

Toward a Predictive Understanding of Developmental Decision-Making

by

Jiaxi Zhao

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Hernan G. Garcia, Chair

Professor Ahmet Yildiz

Professor Murat Arcak

Summer 2023

Toward a Predictive Understanding of Developmental Decision-Making

Copyright 2023

by

Jiaxi Zhao

Abstract

Toward a Predictive Understanding of Developmental Decision-Making

by

Jiaxi Zhao

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Hernan G. Garcia, Chair

The main objective of this thesis is to advance a quantitative understanding of the developmental genetic network. Three projects are presented toward this goal. The first project involves the quantitative dissection of transcription factor dynamics and gene regulation focusing on optogenetics to examine and manipulate transcription factors within a multicellular organism. This project provides a precise understanding of how transcription factors regulate target gene expression.

The second project explores the concept of bistability in autoactivation network motifs, focusing on how feedback loops contribute to binary cell fate decisions. Using a combination of mathematical modeling and experimental validation, this work elaborates on the dynamics governing cellular behavior and its implications for better understanding medical therapies related to cell development.

The final project investigates the flexible design principles governing the binding-site organization within enhancers. This entails unraveling the enhancer sequence's essential role in gene regulation, an insight that could inform the design of synthetic regulatory sequences for biotechnological applications and the development of novel therapeutic strategies.

In conclusion, this thesis provides insights into the developmental genetic network, contributing toward predicting complex biological processes based on specific genetic network parameters. This work may ultimately improve our understanding of organismal development and inform therapeutic strategies for developmental disorders.

To my parents, Jiaxu, with love to Cecilia and my twin boys Lucas, and Liam.

Contents

Contents	ii
List of Figures	v
1 Introduction	1
1.1 Multilayered regulation of a gene regulatory network	3
1.1.1 Individual regulatory arrows as the foundational building blocks of a gene regulatory network	3
1.1.2 Feedback motifs as enablers of complex regulatory dynamics	5
1.1.3 Enhancers as integrators of information from upstream factors	7
1.2 Overview of the PhD research projects	8
1.2.1 Quantitative dissection of transcription factor dynamics and gene regulation	9
1.2.2 Exploring bistability in autoactivation network motifs	10
1.2.3 Investigating design principles of enhancer binding-site organization	10
2 Optogenetic dissection of transcriptional repression in a multicellular organism	13
2.1 Introduction	13
2.2 Results	16
2.2.1 An optogenetics platform for dissecting single-cell repression dynamics in development	16
2.2.2 Repressor concentration dictates transcriptional activity through all-or-none response	17
2.2.3 Rapid export of repressor reveals fast, reversible reactivation kinetics at the single-cell level	21
2.2.4 Knirps acts by inhibiting the initiation of transcription bursts	22
2.3 Discussion	26
2.4 Materials and Methods	27
2.4.1 Cloning and Transgenesis	27
2.4.2 Fly lines	28
2.4.3 Embryo preparation and data collection	29

2.4.4	Image processing	29
2.4.5	Predicting Knirps binding sites	29
2.4.6	Compound-state Hidden Markov Model	29
2.5	Supplementary Figures	31
2.6	Supplementary Tables	46
2.7	Supplementary Text	47
2.7.1	Additional cpHMM inference results	47
2.7.2	Stochastic input-output simulations	48
2.7.3	Implementation of the thermodynamic binding model	55
2.7.4	Comparison to other optogenetic approaches developed for multicellular organisms	56
3	A bistable autoregulatory module in the developing embryo commits cells to binary expression fates	57
3.1	Introduction	58
3.2	Results	62
3.2.1	Binary Ftz expression states are established in early development	62
3.2.2	<i>ftz</i> autoactivation is triggered at a specific developmental time	63
3.2.3	Dynamical systems modeling of <i>ftz</i> regulation	67
3.2.4	Quantifying model parameters shows that the autoregulatory module is bistable	70
3.2.5	Ftz expression state is instructed by the early element rather than by the autoregulatory element	71
3.2.6	Ultimate Ftz fate is robustly specified in half an hour	71
3.3	Discussion	76
3.4	Materials and Methods	79
3.4.1	Fly strains/Genotypes	79
3.4.2	Fly lines	79
3.4.3	Method Details	80
3.4.4	Quantification and statistical analysis	94
3.5	Supplementary Figures	95
3.6	Supplemental Table	104
4	Enhancer design principles revealed by binding-site reorganization and modular dissection	105
4.1	Introduction	105
4.2	Results	107
4.2.1	Successful reconstitution of <i>even-skipped</i> (<i>eve</i>) stripe 2 enhancer	107
4.2.2	Systematic reorganization retains enhancer logic	109
4.2.3	Further minimization of <i>eve</i> stripe 2 enhancer	111
4.3	Discussion	113
4.4	Methods and Materials	114

4.4.1	Cloning and Transgenesis	114
4.4.2	Fly strains	115
4.4.3	Embryo preparation and data collection	115
Bibliography		116

List of Figures

1.1	Seeking a predictive understanding of developmental patterning	2
1.2	The two-state model of transcriptional bursting.	4
1.3	Example of the graphical analysis of stable points and temporal trajectory of protein expression from an autoregulation motif with bistability	6
1.4	Overview of my PhD works	12
2.1	Combining optogenetics and live imaging enables dissection of single-cell repression dynamics in a developing animal	15
2.2	Knirps concentration dictates sharp, switch-like repression	18
2.3	Knirps repression is rapidly reversible and memoryless	20
2.4	Knirps represses through instantaneous modulation of burst frequency	23
2.5	<i>even-skipped</i> expression under homozygous optogenetic Knirps (tagged with LEXY and LlamaTag) qualitatively recapitulates wild-type expression dynamics	31
2.6	Nuclei position calibration based on Knirps expression pattern	33
2.7	Spatiotemporal dynamics of Knirps protein and <i>eve</i> 4+6 transcription	34
2.8	Predicted Knirps binding sites in the <i>eve</i> 4+6 enhancer	35
2.9	Repressor titration results in distinct transcriptional dynamics	36
2.10	Example single-cell traces under no illumination	38
2.11	Response to Knirps perturbations is consistent across multiple embryos	39
2.12	Full cpHMM inference results of Knirps-regulated transcriptional bursting	40
2.13	A computational framework for Knirps-dependent stochastic simulations	41
2.14	Full MCMC results for stochastic input-output model parameters	43
2.15	Model predictions for <i>eve</i> 4+6 reactivation dynamics following Knirps export	44
2.16	Distinguishing homozygous from heterozygous embryos	45
3.1	The bistability of the autoregulatory module can determine cell fate	59
3.2	Binary cell states are rapidly established prior to gastrulation	61
3.3	A sequential transition in Ftz regulation from the early to the autoregulatory element occurs while discrete cell states are established	64
3.4	Two-color live imaging reveals that <i>ftz</i> autoregulation is initiated at a specific developmental time	65
3.5	Quantifying dynamical systems model parameters reveals that <i>ftz</i> autoregulation is bistable	68

3.6	Model reveals that the early element alone is sufficient to accurately predict binary Ftz expression states at gastrulation	72
3.7	<i>In silico</i> analysis reveals the window of commitment of cells to Ftz fate	74
3.8	<i>Ftz</i> regulatory region and binding sites	95
3.9	Comparison of temporal dynamics between autoregulatory element constructs driven by the <i>ftz</i> and <i>eve</i> promoters	96
3.10	Procedure for estimating c , the ratio of the maximum mRNA production rate from the endogenous to the transgene	97
3.11	Two ways to visualize a genetic network as an interconnection of modules	98
3.12	Detailed deterministic simulation results	99
3.13	Detailed stochastic simulation results	101
3.14	Classifying nuclei by the cumulative error in model prediction	103
4.1	Reconstituted <i>eve</i> stripe 2 enhancer recapitulates endogenous stripe features	108
4.2	Enhancer reorganization retains expression pattern	110

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Hernan G. Garcia, for his unwavering support, guidance, and mentorship throughout my doctoral journey. His passion for research, dedication to teaching, and commitment to my growth as a scholar have been truly inspiring. He gave me the complete freedom to explore my own scientific endeavor, and supported me with continued optimism, which turned out to be very fruitful. Without the kind of support he gave me, I don't think I would have been so productive and enjoyed the process so much. I deeply appreciate his support throughout my graduate study and look forward to what is to come.

I am incredibly grateful to my fellow lab mates, who have enriched my experience and contributed significantly to my academic development. I would like to extend a special thanks to Nick C. Lammers, whose collaboration has been nothing short of fantastic. Our shared intellectual pursuits have been vital in driving me forward in my research.

I would also like to acknowledge the outstanding contributions of Simon Alamos and Yang Joon Kim. Their mentorship and collaboration have been invaluable in helping me grow as a researcher and navigate the complex landscape of academia. Simon has been the perfect mentor who taught me everything, from how to design an experiment to making all these crazy transgenic flies with multiple transgenes. I am fortunate to have been able to work alongside such knowledgeable and supportive individuals.

To Mindy Perkins, I extend my heartfelt appreciation for sharing in the triumphs and challenges of this journey. Our shared experience of grappling with drama and relief has fostered a bond that transcends the walls of the lab. Your friendship and solidarity have been a source of strength and motivation throughout this endeavor.

I would also like to extend my deepest gratitude to my thesis committee members, Professor Ahmet Yildiz and Professor Murat Arcak, for their unwavering support, constructive feedback, and unyielding encouragement throughout the course of my thesis journey.

On a more personal note, I am forever grateful to my loving wife, Cecilia, for her unwavering support, patience, and understanding throughout this journey. Your love and encouragement have provided me with the strength and resilience to persevere during the most challenging times. I could not have done this without you.

To my twin boys, Lucas and Liam, thank you for being a constant source of joy and inspiration. Your laughter, curiosity, and boundless energy have reminded me of the importance of maintaining balance and cherishing life's precious moments. You are my greatest motivation to strive for excellence and make a positive impact on the world.

Finally, I would like to thank the entire research community for providing a stimulating and nurturing environment in which I could thrive. Your collective wisdom, encouragement, and support have played a crucial role in my personal and professional growth. I am forever grateful for the opportunity to have been a part of such a vibrant and dynamic community.

Chapter 1

Introduction

One of the holy grails of developmental biology is to predict the ultimate fate of cells within a multicellular organism and how these decisions dictate tissue morphology from knowledge of the underlying gene regulatory networks. It is crucial to understand how developing embryos generate precise spatial patterns and establish long-term developmental memory, which will eventually dictate cell fate decisions.

Gene regulatory networks serve as our primary conceptual framework for elucidating these processes (Figure 1.1A). It is essentially a collection of molecular interactions, showing how genes in an organism interact with each other. Specifically, the gene regulatory network catalogs the key participants, predominantly transcription factors, and maps out the regulatory interplay between them. For example, during the process of fruit fly early development (Figure 1.1A), a gene regulatory network can pinpoint which transcription factors are involved, how they interact, and how these interactions sculpt the segments of fruit fly embryos.

Ultimately, from a quantitative biology perspective, we aspire to employ gene regulatory networks predictively [64]. Given a gene regulatory network, our aim is to foretell the ultimate phenotype—including body patterns—under specific mutations or perturbations. To draw an analogy, our goal is to predict the physical traits of a fruit fly, like its body plan (Figure 1.1B) and morphological features (Figure 1.1C), such as wing and bristle shapes, based on certain alterations in its gene regulatory network.

However, the gene regulatory networks currently utilized in biology often fall short of accurately predicting the precise spatiotemporal dynamics that govern cell fate decisions in embryos [64]. The dominant approach within the field is the use of expansive network diagrams that is qualitative and lack quantitative information. The arrows that form the connections within these diagrams primarily serve a descriptive role, delineating types of regulation—activation or repression—but they do not contain a quantitative description, meaning they don't provide specific values or measurements regarding important quantitative features, such as the sharpness of regulation and its temporal kinetics.

Moreover, complexity arises when we consider that, within the context of development, gene regulation is mediated by regulatory sequences known as enhancers [2]. These enhancers

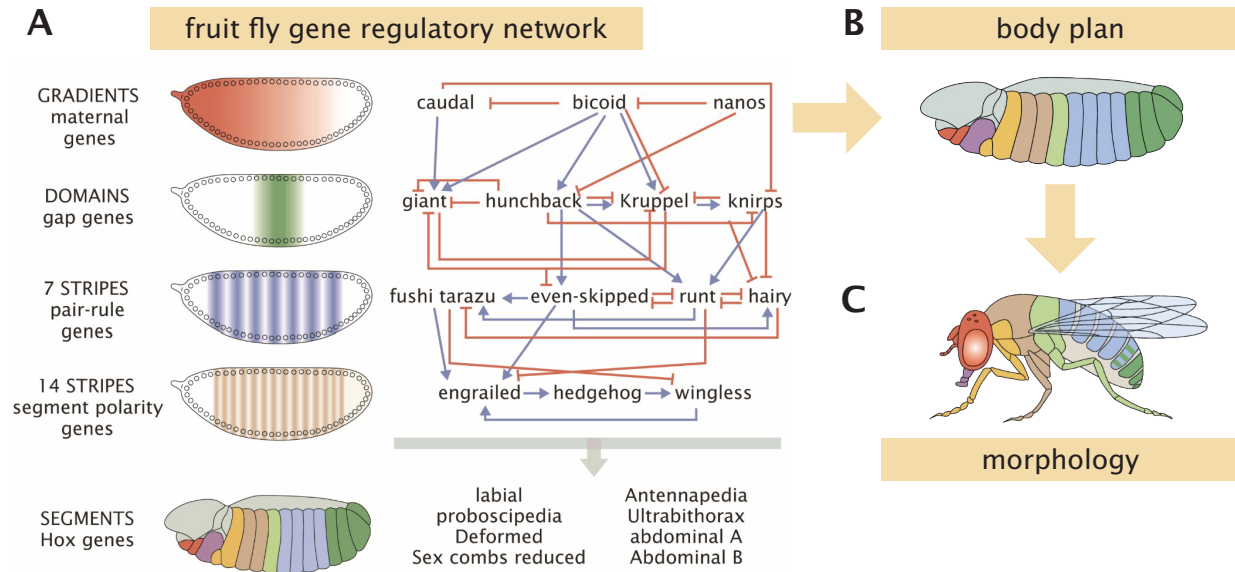


Figure 1.1: **Seeking a predictive understanding of developmental patterning.** (A) Gene regulatory network that drives segmentation in the fruit fly, *Drosophila melanogaster*. (B) The fruit fly body plan is the direct consequence of the body segmentation process, which is dictated by the fly regulatory network. (C) There is one-to-one correspondence between embryo body segments and adult morphology. (Adapted from [64])

act as nodes in the gene regulatory network and can integrate information from upstream transcription factors through binding sites it possesses. As a result, they integrate input from multiple transcription factors and determine whether a gene is transcribed or not [90]. This process can be quite complex, and it's hard to capture these nuances with standard, descriptive diagrams alone.

Finally, the structure of a gene network often deviates from a simple feedforward logic where the consequences could be logically deduced based on the input. For example, under this logic, if one gene activates the other gene, then when input is upregulated, it is easy to predict that there will be a reduced expression of the output. This simple logic does not apply to feedback loops, which many gene networks exhibit. When a network is under the control of feedback loops, concepts such as multistability may arise [3]. Multistability refers to a system that can exist in multiple stable states depending on initial conditions, which adds a layer of complexity to the logic of gene networks. Understanding this multifaceted behavior requires a mathematical approach to navigate the phase space of regulation, which we will discuss in the following sections.

The research projects during my PhD explore different facets of the gene regulatory network that we just elaborated on. Each project represents an effort to understand an important feature of the regulatory network, contributing to a quantitative understanding of

the developmental genetic network. Our ultimate goal, through this collective endeavor, is to advance toward a predictive understanding of developmental decision-making. We strive to move beyond simply identifying the components of these regulatory networks and towards forecasting the outcomes of complex biological processes given a specific genetic network. This predictive understanding could lead to significant advancements, potentially informing therapeutic strategies for developmental disorders and improving our general knowledge of organismal development.

1.1 Multilayered regulation of a gene regulatory network

A developmental gene regulatory network represents the complex relationship between vast factors involved during development. In this section, we discuss the three most important features of this regulatory network and how we can understand them that may contribute to a predictive understanding of the network as a whole.

1.1.1 Individual regulatory arrows as the foundational building blocks of a gene regulatory network

Gene regulatory networks consist of individual regulatory arrows, which symbolize the interactions between genes and transcription factors [3]. These arrows can portray either activation (where a transcription factor enhances the transcription rate of a gene) or repression (where it reduces the transcription rate). These regulatory arrows, representing the foundational connections within the network, dictate the overall topology and dynamics of gene regulation. However, our current understanding of these regulatory arrows is often limited to a descriptive level. We know which genes are activated or repressed, but the precise, quantitative details of these interactions remain largely elusive [64]. Particularly, we often lack specific information about the strength of an interaction, how quickly it takes effect, or how it changes under different cellular conditions [186]. This lack of quantitative data limits our ability to accurately predict the outcome of changes in the network features, which is a key goal in developmental biology.

The regulatory relationship between genes is often characterized by an input-output function that relates the concentration of a regulatory protein (the “input”) to the rate of transcription of the target gene (the “output”) [13]. This relationship is typically sigmoidal or S-shaped, meaning that small changes in the input concentration have little effect on the output when the input is low or high, but can cause substantial changes in the output at intermediate input levels. The sharpness or steepness of this sigmoidal response can be quantified by a parameter known as the Hill coefficient, with larger values indicating a more switch-like response.

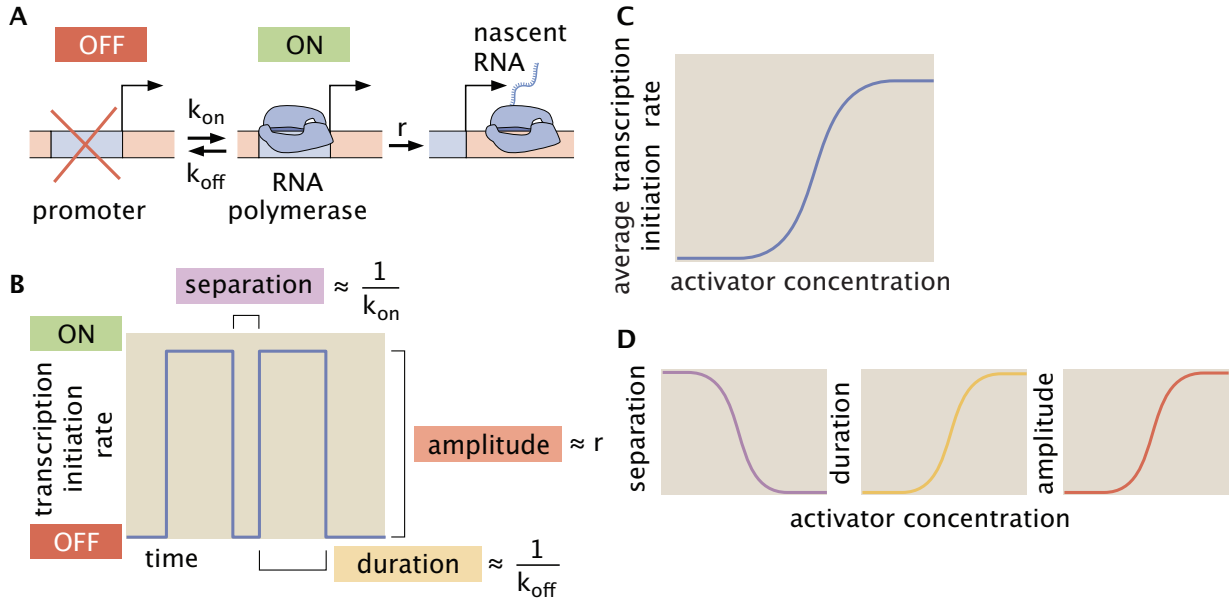


Figure 1.2: **The two-state model of transcriptional bursting.** (A) A two-state model of transcriptional bursting by a promoter switching between ON and OFF states. (B) Mapping the bursting parameters k_{on} , k_{off} , and r to burst duration, separation, and amplitude, respectively. (C) The action 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. Adapted from [101].

The temporal kinetics of gene regulation refer to the timescales over which these regulatory interactions occur. These kinetics can have profound effects on the dynamics of the gene regulatory network as a whole [101]. For instance, a regulatory interaction that takes place on a timescale of seconds will have a very different impact on the network than one that occurs over the course of an hour.

Therefore, when quantifying individual regulatory arrows in a gene network, it's crucial to account for both the shape of the input-output function (as characterized by the Hill coefficient) and the temporal kinetics of the regulatory interactions. This comprehensive approach will allow for a more accurate representation and understanding of the complex dynamics of gene regulatory networks.

Understanding the regulatory relationships and temporal kinetics of gene expression is critical, but it's equally important to consider the molecular mechanisms of transcriptional regulation when constructing mathematical models of these processes [101]. While the molecular mechanisms governing activators and repressors differ significantly, a complete picture of these mechanisms may not be necessary for quantitative descriptions of these processes.

Instead, a mathematical model capturing the key features of eukaryotic gene expression suffices, and recent research has identified a commonality among transcriptional regulations in eukaryotes, namely transcriptional bursting.

Transcriptional bursting is a stochastic phenomenon in which gene expression occurs in short, intense bursts rather than at a steady, constant rate [101, 165]. This leads to significant variability or “noise” in gene expression levels across genetically identical cells in identical environments. To describe this, a two-state model can be used, where the promoter stochastically switches between active (ON) and inactive (OFF) states, as shown in Figure 1.2A. The dynamics of this model can be defined by three key parameters, shown in Figure 1.2B:

1. **Burst frequency:** indicating the frequency of transcription bursts; higher frequency means more frequent active transcription.
2. **Burst duration:** representing the duration at which active transcription occurs.
3. **Burst size:** reflecting the number of mRNA molecules produced during a burst, with larger burst sizes indicating higher mRNA production during each active phase.

Even though transcriptional bursting seems random, it’s regulated [101]. For instance, transcriptional activation can modulate the bursting parameters to fine-tune gene expression levels (Figure 1.2C). This modulation can alter the burst frequency, duration, and amplitude (Figure 1.2D). By using the bursting framework, we can quantitatively describe the transcriptional process, enabling the construction of more accurate mathematical models for the regulatory arrows, the fundamental building blocks of developmental networks.

One of the primary goals of my work is to quantify the regulatory function and kinetics of individual regulatory arrows and to apply the bursting framework to model their behavior. This will ensure that the essential features of the network are captured while maintaining a level of detail that accurately reflects the complex dynamics of gene expression and regulation, such as stochasticity. By employing this approach, we can provide a comprehensive and reliable representation of regulatory processes, enhancing our understanding of developmental networks and enabling us to predict and manipulate their behavior more effectively.

1.1.2 Feedback motifs as enablers of complex regulatory dynamics

Feedback motifs, in which a gene product regulates its own production, are fundamental components of gene regulatory networks [3]. They introduce an added layer of complexity to these networks by fostering intricate regulatory dynamics. These feedback loops may be positive, thereby upregulating the target gene, or negative, thereby downregulating it. Through these mechanisms, gene networks can manifest more intricate behaviors, such as the oscillations observed in circadian rhythms or the multistable states associated with cellular differentiation processes. Consequently, understanding these motifs provides valuable

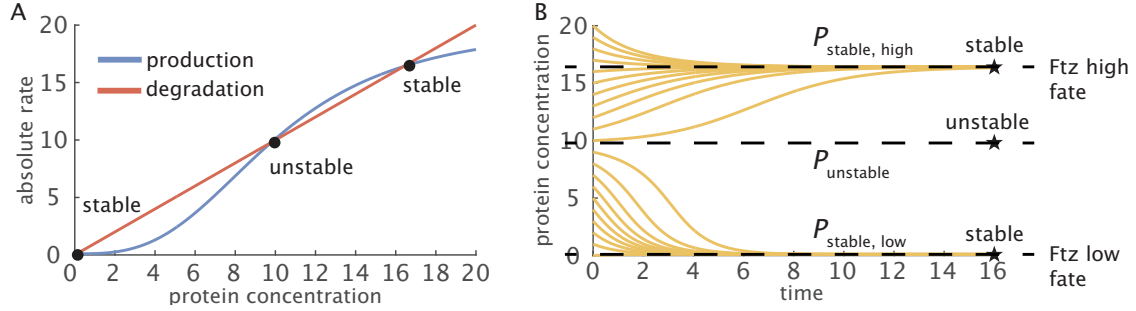


Figure 1.3: **Example of the graphical analysis of stable points and temporal trajectory of protein expression from an autoregulation motif with bistability.** (A) Graphical analysis of the fixed points. The system has two stable fixed points (attractors) and one unstable fixed point. (B) Temporal trajectories with given initial protein P . Bistability guarantees that starting low $P < P_{\text{unstable}}$ leads to the low attractor $P_{\text{stable, low}}$ and starting high $P > P_{\text{unstable}}$ leads to the high attractor P_{high} .

insights into the dynamics of biological systems and establishes a foundation for predicting how system alterations may influence its behavior.

The term ‘multistability’ refers to a system’s capacity to attain various stable states based on its initial conditions [106]. Feedback motifs enable this phenomenon, adding another layer of complexity to the logic underpinning gene regulatory networks. However, feedback motifs often exhibit complex behaviors that are challenging to analyze without appropriate analytical methods. As such, mathematical tools are frequently employed to navigate this complexity and understand the system’s potential outcomes, also known as its phase space.

While feedback has been proposed to contribute to discrete cell fate decisions through multistability, this is not always guaranteed [185]. The quantitative characteristics of the feedback loop, such as the sharpness of regulation and protein half-life, significantly influence this outcome [185]. For instance, it is posited that positive autoregulation could lead to bistability, resulting in binary cell fates [3]. However, this is not an absolute certainty; a distinct regulatory relationship or protein half-life could yield only monostability. Therefore, a careful quantitative examination of these feedback motifs is crucial in accurately predicting and understanding their influence on cell fate decisions.

To further understand the intricacies of feedback motifs and their potential effects on cellular processes, we need to rely on mathematical models and experimental validation [3]. Mathematical models, such as ordinary differential equations (ODEs), can help to predict the behavior of feedback motifs under various conditions. These models enable us to simulate the dynamics of the system over time, providing an in-depth view of how changes in parameters might affect the behavior of the network.

