed throughout this work to enumerate the performance bounds of gene circuit models. We note that this approach is based off of an algorithm previously employed by Eck & Liu et al [77] to explore the behavior of non-equilibrium models of transcription. Figure C.8A illustrates the key steps in this numerical procedure. First, an initial set of gene circuit realizations (typically comprised of 1,000 variants) is generated by sampling random values for each transition rate in the system. We then calculate the performance metrics of interest (S and P for the example in Figure C.8A) for each gene circuit realization. This defines an initial set of points (Figure C.8A, Panel i) that collectively span some region in 2D parameter space with area a_1.

Next (Panel ii), we subdivide parameter space into N different bins along the X and Y axes, with N dictated by the total number of points ($10 \leq N \leq 50$). We then calculate the maximum and minimum point in each X and Y slice (Panel iii). We then randomly select candidate gene circuit models from these boundary points and apply small perturbations to

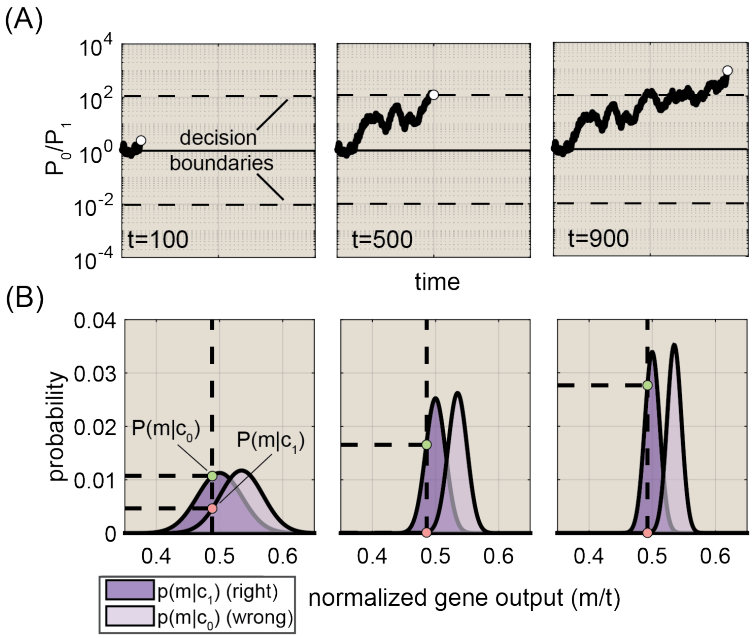

Figure C.7: **The Sequential Probability Ratio Test. (A)** Panels show stochastic trajectory of the relative probabilities of c_0 and c_1 over time, given the observed output of some gene circuit. **(B)** Panels illustrating expected distributions of outputs for each concentration at different time points. Note how the distributions narrow and separate as time progresses.

each transition rate to generate a new set of random variants (iv). In general, these variants will lie close to original model in 2D parameter space and, thus, close to the current outer boundary of parameter space. The key to the algorithm's success is that some of these variants ill lie *beyond* the current boundary (blue points in Figure C.8A, Panel iv). This has the effect of extending the boundary outward, leading to an increase in the surface area spanned by our sample points (panel iv). As a result, cycling through steps ii-iv amounts to a stochastic edge-finding algorithm that will iteratively expand the boundary spanned by sample points outward in 2D parameter space until some analytic boundary is reached.

The panels in Figure C.8B show snapshots of the sweep algorithm's progress exploring sharpness vs. precision parameter space for non-equilibrium realizations of the four state gene circuit (Figure 6.1B). Figure C.8C shows the total area spanned by the sample points for this run as a function of sweep iteration. By eye it appears that most of salient parameter space has been explored by step 10 of the algorithm, but we are quite strict with our convergence criteria. We will only terminating a sweep at step t if $(a_t - a_{t-2})/a_{t-2} \leq 0.001$ and $(a_{t-1} - a_{t-3})/a_{t-3} \leq 0.001$. In this case, this convergence criterion is met following step 25, leading to the final set of sample points shown in Figure C.8D. In general, we run all sweeps until the above criterion is met or some pre-specified maximum number of iterations (usually 50) is reached.

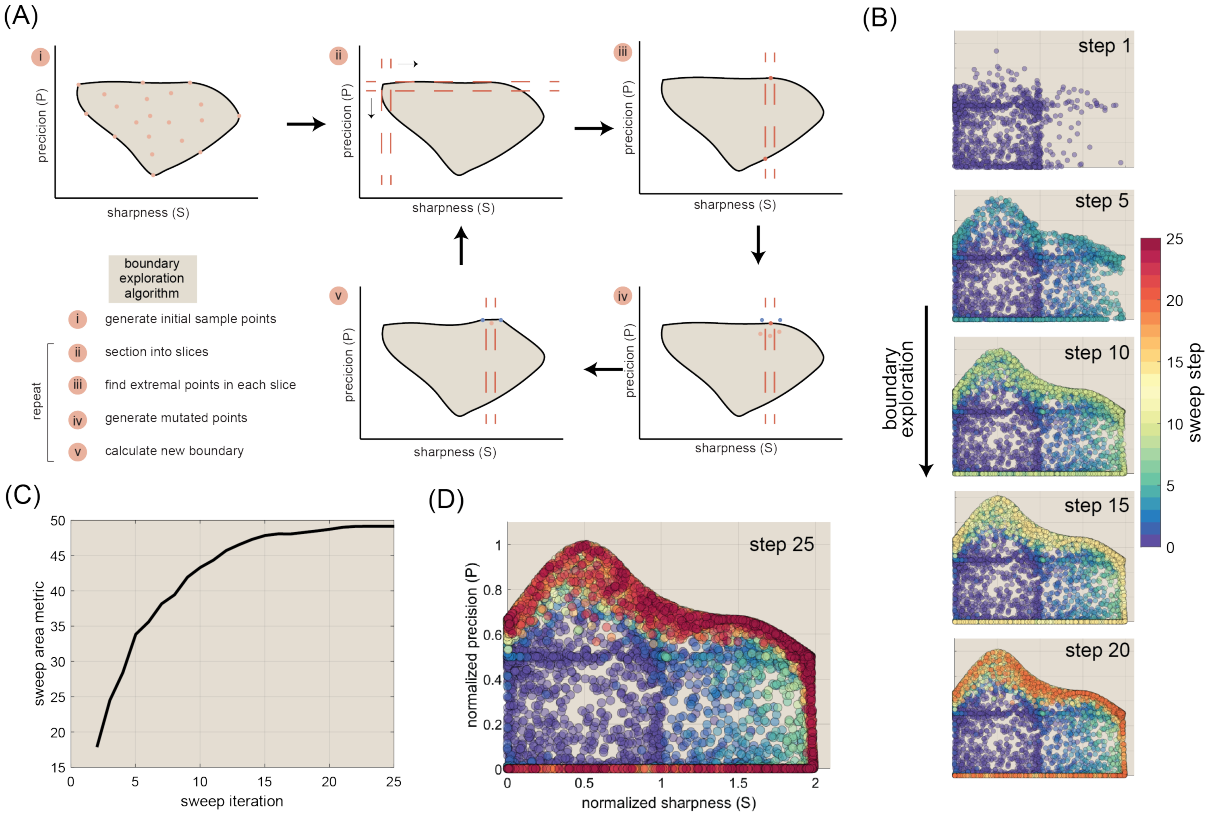

Figure C.8: **A simple edge-finding algorithm for numerical parameter sweeps. (A)** Schematic illustrating key steps in our parameter sweep approach (see text for details). This panel has been adapted with permission from [77]. **(B)** Sequence of snapshots showing progress of sweep algorithm across a single run for the case of normalized sharpness (S) versus normalized precision (P). Circle color indicates the sweep step on which it was generated. **(C)** Plot showing 2D surface area spanned by sample points over time. **(E)** Plot showing final set of sample points obtained by the sweep algorithm.

C.2.6.1 Numeric vs. symbolic metric calculations

The algorithm outline in Figure C.8A depends requires the ability to rapidly calculate performance metric quantities (e.g. S and IR) given a set of transition rate magnitudes. Wherever possible, we use symbolic expressions to perform these calculations; however, this is only feasible for the simple four and six state systems depicted in Figure 6.1B and Figure 6.4B. For more complex models, it is infeasible to perform the symbolic operations required to obtain closed-form symbolic expressions. As a results, we use numerical calculations to arrive at performance metrics for all higher-order models.

C.2.7 Testing the convergence characteristics of the parameter sweep algorithm

Here, we discuss results from a series of tests designed to assess the convergence of our sweep algorithm for key scenarios examined in the main text. This task is the most straight-forward when the algorithm is employed for "two-boundary" sweeps, such as S vs. P (Figure 6.3A) and S_0 vs f (Figure 6.5B), where both parameters examined adhere to finite performance bounds and, thus, where the 2D region of accessible parameter space has a finite area. In this case, our general approach will be to assess whether independent runs of the algorithm (i) converge prior to the 50 run limit and (ii) reach a consistent final estimate for the area 2D space that is attainable for different model architectures. The task becomes more complicated for "one-boundary" sweeps, such as IR vs. Φ (Figure 6.2A, C, and D) and IR vs. w/c, where only a single parameter (IR in each case) has a finite upper bound and the other (Φ and w/c) is limited only be bounds imposed externally as a part of sweep specification. We will begin by assessing convergence for the simpler two-sided case, and will turn thereafter to examining one-sided cases.

C.2.7.1 Sharpness vs. Precision sweeps

Figure 6.3A and Figure C.15A and B show results for parameter sweeps examining trade-offs between normalized sharpness (S) and normalized precision (P) for systems with 1-5 binding sites and 2-5 locus conformations. We note that Figure 6.3C and Figure C.15C also derive from these parameter sweep results. Across the board, we find that nearly all independent runs of the sweep algorithm converge according to the definition laid out above (Figure C.9A and B). Moreover, for simpler architectures, we find that all independent sweep runs converge to essentially the same total area. For instance, Figure C.9B shows normalized area as a function of sweep step for 500 non-equilibrium realizations of the baseline four state model, indicating that all runs terminate near the global maximum found across all runs (dashed line). We take this as strong evidence that the algorithm is consistently exploring the full extend of 2D parameter space.

As might be expected, the task of exhaustively exploring parameter space becomes more difficult as models become more complex. Note the larger spread in outcomes for the non-equilibrium five binding site ($N_B = 5$) and 5 locus conformation ($N_{LC} = 5$) models in FigureFigure C.9C and D, respectively. Nonetheless, we find that a significant number of sweeps converge to a consistent maximum area, even for the most complex models considered. Figure C.9E and F give the total fraction of sweeps having a final area within 95% of the global maximum as a function of binding site number and locus conformation number, respectively. First, we see that *100%* sweeps for equilibrium models uniformly meet this standard for all model architectures considered (squares in Figure C.9E and F). Second, our analysis indicates that, even for the extrema ($N_B = 5$ and $N_{LC} = 5$), 13% and 36% of total runs , respectively (67 and 179 sweeps), still achieve final areas comparable to the global maximum, suggesting that the algorithm still does an adequate job of exploring parameter space in these cases.

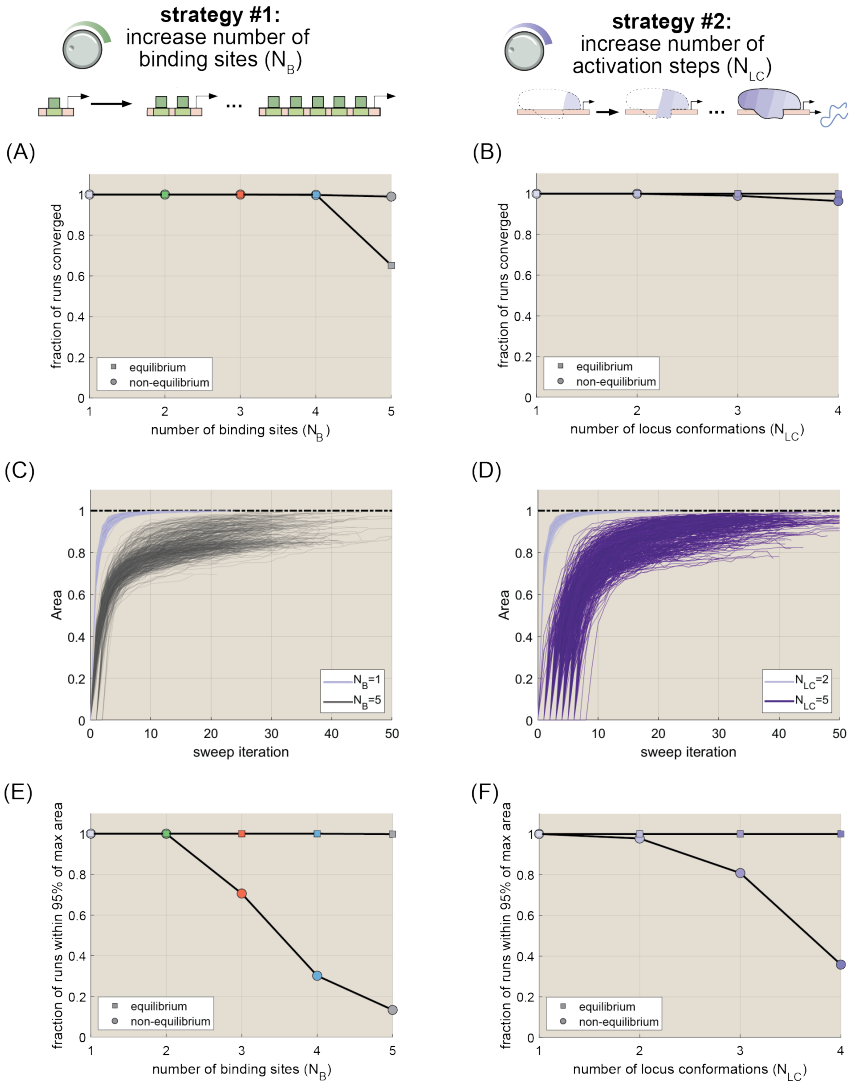

Figure C.9: **Convergence results for S vs. P parameter sweeps. (A-B)** Plots showing fraction of parameter sweep ones that met convergence criteria for multi-binding site and multi-locus conformation models, respectively. Squares indicate results for equilibrium models and circles indicate non-equilibrium models. **(C-D)** Plots of area vs. sweep step for different model architectures. Note that the area corresponding to the first (iteration=1) not recorded by the algorithm, and so has been estimated in each case via linear interpolation. Staggered starts apparent for $N_B = 5$ and $N_{LC} = 5$ models indicate cases where model initialization were aborted one or more times due to an insufficient number of gene circuits meeting quality control criteria. **(E-F)** Fraction of parameter sweeps having a final area within 95% of the global maximum for multi-binding site and multi-locus conformation models, respectively. (All results were calculated using 500 independent runs of the sweep algorithm for each model architecture.)

C.2.7.2 Information vs. energy sweeps

Next, we turn to the one-sided sweeps. First, let's consider the IR vs. Φ sweep results shown in Figure 6.2A, C, and D. Because Φ has not natural barrier in parameter space, the convergence metrics considered above do not provide reliable indicators of model convergence. Instead, we make use of the fact that the IR vs. Φ and the S vs. P parameter sweeps should function (either directly or indirectly) to uncover the maximum achievable non-equilibrium information rate for each model architecture. Thus, as a basic test of sweep performance, we checked for the consistency between IR estimates derived from these different sweep modalities. As illustrated in Figure C.10A, we find excellent agreement between the maximum IR values derived from the S vs. P (x axis) and IR vs. Φ (y axis) parameter sweeps for all model architectures considered. This provides one indication the IR vs. Φ sweeps are fully exploring the relevant parameter space.

As a second check, we compared the IR vs. Φ bounds derived for two separate rounds of parameter sweeps ("round a" and "round b") comprised of 200 and 500 independent parameter sweeps, respectively. We reasoned that, if our algorithm is accurately recovering the true IR vs. Φ bound for each model architecture, this bound (i) should be replicable across different parameter sweep rounds and (ii) should be insensitive to the precise number of sweep runs per round. For each model architecture, w e calculated the maximum IR value returned by sweep rounds a and b for 30 different rates of energy dissipation ranging from $0.1 k_B T$ (close to equilibrium) to $5000 k_B T$ (upper limit of x axis in Figure 6.2D). Figure C.10B and C show the results of this exercise for multi-binding site and multi-locus conformation models, respectively, indicating excellent agreement between different sweep round for all model architectures. This indicates that our information vs. energy bounds are highly replicable across different rounds of sweeps. The consistency across round comprised of significantly different numbers of runs provides further evidence that we are conducting a sufficient number of independent sweeps ($\geq$ 200) per run. Taken together, these results and the results from the preceding paragraph provide strong evidence that our algorithm is robustly recovering accurate IR vs. Φ bounds for all models considered.

C.2.7.3 Information vs. w/c sweeps

Finally, we turn to the parameter sweep results for information (and, correspondingly, decision time) as a function of wrong-to-right activator concentration (w/c) shown in Figure 6.4C-E. We note that the results shown in Figure 6.5A and C are also derived from these sweeps. Like Φ, w/c has no intrinsic boundary in parameter space and, thus, swept area provides a poor indication of convergence. Fortunately, in addition to treating w/c as a sweep parameter, we can also conduct 2D parameter sweeps where w/c is set at a constant value (e.g. $w/c = 1000$ in Figure C.3B). These sweeps *do* converge, with an average of 87% of runs reaching 95% of the global maximum. Thus we cross-validate the IR vs. w/c bounds returned by the sweeps from Figure 6.4 by conducting separate sweeps of IR vs. r (the transcription rate) at different w/c values (illustrated in Figure C.11A).