However, the usefulness of mathematical models depends heavily on the validity of their underlying assumptions and the accuracy of the parameter estimates used. While many of these parameters can be estimated from experimental data, uncertainties remain due to

factors such as measurement error and biological variability. Therefore, predictions made from these models must be experimentally validated to confirm their accuracy [64].

Experimental validation usually involves perturbing the system, for example, by genetically altering the feedback loop and observing the resulting changes in gene expression or cellular behavior. Techniques such as fluorescence microscopy and flow cytometry, combined with reporters like green fluorescent protein (GFP), are often used to measure these changes. This data then provides a means to confirm or challenge the predictions of the mathematical model.

In the context of feedback motifs and multistability, for instance, the system can be perturbed by altering the levels of the gene product involved in the feedback loop. If the system displays bistability, such a perturbation could lead to a shift from one stable state to another. Observing this behavior in a controlled experiment could provide strong evidence supporting the role of feedback in generating multistability.

To summarize, understanding the role of feedback motifs in gene regulatory networks requires a combination of mathematical modeling and experimental validation. By combining these approaches, we can achieve a more comprehensive understanding of the complex dynamics that govern cellular behavior, enabling more precise predictions and offering new avenues for therapeutic intervention in diseases where these dynamics are disrupted.

1.1.3 Enhancers as integrators of information from upstream factors

Enhancers, defined as short DNA segments that can bind transcription factors, hold pivotal roles in orchestrating gene expression [2]. Unconstrained by orientation, distance, or spatial relation to their target gene, enhancers act as hubs for assimilating information from various upstream factors. They translate this multi-dimensional input into a straightforward decision: whether or not to initiate transcription of the target gene, and if so, at what rate. Consequently, enhancers form critical regulatory modules, enabling intricate processing of upstream inputs, thereby facilitating robust developmental patterning in multicellular organisms.

Presently, our understanding posits that enhancers harbor multiple binding sites for various transcription factors. These transcription factors often exhibit spatially distinct expression patterns. The interplay between enhancers and transcription factors is both complex and dynamic; a single enhancer can bind multiple transcription factors, and their collective action governs the eventual outcome of gene transcription. This interplay between enhancers and transcription factors allows for fine-tuned control of gene expression, enabling cells to respond appropriately to various internal and external cues.

The enhancer's DNA sequence holds nearly all the information necessary for creating the desired expression pattern. Upon introducing the enhancer into different genetic locations, they result in identical expression patterns, suggesting that the enhancer sequence alone can theoretically dictate the input-output relationship. Predicting this relationship from the

enhancer sequence forms a key goal within enhancer biology. Nevertheless, this remains a formidable task due to the combinatorial explosion problem: the potential arrangements of binding sites within an enhancer approach infinity [90, 162]. Thus, deciphering simplified governing rules of enhancer organization becomes imperative to circumvent this combinatorial complexity.

The flexibility in the organization of enhancers and how variations in the arrangement of binding sites affect their functionality is a fascinating topic. There are two primary theories that seek to explain this [6, 156]. The first, known as the ‘Enhanceosome’ model, stems from research on the interferon-beta enhanceosome [127]. It proposes that an enhancer only operates efficiently when every component is in a precise configuration. On the other hand, the ‘Billboard’ model believes that the function of an enhancer isn’t tied to a specific arrangement of binding sites, but rather to the mere presence of all necessary elements.

Although the enhanceosome and billboard models appear to be at odds, they might not actually be contradictory. Instead, they could symbolize two ends of a continuum regarding enhancer organization [156]. Some enhancers might necessitate a strict setup of transcription factor binding sites, as the enhanceosome model suggests. In contrast, others might operate efficiently if all essential elements are there, a notion backed by the billboard model. It’s also possible for an enhancer’s position on this continuum to vary based on cellular circumstances and developmental phases, introducing even more intricacy into gene expression regulation.

This flexibility and adaptability of enhancer organization could be key to the ability of multicellular organisms to respond to changes in their environment and to the different needs of their various cell types. It might also play a role in evolution, allowing organisms to experiment with different configurations of enhancer elements and select the ones that provide a survival advantage.

Understanding the rules that govern enhancer function is not only of fundamental biological interest but also has potential applications in biotechnology and medicine. For instance, being able to predict the output of a given enhancer from its sequence would be invaluable in the design of synthetic gene circuits for biotechnological applications. Moreover, since misregulation of gene expression is a common feature of many diseases, including cancer [81], a deeper understanding of enhancer function might help in the development of novel therapeutic strategies.

In conclusion, enhancers are a fascinating topic of research at the interface of biology, chemistry, and physics. Their study is likely to provide new insights into the principles of life and may have wide-ranging implications for biotechnology and medicine.

1.2 Overview of the PhD research projects

The overarching objective of my doctoral research projects is to infuse developmental biology with the accuracy and predictability of quantitative sciences. This aim was pursued through a three-fold approach: first, by employing live imaging, optogenetics, and theoretical modeling, we were able to assign specific quantitative and molecular information to individual

regulatory pathways [186]. Second, using the dynamical systems theory, we investigated a positive feedback motif in development, leading us to show that bistability can facilitate cellular memory [185]. Finally, we unraveled the surprisingly flexible design principles that govern the binding-site organization within enhancers through synthetic reconstitution.

1.2.1 Quantitative dissection of transcription factor dynamics and gene regulation

My initial major PhD project, documented in [chapter 2](#), delves into the mechanisms through which transcription factors influence gene expression in individual cells over time, eventually driving cell-fate decisions in a quantitative manner ([Figure 1.4C](#), 1). Despite its importance for a predictive understanding of developmental decision-making, our knowledge of how transcription factors regulate target gene expression has remained elusive, primarily due to the lack of suitable tools to examine and manipulate transcription factors within their endogenous context in multicellular organisms.

Traditional methods such as fixation and sequencing-based methods, although powerful, fall short in investigating the rapid, single-cell temporal dynamics of transcriptional control characteristic of embryonic development [88, 53, 132, 109, 40, 174, 117, 152, 27, 131]. Therefore, we need innovative experimental methods that allow us to quantify and control gene expression dynamics in multicellular organisms while concurrently assessing output transcriptional activity at the single-cell level.

Optogenetics has shown promise in this regard. However, prior optogenetic methods have primarily concentrated on modulating transcription factor activity rather than concentration, creating a challenge in assessing active transcription factor dynamics.

To overcome this, I developed an experimental framework that combines acute optogenetic manipulation of transcription factor dynamics with concurrent measurement of transcriptional response in a developing fruit fly embryo. Our findings demonstrate that a transcriptional repressor induces abrupt, switch-like transitions into enduring repressive states, which can be rapidly reversed within 1-2 minutes.

As we move forward, it will be essential to validate and refine our mathematical models with experimental data from various biological contexts. Transcriptional bursting parameters can be measured and analyzed in various biological contexts, including different cell types, developmental stages, and environmental conditions. As our models become more precise and robust, they will serve as powerful tools for deciphering the intricate language of gene regulatory networks.

Our approach is a significant stride towards a more comprehensive and quantitative understanding of gene regulatory networks. It enables us not only to dissect gene regulation at the individual pathway level but also to connect these components, thereby offering a holistic view of the network's function and its role in development.

1.2.2 Exploring bistability in autoactivation network motifs

In the second project of my doctoral studies, as presented in [chapter 3](#), I explored the mechanisms through which autoactivation network motifs lead to the establishment of discrete cellular identities ([Figure 1.4C](#), 2). Over many years, the Waddington landscape concept [171] has provided developmental biologists with a theoretical framework to comprehend how cells, like balls rolling down valleys, adopt their definitive fates.

This notion, proposed by Waddington, has frequently been mathematically represented using dynamical systems theory [171]. Within this context, a transcription activator regulating its own expression through positive feedback could potentially create a bistable state. This bistability would theoretically enable cells to adopt one of two distinct fates permanently. The initial activator production is presumed to be controlled by upstream regulatory interactions. Once this production reaches a critical threshold, it triggers the feedback loop, thereby “locking” cells into a high expression state.

Despite the prevalence of the assumption that autoactivation could serve as the foundation for this type of developmental memory, and despite the widespread occurrence of positive feedback loops in gene regulatory networks, positive autoregulation alone doesn’t necessarily lead to bistability. Remarkably, no empirical measurements have so far confirmed the bistability of any autoactivation motifs within the context of a living organism undergoing development.

To address this knowledge gap, in this project, we merged quantitative mathematical modeling with *in vivo* live imaging. This approach allowed us, to the best of our knowledge, to be the first to quantitatively characterize positive autoregulation in single cells while they commit to their ultimate fate within a developing animal. For our case study, we chose the binary pattern (ON or OFF) of the *fushi tarazu* (*ftz*) gene in the fruit fly *Drosophila melanogaster*.

Ftz gene expression is governed by two primary regulatory sequences: an early element, which is responsive to upstream transcription factors and establishes its initial seven-stripe expression pattern [45], and an autoregulatory element that responds to Ftz protein to stimulate its own expression [138, 146, 145]. By distinguishing the initial Ftz expression from its autoregulatory activity, we discovered that *ftz* autoactivation is initiated at a specific developmental timepoint, independent of the Ftz levels driven by the early element. We further elucidated that cells attain binary Ftz states as *ftz* expression swiftly transitions from being regulated by the early element to the autoregulatory element.

1.2.3 Investigating design principles of enhancer binding-site organization

My third project, described in [chapter 4](#), focuses on design principles governing the binding-site organization within developmental enhancers, which contain transcription factor binding sites that determine when, where, and how cell fates are decided ([Figure 1.4C](#), 3).

To probe this question, I first tackled a long-standing challenge in the field of enhancer biology—reconstituting functional developmental enhancers from known binding sites. I was the first to successfully reconstitute the *even-skipped* (*eve*) stripe 2 enhancer—one of the best-studied enhancers in developmental biology [151, 9]—a feat that multiple labs have attempted and failed at during the past two decades. The reconstituted enhancer recaptured the iconic stripe pattern and enabled the complete reorganization of its DNA sequence, which led to a surprising discovery—the enhancer function may be preserved even under a complete random reorganization of its binding sites. My work demonstrated the extraordinary flexibility of enhancer organization, which is the crucial step toward the rational design of functional regulatory sequences.

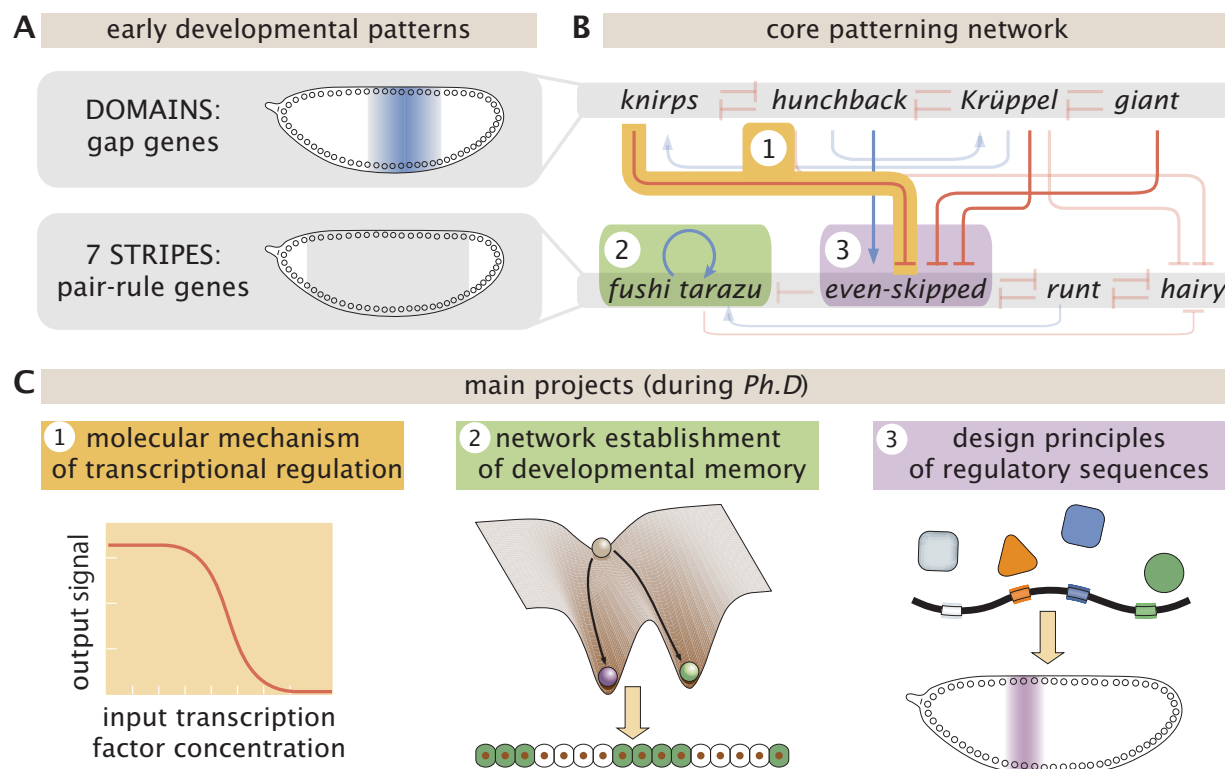


Figure 1.4: **Overview of my PhD works.** (A) My PhD works focus on gap gene and pair-rule gene regulation during fruit fly development. Gap genes give rise to domain-like patterns, while pair-rule genes are expressed in seven-stripe patterns during early fruit fly development. (B) Part of the gene regulatory network of fruit fly early development, describing interactions between gap genes and pair-rule genes. (C) (1) Quantitative information about dynamics, space and concentrations are crucial yet often absent in the network diagrams. In this project, we quantified the input-output relationship and regulation kinetics to reveal the molecular mechanism of transcriptional regulation so that quantitative information is assigned to the network arrow. (2) Feedback network has been proposed to contribute to developmental fates. We pursue a predictive understanding of cell fate decisions through the dynamical systems analysis of developmental networks. (3) How flexible binding site organization within the enhancer has been a key question in the field. We used synthetic reconstitution and rational reorganization and discovered that enhancer organization could be extremely flexible.

Chapter 2

Optogenetic dissection of transcriptional repression in a multicellular organism

This chapter is a reproduction of reference [186]

Abstract

Transcriptional control is fundamental to cellular function. However, despite knowing that transcription factors can repress or activate specific genes, how these functions are implemented at the molecular level has remained elusive, particularly in the endogenous context of developing animals. Here we combine optogenetics, single-cell live-imaging, and mathematical modeling to study how a zinc-finger repressor, Knirps, induces switch-like transitions into long-lived quiescent states. Using optogenetics, we demonstrate that repression is rapidly reversible (~ 1 minute) and memoryless. Furthermore, we show that the repressor acts by decreasing the frequency of transcriptional bursts in a manner consistent with an equilibrium binding model. Our results provide a quantitative framework for dissecting the *in vivo* biochemistry of eukaryotic transcriptional regulation.

2.1 Introduction

Throughout biology, transcription factors dictate gene expression and, ultimately, drive cell-fate decisions that play fundamental roles in development [156], immune responses [84], and disease [22]. Achieving a quantitative and predictive understanding of how this process unfolds over time and space holds the potential both to shed light on the molecular mechanisms that drive cellular decision-making and to lay the foundation for a broad array of bioengineering applications, including the synthetic manipulation of developmental processes [62, 49, 64, 147, 120] and the development of therapeutics [20].

In recent years, great progress has been made in uncovering the molecular mechanism of transcription factor action through cell culture-based methods thanks to the emergence of a wide array of imaging techniques that can query the inner workings of cells in real time, often at the single molecule level (see, for example, [43, 103, 33, 14, 107, 112, 26, 25, 143]). However, progress has been slower in multicellular organisms, where a lack of comparable tools has limited our ability to query the dynamics of transcription factor function in their endogenous context. Indeed, while fixation-based methods, such as immunofluorescence staining, mRNA FISH, and various sequencing-based techniques represent powerful tools for investigating cellular decision-making in animals [88, 53, 132, 109, 40, 174, 117, 152, 27, 131], these methods are mostly silent regarding the single-cell and single-gene dynamics of transcriptional control.

To move beyond these limitations, new experimental techniques are needed that provide the ability to quantify and manipulate input transcription factor concentrations *over time* in multicellular organisms while simultaneously measuring output transcriptional activity. Recently, we and others have developed new technologies to realize this goal through new molecular probes that allow for the direct measurement of protein [18], and transcriptional dynamics [63, 114] in single cells of living multicellular organisms, as well as optogenetic techniques for the light-based modulation of nuclear protein concentration *in vivo* [125, 98].

Here, we combine these technologies to study causal connections between the molecular players that underpin transcriptional control, shedding new light on the molecular basis of transcriptional repression in a developing animal. We use this platform to answer two key questions regarding the kinetic properties of repression. First, despite several studies dissecting repressor action at the bulk level [177, 16, 109, 144, 75], it is not clear whether this repression is implemented in a graded or switch-like fashion at individual gene loci over time (Figure 2.1A, left). Second, the adoption of cellular fates—often dictated by repressors—has been attributed to the irreversible establishment of transcriptional states [104, 159, 84]. However, it is unclear whether the action of repressors is itself reversible—such that sustained repressor binding is required to maintain gene loci in inactive transcriptional states—or if, instead, repression is *irreversible*—such that even transient exposure to high repressor concentrations is sufficient to induce long-lived transcriptional inactivity as a result of, for example, the accumulation of chromatin modifications. To answer these questions, in this work, we examine how the zinc-finger repressor Knirps drives the formation of stripes 4 and 6 of the widely studied *even-skipped* (*eve*) pattern during the early development of the fruit fly *Drosophila melanogaster* (Figure 2.1B) [56, 110, 11].

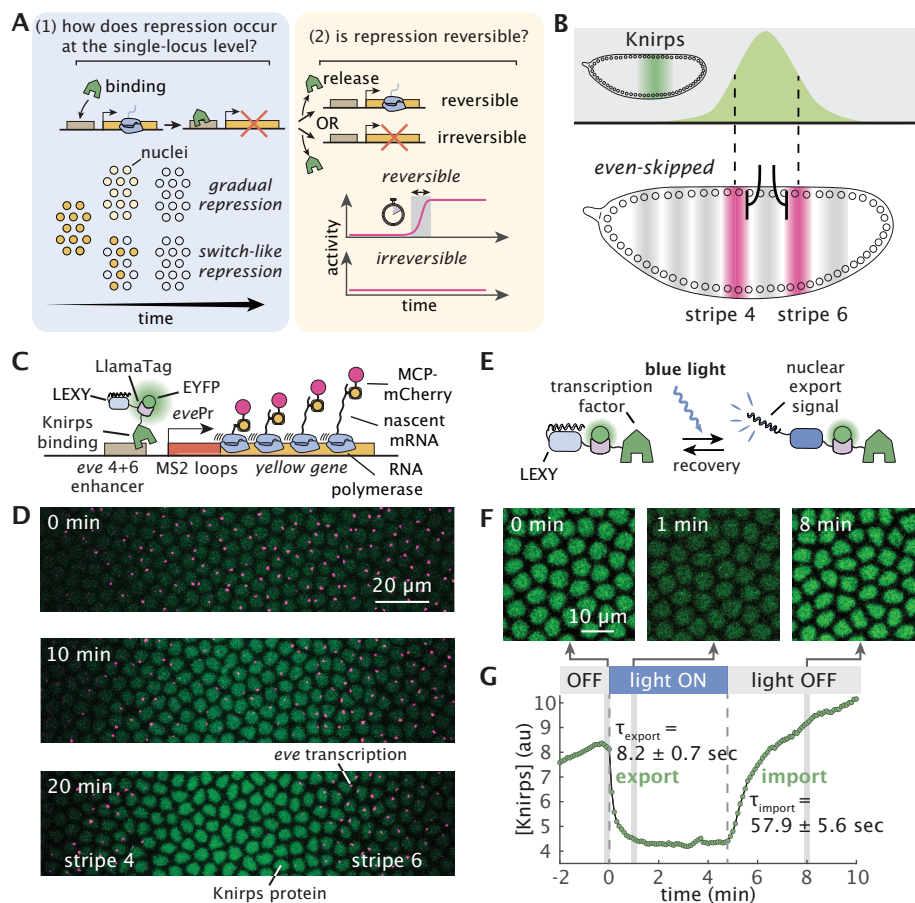


Figure 2.1: **Combining optogenetics and live imaging enables dissection of single-cell repression dynamics in a developing animal.** (A) Key questions regarding transcriptional repression. Left: Whether single-cell repression occurs in a gradual or switch-like fashion over time. Right: Whether repression is reversible. (B) Knirps represses *even-skipped* (*eve*) stripes 4+6 transcription in the fruit fly embryo. Top: Knirps is expressed in a bell-shaped domain during early embryogenesis. Bottom: Knirps specifies the position and sharpness of the inner boundaries of *eve* stripes 4 and 6. (C) Two-color tagging permits the simultaneous visualization of input protein concentration and output transcriptional dynamics *in vivo*. Maternally deposited EYFP molecules bind to Knirps-LlamaTag, resulting in increased nuclear fluorescence, which provides a real-time readout of the nuclear protein concentration. Maternally deposited MS2 coat protein (MCP) binds to MS2 stem-loops in the nascent RNA formed by RNAP molecules elongating along the body of the *eve* 4+6 reporter construct leading to the accumulation of fluorescence at sites of nascent transcript formation. LEXY tag is also fused to Knirps to allow for optogenetic manipulation of its nuclear concentration. (continued on the next page)

Figure 2.1: *(continued)* **Combining optogenetics and live imaging enables dissection of single-cell repression dynamics in a developing animal.** **(D)** Representative frames from live-imaging data. The embryo is oriented with the anterior (head) to the left. Green and magenta channels correspond to Knirps repressor and *eve* 4+6 transcription, respectively. When Knirps concentration is low, *eve* stripe 4+6 is expressed in a broad domain, which refines into two flanking stripes as Knirps concentration increases. **(E)** Optogenetic control of nuclear protein export. Upon exposure to blue light, the nuclear export signal within the LEXY domain is revealed. As a result, the fusion protein is exported from the nucleus. **(F)** Fluorescence images of embryos expressing the Knirps-LEXY fusion undergoing an export-recovery cycle. **(G)** Relative nuclear fluorescence of the repressor protein over time ($n = 55$ nuclei). Half-times for export and recovery processes are estimated by fitting the fluorescence traces with exponential functions.

2.2 Results

2.2.1 An optogenetics platform for dissecting single-cell repression dynamics in development

To measure Knirps protein concentration dynamics, we labeled the endogenous *knirps* locus with a LlamaTag, a fluorescent probe capable of reporting on protein concentration dynamics faster than the maturation time of more common fluorescent protein fusions [18]. Further, we quantified the target transcriptional response using a reporter construct of the *eve* stripe 4+6 enhancer [56], where the nascent RNA molecules are fluorescently labeled using the MCP-MS2 system [12, 63, 114] (Figure 2.1C). The resulting nuclear fluorescence and transcriptional puncta provide a direct readout of input Knirps concentration and output *eve* 4+6 transcription, respectively, as a function of space and time. Our data recapitulate classic results from fixed embryos [148] in dynamical detail: gene expression begins in a domain that spans stripes 4 through 6, subsequently refined by the appearance of the Knirps repressor in the interstripe region.

To enable the precise temporal control of Knirps concentration, we attached the optogenetic LEXY domain [125] to the endogenous *knirps* locus in addition to the LlamaTag (Figure 2.1C). Upon exposure to blue light, the LEXY domain undergoes a conformation change which results in the rapid export of Knirps protein from the nucleus (Figure 2.1E). Export-recovery experiments revealed that export dynamics are fast, with a half-time <10 seconds, while import dynamics are somewhat slower, with a half-time ~60 seconds upon removal of illumination (Figure 2.1F and G). These time scales are much faster than typical developmental time scales [55], allowing us to disentangle rapid effects due to direct regulatory interactions between Knirps and *eve* 4+6 from slower, indirect regulation that is mediated by other genes in the regulatory network. We established stable breeding lines of homozygous optogenetic Knirps flies, demonstrating that the protein tagged with both LEXY and LlamaTag

is homozygous viable. Furthermore, our optogenetic Knirps drives comparable levels of *eve* 4+6 than wild-type Knirps (Figure 2.5). Thus, we conclude that our optogenetics-based approach represents an ideal platform for manipulating transcriptional systems to probe the molecular basis of gene regulatory control without significantly affecting the broader regulatory network and the developmental outcome this network encodes for.

2.2.2 Repressor concentration dictates transcriptional activity through all-or-none response

To understand how Knirps repressor regulates *eve* 4+6 expression, we first analyzed the temporal dynamics of Knirps-LlamaTag-LEXY (hereafter referred to simply as “Knirps”) concentration and *eve* 4+6 expression in the absence of optogenetic perturbations. We generated spatiotemporal maps of input repressor concentration and output transcription by spatially aligning individual embryos according to the peak of the Knirps expression domain along the anterior-posterior axis (Figure 2.6; Figure 2.7). These maps reveal a clear pattern: rising repressor concentrations coincide with a sharp decline in *eve* 4+6 activity at the center of the Knirps domain. Focusing on the central region of the Knirps domain (-2% to 2% of the embryo length with respect to the center of the domain), we observe a clear anti-correlation between Knirps concentration—which increases steadily with time—and the mean transcription rate, which drops precipitously between 10 and 20 minutes into nuclear cycle 14 (Figure 2.2A).

We quantified the regulatory relationship implied by these trends by calculating the Knirps vs. *eve* 4+6 “input-output function”, which reports on the average transcription rate as a function of nuclear repressor concentration (inset panel in Figure 2.2A). This revealed a sharp decline in transcriptional activity across a narrow band of Knirps concentrations, suggesting that *eve* 4+6 loci are highly sensitive to nuclear repressor levels. This finding is consistent with previous observations that Knirps represses *eve* 4+6 [57], and with the discovery of multiple Knirps binding sites in the *eve* 4+6 enhancer region (Figure 2.8) [32]. However, neither our endogenous measurements nor these previous studies can rule out the possibility that other repressors might also play a role in driving the progressive repression of *eve* 4+6 over the course of nuclear cycle 14. Indeed, by themselves, neither live imaging experiments (which are constrained to observing wild-type trends) nor classical mutation-based studies (which are subject to feedback encoded by the underlying gene regulatory network) can rule out the presence of other inputs.

Our optogenetics approach allows us to circumvent these limitations and search for regulatory inputs that impact *eve* 4+6 expression, but are not directly observed in our experiments. Specifically, we used optogenetics to alter Knirps concentration dynamics over the course of nuclear cycle 14. Shortly after the beginning of the nuclear cycle, we exposed embryos to low and high blue light illumination, inducing moderate and strong reductions in nuclear Knirps concentration, respectively, which resulted in distinct transcriptional trends (Figure 2.2B; Figure 2.9). We reasoned that, because we are only altering Knirps con-

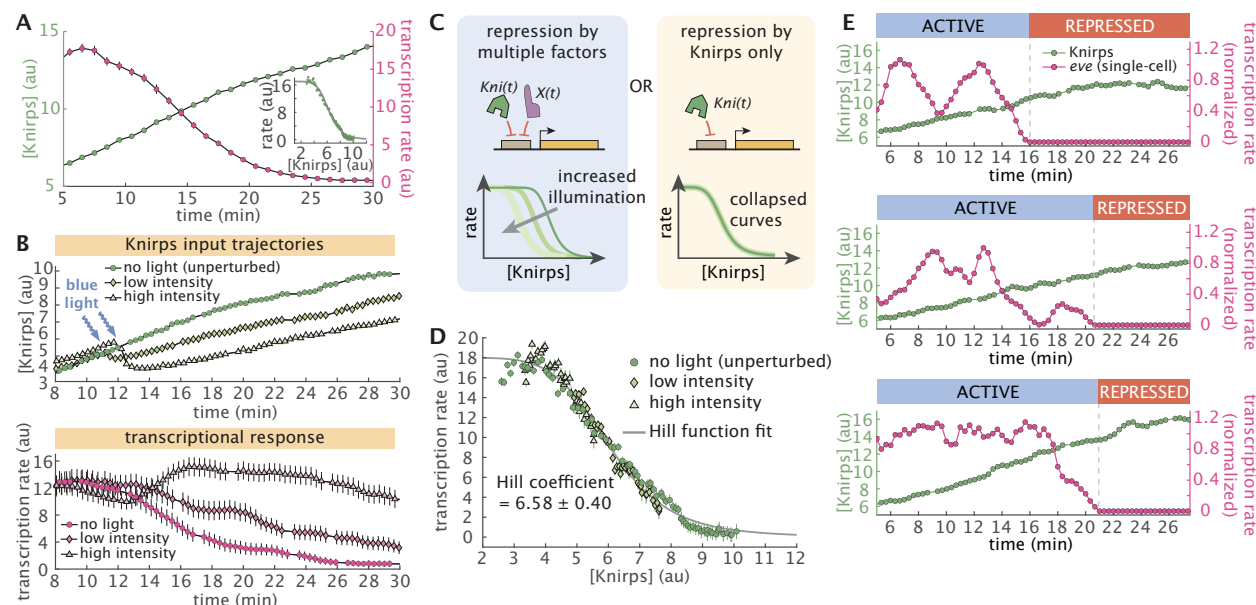


Figure 2.2: Knirps concentration dictates sharp, switch-like repression. (A) Average Knirps concentration (green) and *eve* 4+6 transcription (magenta) over time shows a clear anticorrelation. These dynamics are calculated by averaging the traces over a window of -2% to 2% along the anterior-posterior axis of the embryo and centered around the peak of the Knirps pattern (see Figure 2.6). Target transcription declines sharply as Knirps concentration increases. Inset panel shows the input-output relationship under this no light (unperturbed) condition. (B) Optogenetics allows for titration of protein concentration. Top panel shows the average Knirps concentration for three embryos, each under different illumination intensities. Bottom panel shows the corresponding trends in the *eve* 4+6 transcription rate. The illumination started around 12 minutes into nuclear cycle 14 and continued throughout the experiment. (C) To test whether Knirps is the only repressor whose concentration changes in the system, input-output functions under different illumination conditions can be compared. If there are multiple potentially unknown repressors at play (e.g. the X transcription factor in the figure), then each illumination level should lead to a different input-output function (left). However, if Knirps is the sole repressor, the input-output functions for each condition should collapse onto a single curve (right). (*continued on the next page*)

Figure 2.2: *(continued)* **Knirps concentration dictates sharp, switch-like repression.** **(D)** Average transcription rate as a function of Knirps concentration for each illumination condition (averaged over a window of -2% to 2% along the anterior-posterior axis). All three conditions follow the same trend, suggesting that Knirps is the only repressor regulating target transcription during this developmental stage. The input-output relationship is fitted with a Hill function resulting in a Hill coefficient of 6.36 (95% CI [6.08, 6.64]). (Averaged over $n = 4$ for no light, $n = 4$ for low intensity and $n = 3$ for high intensity embryos.) **(E)** Illustrative single-cell Knirps (green points) and transcriptional dynamics (magenta points) show that repression is switch-like at the single-cell level. Traces are normalized by their maximum transcription rate and smoothened using a moving average of 1 minute. (Error bars in A, B, and D indicate the bootstrap estimate of the standard error. $t = 0$ is defined as the onset of transcription in nuclear cycle 14. Transcription rate reflects the measured MS2 signal, which is an approximation of the *eve* mRNA production rate [63, 102, 18].)

centration dynamics, the presence of other repressors dictating *eve* 4+6 activity together with Knirps should lead to distinct input-output curves across these different illumination conditions (Figure 2.2C, left). Conversely, if Knirps is the sole repressor driving the repression of *eve* 4+6 over time, the transcriptional input-output function should be invariant to perturbations of Knirps concentration dynamics (Figure 2.2C, right).

Comparing the *eve* 4+6 vs. Knirps input-output function for the unperturbed control (inset panel of Figure 2.2A) to that of optogenetically perturbed embryos (Figure 2.2D), we find that all three conditions collapse onto a single input-output curve, providing strong evidence that Knirps is the sole repressor of *eve* 4+6. Moreover, as noted above, we find that Knirps repression occurs in a sharp fashion: *eve* 4+6 loci transition from being mostly active to mostly repressed within a narrow band of Knirps concentrations. To quantify this sharp response, we fit a Hill function to the data in Figure 2.2D (gray line), which yielded a Hill coefficient of 6.58 ± 0.40 . Notably, this is comparable to Hill coefficients estimated for the Bicoid-dependent activation of *hunchback* [71, 180, 131]; another canonical example of sharp gene regulation—in this case, of activation—during developmental patterning which relies on the presence of multiple binding sites for the transcription factor within the enhancer.

The input-output function in Figure 2.2D summarizes the average effect of repressor level on *eve* 4+6 expression, but it cannot alone shed light on *how* this effect is achieved in individual cells. Thus, we next investigated how this sharp *average* decrease in gene expression is realized at the single-cell level. We examined single-cell trajectories of Knirps repressor and corresponding *eve* 4+6 transcription. This revealed that the sharp population-level input-output function illustrated in Figure 2.2D is realized in an all-or-none fashion at the level of individual cells (Figure 2.2E; Figure 2.10). During this process, the gradual rise in Knirps concentration induces an abrupt, seemingly irreversible, transition from active transcription to a long-lived (or even permanent), transcriptionally quiescent state.

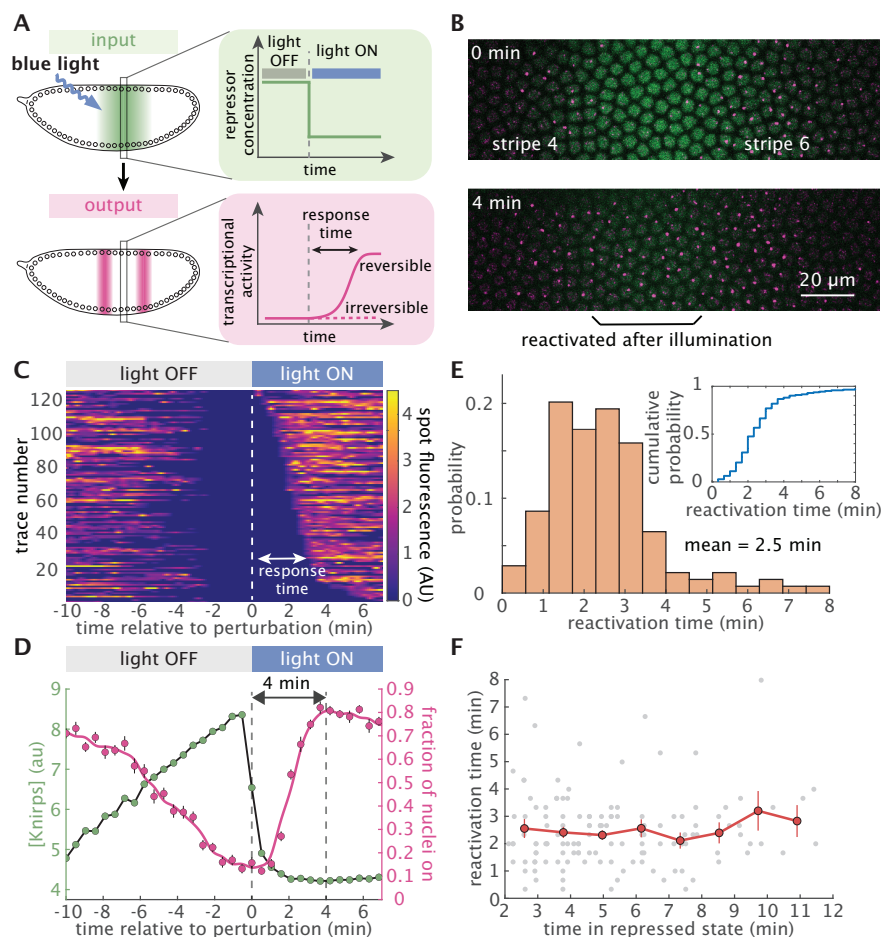


Figure 2.3: **Knirps repression is rapidly reversible and memoryless.** (A) Testing the reversibility of Knirps repression using a step-like optogenetic perturbation. Upon removal of Knirps repressor from the nucleus, transcriptional activity can remain repressed or recover, depending on whether repression is irreversible or reversible. (B) Snapshots from a movie before (top) and after (bottom) the optogenetic export of Knirps protein. Nuclei whose transcription was originally repressed by Knirps fully reactivate after 4 minutes of illumination. (C) Heatmap of single-cell reactivation trajectories sorted by response times. Response time is defined as the interval between the perturbation time and when the MS2 spots reappear. (D) Average repressor concentration (green) and the fraction of actively transcribing cells (magenta) before and after blue light illumination. We find that Knirps repression is rapidly reversible within 4 minutes. ($n = 229$ nuclei from 4 embryos, averaged over a -2% to 2% window along the anterior-posterior axis centered on the Knirps concentration peak). *(continued on the next page)*

Figure 2.3: *(continued)* **Knirps repression is rapidly reversible and memoryless.** (E) Fast reactivation occurs with an average of 2.5 minutes. The reactivation response time is calculated as the interval between the perturbation and when a locus is first observed to resume transcription. ($n = 139$ nuclei from 4 embryos). Inset panel describes the cumulative distribution of reactivation times. To exclude gene loci that were transiently OFF due to transcriptional bursting or missed detections, we focused this analysis on gene loci that were silent for at least 2 minutes before perturbation. (F) Knirps repression is memoryless. Plot showing the reactivation response time of individual loci as a function of the time spent in the repressed state before optogenetic reactivation. The reactivation response time is independent of the repressed duration of the locus. (Error bars in D and F indicate the bootstrap estimate of the standard error.)

2.2.3 Rapid export of repressor reveals fast, reversible reactivation kinetics at the single-cell level

It has been shown that the activity of repressors can have different degrees of reversibility [14, 21]. For example, recruitment of certain chromatin modifiers may silence the locus even if the initial transcription factor is no longer present [14]. The single-cell traces in Figure 2.2E and Figure 2.10 appear to transition into an irreversible transcriptional quiescent state. However, since Knirps concentration keeps increasing after *eve* 4+6 expression shuts off, it is possible that repression is, in fact, reversible and that the observed irreversibility is due only to the monotonic increase of the repressor concentration over time.

To probe the reversibility of Knirps-based repression, we used optogenetics to induce rapid, step-like decreases in nuclear Knirps concentration (Figure 2.3A). Prior to the perturbation, the system was allowed to proceed along its wild-type trajectory until the majority of *eve* 4+6 loci at the center of the Knirps domain were fully repressed. Strikingly, when blue light was applied to export Knirps, we observed a widespread, rapid reactivation of repressed *eve* loci (Figure 2.3B and C). To probe the time scale of reactivation, we calculated the fraction of active nuclei as a function of time since Knirps export (Figure 2.3D, Figure 2.11). This revealed that *eve* loci begin to reactivate in as little as 1 minute following illumination. We obtain a reactivation time distribution from single-cell trajectories with a mean response time of 2.5 minutes (Figure 2.3E) and find that transcription fully recovers within 4 minutes of Knirps export (Figure 2.3D). Thus, Knirps repression is completely reversible.

Previous studies have revealed regulatory “memory” wherein the repressive effect of certain repressors increases with longer exposure [14]. Thus, we reasoned that prolonged exposure to high levels of a repressor could induce the accumulation of specific chemical or molecular modifications that prevent activator binding and, as a result, impede reactivation at the target locus, such as histone modifications [166]. If this process is present, we should expect gene loci that have been repressed for a longer period before optogenetically triggering repressor export to require more time to reactivate. To test this hypothesis, we

used the measured single-cell reactivation trajectories (Figure 2.3C) to calculate the average reactivation time as a function of how long cells had been repressed prior to Knirps export. Interestingly, our analysis reveals that the reactivation time has no dependence on the repressed duration (Figure 2.3F). This, combined with the fact that nearly all (97%) repressed gene loci reactivate upon Knirps export (inset panel in Figure 2.3E), argues against the accumulation of any significant molecular memory amongst repressed gene loci within the ~ 10 minute time scale captured by our experiments. Instead, it points to a model where repressor action is quickly reversible and memoryless.

2.2.4 Knirps acts by inhibiting the initiation of transcription bursts

One of the simplest model that can capture the reversible, memoryless transitions between active and inactive transcriptional states observed in Figure 2.3 is a two-state model, in which the gene promoter switches stochastically between periods of transcriptional activity (“bursts”) and periods of inactivity [187, 102, 11, 58, 17, 130, 111, 180]. Here, the gene promoter switches between active (ON) and inactive (OFF) states with rates k_{on} and k_{off} , and initiates RNAP molecules at a rate r while in the ON state (Figure 2.4A). Consistent with this model, our single-cell transcriptional traces show clear signatures of transcriptional bursting (see, e.g., top two panels of Figure 2.2E; Figure 2.10), suggesting that this two-state framework provides a viable basis for examining how Knirps regulates transcriptional activity at *eve* 4+6 loci.

Within this model, the repressor can act by decreasing burst frequency (decreasing k_{on}), by decreasing the duration of transcriptional bursts (increasing k_{off}), by decreasing the burst amplitude (decreasing r), or any combination thereof as shown in Figure 2.4A. To shed light on the molecular strategy by which Knirps represses *eve* 4+6, we utilized a recently-developed computational method that utilizes compound-state Hidden Markov Models (cpHMM) to infer promoter state dynamics and burst parameter values (k_{on} , k_{off} , and r) from single-cell transcriptional traces as a function of Knirps concentration (Figure 2.4B) [102]. We used data from all three illumination conditions (outlined in Figure 2.2B) and conducted burst parameter inference on 15-minute-long segments of MS2 traces.

To reveal burst parameter dependence on Knirps concentration, we grouped traces based on low ($[\text{Knirps}] \leq 4$ au) and high ($[\text{Knirps}] \geq 6$ au) Knirps concentrations (Figure 2.4B) and conducted cpHMM inference. We find that the repressor strongly impedes locus activation, decreasing the frequency of transcriptional bursts (k_{on}) from 2.3 bursts per minute down to 1.1 burst per minute between low and high Knirps concentrations (Figure 2.4C, left panel). We also find a moderate ($\sim 30\%$) increase in the duration of transcriptional bursts between low and high Knirps concentrations; however this change is smaller than the uncertainty in our inference (Figure 2.4C, middle panel). Finally, we find no significant change in the burst amplitude as a function of Knirps concentration (Figure 2.4C, right panel). Thus, burst parameter inference indicates that Knirps represses *eve* 4+6 loci mainly by interfering with

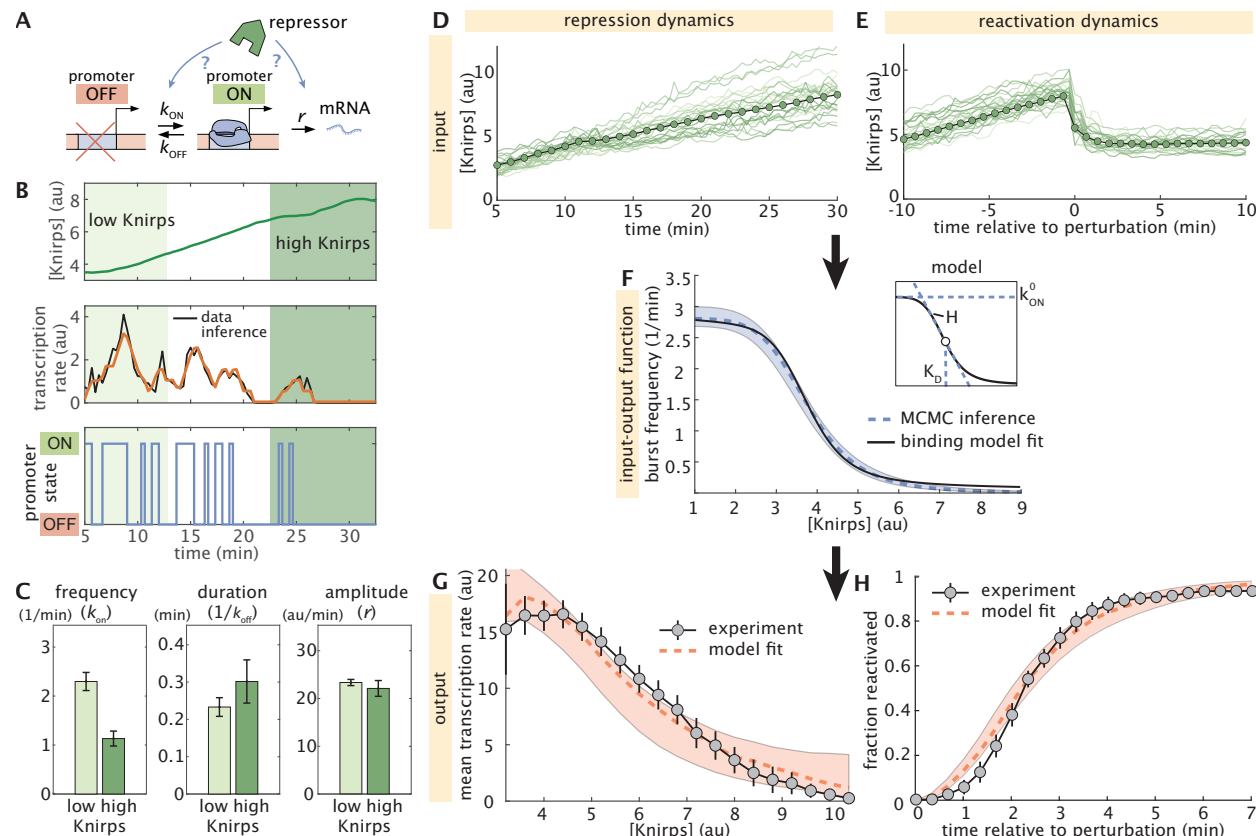


Figure 2.4: **Knirps represses through instantaneous modulation of burst frequency.** (A) Cartoon illustrating the two-state model of transcriptional bursting where a promoter can stochastically transition between active and inactive states. Knirps may regulate *eve* by altering any of the three kinetic rates in the model. (B) A representative experimental trace of Knirps protein (top) and transcription dynamics, along with the best fit (middle) and the corresponding sequence of inferred promoter activity states (bottom) returned by cpHMM inference. (C) Bar plots indicating cpHMM burst parameter inference results for *eve* 4+6 loci subjected to low ($[\text{Knirps}] \leq 4$ au) and high ($[\text{Knirps}] \geq 6$ au) Knirps concentrations. Inference reveals a two-fold decrease in the burst frequency, a moderate (30% though within error bars) increase in burst duration, and no notable change in the burst amplitude between low and high concentrations. (D-H) Summary of stochastic simulation methodology and results. First, we sample real single-cell Knirps concentration trajectories from (i) the three illumination conditions shown in Figure 2.2D and (ii) the reactivation experiments. (D) Illustrative individual (green lines) and average (green circles) nuclear Knirps concentration trajectories as a function of time in unperturbed embryos. *(continued on the next page)*

Figure 2.4: *(continued)* **Knirps represses through instantaneous modulation of burst frequency.** **(E)** Individual and average nuclear Knirps concentrations before and after optogenetic export, which happens at time $t = 0$. **(F)** We take k_{on} to be a Hill function of Knirps concentration, with a shape that is determined by three microscopic parameters: k_{on}^0 , K_D , and H (see inset panel and Equation 2.1). Given some set of microscopic parameters, we can plug Knirps concentration trajectories from (D) and (E) into the corresponding k_{on} input-output function to predict transcriptional outputs. The dashed blue curve indicates the input-output function for the burst frequency trend (k_{on}) corresponding to the best-fitting set of microscopic parameters. Light blue shading indicates the standard error of the mean of the k_{on} input-output trend, as estimated by MCMC inference. To test the possibility that Knirps binding at the *eve* 4+6 enhancer, we fit a simple thermodynamic model to the trend revealed by our input-output simulations. The black line shows the best-fitting curve predicted by this molecular model. The binding model assumes 10 Knirps binding sites. We used the input-output function in (F) to generate a population of simulated MS2 traces that we used to predict. **(G)** the average fluorescence as a function of Knirps concentration and **(H)** the reactivation dynamics. Dashed red line indicates the prediction of the best-fitting model realization. Shaded red regions indicate standard deviation of the mean, as indicated by our MCMC inference. (Error bars in C reflect the standard error of the mean, as estimated from no fewer than 20 bootstrap burst inference replicates. The transcription rate is calculated from the measured MS2 signal, which is an approximation of the mRNA production rate [63, 102, 18].)

the initiation of transcriptional bursts. See Supplementary Text Section 2.7.1 and Figure 2.12 for additional cpHMM inference results.

To our knowledge, Figure 2.4C provides the first simultaneous measurement of transcription factor concentration and burst dynamics in a living multicellular organism. However, these results are, necessarily, a coarse-grained approximation of the true regulatory dynamics. This is because our cpHMM inference has an inherently low temporal resolution, reflecting averages taken across 15-minute periods of time and across large ranges of input Knirps concentrations. However, in principle, our live imaging data—which contains high-resolution time traces of both input repressor concentration dynamics and output transcription rates—should make it possible to move beyond these coarse-grained estimates to recover the true, *instantaneous* regulatory relationship between Knirps concentration and burst dynamics.

To answer these questions, we developed a novel computational method that utilizes stochastic simulations of single-cell transcriptional trajectories to test theoretical model predictions against our experimental measurements and uncover repressor-dependent burst parameter trends (Figure 2.13; Supplementary Text Section 2.7.2). Motivated by the cpHMM inference shown in Figure 2.4C, as well as finer-grained results shown in Figure 2.12, we allow both the burst frequency and the burst duration (but not the burst amplitude) to vary as a function of Knirps concentration. We assume a model in which these parameters are

simple Hill functions of repressor concentration. For the burst frequency (k_{on}), this leads to a function with the form

$$k_{\text{on}}([\text{Knirps}]) = k_{\text{on}}^0 \frac{K_D^H}{[\text{Knirps}]^H + K_D^H}, \quad (2.1)$$

where k_{on}^0 sets the maximum burst frequency value, the Hill coefficient H sets the sharpness of the response, and K_D dictates the Knirps concentration midpoint for the transcriptional response, giving the repressor concentration where k_{on} drops to half its maximum value. Together, these “microscopic” parameters define an input-output function that directly links the burst frequency to Knirps concentration. As noted above, we also allow the burst duration to vary as a function of Knirps concentration (see Equation 2.3 and Supplementary Text Section 2.7.2.1 for further details). However we focus on k_{on} throughout the main text, since it is the only parameter that decreases as a function of Knirps concentration (and, thus, the only parameter that could drive *eve* 4+6 repression).

With our model defined, our procedure is as follows: we start by sampling real single-cell Knirps concentration trajectories from (i) the three illumination conditions shown in Figure 2.2D and (ii) the reactivation experiments shown in Figure 2.3 (Figure 2.4D and E, respectively). Then, we plug these Knirps trajectories into the input-output functions defined in Equation 2.1 (for burst frequency; see also Figure 2.4F) and Equation 2.3 (for burst duration). Next, given a set of microscopic parameters (e.g., H , K_D , and k_{on}^0 for Equation 2.1), we generate time-dependent burst parameter trends (Figure 2.13B). We then use these trends to simulate corresponding ensembles of MS2 traces (Figure 2.13C-F; see also Supplementary Text Section 2.7.2.1). We use these simulated MS2 traces to calculate, first, the predicted Knirps vs. *eve* 4+6 input-output function (Figure 2.4G) and, second, the predicted reactivation cumulative distribution function curve (Figure 2.4H). Finally, we compare these predictions to empirical measurements of the same quantities from our live imaging experiments (see Figure 2.2D and inset panel of Figure 2.3E). Through this process of simulation and comparison, each set of microscopic parameters used to calculate our predictions are assigned a fit score. We then use parameter sweeps and Markov Chain Monte Carlo (MCMC) [65, 140] to search for parameters that most successfully reproduced our live imaging results (see Figure 2.13E-G and Sections 2.7.2.3 and 2.7.2.4).