Figure C.11B and C show the results of this comparison for three different values of w/c:

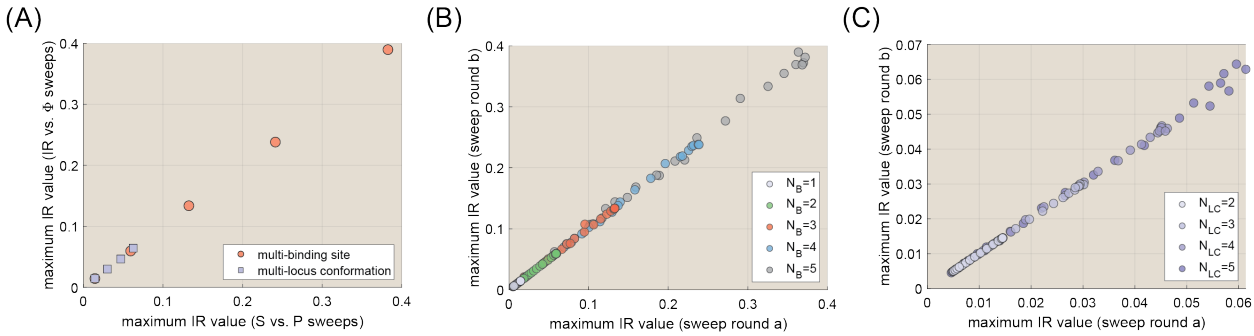

Figure C.10: **Convergence results for IR vs. Φ parameter sweeps. (A)** Scatter plot comparing maximum information rate estimated from S vs. P and from IR vs. Φ sweeps. **(B-C)** Scatter plots comparing results for the upper IR bound at different points along the curves shown in Figure 6.2C and D from two independent rounds of parameter sweeps comprising 200 and 500 separate runs, respectively. Points reflect IR maxima for Φ values ranging from $0.1k_BT$ to $5000k_BT$.

10, 10^2, and 10^3. We focus on the architectures depicted in Figure 6.4, namely equilibrium systems with 1 to five binding sites and non-equilibrium systems with 2-5 locus conformations. In most cases, we find good agreement between the two methods, suggesting that the IR vs. w/c sweeps are generally returning accurate estimates for the IR vs. w/c bound. We do note, however, that IR vs. w/c sweeps appear to underestimate the upper IR bound to a significant degree for the non-equilibrium model with 5 locus conformations when $w/c = 10$ (circle in upper right-hand corner of Figure C.11C). This indicates that the IR vs. w/c sweep is performing sub-optimally in this case. However, since this deviation occurs in the extreme low interference regime and our focus in Section 6.3.5 lies on model performance at higher w/c levels ($w/c \gtrsim 100$), where our sweep algorithm performs reliably, it does not impact any conclusions drawn throughout the course of the main text. Thus, we conclude that the IR vs. w/c sweeps provide a viable basis for the investigations undertaken in this study.

C.2.8 Estimating decision time ranges for different biological systems

Caenorhabditis elegans decision time estimation

A recent study by Lee and colleagues [178] used live imaging to examine Notch-dependent burst dynamics in the *sygl-1* gene in the germ line of young adult nematodes. Their results indicate that the gene exhibits burst cycle times ranging from 60.5 minutes up to 105.3 minutes (see Figure 2 E and F in [178]). Meanwhile, a review article indicated potential values for the cell cycle time for adult germ-line cells in C. *elegans* as ranging from 16 to 24 hours [138]. A separate study examining nonsense-mediated mRNA decay in C. *elegans*

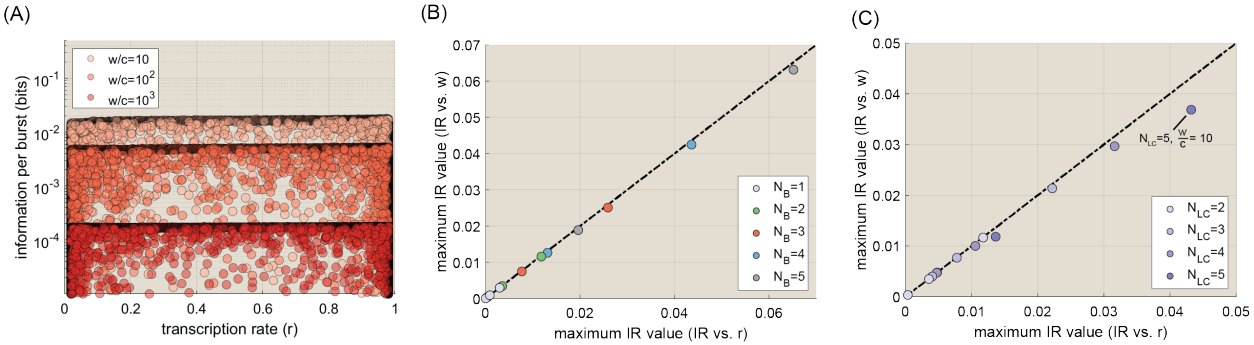

Figure C.11: **Convergence results for IR vs. w/c parameter sweeps. (A)** Illustrative scatter plot showing IR vs. r sweep results for the three binding site model at equilibrium for three different values of w/c. Scatter plot comparing maximum information rate estimated from S vs. P and from IR vs. Φ sweeps. **(B-C)** Scatter plots comparing results for the upper IR bound at three different w/c levels (10, 10^2, and 10^3) derived from IR vs. r sweeps (x axis) and IR vs. w/c sweeps (y axis) for equilibrium multi-binding site and non-equilibrium multi-locus conformation models, respectively.

reported a half life of approximately 6 hours for the *rpl-7A* gene (Figure 4k in [282]). If we take the cell cycle time as the upper time limit for cellular decision-making, this leads to an estimate of $1440/60.5 = 23.8$ burst cycles.

Mus musculus decision time estimation

Burst cycle time estimates were taken from Table A.1 in Appendix A of [166], which indicates times ranging from 30 minutes to a "few hours". mRNA half life estimates were taken from Table 1 of [235], which indicates a range of 30 minutes to 30 hours for mouse cells. To estimate the effective decision time corresponding to an mRNA half-life of 30 hours (1,800 minutes), we recognize that, once mRNA levels have reached a steady state, they will reflect (in effect) an weighted average of preceding transcriptional activity, where weights moving backward in time contribute

$$w(t) = e^{-\frac{t}{\tau_{\mathrm{mRNA}}}}, \tag{C.47}$$

where t indicates temporal distance from the present and τ_{mRNA} is the exponential time constant, given by $\tau_{\mathrm{mRNA}} = t_{\frac{1}{2}}/\ln(2)$. Integrating Equation C.47, we find that τ_{mRNA} time steps are effective present in steady-state mRNA levels. Taking 30 minutes as the lower bound for bursting timescales, this yields an upper bound of $((1800/\log 2)/30 = 86.6$ burst cycles. We note that this estimate is not materially different from the 60 cycle estimate that would be obtained by simply dividing 1,800 by 30.

***Drosophila melanogaster* decision time estimation**

We take the duration of nuclear cycle 14, which follows the thirteenth (and final) round of synchronous cellular divisions in early *Drosophila melanogaster* development, as the relevant timescale for cellular decisions in early fruit fly development. Studies have found that the duration of this developmental period varies along the embryo, with a minimum duration of 65 minutes [90]. To estimate bursting timescales, we use burst inference results from our previous work [168], which indicate a burst cycle time of approximately 2 minutes for the *even-skipped* gene. Thus, we arrive at an upper limit of $65/2 = 32.5$ cycles.

C.2.9 Higher-order molecular models

Here we provide an overview of key modeling assumptions underlying our approach to modeling gene circuits with multiple activator binding sites or multiple locus conformations. For simplicity, we assume that all activator bindings sites are identical. The case of multiple locus conformations is slightly more subtle. Setting $\mathrm{N_{LC}} > 2$ is intended to reflect the reality that multiple distinct molecular reactions—e.g. mediator engagement, PIC assembly, and nucleosome displacement—are necessary preconditions for the onset of transcription. While, in reality, these reactions are likely characterized by hetrogeneous dynamics (see, e.g., [166]) we once again make the simplifying assumption that all steps connecting different locus conformations—referred to as "molecular steps" or "activation steps" are identical.

As a result, from the perspective of model behavior, it does not matter *which* binding sites are bound or which molecular steps have been traversed, it matters only *how many* activators are bound or, similarly, how many activation steps have been traversed. Mathematically, these assumptions mean that the reaction rates (k_i) and interaction terms (η_{ij}) for all additional binding sites and locus conformations are identical with those shown in Figure 6.1B, but that the presence of multiple bound activators of multiple engaged general factors is felt via extra factors of η_{ij}. For instance, if the basal activation rate is k_{a}, then the activation rate n activators bound will be $k_{\mathrm{a}}\eta_{\mathrm{ab}}^{n}$. Similarly, if the basal activator unbinding rate is k_{-}, then the unbinding rate in the mth conformation is given by $k_{-}\eta_{\mathrm{ua}}^{m-1}$.