As illustrated in Figure 2.4F, we find that the best-fitting model features a sharp k_{on} versus Knirps input-output function ($H = 6.05 \pm 0.7$). We also find that k_{on} has a relatively low K_D of $3.7 \text{ au} \pm 0.13$ with respect to the range of Knirps concentrations experienced by *eve* 4+6 loci (see Figure 2.2B, bottom), which implies that gene loci have a low concentration threshold for Knirps repression. As a result of this low threshold, *eve* 4+6 loci are effectively clamped in the OFF state ($k_{\text{on}} \leq 0.1$ bursts per minute) once the Knirps concentration exceeds 6 au, which happens about 12 minutes into nuclear cycle 14 for the average nucleus at the center of the Knirps domain (Figure 2.2B, bottom). See Figure 2.14 and Supplementary Text Section 2.7.2.5 for full inference results. Our findings also demonstrate that a simple two-state model in which Knirps represses *eve* 4+6 by decreasing the frequency of

transcriptional bursts is sufficient to quantitatively recapitulate both the sharp decrease in the average transcription rate with increasing Knirps concentration (Figure 2.4G) and the kinetics of reactivation following Knirps export (Figure 2.4H).

Our simulation results also shed further light on the dynamics of *eve* reactivation following the step-like optogenetic export of Knirps protein from the nucleus (Figure 2.3A). From Figure 2.3E and F, we know that it takes approximately 2-4 minutes following Knirps export for MS2 spots to reappear in our live-imaging experiments. Yet this is the time scale for *detection*—for the amount of time it takes for genes to produce detectable levels of transcription and, hence, MS2 fluorescence—and thus likely overestimates the true *eve* 4+6 response time. So how fast is it really? Our model, which accounts for the fluorescence detection limit, predicts that k_{on} recovers to half of its steady-state value within 30 seconds of the start of the optogenetic perturbation (Figure 2.15). Furthermore, we predict that half of all gene loci switch back into the transcriptionally active (ON) state within 102 seconds (1.7 minutes). Thus, it takes fewer than two minutes for *eve* 4+6 loci to “escape” Knirps repression and re-engage in bursty transcription.

2.3 Discussion

Taken together, our results point to a model wherein the repressor acts upon the gene locus while it is transcriptionally inactive (OFF) to inhibit re-entry into the active (ON) state. Consistent with this picture, we find that the functional relation between k_{on} and Knirps concentration inferred by MCMC inference is well explained by a simple equilibrium binding model where the burst frequency is proportional to the number of repressor molecules bound at the 4+6 enhancer (solid black curve in Figure 2.4F; see Supplementary Text Section 2.7.3 for details).

Our *in vivo* dissection provides important clues toward unraveling the molecular basis of repressor action. We show that Knirps repression is switch-like (Figure 2.2), memoryless (Figure 2.3F), and rapidly reversible (Figure 2.3E). Another key point is that, although our model predicts that gene loci require 1-2 minutes to reactivate and enter the ON state following the optogenetic export of Knirps from the nucleus (Figure 2.15), the model assumes that the burst frequency itself responds *instantaneously* to changing Knirps concentration (see Equation 2.1, blue curve in Figure 2.15). While no reaction can truly be instantaneous, the success of this model in describing repression dynamics points to an underlying mechanism controlling the burst frequency that rapidly reads and responds to changing repressor concentrations, likely within a matter of seconds—a timescale that is consistent with the fast binding and unbinding dynamics reported for eukaryotic transcription factors [101]. Lastly, the success of the two-state bursting model (Figure 2.4A) at recapitulating Knirps repression dynamics (Figure 2.4G and H) suggests that the same molecular process may be responsible for both the short-lived OFF periods between successive transcriptional bursts (see, e.g., the middle panel of Figure 2.4B) and the much longer-lived periods of quiescence observed in repressed nuclei (e.g., Figure 2.3C), and that there may be no need to invoke an “extra”

repressor-induced molecular state outside of the bursting cycle [34, 47, 108].

Previous work has established that Knirps plays a role in recruiting histone deacetylase [161] and that Knirps repression coincides with increased histone density at target enhancers such as the one dissected here [109]. This suggests a model in which the repressor modulates the longevity of the OFF state by tuning the accessibility of enhancer DNA, which would impact activator binding [109]. It is notable, however, that the 1-2 minute reactivation time scales revealed (Figure 2.3; Figure 2.15) are faster than most chromatin-based mechanisms measured *in vivo* so far [182, 79, 14, 21, 101]. This rapid reversibility, along with the memoryless nature of Knirps repression, indicates that whatever the underlying mechanism, Knirps binding at the locus is *necessary* in order to maintain the gene in a transcriptionally inactive state at the stage of development captured by our live imaging experiments. Interestingly, we found that the modulation of burst frequency by Knirps can be recapitulated by a simple thermodynamic model predicting Knirps DNA occupancy (black line in Figure 2.4F; see Supplementary Text Section 2.7.3 for further details). This suggests that the wide repertoire of theoretical and experimental approaches developed to test these models (see, for example, [137]) can be used to engage in a dialogue between theory and experiment aimed at dissecting the molecular mechanism underlying the control of transcriptional bursting.

Critically, none of these molecular insights would have been possible without the ability to measure and acutely manipulate input transcription factor concentrations in living cells. Thus, by building on previous works using the LEXY technology in different biological contexts [125, 98, 164, 150, 124], our work demonstrates the power of the LEXY system for simultaneously manipulating—and measuring—nuclear protein concentrations and the resulting output transcriptional activity. This capability can serve as a quantitative platform for dissecting gene-regulatory logic *in vivo*. Moreover, the LEXY system improves upon many previously reported methods of optogenetic control in embryos [175, 24, 142, 85, 92, 86, 119, 93, 95, 133] (see Supplementary Text Section 2.7.4 for further discussions).

Looking ahead, we anticipate that our live imaging approach, along with the quantitative analysis framework presented in this work, will provide a useful foundation for similar *in vivo* biochemical dissections of how the transcription factor-mediated control of gene expression dictates transcriptional outcomes, opening the door to a number of exciting new questions relating to transcriptional regulation, cell-fate decisions, and embryonic development that span multiple scales of space and time.

2.4 Materials and Methods

2.4.1 Cloning and Transgenesis

The fly lines used in this study were generated by inserting transgenic reporters into the fly genome or by CRISPR-Cas9 genome editing, as described below. See Table 2.1 for detailed information on the plasmid sequences used in this study.

2.4.1.1 Creation of tagged *knirps* loci using CRISPR-Cas9

To tag endogenous the *knirps* locus with the EGFP-LlamaTag and LEXY modules, we used CRISPR-mediated homology-directed repair with donor plasmids synthesized by Genscript. gRNA was designed using target finder tool from flyCRISPR (<https://flycrispr.org>), and cloned based on the protocol from [70]. A *yw*;nos-Cas9(II-attP40) transgenic line was used as the genomic source for Cas9, and the embryos were injected and screened by BestGene Inc.

2.4.1.2 Creation of *eve* 4+6 reporter

The *eve* 4+6 enhancer sequence is based on 800bp DNA segment described in [57]. The *eve* 4+6 reporter was constructed by combining the enhancer sequence with an array of 24 MS2 stem-loops fused to the *D. melanogaster yellow* gene [18]. The *eve*4+6-MS2-Yellow construct was synthesized by Genscript and injected by BestGene Inc into *D. melanogaster* embryos with a Φ C31 insertion site in chromosome 2L (Bloomington stock #9723; landing site VK00002; cytological location 28E7).

2.4.1.3 Transgenes expressing EYFP and MCP-mCherry

The fly line maternally expressing MCP-mCherry that is attached to a nuclear localization signal (chromosome 3) was constructed as described in [18]. The fly line maternally expressing EYFP (chromosome 2) was constructed as previously described in [96]. To simultaneously image protein dynamics using LlamaTags and transcription using MCP-MS2 system, we combined the *vasa*-EYFP transgene with MCP-mCherry to construct a new line (*yw*; *vasa*-EYFP; MCP-mCherry) that maternally expresses both proteins.

2.4.2 Fly lines

To measure the Knirps pattern and corresponding *eve* 4+6 transcription simultaneously, we performed crosses to generate virgins carrying transgenes that drive maternal EYFP, MCP-mCherry, along with LlamaTag-LEXY tagged Knirps locus (*yw*; *vasa*-EYFP; MCP-mCherry/Knirps-LlamaTag-LEXY). These flies were then crossed with males having both the *eve* 4+6 reporter and LlamaTag-LEXY tagged Knirps locus (*yw*; *eve*4+6-MS2-Yellow; Knirps-LlamaTag-LEXY). This resulted in embryos homozygous or heterozygous for the tagged Knirps locus also carrying maternally deposited EYFP, MCP-mCherry, and a *eve* 4+6 reporter. Embryos homozygous for tagged Knirps can be differentiated from heterozygous embryos through a comparison of their nuclear fluorescence levels as shown in Figure 2.16. All the fly lines used in this work can be found in Table 2.2

2.4.3 Embryo preparation and data collection

The embryos were prepared following procedures described in [63, 18, 102]. Embryos were collected and mounted in halocarbon oil 27 between a semipermeable membrane (Lumox film, Starstedt, Germany) and a coverslip. Confocal imaging on a Zeiss LSM 780 microscope was performed using a Plan-Apochromat 40x/1.4NA oil immersion objective. EYFP and MCP-mCherry were excited with laser wavelengths of 514 nm (3.05 μ W laser power) and 594 nm (18.3 μ W laser power), respectively. Modulation of Knirps nuclear concentration was performed by utilizing an additional laser with a wavelength of 458nm, with laser power of 0.2 μ W (low intensity in Figure 2.2) or 12.2 μ W (high intensity in Figure 2.2 and Figure 2.3). Fluorescence was detected using the Zeiss QUASAR detection unit. Image resolution was 768×450 pixels, with a pixel size of 0.23 μ m. Sequential Z stacks separated by 0.5 μ m were acquired with a time interval of 20 seconds between each frame, except for the export-recovery experiment in Figure 2.1, in which we used 6.5 seconds.

2.4.4 Image processing

Image analysis of live embryo movies was performed based on the protocol in [63, 139], which included nuclear segmentation, spot segmentation, and tracking. In addition, the nuclear protein fluorescence of the Knirps repressor was calculated based on the protocol in [96]. The nuclear fluorescence of Knirps protein was calculated based on a nuclear mask generated from the MCP-mCherry channel. Knirps concentration for individual nuclei was extracted based on the integrated amount from maximum projection along the z-stack. The YFP background was calculated based on a control experiment and subsequently subtracted from the data.

2.4.5 Predicting Knirps binding sites

To dissect Knirps binding to the *eve* 4+6 enhancer, we used Patser [80] with already existing point weight matrices [52] to predict Knirps binding sites. The predicted binding sites with scores higher than 3.5 are shown in Figure 2.8.

2.4.6 Compound-state Hidden Markov Model

To obtain the inference results shown in Figure 2.4C, transcriptional traces were divided into 15 minute-long segments. Each trace segment was then assigned to an inference group based on the average nuclear Knirps concentration over the course of its 15-minute span. Trace segments with an average Knirps concentration of less than or equal to 4 arbitrary fluorescence units (au) were assigned to the “low” group and segments with a Knirps concentration greater than or equal to 6 au were assigned to the “high” group. Parameter estimates for each group were estimated by taking the average across 25 separate bootstrap samples of the “high” and “low” trace segment groups. Each bootstrap sample contained a minimum of

6,027 and 10,000 time points for the high and low groups, respectively. Inference uncertainty was estimated by taking the standard deviation across these bootstrap replicates. We used a model with two burst states (OFF and ON) and an elongation time of 140 seconds (equal to seven time steps; see [\[102\]](#)).

2.5 Supplementary Figures

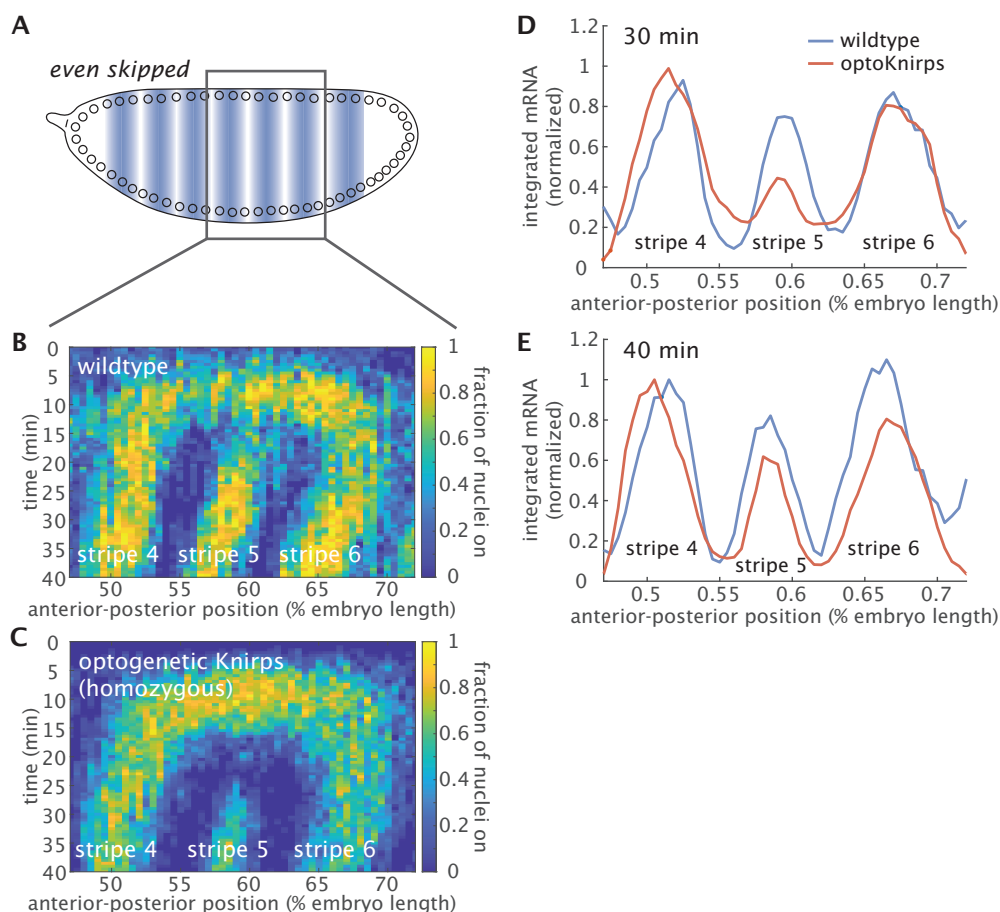


Figure 2.5: *even-skipped* expression under homozygous optogenetic Knirps (tagged with LEXY and LlamaTag) qualitatively recapitulates wild-type expression dynamics. (A) To understand whether and to what degree the *eve* expression pattern is impacted in the homozygous optogenetics Knirps background, we imaged the dynamics of a previously published *eve*-MS2-BAC reporter containing the full endogenous *eve* locus [11] in the wild-type and optogenetics Knirps backgrounds. (B-C) The expression pattern of *even-skipped* as reported by the fraction of detectable MS2 transcription spots is similar under wild-type Knirps (B) and optogenetics Knirps (C) except for a weaker stripe 5. (*continued on the next page*)

Figure 2.5: *(continued) even-skipped* expression under homozygous optogenetic Knirps (tagged with LEXY and LlamaTag) qualitatively recapitulates wild-type expression dynamics. (D-E) Comparison of the amount of mRNA present at 30 minutes into nuclear cycle 14 (as obtained by integrating the MS2 fluorescence signal) and at 40 minutes shows that stripe 5 expression is weaker under homozygous optogenetics Knirps at 30 minutes. The integration was performed assuming a mRNA half-time of 7 min. (D) Stripe 4 and 6 expression is slightly wider than under the wild-type condition at 30 min, suggesting that optogenetics Knirps is a slightly weaker repressor compared to the wild-type Knirps. (E) Stripe 5 expression continues to increase as it reaches a similar level compared to the wild-type around 40 minutes. The anterior-posterior position is aligned based on the center of stripe 5. The plots are normalized according to the peak of stripe 4 at 40 minutes and smoothened using a moving window of 1.5% range along the anterior-posterior axis. (Data from a single embryo is shown for each condition. $t = 0$ is defined as the onset of transcription in nuclear cycle 14.)

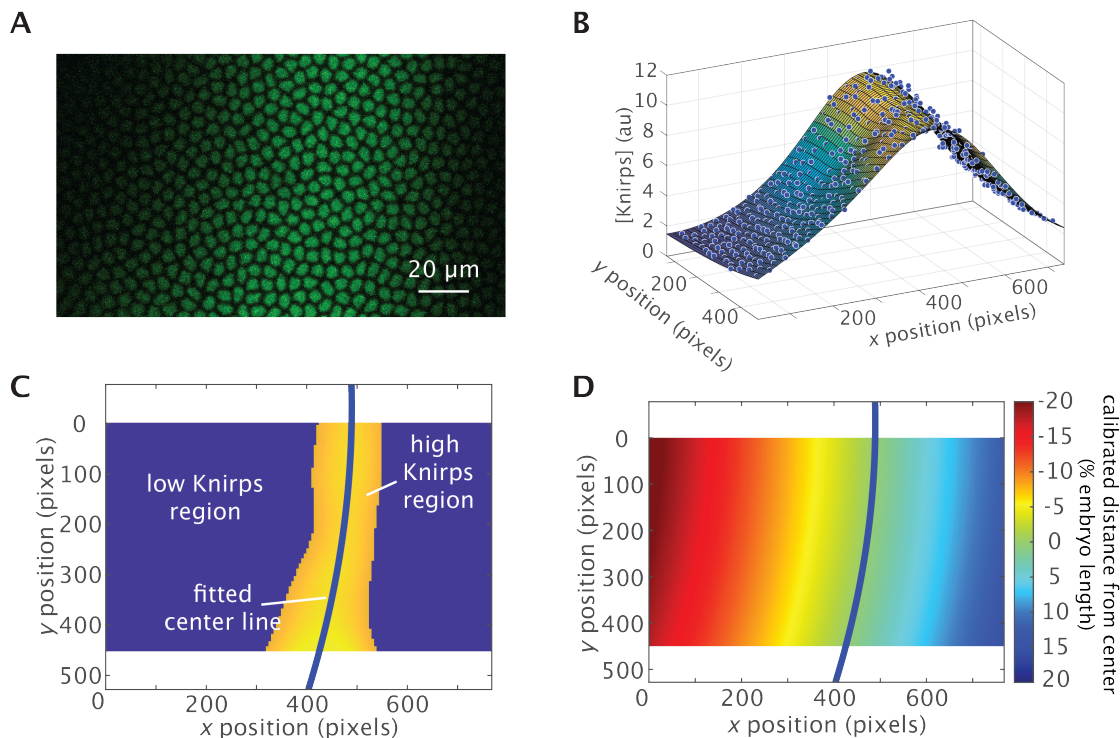


Figure 2.6: **Nuclei position calibration based on Knirps expression pattern.** The Knirps pattern of each individual embryo is used to align embryos along their anterior-posterior position axis. **(A)** Snapshot of the Knirps pattern used to calibrate nuclei position. **(B)** Extracted nuclear fluorescence is smoothed by local quadratic regression. **(C)** The region with high Knirps expression (yellow region) is extracted with a single threshold. Then, a quadratic function is fitted to the nuclei with high Knirps expression (yellow region) to extract the center line of Knirps expression (blue line). **(D)** Calibrated positions relative to the Knirps expression peak are calculated based on the distance to the extracted center line.

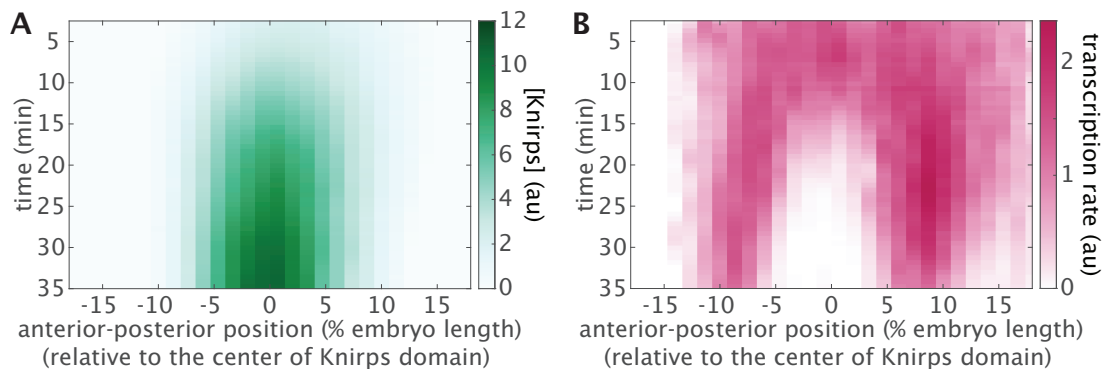


Figure 2.7: **Spatiotemporal dynamics of Knirps protein and *eve* 4+6 transcription.** Nuclei were binned based on their positions relative to the center of the Knirps domain (Figure 2.6, Materials and Methods) and their corresponding (A) Knirps protein concentration reported by LlamaTag fluorescence and (B) transcription reported by MS2 fluorescence were quantified over time.

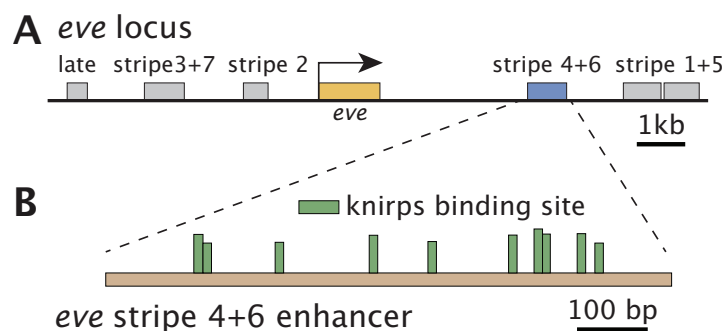


Figure 2.8: **Predicted Knirps binding sites in the *eve* 4+6 enhancer.** (A) The *eve* 4+6 enhancer is an 800 bp segment from the endogenous *eve* locus. (B) Ten Knirps binding sites are predicted within the *eve* 4+6 enhancer using PATSER [80] and Knirps position weight matrices from [52]. Only binding motifs with PATSER scores higher than 3.5 are shown. The bar height of each binding site is proportional to the PATSER score.

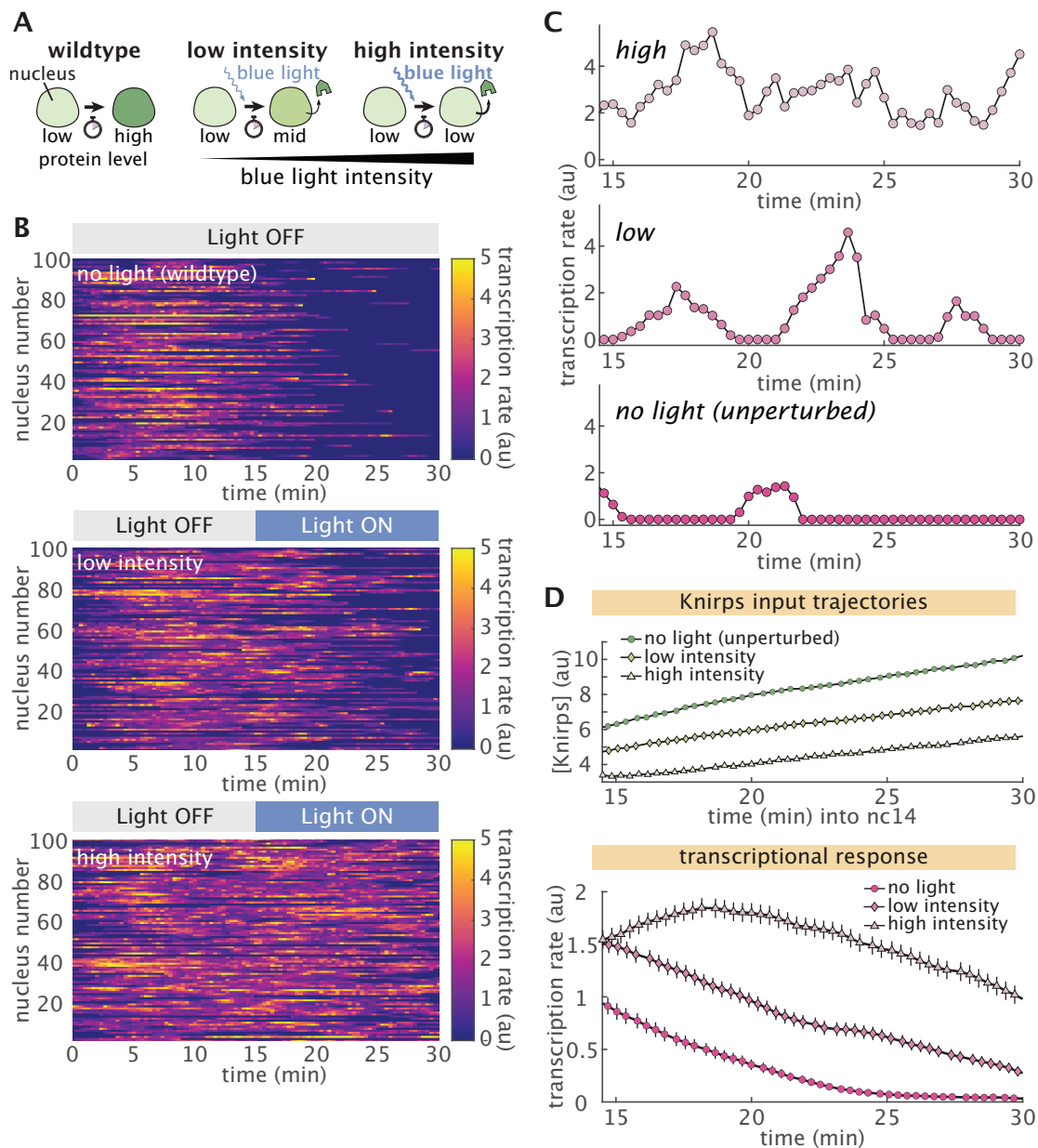


Figure 2.9: **Repressor titration results in distinct transcriptional dynamics.** (A) Optogenetic titration of protein concentration. Cartoon schematics for three different illumination conditions. Left: No illumination results in a negligible export of nuclear Knirps over time (green). Middle: Low dosage of blue light induces weak export of repressor from nuclei. Right: high intensity of blue light results in a strong export of repressor. (*continued on the next page*)

Figure 2.9: *(continued)* **Repressor titration results in distinct transcriptional dynamics.** **(B)** Single-cell traces for embryos with different Knirps export levels show distinct transcriptional dynamics. **(C)** Representative single-cell transcriptional dynamics under different illumination conditions show distinct responses. **(D)** Mean protein (top) and transcription rates (bottom) under different illumination conditions. Averaged over $n = 4$ (no light), $n = 4$ (low intensity) and $n = 3$ (high intensity) embryos. (Error bars in D indicate the bootstrap estimate of the standard error over multiple embryos.)