In all cases, we retain the same basic MWC-like architecture outlined above: no number of bound activators is alone sufficient for mRNA production, but each contributes an extra factor of η_{ab} and η_{ib} to impact locus activation dynamics (Figure 6.1B and C). In models that feature $\mathrm{N_{LC}}$ locus conformations (where $2 \leq \mathrm{N_{LC}} \leq 5$), it is assumed that only the $\mathrm{N_{LC}}$th conformation is transcriptionally active, in keeping with the nature of processes like mediator engagement, PIC assembly, and nucleosome displacement; all of which are necessary preconditions for the onset of transcription. This leads to models with multiple molecular steps along the path toward transcriptional activation, with the kinetics of each conditioned by the number of activators bound (as stipulated above) and, likewise, capable of impacting activator binding dynamics via the parameters η_{ba} and η_{ua}.

Finally, to fully describe these higher-order models we require three additional parameters: one that captures cooperative interactions between bound activators (η_m), one for cooperative interactions between general factors that are engaged with the locus (η_i), and

one that captures the capacity for one engaged general factor to impact the recruitment of another to the gene locus (η_a).

C.2.10 Deriving normalized sharpness and precision metrics

In Figure 6.1A, the transcriptional sharpness, s, is defined as the first derivative of the transcriptional input-output function multiplied by the activator concentration, c^*, such that

$$s = \frac{dr}{dc}c^*. \tag{C.48}$$

The transcriptional precision, p, is defined as the inverse of the intrinsic noise in the transcriptional input-output function:

$$p = \frac{1}{\sigma}, \tag{C.49}$$

where σ is as defined in Equation C.24 in the main text. Under these definitions, a key challenge in comparing sharpness and precision levels across different gene circuits is that the upper bounds on both s and p depend not only on the microscopic transition rates within a system, but also on the fraction of time the system spends in the transcriptionally active state, $\overline{a}$ (defined in Equation C.23). Figure C.12A and B illustrate this $\overline{a}$-dependence for equilibrium and non-equilibrium realizations of the four state system defined in Figure 6.1A. As an example: the equilibrium bound on s is 0.25 when $\overline{a} = 0.5$, but only 0.09 when $\overline{a} = 0.1$ (Figure C.12A). Since we allow gene circuits to take on different transcription rates ($r = \overline{a}\, r_0$) at $C = c^*$, this $\overline{a}$-dependence thus confounds our efforts to understand how the molecular architecture of gene circuits—the number of binding sites, number of molecular steps, and presence or absence of energy dissipation—dictates transcriptional performance.

To overcome this issue, we need to normalize s and p such that they are independent of $\overline{a}$. Focusing first on sharpness, we were inspired by previous works [82, 113] to leverage Hill Function as a flexible conceptual tool for extracting generic sharpness measures. The Hill function is defined as:

$$\overline{a} = \frac{c^S}{c^S + K_d^S}, \tag{C.50}$$

where c is the activator concentration, S is the Hill coefficient, and K_d is a constant that dictates the location of the function's half-max point. In general, the input-output functions generated by our model gene circuits will have more complex functional forms, but nonetheless, Equation C.50 indicates that we can relate these more complex functions to the Hill function via the shared parameters $\overline{a}$ and c.

The sharpness of the Hill function has the form:

$$s_H = S\frac{c^S K_d^S}{(c^S + K_d^S)^2}. \tag{C.51}$$

To better relate this to our input-output function, we need to re-express K_d in terms of C and $\overline{a}$. Solving Equation C.50 for K_d yields

$$K_d = c\Big(\frac{1-\overline{a}}{\overline{a}}\Big)^{\frac{1}{S}}. \tag{C.52}$$

Plugging this in to Equation C.51 we obtain, after simplification:

$$s = \overline{a}(1-\overline{a})S. \tag{C.53}$$

This expression tells us that the sharpness (s) of a Hill function with activity level $\overline{a}$ at $C = c^*$ is equal to the Hill coefficient, S, multiplied by the term $\overline{a}(1-\overline{a})$. By rearranging, we can obtain the Hill coefficient as a function of s and $\overline{a}$

$$\mathrm{S} = \frac{s}{\overline{a}(1-\overline{a})}. \tag{C.54}$$

Thus, for a generic gene circuit input-output function with sharpness s and expression level $\overline{a}$ at $C = c^*$ we can invoke Equation C.54 to calculate the Hill coefficient for the equivalently sharp Hill function (Figure C.12C). This provides us with a generic measure of transcriptional sharpness that is independent of $\overline{a}$ and thus can facilitate comparisons across gene circuits that drive differing activity levels at $C = c^*$ (Figure C.12D). We refer to this independent sharpness metric as the "normalized sharpness" in the main text, and denote it with the variable S.

This leads us to the question of transcriptional precision. The two key considerations in defining the normalized precision metric, P, is that (a) we want it to yield a quantity proportional to the information rate (IR) when multiplied with S and (b) we want it to adhere to constant bounds across the induction curve. There is only one definition that satisfies the first constraint:

$$\mathrm{P} = p\big(\overline{a}(1-\overline{a})\big) = \frac{\overline{a}(1-\overline{a})}{\sigma}. \tag{C.55}$$

Happily, Equation C.55 exhibits consistent upper bounds for all $\overline{a}$ values, and thus satisfies our second constraint (Figure C.12E).

C.2.11 Optimal equilibrium four state gene circuits behave like effective two state systems

In this section, we calculate the normalized sharpness (S) and precision (P) for a simple 2 state gene circuit (Figure C.13A) with one ON state and one OFF state and two transition rates, k_{off} and k_{on}. We assume that activator binding dictates fluctuations into and out of the ON state, such that k_{on} is proportional to c ($k_{\mathrm{on}} = ck^0_{\mathrm{on}}$). For this simple system, the rate of transcription is given by

$$\overline{r} = r_0 \frac{ck^0_{\mathrm{on}}}{ck^0_{\mathrm{on}} + k_{\mathrm{off}}} = r_0\overline{a}. \tag{C.56}$$

Differentiating this expression with respect to c and setting $r_0 = 1$ (as in main text), we find that

$$s = \frac{ck^0_{\mathrm{on}}k_{\mathrm{off}}}{(ck^0_{\mathrm{on}} + k_{\mathrm{off}})^2}. \tag{C.57}$$

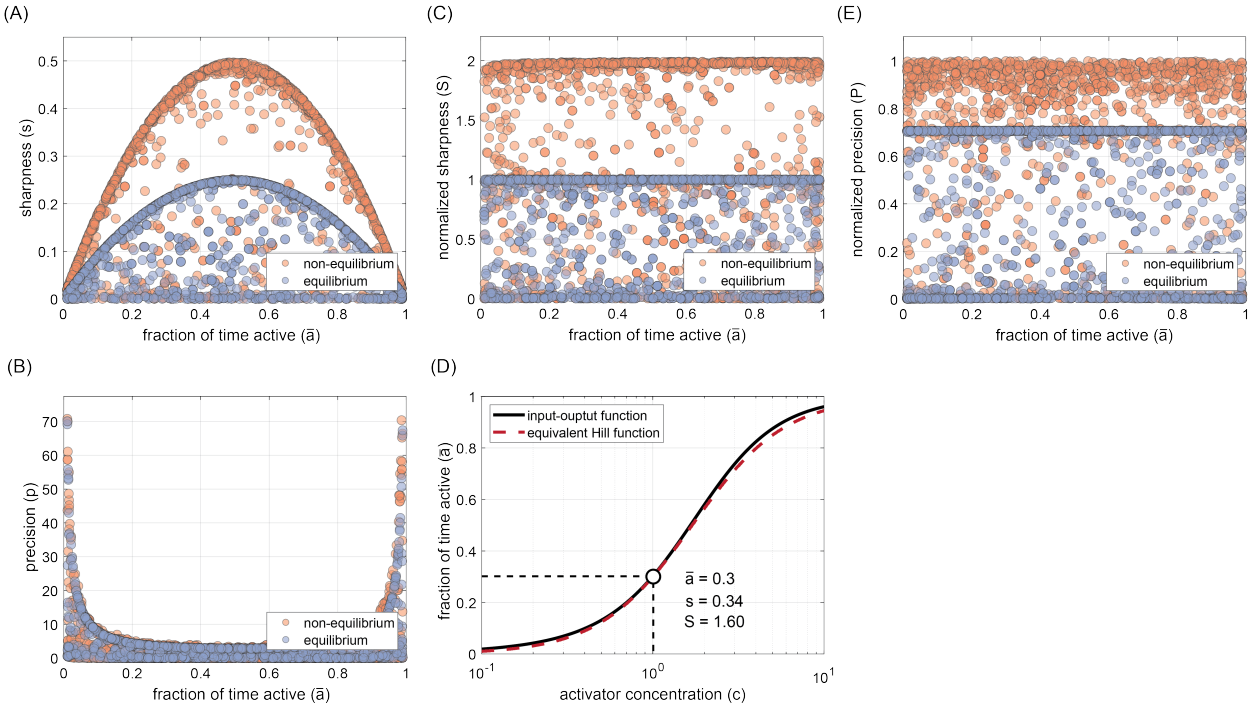

Figure C.12: **Defining normalized sharpness and precision.** **(A)** Plot depicting the upper sharpness limit for equilibrium (blue) and non-equilibrium (red) realizations of the four state system depicted in Figure 6.1B. The upper limit depends on the fraction of time spent in the active state, $\bar{a}$. **(B)** Plot of precision as a function of the transcription rate. Here again, the upper bounds depend on $\bar{a}$. **(C)** Plot of normalized sharpness as a function of the transcription rate. In the case, the upper limits are invariant. **(D)** Illustration of normalized sharpness concept. For a given input-output curve, we identify normalized sharpness, S, as the Hill coefficient of an equivalently sharp Hill function with the same expression level at $C = c^*$. **(E)** On the other hand, the normalized precision, P, exhibits invariant performance bounds.