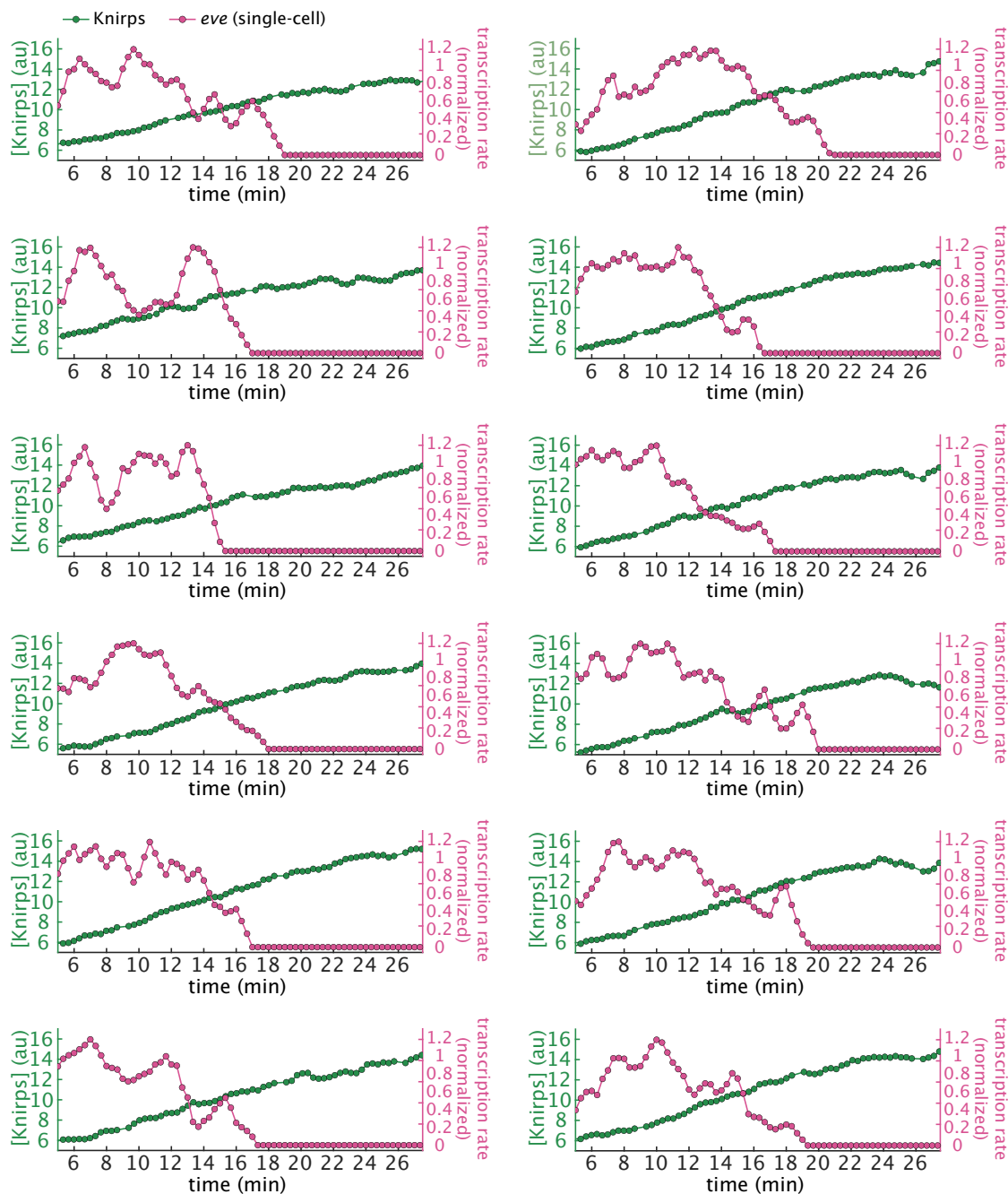


Figure 2.10: **Example single-cell traces under no illumination.** Single-cell input Knirps and output transcriptional dynamics traces show clear signs of transcriptional bursting, and that repression is switch-like. Traces are normalized by their maximum transcription rate and smoothed using a moving average of 1 minute.

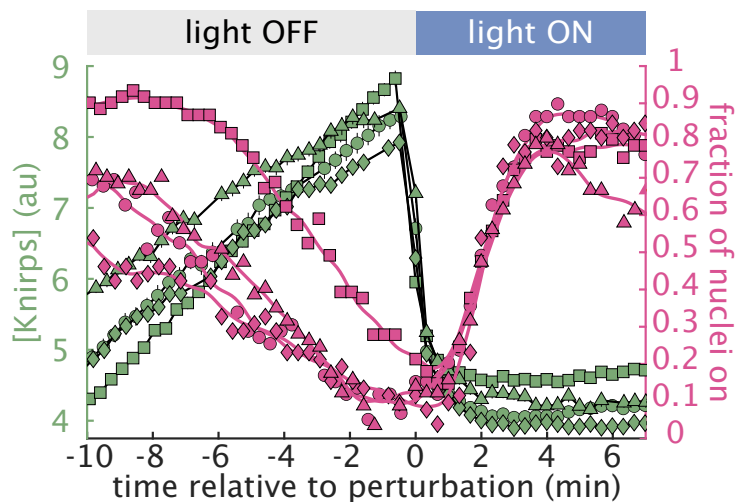


Figure 2.11: **Response to Knirps perturbations is consistent across multiple embryos.** Plot showing input Knirps concentration and output transcriptional activity for four individual embryos. All embryos display similar responses to Knirps export upon light exposure. Each marker shape corresponds to one embryo. (Error bars indicate the bootstrap estimate of the standard error.)

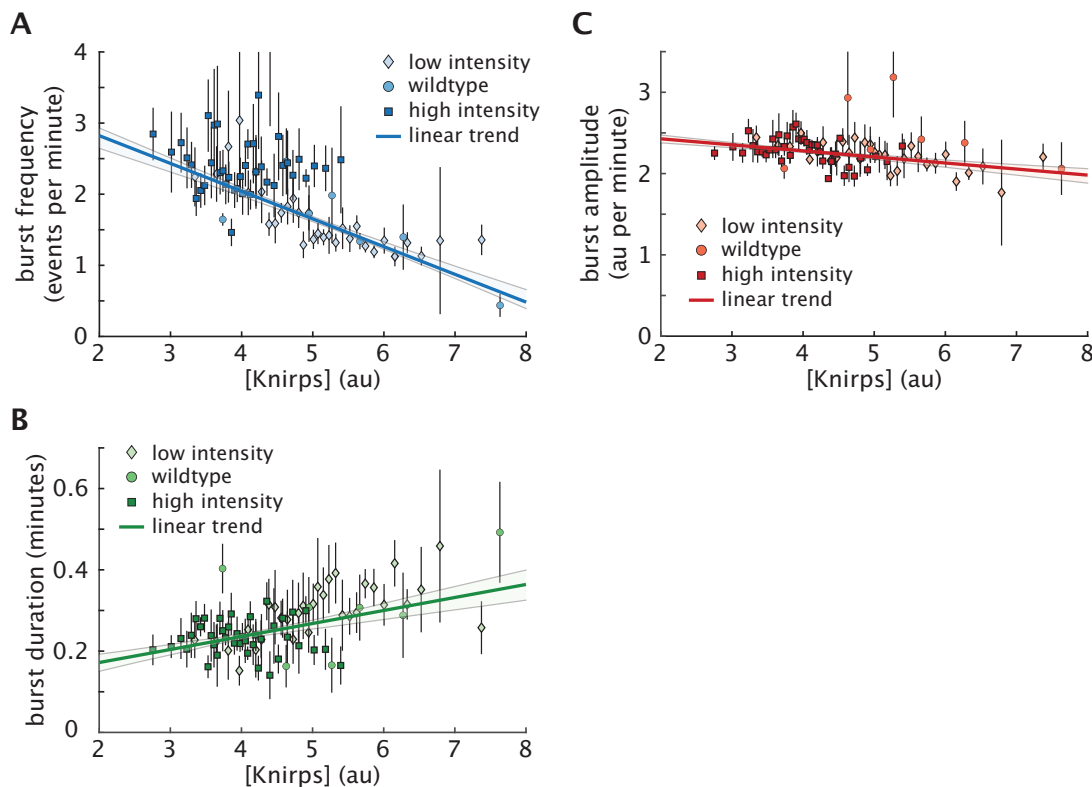


Figure 2.12: **Full cpHMM inference results of Knirps-regulated transcriptional bursting.** (A) We find that the burst frequency (k_{on}) decreases significantly as a function of Knirps concentration. (B) We also find a moderate increase in burst duration ($1/k_{\text{off}}$) with Knirps concentration, (C) while burst amplitude (r) remains approximately constant. Lines in A, B and C indicate the best linear fit to data. Circles, diamonds, and squares indicate data points from no light (unperturbed), low illumination, and high illumination experiments, respectively, as described in Figure 2.2B.

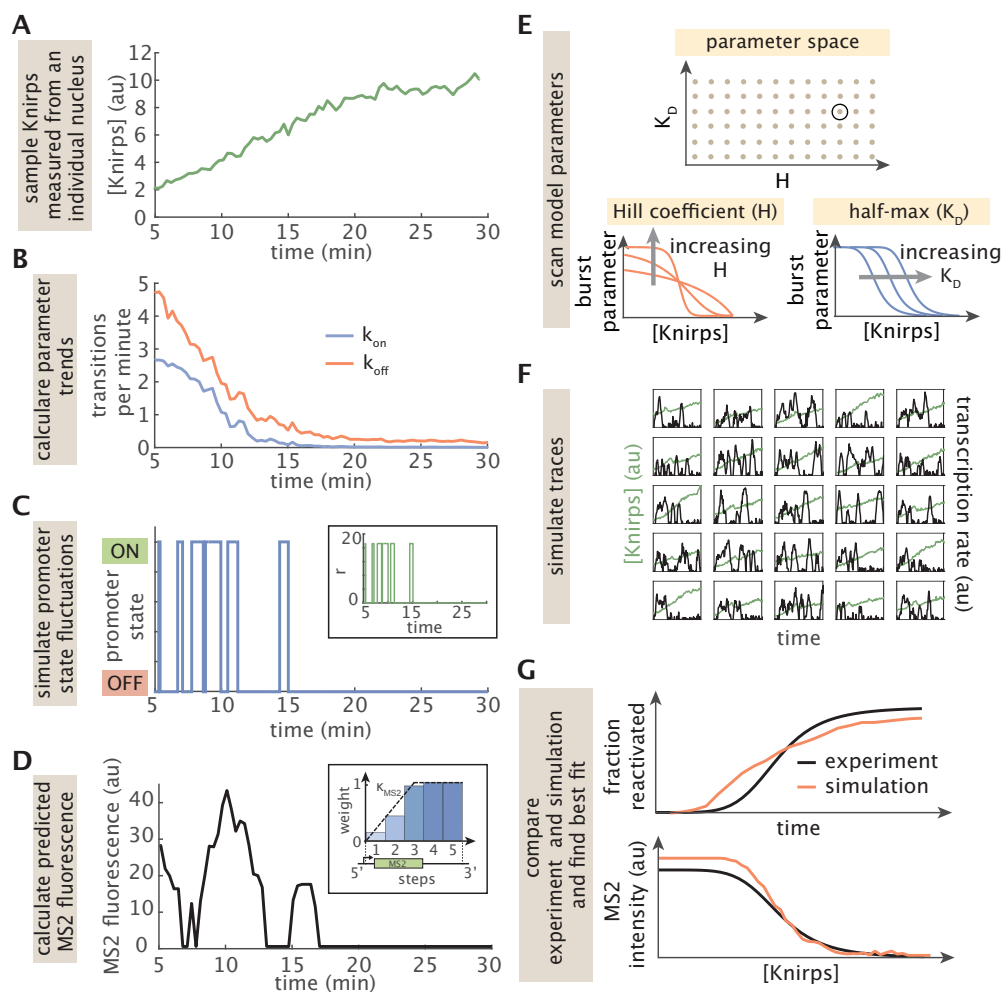


Figure 2.13: A computational framework for Knirps-dependent stochastic simulations. (continued on the next page)

Figure 2.13: *(continued)* **A computational framework for Knirps-dependent stochastic simulations.** **(A-D)** Schematic showing process for simulating stochastic transcription time traces. **(A)** We first sample an empirical time trace of Knirps concentration from a nucleus in our live imaging dataset. **(B)** Next, we plug this Knirps trace into the input-output functions for k_{on} (Equation 2.2) and k_{off} (Equation 2.3) to generate time-dependent burst parameter trends. **(C)** We then use a discrete implementation of the Gillespie Algorithm to simulate a stochastic time-series of promoter activity that reflects the time-dependent parameter trends. Inset panel shows corresponding initiation rate time series. **(D)** Finally, we use this promoter time series to calculate the predicted MS2 fluorescence at each time point. We assume an initiation rate of 21.5 au when the promoter is in the ON state and a basal rate of 0.6 au when the promoter is OFF. **(E-G)** Schematic illustrating the parameter sweep algorithm. **(E)** We use a simple gridded search to sweep a broad space of values for key parameters in Equations 2.2 and 2.3. Cartoon illustrates case for a 2D search for k_{on} -related parameters. In reality, we also scan the analogous k_{off} parameters, leading to a 4D gridded search. For each iteration of the sweep algorithm, we select a new combination of parameters (black circle in top panel). **(F)** Then, we use the process illustrated in A-D to simulate an ensemble of MS2 traces that reflect these parameter values. We generate one simulated MS2 trace for each experimental Knirps trace in our dataset **(G)** Finally, we use these simulated traces to calculate dynamics of the fraction of reactivated and MS2 fluorescence as a function of Knirps concentration for comparison with our experimental results. The mean squared error is used to assess agreement between prediction and experimental data and to identify the set of microscopic parameters that best describes the data.

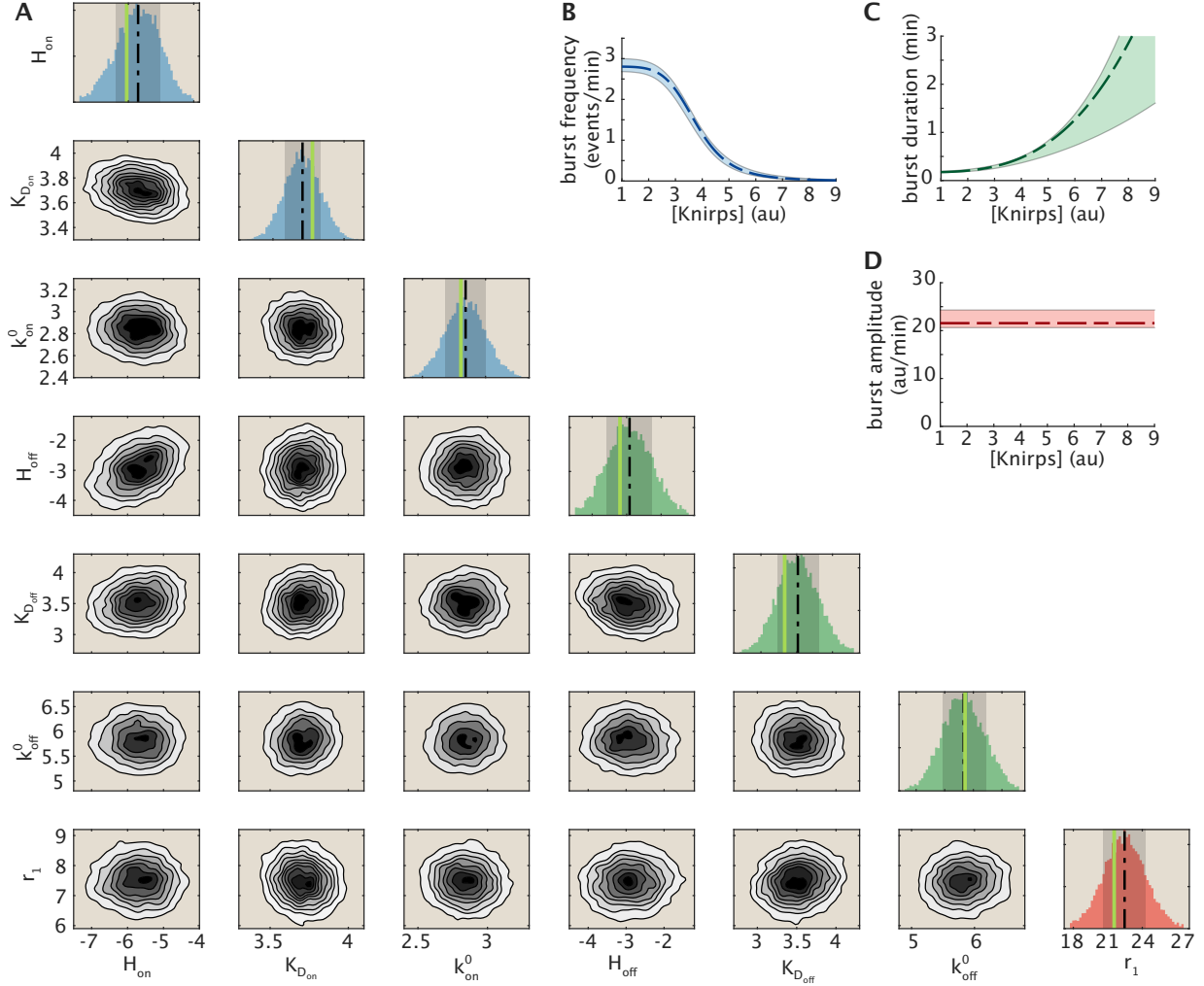


Figure 2.14: **Full MCMC results for stochastic input-output model parameters.** (A) Univariate and bivariate density plots. Vertical green lines in histograms indicate the mean parameter value taken across the 25 best-fitting model realizations. Dashed black lines indicate average parameter values taken across all MCMC samples; i.e. the full distribution shown in each histogram. Shaded regions in histograms indicate 1 standard deviation above and below the mean. (B) Inferred trends for the burst frequency (k_{on}), (C) burst duration ($1/k_{off}$) and (D) burst amplitude (r_1). k_{off} was modeled as a Hill function of Knirps (see Equation 2.3) and r was assumed to be invariant relative to Knirps concentration.

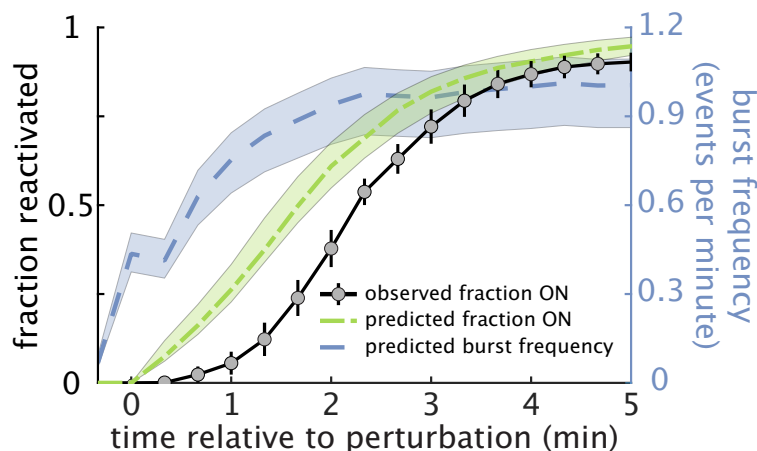


Figure 2.15: **Model predictions for *eve4+6* reactivation dynamics following Knirps export.** Blue curve shows the predicted recovery of burst frequency (calculated from Equation 2.2) following the optogenetic perturbation of Knirps. The green curve indicates the corresponding cumulative fraction of loci that are predicted to have reentered the ON state as a function of time since the perturbation. Black curve is identical to the one shown in Figure 2.4H and corresponds to the measured fraction of loci that have reentered the ON state. We observe a lag between the cumulative fraction of ON loci and the experimentally observed fraction because recently reactivated gene loci typically require multiple time steps to accumulate sufficient fluorescent MS2 signal in order to be experimentally detected.

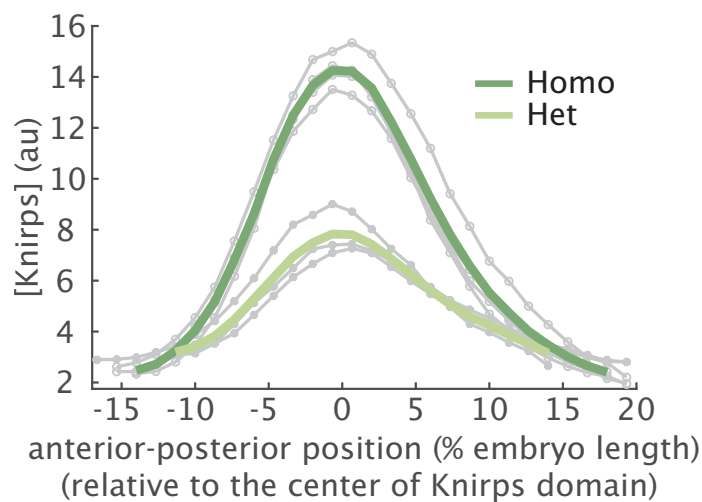


Figure 2.16: **Distinguishing homozygous from heterozygous embryos.** Homozygous embryos ($n = 4$) can be easily distinguished from heterozygous embryos ($n = 3$) by comparing Knirps concentration at 30 min into nc14.

2.6 Supplementary Tables

Table 2.1: List of plasmids used in this study.

Name	Function
pCasper-vasaPr-EYFP	P-element insertion plasmid for vasa promoter driven EYFP
pBPhi-eve4+6-evePr-MS2-Yellow	<i>eve</i> 4+6 reporter
pHD-Kni-LlamaTag-LEXY-dsRed	Donor plasmid for Knirps-LlamaTag-LEXY CRISPR knock-in fusion
pU6-3-gRNA-Knirps-1	guide RNA 1 for Knirps-LlamaTag-LEXY CRISPR knock-in fusion
pU6-3-gRNA-Knirps-2	guide RNA 2 for Knirps-LlamaTag-LEXY CRISPR knock-in fusion

Table 2.2: List of fly lines used in this study.

Genotype	Usage
<i>yw; vasa-EYFP; +</i>	Maternally deposit ubiquitous EYFP
<i>yw; +; MCP-mCherry/TM3,Sb</i>	Maternally deposit MCP-mCherry protein
<i>yw; eve4+6-evePr-MS2-Yellow; +</i>	MS2 reporter for <i>eve</i> 4+6 enhancer
<i>yw; +; Kni-LlamaTag-LEXY</i>	CRISPR knock-in of LlamaTag and LEXY at Knirps C-terminal
<i>yw; vasa-EYFP; MCP-mCherry/TM3,Sb</i>	Maternally deposit both ubiquitous EYFP and MCP-mCherry proteins
<i>yw; vasa-EYFP; Kni-LlamaTag-LEXY</i>	Maternally deposit ubiquitous EYFP, and expresses Knirps protein labeled with LlamaTag and LEXY
<i>yw; eve4+6-evePr-MS2-Yellow; Kni-LlamaTag-LEXY</i>	MS2 reporter for <i>eve</i> 4+6 enhancer with endogenous <i>knirps</i> locus labeled with LlamaTag and LEXY

Table 2.3: **List of parameter ranges used for parameter sweeps.** Brackets denote inclusive ranges. Parameters with a single value appearing in the “range” column were held fixed during the sweeps. Parameters with two values were sampled at 15 equally spaced points bounded by the values indicated in the brackets.

Parameter	Range
burst frequency Hill Coefficient (H_{ON})	[3.15, 12.6]
burst frequency half-maximum (K_{DON})	[2.5, 10.2] (au)
max burst frequency (k_{on}^0)	2.85 (events per min)
off rate Hill Coefficient (H_{OFF})	[0, 4]
off rate half-maximum (K_{DON})	[2, 6] (au)
max off rate (k_{off}^0)	5.81 (events per min)
ON state initiation rate (r_1)	22.76 (au per min)
OFF state initiation rate (r_0)	0.6 (au per min)

Table 2.4: **List of parameter priors used for MCMC sampling.**

Parameter	Prior distribution
burst frequency Hill Coefficient (H_{ON})	$\mathcal{N}(5.7, 0.8)$
burst frequency half-maximum (K_{DON})	$\mathcal{N}(3.7, 0.15)$ (au)
max burst frequency (k_{on}^0)	$\mathcal{N}(2.84, 0.17)$ (events per min)
off rate Hill Coefficient (H_{OFF})	$\mathcal{N}(3.1, 0.8)$
off rate half-maximum (K_{DON})	$\mathcal{N}(3.5, 0.3)$ (au)
max off rate (k_{off}^0)	$\mathcal{N}(5.8, 0.4)$ (events per min)
initiation rate (r_1)	$\mathcal{N}(22.8, 2.1)$ (au per min)

2.7 Supplementary Text

2.7.1 Additional cpHMM inference results

In this section, we briefly describe additional cpHMM inference results. In addition to the binary inference results shown in Figure 2.4C that examine burst parameter values at high and low Knirps values, we also conducted finer-grained cpHMM inference runs, in which we queried burst parameter values across the full range of Knirps concentrations observed in our experiments. The plots in Figure 2.12 summarize our results. As with the results in the main text, this inference was conducted on 15-minute-long fragments of transcriptional traces. Multiple such fragments were generated from each transcription trace by sliding a 15-minute window along each and sampling in 1 minute increments. This produced a dataset of transcriptional “reads” that were then grouped by average Knirps concentration. In addition,

we grouped transcriptional reads by experiment type (as defined in Figure 2.2B and D): no light (circles in Figure 2.12), low intensity (diamonds), and high intensity (squares).