Finally, dividing through by $b = \bar{a}(1 - \bar{a})$ yields the normalized sharpness, which is simply given by

$$S = 1. \tag{C.58}$$

Thus, we see that the two state model is constrained to a normalized sharpness level that represents the upper performance limit for the four state model operating at equilibrium (blue circles in Figure 6.3A).

Next, we turn to precision. From Equation C.24, we find that

$$\sigma^2 = \frac{2ck_{\text{on}}^0 k_{\text{off}}}{(ck_{\text{on}}^0 + k_{\text{off}})^3}. \tag{C.59}$$

Inverting and multiplying by b^2 gives

$$P^2 = \frac{ck_{\text{on}}^0 + k_{\text{off}}}{2(ck_{\text{on}}^0 + k_{\text{off}})}. \tag{C.60}$$

Finally, multiplying through by τ_b (Equation C.30) and taking the square root gives

$$P = \frac{1}{\sqrt{2}}, \tag{C.61}$$

which, again, is equivalent to the upper limit of the four state gene circuit at equilibrium (Figure 6.3A).

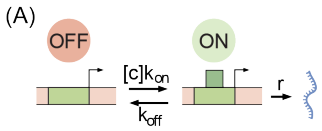

Figure C.13: **A simple 2 state model of transcription. (A)** Cartoon of a simple 2 state gene circuit model in which activator binding and unbinding dictate transitions into and out of a transcriptionally active state.

C.2.12 Sharp and precise non-equilibrium networks exhibit distinct and incompatible microscopic topologies

One simple way to probe the microscopic architectures of different gene circuits is to measure the degree of heterogeneity (or dispersion) in (a) transition rates and (b) state probabilities. We developed entropy-based dispersion metrics ranging from 0 to 1 to quantify how uniform (0) or heterogeneous (1) transition rates and state probabilities were for different realizations of the four state network shown in Figure 6.1A. While crude, these measures can provide useful microscopic insights. For instance, in gene circuits with a state probability score of 0 each microscopic state must be equiprobable ($p_1 = p_2 = p_3 = p_4 = 1/4$), while those with a 1 are maximally heterogeneous, with all probability concentrated in just two of the four states. Identical considerations hold for the transition rate axis. We conducted parameter sweeps to explore the space of achievable dispersion values for 10,000 non-equilibrium gene circuits (gray circles in Figure C.14A). For simplicity, we focused on gene circuits at half-max expression ($\overline{a} = 0.5$).

From Figure C.14A, we can see immediately that precise and sharp gene circuits occupy opposite extremes of dispersion space. Specifically, precise systems exhibit highly uniform state probability and transition rate values, while sharp networks are highly heterogeneous, both with respect to the fraction of time spent in each state and the relative magnitudes of their transition rates. These stark differences, as well as the tight clustering of each motif, suggest that sharpness and precision arise from distinct and non-overlapping microscopic topologies.

Detailed examination of maximally precise gene circuits from our parameter sweeps indicated that these systems exhibit highly uniform molecular architectures wherein each microscopic state is equiprobable, all clockwise transition rates are uniform, and all counterclockwise rates are negligible. This results in a "clock-like" system that maximizes the regularity of molecular transitions. Maximally sharp gene circuits, on the other hand, exhibit an all-or-none character, behaving as effective two state systems that spend most of their time either activator-bound and active (0), or unbound and inactive (1), and which have effective ON and OFF rates that are concentration-dependent (see Appendix C.2.13 for further details).

C.2.13 A hierarchy of microscopic transition rates underpins non-equilibrium sharpness gain

Figure 6.3A shows that energy dissipation opens up a broad spectrum of S and P values that are not attainable at equilibrium. It is difficult to formulate general statements that apply to all gene circuit models inhabiting these spaces beyond the upper equilibrium limit; however, we can learn much by examining the architecture of gene circuits lying at the outer limits of non-equilibrium performance, since these systems tend to distil the logic underpinning of non-equilibrium performance gains into relatively simple regulatory motifs.

Such is the case for the IR-optimized non-equilibrium four state systems depicted as black circles in Figure 6.3A. In Section 6.3.4, we found that the driver of this IR is a twofold increase in sharpness relative to the upper equilibrium limit. To realize this twofold sharpness gain, we find that non-equilibrium driving is harnessed to facilitate effective one-way transitions between the active and inactive conformations—specifically, from states 1 to 2 and 3 to 0 in Figure 6.1B—ensuring that the system will have a strong tendency to complete transcriptional cycles in the clockwise direction ($J > 0$).

In addition to this non-equilibrium driving, sharpness maximization places strict constraints on the relative magnitudes of microscopic transition rates within the network. To understand these constraints, it is instructive to consider a coarse-grained representation of our network with a single ON state (2), a single OFF state (0). We can obtain expressions for the two effective transition rates in the network by recognizing that they are equal to the inverse of the mean first passage times between states 2 and 0, which we can calculate using Equation C.29 from Appendix C.2.4.

If we neglect the energetically disfavored transitions from 2 to 1, the effective ON rate (k^*_{on} in Figure 6.3D) takes on a relatively simple form

$$k^*_{\text{on}} = \frac{[c]k^i_+ k_{\text{a}} \eta_{\text{ab}}}{[c]k^i_+ + \eta_{\text{ab}} k_{\text{a}} + k_-}. \tag{C.62}$$

From Equation C.62, we see that the effective ON rate becomes proportional to the concentration, c, when the factor of $[c]k_+$ becomes negligible in the denominator. The limit where $k_- \gg [c]k_+, \eta_{\text{ab}} k_{\text{a}}$ represents a scenario in which the activator K_d is large when the network is in the inactive conformation such that the activator must bind multiple times (on average) before it succeeds in driving the system into the active conformation. The other limit, when

$\eta_{\mathrm{ab}} k_{\mathrm{a}} \gg [c]k_+, k_-$, corresponds to a system where locus activation happens rapidly upon activator binding.

In similar fashion, the effective OFF rate can be expressed as the inverse of the first passage time from 3 to 1

$$k^*_{\mathrm{off}} = \frac{\eta_{\mathrm{ua}} k_- k_{\mathrm{i}}}{\eta_{\mathrm{ba}}[c]k_+ + k_{\mathrm{i}} + \eta_{\mathrm{ua}} k_-}. \tag{C.63}$$

Interestingly, we see that k_{off} becomes *inversely* proportional to c when activator binding rate exceeds both the unbinding rate and the rate of locus deactivation ($\eta_{\mathrm{ba}}[c]k_+ \gg k^a_-, k^-_{\mathrm{off}}$). This imbalance causes the system to become kinetically trapped in the active conformation for multiple cycles of activator unbinding and rebinding, with an average duration inversely proportional to $\eta_{\mathrm{ba}}[c]k_+$.

Thus, when the proper hierarchy of microscopic rates is realized, our four state network behaves *as though* it were a two state system in which both the on and off rates are concentration dependent, such that

$$r \approx \frac{[c]k^*_{\mathrm{on}}}{[c]k^*_{\mathrm{on}} + \frac{1}{[c]}k^*_{\mathrm{off}}}. \tag{C.64}$$

Repeating the calculations from Appendix C.2.11 for the above effective two state system will yield an S value of 2 and a P value of 1, in agreement with our numerical results from Figure 6.3A. We propose that this doubled concentration dependence can be conceptualized as a kind of "on rate-mediated" proofreading. In contrast to classical kinetic proofreading, which works by amplifying intrinsic differences in ligand off rates [136, 226], sharp networks amplify the concentration-dependence carried by binding rates, effectively "checking" C twice per cycle.

C.2.14 Non-equilibrium gains in sharpness drive IR increases in more complex regulatory architectures

This appendix section contains additional discussion relating to sharpness-precision tradeoffs for higher-order model architectures with multiple binding sites or multiple locus conformations.