We find that the inference results are consistent with the trends indicated in Figure 2.4C. We once again see that the burst frequency decreases with increasing Knirps concentration, though it is notable that the increased dynamic range of our inference reveals a more dramatic dependency, with burst frequency (k_{on}) dropping by a factor of 6 across the range of concentrations examined (Figure 2.12A). Additionally, we see that the burst duration ($1/k_{\text{off}}$) increases with increasing Knirps and that burst amplitude (r) remains roughly constant (Figure 2.12B and C). We note that, on its own, the Knirps-dependent increase in burst duration would actually lead to *activation*. Thus, although the burst duration exhibits Knirps-dependence, the burst frequency is the only parameter that is modulated in a manner consistent with the reduction in transcription as a result of repressor action.

However, while these findings paint a more detailed picture of how Knirps regulates transcriptional dynamics than the binary results presented in the main text, their resolution is nonetheless still limited by the fact that we must use 15-minute fragments for cpHMM inference. As a result, this approach is not suitable for recovering the true, instantaneous input-output functions that dictate how Knirps dictates burst parameter values. To make progress toward this goal, we developed a simulation-based computational framework for input-output function inference. We provide further details on this approach in the following sections.

2.7.2 Stochastic input-output simulations

Here we provide further details regarding the implementation of the simulation-based computational method that was utilized to produce the results featured in Figure 2.4F-H of the main text. Our aims in developing this method were two-fold: first, we sought to use our live imaging data to uncover burst parameter input-output functions and, second, we sought to assess whether a simple two-state model of transcriptional control based on our inference results in Figure 2.4C is *sufficient* to explain both the sharp input-output function (Figure 2.2D) and rapid reactivation dynamics (Figure 2.3D-E) revealed by our experiments.

2.7.2.1 Model specification

Our coarse-grained cpHMM burst inference results indicate that both burst frequency (k_{on}) and burst duration ($1/k_{\text{off}}$) vary as functions of Knirps concentration (Figure 2.4C). Accordingly, we employed a modeling framework in which both of these parameters vary as a function of Knirps concentration. Specifically, we model k_{on} and k_{off} as simple Hill functions of nuclear Knirps concentration (see inset panel of Figure 2.4F), such that:

$$k_{\text{on}}([\text{Knirps}]) = k_{\text{on}}^0 \frac{K_{D_{\text{on}}}^{H_{\text{on}}}}{[\text{Knirps}]^{H_{\text{on}}} + K_{D_{\text{on}}}^{H_{\text{on}}}}, \quad (2.2)$$

and

$$k_{\text{off}}([\text{Knirps}]) = k_{\text{off}}^0 \frac{K_{D_{\text{off}}}^{H_{\text{off}}}}{[\text{Knirps}]^{H_{\text{off}}} + K_{D_{\text{off}}}^{H_{\text{off}}}}. \quad (2.3)$$

where k_{on}^0 and k_{off}^0 set the upper limits for on and off rates, respectively; where the Hill coefficient H_{on} and H_{off} set the sharpness of each parameter's response to increasing Knirps concentration; and where $K_{D_{\text{on}}}$ and $K_{D_{\text{off}}}$ dictate the half-max points for the k_{on} and k_{off} input-output curves. Finally, we assume that the burst amplitude, r , takes on a fixed value that does not vary as a function of Knirps concentration.

2.7.2.2 Stochastic simulations

We can use Equations 2.2 and 2.3 to generate simulated fluorescent traces with burst dynamics that vary as a function of nuclear Knirps concentration. To do this, we first sample real single-cell Knirps concentrations from (i) the three illumination conditions shown in Figure 2.2B and (ii) the reactivation experiments shown in Figure 2.3B-D (see also Figure 2.4D and E), and use these to generate time-dependent burst parameter trends. Figure 2.13A shows an illustrative time trace of Knirps concentration and panel Figure 2.13B shows the corresponding k_{on} (blue curve) and k_{off} (red curve) trends generated by plugging that trace into Equations 2.2 and 2.3. Note that the burst duration can be obtained simply by taking the inverse of the k_{off} trend. These burst parameter trends are used to simulate an ON/OFF promoter trajectory (Figure 2.13C), which, in turn, is used to generate a predicted MS2 trace (Figure 2.13D) with Knirps-dependent burst dynamics.

To simulate promoter trajectories with concentration-dependent burst parameters, we used a discrete implementation of the widely used Gillespie Algorithm [66], in which the promoter state is sampled with a time resolution of 1 second. We provide a brief overview of the approach here, and direct readers to the [Github](#) repository accompanying this work for further details regarding the algorithm's implementation. Consider the time-varying burst parameter trends shown in Figure 2.13B, along with the simulated ON/OFF promoter trajectory in Figure 2.13C. At 11 minutes, we see that the promoter switches into the OFF state. In a standard Gillespie simulation with constant burst parameters, we would obtain the time until the next transition, τ_{OFF} , by drawing a random sample from an exponential distribution with rate parameter $\lambda = k_{\text{on}}$, such that

$$\tau_{\text{OFF}} \sim \text{Exp}(k_{\text{on}}). \quad (2.4)$$

At time $11 + \tau_{\text{OFF}}$, the promoter would then transition out of the OFF state and into the ON state.

Our case is more complicated, however, since k_{on} may change over time as the nuclear Knirps concentration changes. One simple way to capture this time-dependence is to adopt a discrete approach to promoter state simulations. In this approach, we designate some finite simulation time resolution, Δt . Starting again at $t = 11$ minutes (with the promoter in the OFF state), the algorithm proceeds as follows:

1. Use Equation 2.2 to calculate k_{on} based off of the current Knirps concentration
2. Sample an expected jump time τ
 - if promoter is OFF, sample τ from an exponential distribution with rate parameter k_{on}
 - else, sample τ from an exponential distribution with rate parameter k_{off}
3. Compare τ to Δt
 - if $\tau \geq \Delta t$: the promoter state remains unchanged
 - else, if $\tau < \Delta t$: change the promoter state (OFF to ON in our case)
4. Increment the time variable such that $t = t + \Delta t$, and return to (1).

To understand why we see this discrete rejection procedure for sampling the jump time τ is valid, consider the probability that the promoter remains in the OFF state for longer than n time steps ($P(\tau_{\text{OFF}} > n\Delta t)$). If we were sampling τ_{OFF} directly from the exponential distribution—as is the case for the standard Gillespie Algorithm—the probability of this outcome would be given by:

$$P(\tau_{\text{OFF}} > n\Delta t) = e^{-nk_{\text{on}}\Delta t}. \quad (2.5)$$

In our discrete rejection-based approach, the probability that $\tau_{\text{OFF}} > n\Delta t$ is given by the joint probability that independently sampled values of τ , drawn at each iteration, are less than the sampling time resolution Δt . The fact that each sample is independent means that the joint probability takes the form of a product:

$$P_{\text{samp}}(\tau_{\text{OFF}} > n\Delta t) = P(\tau_{\text{OFF}} > \Delta t)_{t_1} \times P(\tau_{\text{OFF}} > \Delta t)_{t_2} \times \dots P(\tau_{\text{OFF}} > \Delta t)_{t_n}. \quad (2.6)$$

Simplifying, we see that the discretely sampled probability exactly equals the true probability

$$P_{\text{discrete}}(\tau_{\text{OFF}} > n\Delta t) = [P(\tau_{\text{OFF}} > \Delta t)_{t_1}]^n = e^{-nk_{\text{on}}\Delta t}. \quad (2.7)$$

The main advantage of our discrete approach relative to the standard Gillespie Algorithm is that we are able to change the rate parameter (k_{on} or k_{off}) at each sampling step to reflect changing Knirps concentrations. This leads to sampled jump time distributions of the form:

$$P_{\text{samp}}(\tau_{\text{OFF}} > n\Delta t) = e^{-k_{\text{on}}([Knirps]_{t_0})} \times e^{-k_{\text{on}}([Knirps]_{t_1})} \times \dots e^{-k_{\text{on}}([Knirps]_{t_n})} \quad (2.8)$$

that reflect the effects of dynamic transcription factor concentrations.

Thus, by following our discrete sampling procedure, we obtain a discrete time trace of promoter activity, $\mathbf{p}(t)$, that reflects time-dependent changes to the transition rates k_{on} and k_{off} due to changes in Knirps concentration. We set $\Delta t = 1$ second, such that the

resolution of our discrete sampling is significantly faster than the promoter burst dynamics being simulated (defined by k_{on} and k_{off} ; see Figure 2.4C). By enforcing this separation of timescales, we ensure that our discretely sampled time trace is a good approximation of a continuous Knirps-dependent trajectory.

Unlike k_{on} and k_{off} , we assume that the initiation rates, r_0 and r_1 , which encode the rate of Pol II initiation in the OFF and ON states, respectively, are Knirps-independent. Note that, for simplicity, we refer to r_1 simply as “ r ” in the main text, and do not discuss results for r_0 since $r_0 \approx 0$. Thus, to obtain a predicted time series of initiation rates, $\mathbf{r}$ from promoter states $\mathbf{p}$, we simply, set $\mathbf{r} = r_0$ for all time points when the promoter is OFF and $\mathbf{r} = r_1$ for all time points when the promoter is ON (see inset panel of Figure 2.13C). Finally, we obtain a predicted MS2 trace shown in Figure 2.13D by convolving $\mathbf{r}$ with the kernel κ_{MS2} (Figure 2.13D, inset), which has the effect of taking a moving sum of past initiation rates over a time window defined by the time required for which nascent polymerase molecules remain on the gene body (set to 140 seconds throughout this work). This procedure also accounts for the finite amount of time needed for newly initiated Pol II molecules to transcribe the MS2 cassette and become fluorescent. We direct readers to Appendix D of [102] for further details.

2.7.2.3 Parameter sweeps

We used parameter sweeps to systematically test model performance across a broad range of plausible parameter values. As illustrated in Figure 2.13E, we performed a gridded sweep across 15 different values for $K_{D_{\text{on}}}$ and H_{on} from Equation 2.2. In addition we sampled 15 values each for $K_{D_{\text{off}}}$ and H_{off} (not pictured) from Equation 2.3, making for a total of $15^4 = 60625$ distinct parameter combinations. The remaining parameters, namely k_{on}^0 , k_{off}^0 , r_0 , and r_1 were held fixed at their average values as calculated from the Knirps-dependent inference results shown in Figure 2.12A-C. Table 2.3 specifies the values and value ranges used for this procedure.

For each combination of parameter values, the procedure outlined in Figure 2.13A-D was used to generate ensembles of simulated fluorescent traces with realistic Knirps-dependent burst parameters using real experimental measurements of Knirps concentration over time (Figure 2.13F). We could then use these trace ensembles to calculate predictions for the fluorescence vs. [Knirps] input-output function and reactivation cumulative distribution function (CDF, Figure 2.13G). By comparing our model predictions to our experimental results (Figure 2.13G), it was possible to assess whether a given set of model parameters was sufficient to recapitulate these key features of Knirps repression.

We used the mean-squared error to assess model fits to the input-output function and reactivation CDF. In each case, deviations were normalized by the mean of the experimental curve to ensure comparable scaling between the fluorescence input-output errors (which are natively in arbitrary units) and CDF errors (which are probabilities). For the fluorescent input-output function (Figure 2.4G) this gives

$$\delta_{io}^2 = \frac{1}{N_k} \sum_{k=1}^{N_k} \left(\frac{f_k - \hat{f}_k}{\mu_f} \right)^2, \quad (2.9)$$

where N_k is the number of Knirps concentration bins for which the average was calculated, μ_f is the average fluorescence of the experimental curve in [Figure 2.4G](#) taken across all N_k points, and where f_k and $\hat{f}_k$ are the observed and predicted fluorescent values for Knirps concentration group k . Similarly, for the reactivation CDF we have

$$\delta_{ra}^2 = \frac{1}{N_t} \sum_{t=1}^{N_t} \left(\frac{p_t - \hat{p}_t}{\mu_p} \right)^2, \quad (2.10)$$

where N_t is the number of time points post-reactivation that were considered, μ_p is the average probability taken across the CDF in [Figure 2.4H](#), and where p_t and $\hat{p}_t$ are the observed and predicted fraction of reactivated nuclei at time point t post Knirps export.

We defined the total error in model fit as the weighted sum of δ_{io}^2 and δ_{ra}^2 , such that

$$\delta^2 = (w_{io}\delta_{io}^2 + w_{ra}\delta_{ra}^2)(N_k + N_t), \quad (2.11)$$

where the sum $(N_k + N_t)$ up-weights δ^2 according to the total number of data points considered, and where w_{io} and w_{ra} are weight parameters that tune the relative impact of δ_{io}^2 and δ_{ra}^2 to the total loss, δ^2 . These weights can be adjusted to navigate tradeoffs between the minimization of input-output and reactivation CDF fitting loss. In our case, we find that values of $w_{io} = 1/4$ and $w_{ra} = 3/4$ lead to the best visual alignment between model predictions and experimental observations.

2.7.2.4 Estimating uncertainty bounds with MCMC

The parameter sweep procedure outlined above produced a δ^2 estimate for each of the 60625 parameter combinations considered. In principle, the model realization corresponding to the lowest δ_t^2 could be selected to obtain an approximate point estimate for the optimal K_{Don} , H_{on} , K_{Doff} , and H_{off} values; however the parameter sweep results are not alone sufficient to obtain uncertainty bounds, nor do they provide insights into the remaining parameters not included in the sweep. To obtain this information, we employed Markov Chain Monte Carlo (MCMC) to sample the posterior distributions of our model parameters, conditional on our experimental data. MCMC is a widely used class of algorithms that are capable of efficiently sampling high-dimensional probability distributions [\[65\]](#).

As a first step in this process, we utilize information from the parameter sweeps to obtain parameter priors that are used to initialize and constrain MCMC sampling. To do this, we generate a weight vector, $\mathbf{w}$, comprised of terms with the form

$$w_i = e^{-\delta_i^2}, \quad (2.12)$$

where δ_i^2 is the total loss from Equation 2.11 for the i th set of parameter values. If we assume that model errors are approximately Gaussian-distributed, then each w_i can be interpreted as an unnormalized probability that is proportional to the likelihood of the data $\mathbf{x}$ (the input-output and reactivation curves) conditional on the i th parameter set $\boldsymbol{\theta}_i$:

$$w_i \propto P(\mathbf{x}|\boldsymbol{\theta}_i). \quad (2.13)$$

Moreover, from Bayes' Theorem we have that

$$w_i \propto P(\mathbf{x}|\boldsymbol{\theta}_i)P(\boldsymbol{\theta}_i) = P(\boldsymbol{\theta}_i|\mathbf{x})P(\mathbf{x}). \quad (2.14)$$

From here, we see that if we take a uniform prior across all $\boldsymbol{\theta}_i$ values (such that $P(\boldsymbol{\theta}_i)$ is a constant), then the weight w_i will be proportional to the likelihood of each set of parameter values, conditional on the experimental data:

$$w_i \propto P(\boldsymbol{\theta}_i|\mathbf{x}). \quad (2.15)$$

Motivated by this observation, we resampled the parameter values, $\boldsymbol{\theta}$, surveyed in the parameter sweep according to the weight vector $\mathbf{w}$. This leads to a new set of parameter values, $\boldsymbol{\theta}^*$, where the frequency of a given parameter vector, $\boldsymbol{\theta}_i$, is proportional to its likelihood. As a result, the best-fitting parameter sets will appear more frequently in $\boldsymbol{\theta}^*$, and the worst-fitting are unlikely to appear at all. We calculate prior distributions for K_{Don} , H_{on} , K_{Doff} , and H_{off} (assumed to be Gaussian) by taking the mean and standard deviation of each parameters values across $\boldsymbol{\theta}^*$. The prior distributions for k_{on}^0 , k_{off}^0 , and r_1 were initialized using the Knirps-dependent cpHMM inference results shown in Figure 2.12A-C. Specifically, the mean and standard deviation for k_{on}^0 and k_{off}^0 were estimated using the mean and standard deviations of the intercepts of the linear fits shown in Figure 2.12A and B, which we reasoned should provide reasonable estimates for the upper limit of each parameter. Given the lack of strong Knirps-dependence in the burst amplitude, the mean and standard deviation for the r_1 prior were calculated by taking the mean and standard deviation of all cpHMM results shown in Figure 2.12C. The initiation rate when the system is in the OFF state, r_0 , was not subject to MCMC sampling, and was held fixed at its mean value from cpHMM inference. See Table 2.4 for the precise values used for each parameter prior.

With our prior distributions established, we conducted MCMC sampling to obtain estimates for the posterior distribution of each parameter. We conducted 24 independent MCMC simulations, each of which was run for 2500 total steps. We used standard Metropolis Hastings [123] updates during sampling. The procedure for each step was as follows:

1. At the t th step in the simulation, a new proposal for the parameter vector, $\boldsymbol{\theta}'_t$, was generated by sampling from a multivariate normal distribution centered at the parameter values from the previous step, such that

$$\boldsymbol{\theta}'_t \sim \mathcal{N}(\boldsymbol{\theta}_{t-1}, \Sigma). \quad (2.16)$$

The covariance matrix, Σ , dictates how large or small the randomly proposed jumps tend to be relative to the previous parameter values. We assumed Σ to be a diagonal matrix and set each component, σ_i , to be equal to 15% of the standard deviation of the corresponding parameter's prior distribution.

2. Next, we used the proposed parameters, θ'_t , to simulate populations of MS2 traces and calculate predictions for the fluorescence vs. Knirps curve (Figure 2.4G) and reactivation CDF (Figure 2.4H) as outlined in the preceding sections.
3. We then calculated the total likelihood of the new parameters, defined as

$$P(\theta'_t|\mathbf{x}) = P(\mathbf{x}|\theta'_t)P(\theta'_t). \quad (2.17)$$

Here the first term on the right-hand-side is as defined in Equations 2.12 and 2.13, and functions to penalize proposals that produce curves that deviate too far from experimental measurements. The second component is the prior probability, and has the effect of penalizing proposals that deviate too far from our priors regarding parameter values.

4. Finally, we perform the standard Metropolis-Hastings move [123, 140]. We calculate a probability, p , that takes the form

$$p = \min\left\{\frac{P(\theta'_t|\mathbf{x})}{P(\theta_{t-1}|\mathbf{x})}, 1\right\}, \quad (2.18)$$

where $P(\theta_{t-1}|\mathbf{x})$ is the likelihood of the previous set of parameter values. Next we draw a random number, z , from the uniform distribution ($z \sim \mathcal{U}[0, 1]$). If $p \geq z$: $\theta_t = \theta'_t$. Otherwise: $\theta_t = \theta_{t-1}$.

2.7.2.5 Additional MCMC results

Figure 2.14 contains bivariate density plots and univariate histograms illustrating the results of MCMC sampling for each of the seven parameters examined. The results for the burst frequency (k_{on}) are as quoted in the main text. We find that, like k_{on} , k_{off} has a negative dependence on ($H_{\text{off}} = 3.2 \pm 0.65$). This translates to a burst duration that is predicted to *increase* as a function of increasing Knirps concentration (Figure 2.14C). On its own, this trend would *increase eve* 4+6 activity; however, this effect is dominated by the stronger Knirps-dependent decrease in k_{on} , leading to a strong overall repressive effect (see Figure 2.4G). Additionally, our sampling returns a burst amplitude (r_1) value of 21.6 ± 1.9 au/min.

2.7.3 Implementation of the thermodynamic binding model

Here we provide a brief description of the theoretical underpinnings of the binding model that was used to generate the solid black curve in [Figure 2.4F](#). The core assumption of this model is that k_{on} is proportional to the number of Knirps molecules bound to the locus, such that

$$k_{\text{on}} = k_{\text{on}}^0 \left(1 - \frac{n_b}{N}\right), \quad (2.19)$$

where k_{on}^0 is the maximum burst frequency value (set to the 2.8 min^{-1} value returned by MCMC inference), n_b is the number of Knirps molecules bound, and N is the total number of binding sites along the enhancer. Using PATSER scores for the *eve* 4+6 enhancer, we assess that there are 10 Knirps binding sites along the enhancer, such that $N = 10$ (see [Figure 2.8](#)). Thus, in this model $k_{\text{on}} = 0$ when N sites are bound and $k_{\text{on}} = k_{\text{on}}^0$ when 0 sites are bound.

Knirps-dependence enters into [Equation 2.19](#) through n_b , which should vary as a function of Knirps concentration. Note that, for ease of notation, we denote Knirps concentration by $[R]$ (as opposed to $[Knirps]$) throughout this appendix. To model n_b , we adopt the simple equilibrium chain model developed in [\[101\]](#). Briefly, this model assumes that all binding sites are identical, such that there are only $N + 1$ distinct binding states in which the enhancer can exist, ranging from 0 sites bound to all N sites bound. In this model, the Knirps concentration will induce a probability distribution over the set of possible binding states. Each state's probability is given by

$$p(n_b; [R]) = \frac{W(n_b) \omega^{\frac{n_b(n_b-1)}{2}} \left(\frac{K_D}{[R]}\right)^{-n_b}}{Z}, \quad (2.20)$$

where K_D is the dissociation constant for Knirps binding to specific sites at the locus. We assume that any pair of Knirps molecules can interact with a cooperativity factor ω . Given n_b bound Knirps molecules, there are $\frac{n_b(n_b-1)}{2}$ such pairwise interactions and, hence, a cooperativity contribution of $\omega^{\frac{n_b(n_b-1)}{2}}$. Further, $W(n_b)$ accounts for the number of different microscopic binding configurations that correspond to each macroscopic binding state (i.e., in how many different configurations can n_b Knirps molecules be bound?):

$$W(n_b) = \frac{N!}{(N - n_b)! n_b!}. \quad (2.21)$$

Lastly, we note that the denominator Z in [Equation 2.20](#) is a normalizing factor equal to the sum of all $N + 1$ numerators. We direct the reader to Appendix C.3 of [\[101\]](#) for a detailed derivation of [Equation 2.20](#).

We then use this expression for $p(n_b)$ to calculate the average expected number of bound Knirps molecules as a function of nuclear Knirps concentration, such that

$$\langle n_b([R]) \rangle = \sum_{n_b=0}^N p(n_b; [R]) n_b. \quad (2.22)$$

Finally, we plug Equation 2.22 into Equation 2.19 for the experimentally observed range of Knirps concentrations to produce a predicted burst frequency vs. Knirps input-output curve like the one shown in Figure 2.4F.

Using this approach, we conduct nonlinear least squares fitting to identify the values of ω and K_D that best fit the blue k_{on} trend in Figure 2.4F that was returned by our MCMC inference. Our fit indicates that K_D and ω values of 70.8 au and 1.9, respectively, produce the optimal fit. We find that the k_{on} trend generated by these parameter values (black in Figure 2.4F) is in close agreement with our MCMC inference result (blue curve), demonstrating that simple equilibrium binding could explain Knirps regulation of the burst frequency.

2.7.4 Comparison to other optogenetic approaches developed for multicellular organisms

In this work, we build on previous works using the LEXY technology [125, 98, 164, 150, 124] and demonstrate the power of the LEXY system for modulating protein dynamics inside developing embryos. The LEXY tag-based method addresses several key limitations faced by many previously reported methods.

First, some optogenetic tools are designed for specific signaling pathways [175, 86, 92, 93, 95, 133], and receptor [142] targets, and as a result, are not readily generalizable. In contrast, LEXY can be directly attached to any protein (though issues of genetic rescue [150] and its modulation strength [98] remain).

Second, many optogenetic tags do not act through concentration modulation, which makes it difficult to draw quantitative conclusions from the results. For example, the blue light-induced dimerization of *Arabidopsis* cryptochrome 2 (CRY2) controls downstream transcription by disrupting the function of the tagged protein through multimerization without affecting its concentration [24, 85, 119]. On the other hand, LEXY controls transcriptional activity through direct modulation of the protein concentration within the nucleus, allowing for easy quantification and straightforward interpretation.

Chapter 3

A bistable autoregulatory module in the developing embryo commits cells to binary expression fates

This chapter is a reproduction of reference [\[185\]](#)

Abstract

Bistable autoactivation has been proposed as a mechanism for cells to adopt binary fates during embryonic development. However, it is unclear whether the autoactivating modules found within developmental gene regulatory networks are bistable, unless their parameters are quantitatively determined. Here, we combine *in vivo* live imaging with mathematical modeling to dissect the binary cell fate dynamics of the fruit fly pair-rule gene *fushi tarazu* (*ftz*), which is regulated by two known enhancers: the early (non-autoregulating) element and the autoregulatory element. Live imaging of transcription and protein concentration in the blastoderm revealed that binary Ftz fates are achieved as Ftz expression rapidly transitions from being dictated by the early element to the autoregulatory element. Moreover, we discovered that Ftz concentration alone is insufficient to activate the autoregulatory element, and that this element only becomes responsive to Ftz at a prescribed developmental time. Based on these observations, we developed a dynamical systems model and quantitated its kinetic parameters directly from experimental measurements. Our model demonstrated that the *ftz* autoregulatory module is indeed bistable and that the early element transiently establishes the content of the binary cell fate decision to which the autoregulatory module then commits. Further *in silico* analysis revealed that the autoregulatory element locks the Ftz fate quickly, within 35 min of exposure to the transient signal of the early element. Overall, our work confirms the widely held hypothesis that autoregulation can establish developmental fates through bistability and, most importantly, provides a framework for the quantitative dissection of cellular decision-making.

3.1 Introduction

One of the central questions in developmental biology concerns how cells precisely and irreversibly adopt distinct cellular fates. It has been argued that cells assume their unique gene expression profiles through a sequence of decisions among branching paths [183, 153], famously encapsulated by “Waddington’s landscape” of peaks and valleys delineating the possible trajectories that a cell can follow [171]. Genetic networks that lock a cell into one of these trajectories may be thought of as “memory modules” that guide cells through valleys in the landscape to their ultimate fates. In the simplest case, where a decision is made between two alternative developmental fates, the memory module is binary and referred to as a switch. The state of the switch is set by the action of transient upstream regulatory signals.

Several genetic motifs, including autoactivation and mutual repression, are capable of maintaining binary cell fates [4, 134]. However, the mere presence of these motifs is insufficient to guarantee that a network can remember its expression state once upstream regulators have degraded. The ability to lock onto high or low expression levels results from bistability (see “A primer on bistability” in Materials and Methods), a systems-level property that depends upon the quantitative details of the kinetics of the involved chemical reactions [54, 5, 69].