Sharpness maximization remains optimal for systems with multiple binding sites

To assess whether sharpness-maximization remain the optimal strategy for more complex architectures featuring multiple activator binding sites, we employed parameters sweeps to examine the space of achievable S and P values for gene circuits with 1-5 activator binding sites (and N_B fixed at 1). Figure C.15A shows the results of this analysis. For ease of comparison across different models, we plot the relative gains in S and P for each model with respect to their maximum equilibrium values. For instance, the maximum equilibrium

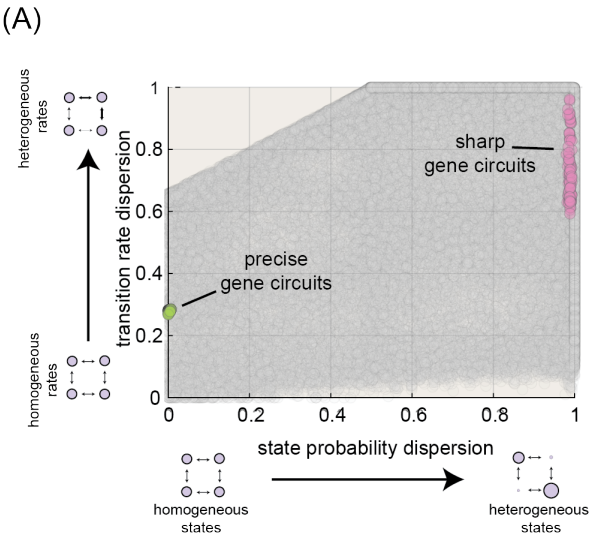

Figure C.14: **Sharp and precise non-equilibrium networks exhibit distinct and incompatible microscopic topologies. (A)** Plot showing dispersion scores for state probabilities and transition rates for 10,000 non-equilibrium networks. Here, a score of 0 indicates maximal uniformity (all rates or probabilities are equal) and a 1 indicates maximal heterogeneity. Green and pink circles indicate the scores for the 100 sharpest and most precise gene circuits, respectively.

S value for the $N_B = 2$ model is 2, so a non-equilibrium gene circuit model with two binding sites that exhibits an S value of 2.5 will be calculated to have a sharpness gain of $2.5/2 = 1.25$.

Figure C.15A reveals that the sharpness-precision tradeoff observed for the one-binding site model persists and, indeed, becomes more severe for systems with additional activator binding sites. We see that the non-equilibrium gain in S is fixed at approximately 2. And while the non-equilibrium gain in P increases from $\sqrt{2}$ for $N_B = 1$ to approximately 2.25 for $N_B = 5$, these P maxima (peaks in the upper left quadrant of Figure C.15A) occur at lower and lower values of S, which renders them more and more disadvantageous from an IR perspective. As a result, when we plot IR-optimal gene circuits for each value of N_B (colored circles in Figure C.15A), we find that they are invariably located in regions where $S/S_{eq} \approx 2$ and $P/P_{eq} \approx 1$. These results demonstrate that spending energy to maximize sharpness remains the key to maximizing transcriptional information transmission, irrespective of the number of activator binding sites.

Multiple locus conformations increases upper sharpness bound away from equilibrium

Figure C.15B shows the range of achievable non-equilibrium gains in S and P for systems with 2-5 locus conformations (and $N_B = 1$). Once again we observe a strong tradeoff between sharpness and precision, which suggests that this incompatibility is a general feature of

transcriptional systems. And, once again, we find that IR-maximizing gene circuits (colored circles) lie at or near the right-most edge of achievable parameter space, indicating that dissipating energy to enhance transcriptional sharpness (rather than precision) remains the best strategy for maximizing the information rate.

Yet unlike the systems examined in Figure C.15A, Figure C.15B reveals that the non-equilibrium gain in transcriptional sharpness (S) is not fixed but, rather, increases with the number of molecular steps from a factor of two when $N_{LC} = 1$ to a factor of *five* when $N_{LC} = 4$. This indicates that increasing the number of dissipative molecular steps in the activation pathway raises the upper limit on the sharpness of the transcriptional input-output function, even when the number of binding sites is held constant.

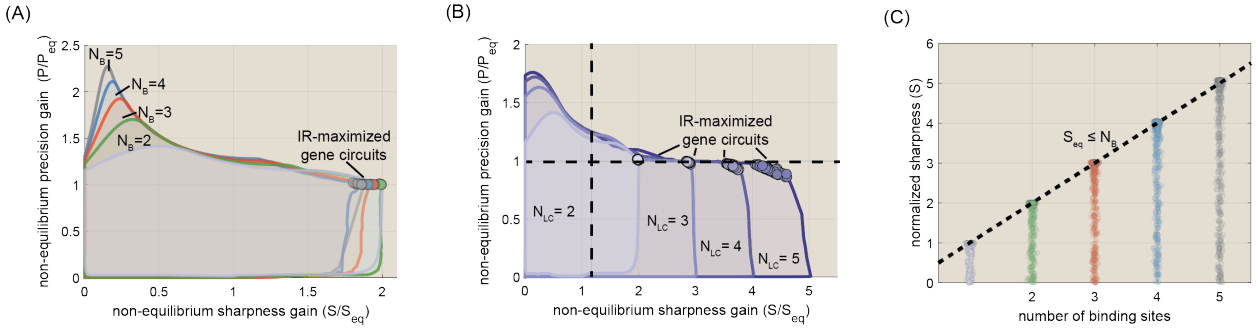

Figure C.15: **Tradeoffs between sharpness and precision persist for more complex gene regulatory architectures.** **(A)** Non-equilibrium gains in sharpness and precision for gene circuits with different numbers of activator binding sites (N_B) and two locus conformations. Shaded regions indicate achievable regimes for each system, as determined by no fewer 10,000 unique simulated gene circuits. **(B)** Non-equilibrium gains in sharpness and precision for gene circuits with different numbers of locus conformations (N_{LC}) and one activator binding site. **(C)** Scatter plots indicate sharpness levels for equilibrium gene circuits as a function of the number of binding sites. Bounding line is for a function of the form $S = N_B$.

C.2.15 Specificity definitions and details

This Appendix Section uses a simple two state gene circuit model to compare and contrast the specificity definition employed in this work, which focuses on cognate and non-cognate factors competing to activate a single locus, and the definition employed in two recent works [272, 113], which instead compare how a single factor activates two gene loci: a target locus with specific binding sites, and a non-cognate locus that lacks binding sites.

A detailed comparison of specificity definitions for a simple 2 state model of transcription

Figure C.16A illustrates this scenario for the case of a simple two state network with a single binding site and no possibility of a conformation change at the locus; however the same idea applies equally well for the 4 state network we considered above, as well as more complicated architectures. Here transcriptional specificity is then defined as the ratio of the average steady state transcription rates at on- and off-target gene loci:

$$f_p = \frac{r_r}{r_w}, \tag{C.65}$$

where f_p is the specificity under the parallel framing of the problem, and r_r and r_w indicate the transcription rates at the cognate (right) and non-cognate (wrong) loci, respectively. In [272], the authors show that specificity for the two state system shown in Figure C.16A is given by:

$$f_p = \frac{\alpha k_- + [c]k_{\text{on}}}{k_+ + [c]k_{\text{on}}}. \tag{C.66}$$

From Equation C.66, we see that the activator specificity is bounded from above by α. Moreover, this upper performance limit is achieved only in an off rate-dominated regime where $k_- >> [c]k_+$, which the authors in [272] note leads to a runaway increase in transcriptional noise with increasing specificity under the constraint that the mean transcription rate must remain constant. As a result, the authors conclude that non-equilibrium network architectures are necessary in order to improve specificity and minimize transcriptional noise [272].

In analogy to the parallel case outlined above, we employ a "gene-centric" definition (Figure C.16B), which takes specificity as the ratio of the average number of cognate and non-cognate factors bound while the locus is in a transcriptionally productive state, normalized by concentration:

$$f = \frac{w}{c}\frac{n_c}{n_w}. \tag{C.67}$$

In the case of the two state model shown in Figure C.16B, this is simply given by the ratio of fractional occupancies of states 1 and 1*:

$$f = \frac{w}{c}\frac{\pi_2}{\pi_{2^*}}, \tag{C.68}$$

which simplifies to

$$f = \alpha. \tag{C.69}$$

From Equation C.69, we see that f is simply equal to the binding specificity factor α for our three state network, *irrespective of binding kinetics*. Thus, in contrast to [272], we find that equilibrium gene circuits need not shift towards a noisy, off rate-dominated regime to achieve maximum fidelity; indeed *all* systems necessarily achieve precisely $f = \alpha$. Intuitively, this difference stems from the fact that our model captures the effects of kinetic competition

between cognate and non-cognate activators: whenever the cognate activator (green square in Figure C.16B) is bound, non-cognate factors cannot bind.

A key limitation of this approach is that it neglects the presence of non-specific stretches of regulatory DNA, even at cognate gene enhancers. Thus, to more accurately reflect the specificity challenges faced by real gene loci, a synthesis of the two approaches summarized above will be necessary, which considers competition between cognate and non-cognate factors to bind and activate a gene locus that features both specific binding sites (which favor the cognate activator) and neutral sites (to which all activator species bind non-specifically). One expectation for such a scenario is that the simple equality stated in Equation C.69 will no longer hold, and tradeoffs similar to those observed in [272] will again emerge; although, this time, the severity of these tradeoffs will depend on w/c.

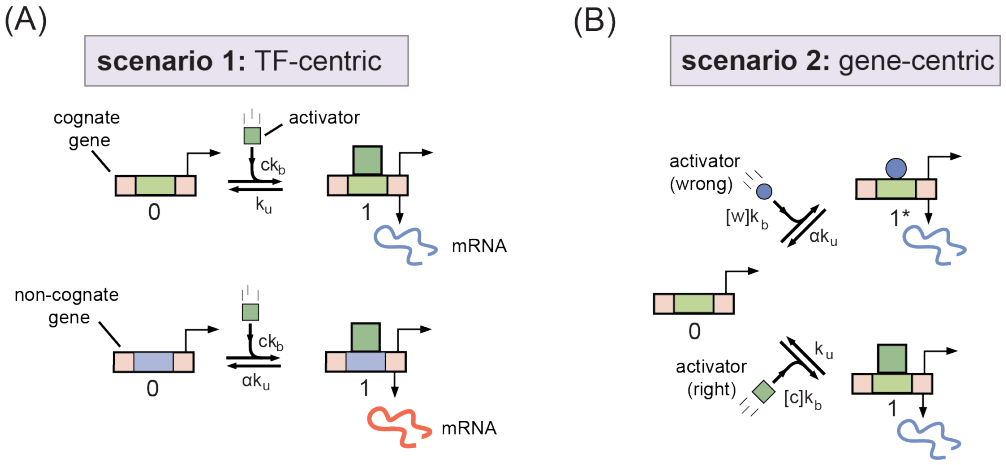

Figure C.16: **Accounting for the influence of off-target activation.** **(A)** An illustration of the parallel definition of activation fidelity. This approach considers the relative amounts of transcription driven by a transcriptional activator at its target locus and at an off target locus. **(B)** Cartoon illustrating "gene-centric" specificity definition, which considers competition between cognate and non-cognate factors to bind and activate a single gene locus.

C.2.16 Deriving non-equilibrium tradeoff bound between intrinsic sharpness and specificity

In this section, we lay out the key steps in deriving the non-equilibrium tradeoff bound between sharpness and specificity given in Equation 6.7 in the main text. To do so, we make use of insights gained in Appendix C.2.13, where we use first passage times to examine the key microscopic conditions for the twofold gain in sharpness away from equilibrium observed in Figure 6.3A. Even for the simple six state system illustrated in Figure 6.4B, our system has eight degrees of freedom when operating away from equilibrium. As such,

a key part of our approach will be to first reduce this complexity as much as possible while preserving the salient behaviors, namely the possibility for non-equilibrium gains in sharpness and specificity. After this, we identify a tuning parameter, β, that can be used to interpolate between maximally sharp to maximally specific non-equilibrium gene circuit architectures. Since the expressions for non-equilibrium gene circuits are, in general, quite complex, we sketch the key steps here and direct the reader to the Mathematica notebook entitled "sharpness_specificity_bound_derivation.nb" on the project git repository (`https://github.com/nlammers371/noneq-gene-regulation.git`) for additional details. Note that we work in units of c throughout, such that $c = 1$.

To begin, we strip unnecessary dimensions from our system. We set $\eta_{\mathrm{ib}} k_{\mathrm{i}}$ and k_{a} to the same generic rate, k_1. Next, we set $\eta_{\mathrm{ab}} k_{\mathrm{a}}$, $\eta_{\mathrm{ua}} k_-$, and k_+ to a second rate parameter, k_2. Finally, we set k_{i} equal to $\beta \eta_{\mathrm{bs}} k_+$, where β is our interpolation parameter. This leaves us with a system with five free parameters, rather than eight.

In Appendix C.2.13, we saw that maximally sharp non-equilibrium gene circuits (i) only switch into the active transcriptional conformation when the activator is bound and (ii) only switch *out* of the ON states when the activator is unbound. This amounts to effective one-way transitions from states $1 \rightarrow 2$ (equivalently, $5 \rightarrow 6$) and $3 \rightarrow 1$. We impose this condition by taking the limit where $k_1 \rightarrow 0$. Next, we impose the condition uncovered by examination of Equation C.62,

$$k_- \gg [c] k_+, \eta_{\mathrm{ab}} k_{\mathrm{a}}, \tag{C.70}$$

by taking the limit where k_- approaches infinity.

These limits lead to a further simplified system that can be used to investigate fundamental tradeoffs between intrinsic sharpness and specificity. For this stripped-down system, we find that the expression for specificity, f, is quite simple:

$$f = \frac{\alpha(\alpha + \alpha\beta + w)}{\alpha + \beta + w}, \tag{C.71}$$

where we see that all dependence on microscopic transition rates has dropped out, with the exception of our interpolation parameter, β. Furthermore, tuning β causes Equation C.71 to shift from equilibrium levels ($f = \alpha$ when $\beta = 0$) to the non-equilibrium limits revealed by Figure 6.5B ($f = \alpha^2$ when $\beta \gg \alpha, w$).

The normalized sharpness, S, has a slightly more complicated functional form, given by

$$S = \frac{\alpha\big[\alpha(k_2 + (2+\beta)\eta_{\mathrm{ba}} k_+) + 2\eta_{\mathrm{ba}} k_+ w\big]}{\alpha^2(k_2 + \eta_{\mathrm{ba}} k_+ + \beta\eta_{\mathrm{ba}} k_+) + \alpha w(k_2 + 2\eta_{\mathrm{ba}} k_+) + w\eta_{\mathrm{ba}} k_+(\beta + w)}. \tag{C.72}$$

To obtain an expression for the intrinsic sharpness, S_0, we divide through by the specificity prefactor from Equation 6.7:

$$\mathrm{S}_0 = \frac{f + w}{f} S. \tag{C.73}$$

Simplifying and applying the condition that $k_2 \approx 0$ leads to

$$\mathrm{S}_0 = 2 - \frac{\alpha\beta}{\alpha + \alpha\beta + w}. \tag{C.74}$$

Here again, as with Equation C.71, we see that all dependence on rate parameters drops away. Further, it is easy to see that this expression goes to 2 when $\beta = 0$ and 1 when $\beta \gg \alpha, w$. Thus, when β is small, our system exhibits equilibrium levels of specificity and non-equilibrium levels of intrinsic sharpness and, when β is large, it exhibits non-equilibrium specificity and equilibrium sharpness levels. Thus, we have succeeded in our initial aim to establish a simplified model that can capture the tradeoffs between sharpness and specificity revealed by our numerical parameter sweeps (Figure 6.5B).

As a final step, we can solve Equation C.71 to obtain an expression for β in terms of f:

$$\beta = \frac{(f-\alpha)(\alpha+w)}{\alpha^2 - f}. \tag{C.75}$$

Plugging this expression into Equation C.73 and simplifying yields an expression for S_0 as a function of f:

$$\mathrm{S}_0 = \frac{\alpha^2 + \alpha f - 2f}{\alpha f - f}, \tag{C.76}$$

where we assume that $\alpha \leq f \leq \alpha^2$. Thus, we have obtained the final S_0 expression depicted in Equation 6.7. Observe that $\mathrm{S}_0 \approx 2$ when $f = \alpha$ and $\mathrm{S}_0 \approx 1$ when $f = \alpha^2$. Equation C.76 gives the dashed black curve bounding f vs. S_0 sweep results shown in Figure 6.5B, confirming that it represents the limiting behavior of intrinsic sharpness and specificity for non-equilibrium realizations of the six state model from Figure 6.4B.

Appendix D

SI for Fast, flexible inference of transcriptional dynamics with MCMC

D.1 Defining the MS2 kernel, κ

If we define $n_{\mathrm{MS2}} = \tau_{\mathrm{MS2}}/\Delta\tau$ as the number of time steps (not necessarily an integer) necessary for transcribing the MS2 sequence, the mean fraction of the full MS2 stem loop cassette transcribed by an elongating RNAP at the d^{th} time step of the elongation window will be given by

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

where $\lceil n_{\mathrm{MS2}} \rceil$ and $\lfloor n_{\mathrm{MS2}} \rfloor$ are the ceiling and the floor of n_{MS2}, respectively.