Though bistability is widely invoked to explain cell fate decisions [134], relatively little quantitative data exist to confirm bistability in gene expression modules within developing embryos. Previous studies in cell culture and fixed embryos have provided evidence for bistability in hematopoietic differentiation [105, 99], the Shh network [100], the vertebrate hindbrain [19], between the BMP and FGF morphogens [158], and within the Notch-Delta signaling system [157]. Quantitative evidence for multistability in fruit fly embryos has also been derived from fitting the parameters of high-dimensional network models to fixed tissue measurements [44, 89, 113, 116, 129, 167]. While these models are capable of reproducing the observed phenomenology, there is no guarantee that the optimal set of inferred parameter values reflects actual biophysical quantities [73, 37, 168, 178]. Thus, it is important to verify that the conclusions drawn from computational modeling and *in vitro* experiments apply to developmental systems *in vivo* in the context of models that quantitatively capture the molecular interactions that underlie cellular decision-making. To the best of our knowledge, evidence for the bistability of a genetic module based on these molecular interactions in an intact multicellular organism has not yet been demonstrated.

The early development of the fruit fly *Drosophila melanogaster* is an ideal model system for studying binary cell fate determination, due to the presence of pair-rule genes such as *fushi tarazu* that form discrete stripes at the cellular blastoderm stage prior to gastrulation (2.5 – 3.5 hours after fertilization; [126, 74, 172, 176, 83]). The expression of *ftz* is regulated by two main enhancers: the early, or zebra, element and the autoregulatory element (Figure 3.1A; Figure 3.8; [82]). The early element responds to upstream transcription factors such as the gap genes to establish the initial expression pattern of seven stripes [45]. This element is functionally distinct from the autoregulatory element, which contains multiple

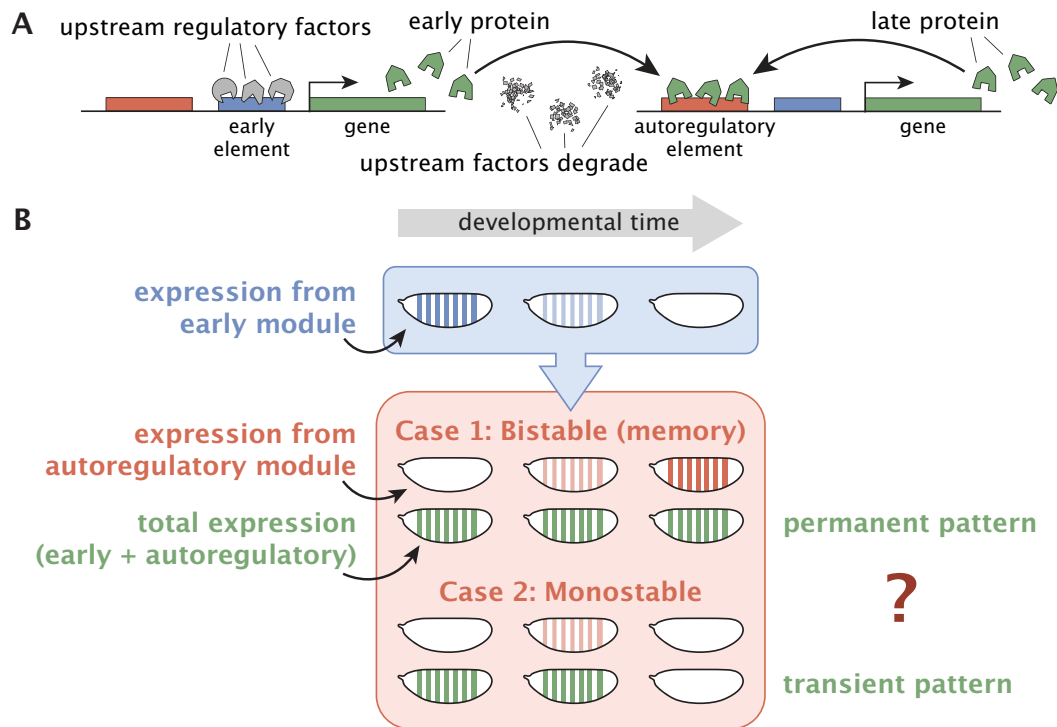


Figure 3.1: **The bistability of the autoregulatory module can determine cell fate.** (A) The autoregulatory architecture consists of an early element to which upstream factors bind to transiently upregulate gene expression of an activator, and an autoregulatory element to which the activator binds to, under certain circumstances, promote its own expression even once the upstream factors have degraded. (B) The total activator concentration (green) is a sum of the protein produced by the early (blue) and autoregulatory (red) modules. If the autoregulatory module is bistable, it possesses binary memory that permits transiently high concentrations of early protein to be locked into permanent high expression levels (high cell fate) as shown in Case 1. If, in contrast, the autoregulatory module is monostable, then it may transiently boost protein levels from the early module, but over time all cells will ultimately revert to the same low fate as depicted in Case 2.

Ftz binding sites that allow Ftz to activate its own expression (Figure 3.8; [138, 146, 145]). This autoactivation network motif is theoretically capable of exhibiting bistability and has therefore been hypothesized to act as a binary memory module [4, 179].

Whether a cell possesses a memory module determines whether observed states of gene expression are transient in the absence of continued external signaling, or whether these states can be locked into permanent cell fates that can be maintained without further intervention. Specifically, if the autoregulatory module is bistable, then it maintains high *ftz* expression driven by the transient presence of upstream factors (Figure 3.1B, case 1), even once those factors degrade (or until further regulatory mechanisms intervene). If, instead, the autoregulatory element is monostable, then the observed separation of Ftz concentration into high and low levels persists only as long as upstream factors are present to regulate expression. In their absence, Ftz expression would revert to a single fate for all cells (Figure 3.1B, case 2). It is important to note, however, that in this case, the transiently high or low trajectory of Ftz concentration could still be instructive for regulating downstream genes.

Here, we characterize the *ftz* autoregulatory module *in vivo* through quantitative real-time measurements in living fruit fly embryos. Focusing on stripe 4 expression, we observe that Ftz expression separates into discrete high and low levels at the blastoderm stage during the 20 min prior to gastrulation, concurrent with a transition in regulatory control from the early to the autoregulatory element. We discover that autoregulation is triggered at a specific time point in development—presumably through the action of “timer genes” [29, 31, 30]—rather than through a readout of Ftz concentration alone. Based on these observations, we develop a dynamical systems model and quantitate its parameters from simultaneous real-time measurements of *ftz* transcription and Ftz protein dynamics in single cells of living embryos. Our model predicts binary Ftz expression levels at gastrulation with high accuracy and demonstrates that, indeed, the *ftz* autoregulatory module is bistable. We conclude that the *ftz* autoregulatory element acts as a memory module to commit cells to binary fates that are otherwise transiently defined by the early element, thereby validating a long-standing hypothesis in developmental and systems biology that bistability underlies cell fate determination. Simulations further make it possible to quantitatively define a developmental commitment window, which shows that the autoregulatory module requires about half an hour to establish a memory of the transient signal from the early module. Thus, our work provides a framework for the dissection of other regulatory modules in the gene regulatory networks that dictate development based on this interplay between dynamical systems models and real-time experiments.

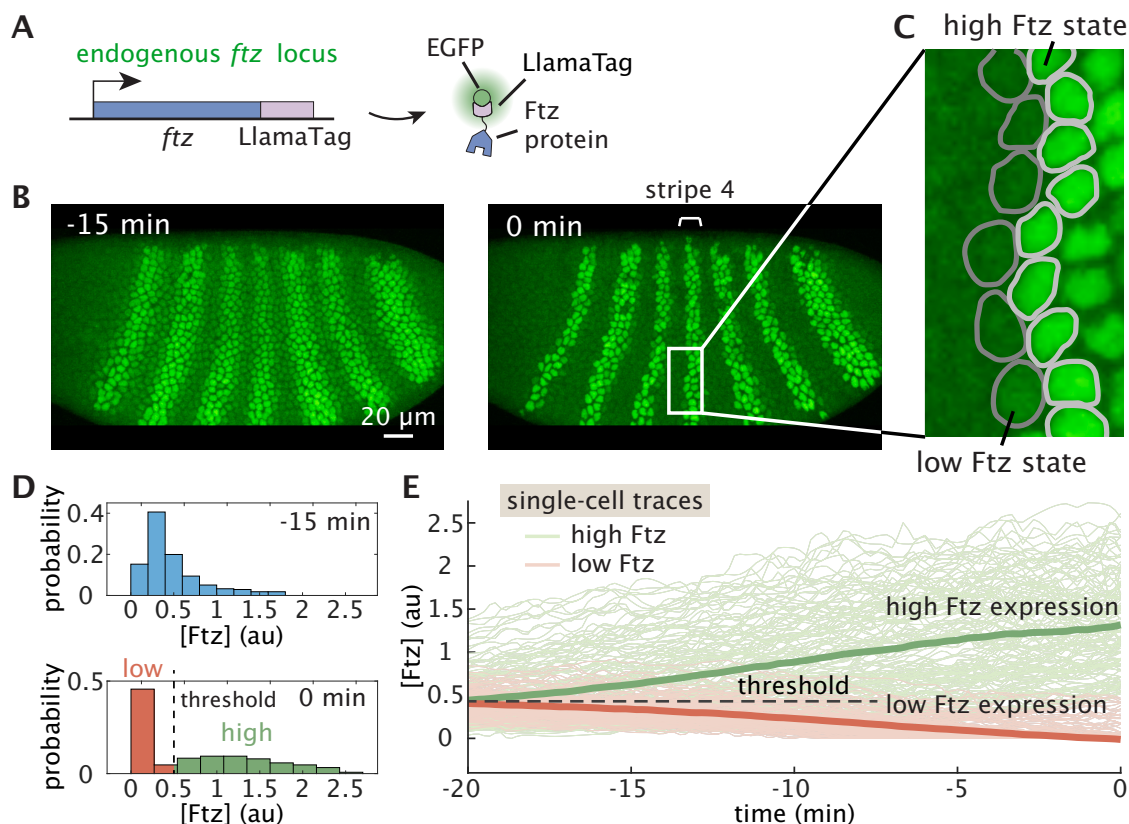


Figure 3.2: **Binary cell states are rapidly established prior to gastrulation.** (A) A fusion of endogenous Ftz to a LlamaTag makes it possible to visualize the highly dynamic Ftz protein pattern in the early fly embryo. Once Ftz protein is translated in the cytoplasm, LlamaTag binds to maternally deposited EGFP and is transported into the nucleus, increasing nuclear fluorescence to produce a direct readout of Ftz protein concentration. (B) Snapshots from a movie capturing Ftz protein concentration dynamics. The anterior side of the embryo is oriented towards the left, and the time is given relative to gastrulation. (C) Ftz expression along the anterior boundary of stripe 4 shows a discrete transition between distinct cell states. (D) Histograms of single-nucleus fluorescence values at different developmental time points show that a single threshold can be used to classify cells into high Ftz and low Ftz cell states at gastrulation. (E) Single-cell trajectories of nuclei at the anterior boundaries of Ftz stripe 4. Green and red lines are averages for nuclei determined to ultimately have high and low Ftz levels at gastrulation, respectively, as defined in (D).

3.2 Results

3.2.1 Binary Ftz expression states are established in early development

To understand the role of autoactivation in deciding Ftz expression levels, we first visualized Ftz protein dynamics over time. We used CRISPR-mediated recombination [70] to fuse a LlamaTag, a fluorescent probe that reports on the fast protein dynamics that characterize early embryonic development, to the C-terminus of the endogenous Ftz (Figure 3.2A; [18]). An examination of the fluorescently labeled Ftz in the early embryo shows that, around 15 min before gastrulation, Ftz is expressed in a seven-stripe pattern with clear, smooth boundaries (Figure 3.2B, left). This expression pattern refines over the following 15 min into sharp stripe boundaries by the start of gastrulation (Figure 3.2B, right). The result shows that cells express either high or low levels of Ftz, as pictured in Figure 3.2C for the anterior boundary of stripe 4, consistent with results from previous studies [148, 28, 18]. Note that, throughout this work, we focus on the anterior boundary of stripe 4 because it is the only Ftz stripe that has been demonstrated to be driven exclusively by the early and autoregulatory elements and not by other enhancers in the gene’s vicinity (Figure 3.8; [148]).

Our live imaging measurements allowed us to quantitatively examine the dynamics with which binary cell states are established by calculating the Ftz protein distribution in individual nuclei at different time points in development (Figure 3.2D). Our analysis revealed that the expression level is initially unimodal across all cells (Figure 3.2D, top and E) and then evolves into a bimodal distribution within 15 min (Figure 3.2D, bottom and E). Consistent with our qualitative observations in Figure 3.2B and C, cells at these later times can be quantitatively classified into distinct “high-Ftz” and “low-Ftz” cell states using a single threshold chosen by a visual inspection of the distribution (Figure 3.2D, bottom and E), which indicates that binary cell states are already established prior to the onset of gastrulation. Moreover, though the two cell states are clearly distinguishable from each other, there is significant cell-to-cell variability within the high-Ftz state (Figure 3.2D, bottom and E).

Previous studies have established that autoregulation plays a key role during *ftz* expression: a lack of the *ftz* autoregulatory element or mutated Ftz binding sites within the element results in the loss of Ftz expression at later developmental stages [82, 146]. However, it is unclear at what developmental time *ftz* autoregulation is initiated in response to the Ftz expression stemming from the early element. Specifically, is the autoregulatory element active before stripes of Ftz expression emerge, or is autoregulation invoked after the early element has already established this pattern? To distinguish between these two scenarios, we decoupled the transcriptional dynamics driven by the early and autoregulatory elements by creating two separate reporter constructs, each containing only the early or autoregulatory elements followed by MS2 stem-loops that enable the direct visualization of transcriptional dynamics (Figure 3.3A; Figure 3.8; [63, 114]). All constructs were inserted in the same genomic location.

We observed that the early element already drives a relatively constant gene expression

level around 20 min prior to gastrulation (Figure 3.3B and C). Then, at 15 min before gastrulation, its transcriptional activity decreases significantly, resulting in a 60% reduction within the next 20 min of development (Figure 3.3B and C). Conversely, autoregulation is initiated 20 min prior to gastrulation, with its activity increasing until gastrulation starts (Figure 3.3C). This transition between the early and autoregulatory elements occurs while binary cell states are being established (Figure 3.2).

3.2.2 *ftz* autoactivation is triggered at a specific developmental time

Previous studies have established that *ftz* is autoactivated through a specialized enhancer referred to as the autoregulatory element [82, 138, 146, 145]. The autoregulatory element cannot drive transcription in the absence of Ftz, but it is unclear whether the mere presence of high Ftz concentrations is sufficient to initiate autoactivation, or whether other factors must also be present to trigger the observed transition in *ftz* expression between the early and autoregulatory elements (Figure 3.3C).

To further investigate the nature of the regulatory transition that initiates autoactivation, we quantified the gene regulatory function of the *ftz* autoregulatory element at distinct developmental times. The gene regulatory function—also called the input-output function—describes the relationship between Ftz protein concentration and the rate of *ftz* transcription driven by the autoregulatory element. Since Ftz is autoactivating, we expect the gene regulatory function to be S-shaped (i.e., sigmoidal, see Figure 3.4E, inset), with no transcription in the absence of Ftz, a steep increase in transcription as Ftz levels approach a “threshold” level, and maximal transcription for high concentrations of Ftz. If Ftz is the only factor that regulates transcription levels from the autoregulatory element, then the gene regulatory function should remain unchanged even as the concentrations of other regulatory factors evolve over time; that is, a fixed concentration of Ftz should correspond to the same transcription rate regardless of the time of measurement. Alternatively, if other time-varying factors also contribute to *ftz* expression from the autoregulatory element, then we expect the gene regulatory function to change over time such that the same level of Ftz will drive different levels of transcription depending on developmental time.

We measured the gene regulatory function of the autoregulatory element at different time points by constructing an experimental system that permits simultaneous monitoring of input Ftz concentration and corresponding output transcriptional dynamics driven by the autoregulatory element. We used tagged endogenous Ftz as the input and introduced a transgenic reporter with MS2 loops under the control of the *ftz* autoregulatory element as the output (Figure 3.4A). To prevent the reporter from expressing unlabelled Ftz, which could feed back into the autoregulatory element and interfere with the measurement of the gene regulatory function, we created a new autoregulatory construct by combining the *ftz* autoregulatory element with the coding region of the *yellow* gene. Control experiments confirmed that the autoregulatory constructs used in Figure 3.3 and Figure 3.4 show the

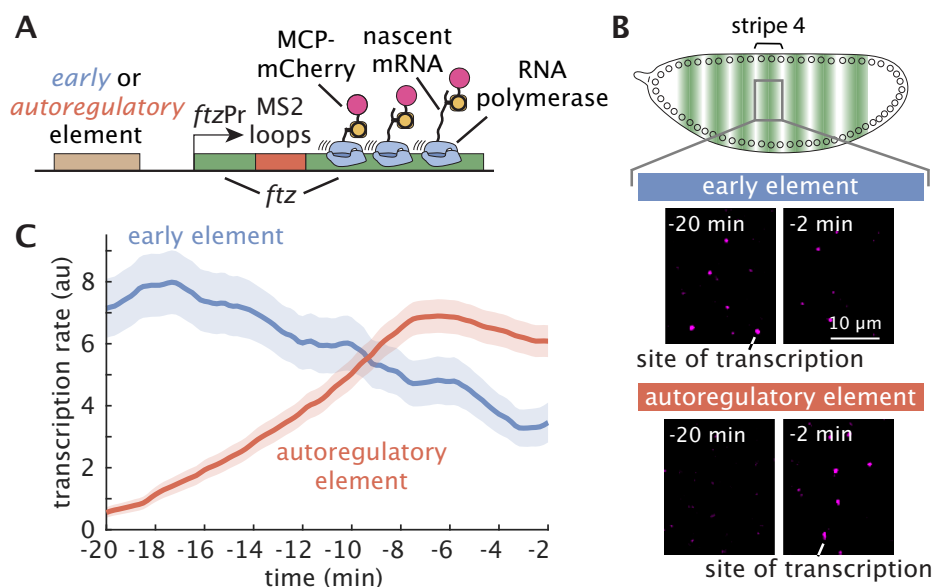


Figure 3.3: **A sequential transition in Ftz regulation from the early to the autoregulatory element occurs while discrete cell states are established.** (A) Imaging transcriptional dynamics of the early and autoregulatory elements using the MS2 system. Maternally deposited MS2 coat protein (MCP) fused to mCherry binds to MS2 stem-loops in the nascent RNA of the reporter construct. (B) Snapshots of sites of nascent transcript formation labeled by MS2 from reporters of the early and autoregulatory elements at different time points. Transcription from the early element significantly decreases, and transcription from the autoregulatory element increases shortly before gastrulation. (C) Quantification of the transcriptional activity reported by the MS2 fluorescence from the early ($N = 3$ embryos) and autoregulatory ($N = 6$ embryos) elements as a function of time confirms that, within 20 min, *ftz* gene expression transitions from originating mainly from the early element to being dominated by the autoregulatory element. MS2 traces are smoothed using a moving average of 5 min. The transcription rate is calculated from the measured MS2 signal, which is an approximation of the mRNA production rate [63, 102, 18]. (Shaded region indicates standard errors over multiple embryos.) See also Figure 3.8.

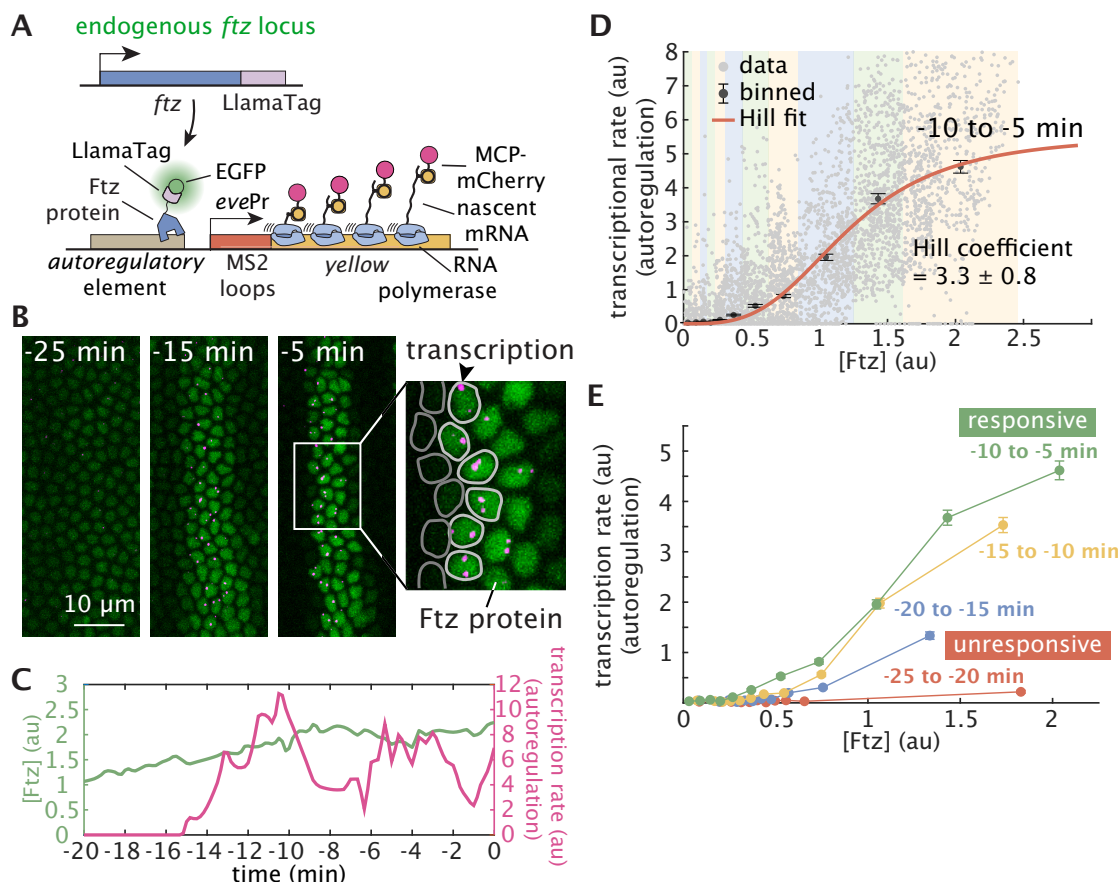


Figure 3.4: **Two-color live imaging reveals that *ftz* autoregulation is initiated at a specific developmental time.** (A) Two-color tagging permits *in vivo* simultaneously visualizes input Ftz concentration using LlamaTag and output autoregulatory transcriptional dynamics using a reporter carrying the MS2 system. (B) Representative frames from live-imaging data. Green and magenta channels correspond to Ftz concentration and transcriptional output from the *ftz* autoregulatory element, respectively. Outlines indicate cells expressing high (light gray) or low (dark gray) levels of Ftz protein. (C) Illustrative single-cell trace of Ftz and autoregulatory activity. Green and magenta lines correspond to the Ftz protein and transcriptional activity of the autoregulatory element, respectively. Both protein and MS2 traces are smoothed using a moving average of 1 min. (D) Experimentally measured gene regulatory function of the *ftz* autoregulatory element between -10 min to -5 min relative to gastrulation. Grey points correspond to simultaneous measurements of Ftz and MS2 fluorescence at individual time points from single-cell traces at the anterior boundary of stripe 4 ($N = 211$ nuclei). These points were grouped into quantiles, and a Hill function (red line) was fit to the quantile means. (continued on the next page)

Figure 3.4: (*continued*) **Two-color live imaging reveals that *ftz* autoregulation is initiated at a specific developmental time.** (E) The autoregulatory input-output function evolves over time, as the *ftz* autoregulatory element transitions from an unresponsive to a responsive state within 15 min, indicating that *Ftz* autoregulation is initiated through a developmental time-based mechanism. The transcription rate is calculated from the measured MS2 signal, which is an approximation of the mRNA production rate [63, 102, 18]. (Error bars shown indicate standard errors. All data are from $N = 7$ embryos.) See also Figure 3.9.

same temporal dynamics prior to gastrulation (Figure 3.9). Live imaging of the anterior boundary of *Ftz* stripe 4 showed that, initially, around 25 min prior to gastrulation, both *Ftz* expression and the autoregulatory response were relatively low, and later increased as development progressed (Figure 3.4B). Just before gastrulation, the *Ftz* protein pattern refined into a discrete boundary, with the *ftz* autoregulatory response clearly following the stripe boundary (Figure 3.4B).

To calculate the regulatory function, we restricted our analysis to the cells at the anterior boundary of stripe 4 as a means to minimize the influence of other position-dependent transcription factors that might also contribute to *ftz* autoactivation [145]. We first extracted two rows (high and low) of boundary cells. Then, we separated the input *Ftz* concentration and output transcription from the autoregulatory element in the data corresponding to each individual cell (Figure 3.4C) into ten quantiles and fit a Hill function to the quantile averages to get the gene regulatory function of *ftz* autoregulation within a defined temporal range (for example, -10 min to -5 min for Figure 3.4D). Our analysis revealed a sharp regulatory relationship between *Ftz* concentration and autoregulatory response shortly before gastrulation, with a Hill coefficient of 3.3 ± 0.8 (Figure 3.4D). Such Hill coefficient is comparable to those estimated in the context of autoactivation in vertebrate hindbrain development [19] as well as those observed in simpler regulatory motifs that do not feature feedback [72].

We repeated the process described above at multiple developmental times to measure the temporal dynamics of the regulatory function for *ftz* autoactivation. The results, shown in Figure 3.4E, revealed that the regulatory function is clearly distinct at different time points. Around -25 min to -20 min, the *ftz* autoregulatory element drives almost no transcription regardless of *Ftz* concentration, as can be appreciated by the fact that the gene regulatory function is almost flat near 0 during this time interval (Figure 3.4E, red line). However, within 15 min, *Ftz* levels above a threshold of about 1 au are sufficient to drive *Ftz* transcription from the autoregulatory element, with higher input *Ftz* concentrations resulting in progressively higher transcription rates at later times (Figure 3.4E, green line). The fact that *Ftz* levels above a threshold result in high transcription rates at gastrulation but produce no transcription -20 min earlier is a clear indicator that the autoregulatory element is not always equally responsive to input *Ftz*. Instead, its capacity to drive *ftz* expression is enabled at a specific developmental time, presumably by upstream transcription factors.