D.2 Sampling hidden promoter states

For a given set of model parameters, $\boldsymbol{\theta}_k$, we wish to obtain an estimate for the sequence of promoter states, $\boldsymbol{z}$, that corresponds to the observed MS2 trace, $\boldsymbol{y}$. Unlike the algorithm employed in Chapter 7, which estimated the probability of all possible promoter states at each time step, *burstMCMC* works by sampling discrete promoter state values at each time step, such that the final outcome is a single sequence of promoter states, $\boldsymbol{z}_k$, drawn from the distribution $p(\boldsymbol{z}|\boldsymbol{\theta}_k, \boldsymbol{y})$. For each inference step, k, each of the T promoter states is updated by drawing from $z_t \in \{1, 2, ...K\}$ based off of the marginal likelihood of each promoter state value, conditioned on the system parameters ($\boldsymbol{\theta}_k$), the observed fluorescent values ($\boldsymbol{y}$), and the remainder of the hidden promoter state sequence ($\boldsymbol{z}_{\varnothing t}$): $p(z_t|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \boldsymbol{A}, \boldsymbol{y})$. We can express this as the product of two terms, such that

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \boldsymbol{A}, \boldsymbol{y}) = p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{A}) p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \sigma, \boldsymbol{y}), \tag{D.2}$$

where the first term on the right-hand side captures the component of the probability derived from promoter state transitions and the second term captures the component relating to the fluorescent emission probabilities. It is helpful to consider these two components separately.

First, because of the Markovian nature of promoter switching, the transition component of the likelihood (first term on the right-hand side of Equation D.2) depends only the promoter states in the time points immediately preceding and following t. Thus, if $z_{t-1} = j$ and $z_{t+1} = l$, then we have that

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{A}) = a_{ij}a_{li}, \tag{D.3}$$

where a_{mn} denotes the probability of switching from state n to state m.

Next we turn to the fluorescence component of Equation D.2. The first step here is to recognize that, by Bayes' Theorem,

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \sigma, \boldsymbol{y}) \propto p(y(i)|z_t = i, \boldsymbol{v}, \sigma). \tag{D.4}$$

The right-hand side of this expression is straight-forward to calculate, so we will proceed with it moving forward. Since the promoter state at time t impacts the predicted fluorescence, $\hat{\boldsymbol{y}}$, for time points up to and including $t - \lfloor n_{\text{MS2}} \rfloor$. In this case, we have

$$p(z_t = i|\boldsymbol{z}_{\varnothing t}, \boldsymbol{v}, \sigma, \boldsymbol{y}) \propto (2\pi\sigma^2)^{-\frac{\lceil w \rceil}{2}} \prod_{n=t}^{t+\lfloor w \rfloor} e^{-\frac{(\hat{y}_n - y_n)^2}{2\sigma^2}}. \tag{D.5}$$

Plugging Equations D.3 and D.5 in Equation D.2 allows us to calculate the conditional probability of each promoter state for a given time point t. We then update $\boldsymbol{z}$ by randomly selecting time points without replacement from the set $t \in \{1, 2, ...T\}$ and drawing a new promoter state for each based off of Equation D.2.

D.3 Update expressions for model parameters

Once we have drawn a new sample promoter sequence, we can use this vector of hidden states to sample the posterior distributions of our core model parameters, $\boldsymbol{\theta} \in \{\boldsymbol{A}, \boldsymbol{v}, \sigma\}$. Here we quote well established update equations for each parameter. We direct readers to the original sources cited along the way for more detailed derivations.

We begin with the transition probability matrix, $\boldsymbol{A}$. As stipulated in the main text, we assume that columns of $\boldsymbol{A}$ ($\boldsymbol{a}^i$) are Dirichlet-distributed. Given an inferred latent state vector at iteration k, $\boldsymbol{z}_k$, the posterior distribution for each $\boldsymbol{a}^i$ is simply:

$$p(\boldsymbol{a}^i_{k+1}|\boldsymbol{z})p_0(\boldsymbol{a}^i) = Dir(\boldsymbol{n}^i_k + \boldsymbol{\alpha}^i_0), \tag{D.6}$$

where $\boldsymbol{\alpha}^i_0$ is the vector of prior pseudo-counts, and $\boldsymbol{n}^i_k$ is a vector of length K, where the jth element indicates the number of transitions *from* state i *to* state j in the state vector $\boldsymbol{z}_k$. From Equation D.6 it is easy to see how our prior factors into the posterior estimate: if the number of elements in $\boldsymbol{\alpha}_i$ is much less than the number of transitions out of state i, then our prior will have little impact, and the information gleaned from the data (the

fluorescent trace), will dominate. Conversely, if the elements of $\boldsymbol{\alpha}^i$ are of order with (or larger than) elements of $\boldsymbol{n}^i$, then our prior assumptions regarding the relative magnitudes of the transition probabilities our of state i will strongly influence our final posterior estimate of $\boldsymbol{a}^i$.

In developing update equations for the emission rate vector $\boldsymbol{v}$ and fluorescent noise parameter σ, we follow a similar path to that outlined in Appendix A.3.4. The first step is to calculate $\boldsymbol{F}$, a $T \times K$ matrix of counts that indicates the total number of time steps (potentially a fractional quantity) of each $z \in \{1, 2, ...K\}$ promoter state are reflected in the observed fluorescence at each time point t. For instance, if $w = 6$ and states $z_{i-5:i} = \{2, 1, 2, 1, 3, 2\}$, then (if we neglect τ_{MS2}), the ith row of $\boldsymbol{F}$ will read $\{2, 3, 1\}$. More generally, for some gene with MS2 kernel κ, the jth column of $\boldsymbol{F}$, $\boldsymbol{f}_j$, can be expressed as

$$\boldsymbol{f}_j = \kappa * \boldsymbol{1}_j(\boldsymbol{z}), \tag{D.7}$$

where $\boldsymbol{1}_j(\boldsymbol{z})$ is an indicator vector with T elements that are equal to 1 at time points when $z_t = j$ and 0 otherwise. Given $\boldsymbol{F}$, the problem of solving for $\boldsymbol{v}$ and σ can be framed as a multivariate linear regression:

$$\boldsymbol{y} = \boldsymbol{v}\boldsymbol{F}^T + \boldsymbol{\epsilon}, \tag{D.8}$$

where $\boldsymbol{\epsilon}$ is a T-element noise vector distributed as $\mathcal{N}(0, \sigma)$.

Now, we are ready to examine the update expressions. We note that we simply quote the results here. See Chapter 3 of [49] for further details. Let's first examine the posterior for the emission (slope) coefficients: $p(\boldsymbol{v}|\boldsymbol{y}, \boldsymbol{z}, \sigma)$. As stipulated above, we assume that $\boldsymbol{v}$ follows the multivariate normal distribution:

$$p(\boldsymbol{v}) \sim \mathcal{N}(\overline{\boldsymbol{v}}, \boldsymbol{\Sigma}), \tag{D.9}$$

such that we need to update both the mean vector ($\overline{\boldsymbol{v}}$) and covariance matrix ($\boldsymbol{\Sigma}$). To calculate these quantities, we need a helper matrix, $\boldsymbol{M}$, given by

$$\boldsymbol{M} = \boldsymbol{F}^T * \boldsymbol{F},$$

and a helper vector, $\boldsymbol{b}$, given by

$$\boldsymbol{b} = \boldsymbol{F}^T * \boldsymbol{y}.$$

Using these two quantities, we can calculate the update expression for $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \sigma^2\big(\boldsymbol{M} + \boldsymbol{M_0}^{-1}\big), \tag{D.10}$$

where $\boldsymbol{M_0}$ is our prior on $\boldsymbol{M}$, and is essentially an array of weights indicating confidence in these prior guesses regarding the values of the emission rate vector $\boldsymbol{v}$. In practice, we keep these weights small, such that the influence of $\boldsymbol{M_0}$ is negligible¿ A key exception to this rule is state 0, which we assume to be the OFF state. In this case, we set the first row and column of $\boldsymbol{M_0}$ to 10^5, effectively fixing $v(0) = 0$.

Now, for $\boldsymbol{v}$, we have that

$$\overline{\boldsymbol{v}} = \frac{\boldsymbol{\Sigma}}{\sigma^2} * \big(\boldsymbol{M_0}^{-1}\boldsymbol{v}_0 + \boldsymbol{b}\big), \tag{D.11}$$

where $\boldsymbol{v}_0$ prior on the emission rate vector.

Moving on to σ, recall that we have assumed that this parameter follows the inverse gamma distribution:

$$p(\sigma^2) \sim \text{Inverse-Gamma}(k, \theta), \tag{D.12}$$

where k is simply given by

$$k = k_0 + \frac{n}{2}, \tag{D.13}$$

with n being the number of observations in the sample used for inference. θ (or, rather, its inverse) is given by

$$\theta^{-1} = \theta_0^{-1} + \frac{1}{2}(\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}}) + \frac{1}{2}(\boldsymbol{v} - \boldsymbol{v}_0)^T \boldsymbol{M}_0^{-1}(\boldsymbol{v} - \boldsymbol{v}_0). \tag{D.14}$$