3.2.3 Dynamical systems modeling of *ftz* regulation

The fact that *Ftz* can exhibit a high state at one point in time does not imply that this high state will persist in the absence of upstream regulation (Figure 3.1B). Determining whether the autoregulatory module is bistable requires turning the schematic in Figure 3.1A (including both early and autoregulatory elements) into an explicit dynamical systems model with empirically determined parameter values.

We begin by mathematically describing the expression dynamics driven by the *ftz* early enhancer. We assume that upstream regulators binding to the early enhancer dictate the transcription rate $r(t)$ (Figure 3.5A). Our measurements (Figure 3.3C; Figure 3.5I and J; Materials and Methods “Early element transcriptional activity decay rate β ”) revealed that, after the onset time for the autoregulatory enhancer $t_{on} = -20$ min (Figure 3.4E), $r(t)$ follows an approximately exponential decay with a decay constant $1/\beta$. This allows us to represent $r(t)$, the mRNA $R_{early}(t)$, and protein $P_{early}(t)$ produced from the early module as

$$\begin{cases} \frac{dr}{dt} &= -\beta r(t) \\ \frac{dR_{early}}{dt} &= \underbrace{r(t)}_{\text{synthesis}} - \underbrace{\gamma_R R_{early}(t)}_{\text{degradation}} \\ \frac{dP_{early}}{dt} &= \underbrace{\alpha R_{early}(t)}_{\text{synthesis}} - \underbrace{\gamma_P P_{early}(t)}_{\text{degradation}}, \end{cases} \quad t \geq t_{on}, \quad (3.1)$$

where α is the translation rate and γ_R , γ_P are the mRNA and protein degradation rates, respectively. Regardless of the choice of parameter values or initial conditions, all three quantities $r(t)$, $R_{early}(t)$, and $P_{early}(t)$ converge to 0 as t goes to infinity (Materials and Methods “Modeling - Early element”), indicating that the early module is monostable and signaling from upstream factors is transient. Since embryos at this stage in *nc14* are still undergoing cellularization and cell membranes have not fully formed, we assume that diffusion of mRNA and protein is negligible [89, 118, 167].

As shown in Figure 3.4E, after the onset time t_{on} , the autoregulatory element becomes responsive to *Ftz*. Building on the simple model introduced in Materials and Methods “A primer on bistability”, we describe the mRNA ($R_{late}(t)$) and protein ($P_{late}(t)$) dynamics dictated by this element as

$$\begin{cases} \frac{dR_{late}}{dt} &= \underbrace{cf(P_{early}(t) + P_{late}(t))}_{\text{synthesis}} - \underbrace{\gamma_R R_{late}(t)}_{\text{degradation}} \\ \frac{dP_{late}}{dt} &= \underbrace{\alpha R_{late}(t)}_{\text{synthesis}} - \underbrace{\gamma_P P_{late}(t)}_{\text{degradation}} \end{cases}, \quad t \geq t_{on}. \quad (3.2)$$

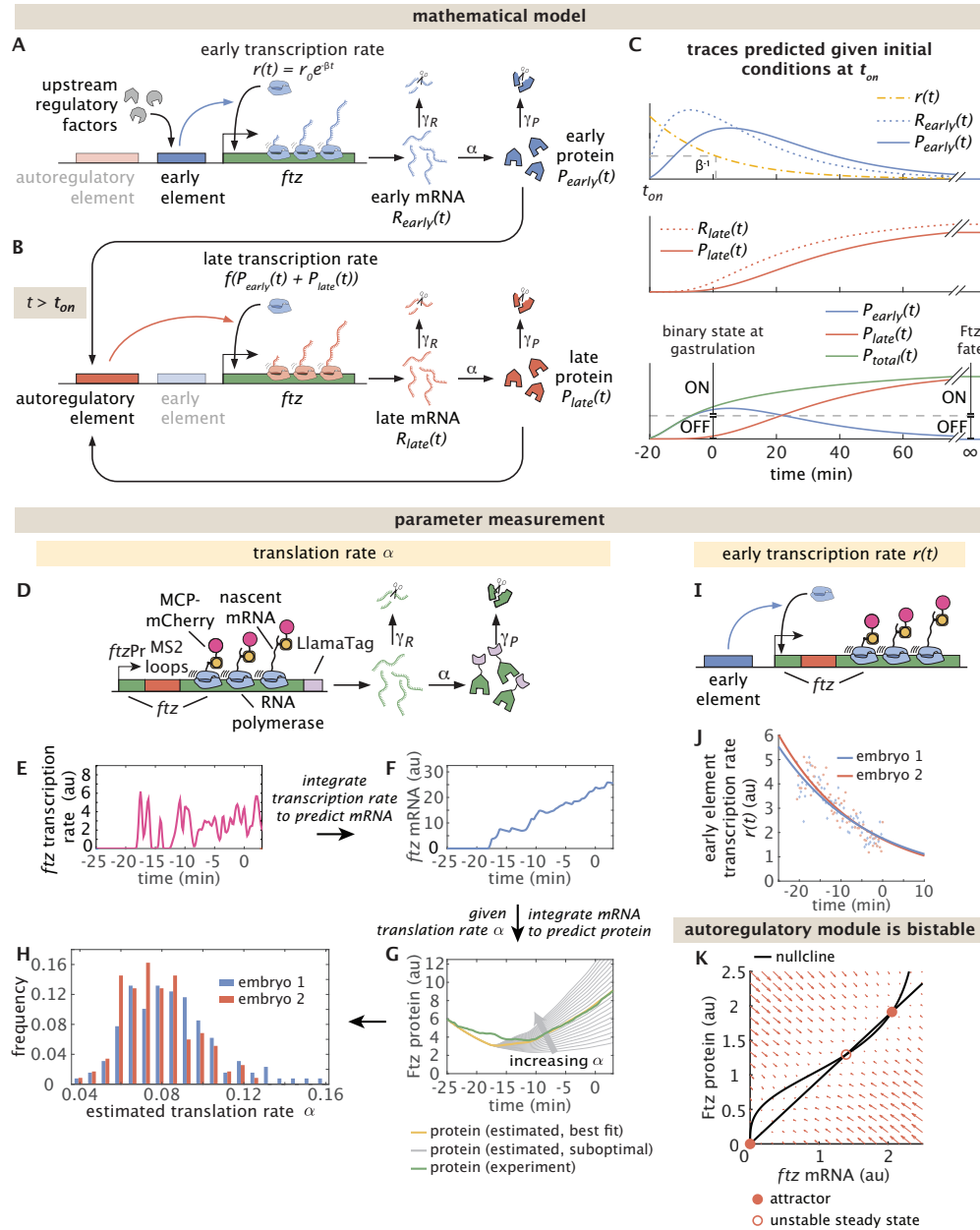


Figure 3.5: Quantifying dynamical systems model parameters reveals that *ftz* autoregulation is bistable. (continued on the next page)

Figure 3.5: **(continued) Quantifying dynamical systems model parameters reveals that *ftz* autoregulation is bistable.** (A) The early element drives *ftz* transcription at a rate $r(t)$ that drops exponentially over time with rate β , producing early mRNA $R_{early}(t)$ that is translated at rate α into Ftz protein $P_{early}(t)$. (B) The autoregulatory element drives Ftz expression starting at time t_{on} with a rate determined by the gene regulatory function $f(P)$, where P is the total Ftz produced from both the early and autoregulatory modules, $P_{early} + P_{late}$. (C) The model can be used to predict traces for the mRNA and protein concentrations over time based on initial transcription rates and molecular concentrations at t_{on} . We use the term “state” to refer to the whether Ftz expression is high or low at gastrulation, and “fate” with regard to the total Ftz concentration at long time scales ($t = \infty$). (D-H) Estimating the translation rate α . (D) We used the Ftz-MS2-LlamaTag transgenic construct for estimating the translation rate α (E) Illustrative example of a single-cell raw MS2 trace that reports on the instantaneous *ftz* transcription rate. MS2 traces are smoothed using a moving average of 1 min. Ftz traces are smoothened using a moving average of 5 min. (F) We integrate the MS2 traces to estimate the mRNA dynamics in individual nuclei. (G) For a given value of the translation rate α , and known degradation rates γ_R and γ_P [51, 18], we integrate the estimated mRNA dynamics to predict Ftz dynamics in individual nuclei. We then choose the value of α that leads to the best agreement between our prediction (gold line) and experiment (purple line). Gray lines show suboptimal fits. (H) Histogram distribution of best-fitted α values for individual cells within two embryos. (I-J) Inferring the decay rate of the early element transcription rate β . (I) The early element regulates the transcription rate $r(t)$. (J) The exponential fit of early element transcription ($n = 2$ embryos). Gray dots represent the averaged early element transcription rate at individual time points from a single embryo. Red and blue lines are the exponential fits that are used to estimate the early element decay rate β . (K) Given the measured model parameters, our model shows that the autoregulatory module is bistable. A phase portrait (as described in Box 1) illustrates that the autoregulatory module represented by Equation 3.2 is bistable in the absence of early protein. This implies that the autoregulatory enhancer is capable of maintaining high or low levels of Ftz indefinitely. The transcription rate is calculated from the measured MS2 signal, which is an approximation of the mRNA production rate [63, 102, 18]. The numerical values of experimentally measured parameters are reported in Table 3.1. See also Figure 3.10 and Figure 3.11.

Here, $f(P)$ is the gene regulatory function describing the output rate of mRNA production as a function of the total input Ftz concentration (the sum of the protein produced by the early and autoregulatory elements) as measured for the transgene in Figure 3.5B. Because $f(P)$ was measured for a transgene and not for the endogenous autoregulatory element, we also assume a scaling factor c . As indicated in Equation 3.2, we assume that the autoregulatory element is only active for $t \geq t_{on}$. Our model produced nearly identical results whether we assumed that this transition to full responsiveness occurred instantaneously at time t_{on} as

shown here, or whether we assumed a gradual increase in responsiveness over time (Materials and Methods “Simulations - Best predicted nuclei”).

Given quantitative parameter values and initial conditions ($r(t_{on})$, $R_{early}(t_{on})$, and $P_{early}(t_{on})$), the full model for *ftz* regulation described by Equation 3.1 and Equation 3.2 can be used to simulate the trajectory of total Ftz concentration in a nucleus over time. Specifically, we calculate Ftz expression levels (low or high) at gastrulation as well as at long times (i.e., $t \rightarrow \infty$). We will refer to the predicted stable concentration of Ftz at long times as the ultimate *fate* of the cell, as distinct from the (possibly transient) binary expression state at gastrulation (Figure 3.5C).

3.2.4 Quantifying model parameters shows that the autoregulatory module is bistable

The first step toward testing our model is to obtain numerical values for the parameters. The mRNA and protein decay rates γ_R and γ_P were drawn from existing measurements in the literature [51, 18]. The gene regulatory function $f(P)$ was already measured as described in the previous section (Figure 3.4D). The remaining free parameters (β , α , and c) were directly measured through a set of independent experiments.

First, we determined the decay rate β for transcription from the early element by fitting an exponential function to the average transcription dynamics of the early element construct within individual embryos (Figure 3.5I and J). Averaging the resulting decay rates per embryo resulted in $\beta = 0.048 \pm 0.0021$ 1/min. Second, we measured the translation rate α using a reporter transgene that combined MS2 and LlamaTag such that *ftz* transcription rate and Ftz concentration could be simultaneously measured (Figure 3.5D). We estimated $R(t)$, the total amount of mRNA produced by individual nuclei up until time t , by integrating the MS2 fluorescence over time (Figure 3.5E and F; [63]). Next, we integrated $R(t)$ for different values of the translation rate α to estimate protein dynamics $P(t)$ (Figure 3.5G). We calculated the best-fitted α for each nucleus and averaged them for each embryo (Figure 3.5H). Averaging α over two embryos yielded $\alpha = 0.082 \pm 0.004$ protein/(mRNA·min) (Table 3.1).

Using these parameters, together with the inferred scaling factor c (Figure 3.10; see Materials and Methods “Modeling - Intersection test for bistability” for details), we analyzed the autoregulatory module and determined that it is indeed bistable once upstream regulatory factors have degraded and the early element ceases to contribute to transcription ($P_{early} = 0$). As illustrated by the phase portrait in Figure 3.5K (and introduced in Materials and Methods “A primer on bistability”), the nullclines intersect at three points that define two attractors separated by an unstable steady state. Thus, our model indicates that the *ftz* autoregulatory module is capable of maintaining high or low levels of *ftz* expression indefinitely, and hence that there are two fates for Ftz on long time scales. Which fate a cell ultimately adopts will therefore depend on the protein initially produced by the early module.

3.2.5 Ftz expression state is instructed by the early element rather than by the autoregulatory element

Our model does not only predict that the Ftz autoregulatory module acts as a bistable switch. Because no free parameters remain, given initial conditions to the system, the model can also predict how Ftz concentration changes over time. Therefore, a critical test for the accuracy of the model is whether it can predict Ftz expression state at gastrulation.

To predict Ftz dynamics, we measured the initial transcription rate ($r(t_{on}) = r_{-20}$), mRNA level ($R_{early}(t_{on}) = R_{-20}$), and protein level ($P_{early}(t_{on}) = P_{-20}$) from the early module at the onset time $t_{on} = -20$ min (Figure 3.6A) in embryos expressing endogenous Ftz labeled with LlamaTag (Figure 3.2A, Materials and Methods “Extracting initial conditions”). Since, prior to onset, the autoregulatory element has not yet produced any mRNA or protein, knowing these three initial conditions is sufficient to calculate the trajectory for total Ftz concentration $P_{total}(t) = P_{early}(t) + P_{late}(t)$ within the corresponding nucleus using our model (Figure 3.6B).

For each simulated trajectory, we determined binary Ftz expression state at gastrulation based on whether total Ftz concentration exceeded the threshold empirically determined in Figure 3.2D, as shown in Figure 3.6B. Comparing our predictions and experimental measurements revealed that our model predicts binary Ftz expression state at gastrulation with an accuracy of 86.4% (102 of 118 nuclei) across $N = 3$ embryos (Figure 3.6C). The majority of classification errors were derived from a single embryo, with stochastic simulations suggesting the remainder may be attributed to noisy gene expression dynamics (Materials and Methods “Stochastic simulations”, Figure 3.13). Thus, our model appears to capture the essential deterministic components of Ftz dynamics.

While it is clear that the autoregulatory element establishes developmental memory of Ftz expression over long time scales, our simulations show that a stripe pattern is already evident at gastrulation from the contribution of the early protein alone, ignoring the autoregulatory contribution (Figure 3.6D). Indeed, predictions of Ftz state at gastrulation based on thresholding the early protein concentration alone were 84.7% accurate (100 of 118 nuclei)—only 1.7% less accurate than when the autoregulatory contribution is also taken into account (Figure 3.6C). Thus, it appears that the anterior boundary of stripe 4 at gastrulation is predominantly defined by the regulatory activity of upstream factors binding the early element independent of autoregulatory activity. This result supports the conclusion that the autoregulatory module primarily serves to commit cells to fates predetermined by the early element.

3.2.6 Ultimate Ftz fate is robustly specified in half an hour

Although the autoregulatory element becomes active 20 min before gastrulation, our model suggests that Ftz expression state at gastrulation is nevertheless primarily determined by the early element. It is therefore conceivable that some cells transiently express high Ftz at gastrulation without committing to stably expressing high Ftz in the long term. To ex-

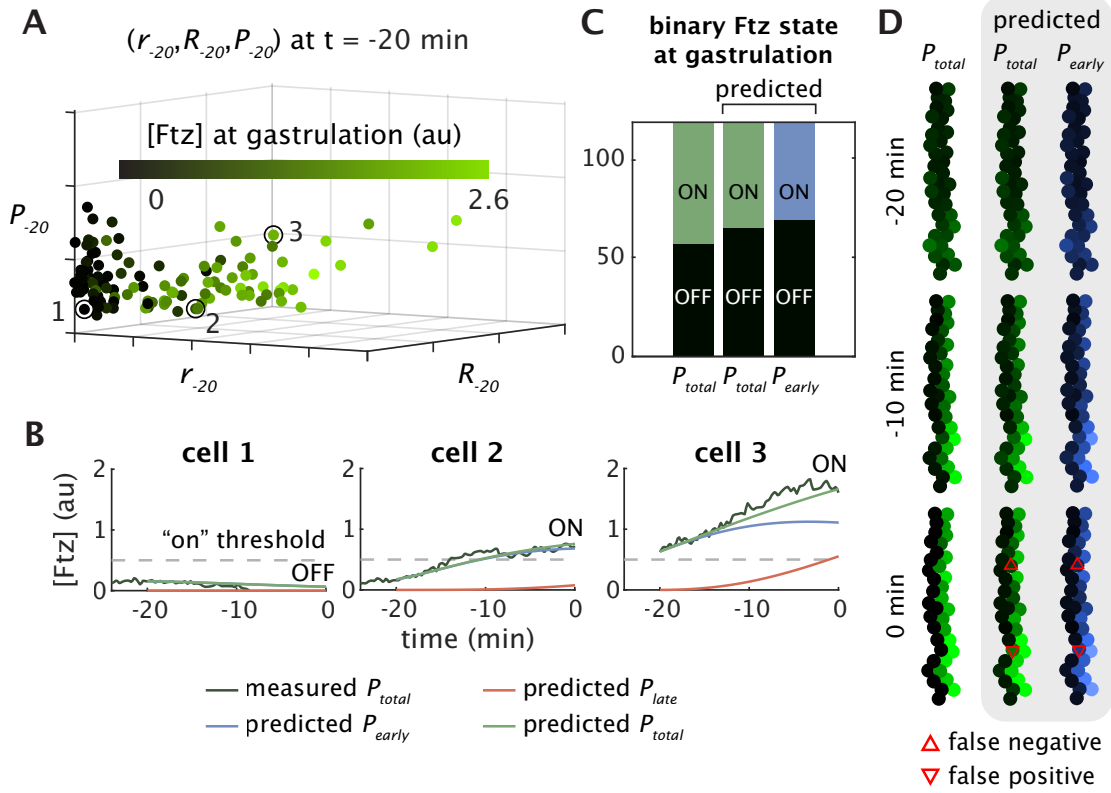


Figure 3.6: Model reveals that the early element alone is sufficient to accurately predict binary Ftz expression states at gastrulation. (A) The mRNA production rate r_{-20} , mRNA concentration R_{-20} , and protein concentration P_{-20} from the early element are measured at 20 min before gastrulation for each nucleus (filled circle). These measurements are used to set the initial conditions for simulations of Ftz concentration at all times after t_{on} in each individual cell. Circle color corresponds to the measured total Ftz concentration at gastrulation. (B) Measured (dark green) vs. simulated (light green) total Ftz for three representative cells corresponding to the circled points in panel A. We can classify a nucleus as in the high Ftz state (ON) or the low Ftz state (OFF) at gastrulation ($t = 0$ min) depending on whether the total Ftz levels are above or below the empirically identified threshold (gray dashed line) (Figure 3.2D). Predicted simulated traces are also shown for early (blue) and late (red) Ftz contributions. Note that, for cell 2, the blue and green curves substantially overlap. (*continued on the next page*)

Figure 3.6: *(continued)* Model reveals that the early element alone is sufficient to accurately predict binary Ftz expression states at gastrulation. (C) Measured fraction of cells in the ON or OFF expression state at gastrulation (left) compared to predictions based on thresholding total Ftz (center) or early Ftz alone (right). Predictions using total Ftz concentration or Ftz produced by the early element lead to similar accuracy in comparison to empirical results (total Ftz: 102 of 118 nuclei; early Ftz: 100 of 118 nuclei). (D) Results from a representative embryo show that the experimentally measured stripe pattern (left) is recapitulated by simulation (middle). A stripe pattern is still evident at gastrulation even from the predicted early Ftz concentration alone (right). Nuclear intensities at all time points are normalized to the predicted steady-state high Ftz concentration. Red triangles denote nuclei predicted to be ON that were OFF (downward) and nuclei predicted to be OFF that were empirically ON (upward). Parameters for all simulations were experimentally measured as reported in the main text; numerical values are reported in Table 3.1. See also Figure 3.12 and Figure 3.13.)

to explore this possibility, we compared the binary classification of nuclei at gastrulation with the final high or low expression fate predicted by the model at long time scales. Specifically, we analyzed a subset of nuclei, which we call the “best predicted nuclei”, whose quantitative expression levels—including their expression state at gastrulation—were successfully predicted using our model (Figure 3.6C and D; Materials and Methods “Simulations - Best predicted nuclei”). We found that 93.7% (74 of 79 nuclei) were predicted to maintain their binary state at gastrulation and adopt the corresponding Ftz expression fate in steady state (e.g., cell 3 in Figure 3.7A). The remaining 6.3% of nuclei (5 of 79), despite having a high Ftz expression state at gastrulation, were predicted to drop to a low Ftz fate (e.g., cell 2 in Figure 3.7A). No nuclei classified as low Ftz at gastrulation were predicted to express high Ftz after gastrulation. Thus, in principle, the autoregulatory element ensures that the vast majority of cells adopt a fate matching the transient state at gastrulation (Figure 3.7B).

What are the features of the transient early Ftz dynamics that determine whether the autoregulatory module commits to a high or low fate? We can address this question by looking at the relationship between the initial conditions, which dictate the dynamics of the early module, and the fate predicted by the model. Specifically, because increasing any one of the three initial conditions can only increase Ftz concentration at subsequent points in time, the regions of parameter space (r_{-20} , R_{-20} , P_{-20}) resulting in high fates are separated from the region of parameters resulting in low fates by a surface called the switching separatrix [155], so named because if the initial conditions of a cell are above the surface, then the bistable autoregulatory module will switch on.

Figure 3.7C illustrates the initial conditions for all nuclei at the anterior of stripe 4 alongside the switching separatrix (green surface). The separatrix indicates that a combination of all three initial conditions define Ftz fate. Particularly noteworthy is that the initial amount of early protein P_{-20} is not the strongest predictor of fate. Cells with comparable levels

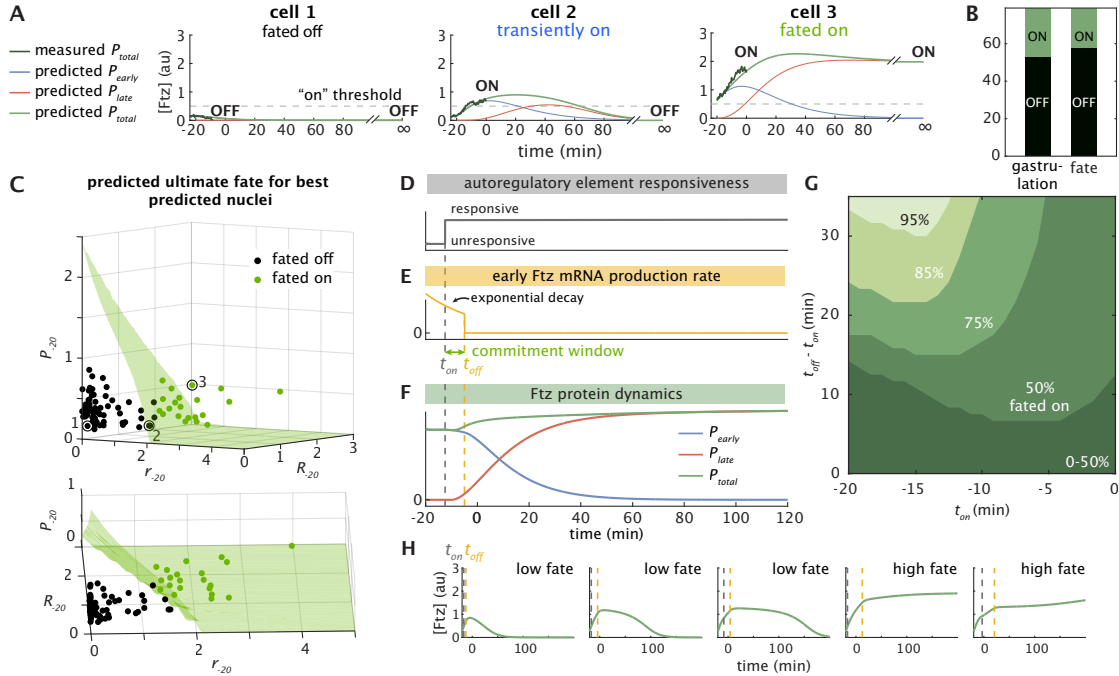


Figure 3.7: *In silico* analysis reveals the window of commitment of cells to Ftz fate. (A) Simulated Ftz concentration over long time scales (total: light green, early: blue, late: red) for the same cells shown in Figure 3.6B (circled again here in panel C). Empirical measurements until gastrulation are plotted in dark green for comparison. (B) Bar charts for the best predicted subset of nuclei ($N = 79$; see Materials and Methods “Simulations - Best predicted nuclei”) show that most analyzed cells are predicted to adopt an ultimate fate (right) that matches the measured binary Ftz expression state at gastrulation (left). (C) The switching separatrix (green surface) separates nuclei predicted to adopt the high Ftz fate (green circles above the surface) from those predicted to adopt the low Ftz fate (black circles below the surface) given their initial conditions. Top and bottom subpanels show the same plot from two different viewing angles. See also Figure 3.12. (D–E) Schematic illustrating (D) t_{on} , the time the autoregulatory element becomes responsive, and (E) t_{off} , the time the early element ceases production. We define t_{off} such that $r(t \geq t_{off}) = 0$, where $r(t)$ is the mRNA production rate of the early element. Before t_{off} , $r(t)$ follows the usual exponential decay observed experimentally with rate $\beta^{-1} = 21$ min. The commitment window $t_{off} - t_{on}$ is the timespan during which both upstream factors and autoactivation dictate Ftz expression, and serves as an estimate for how long the autoregulatory module has to establish a memory of the signal from the early module. (continued on the next page)

Figure 3.7: (*continued*) ***In silico* analysis reveals the window of commitment of cells to Ftz fate.** (F) Simulated Ftz concentrations resulting from the timing pictured in C and D. (G) Contour plot showing what percentage of the subset of analyzed nuclei ($N = 21$; Materials and Methods “Simulations - Best predicted nuclei”) that reach the high Ftz fate under wild-type conditions still reach that fate as the commitment window $t_{off} - t_{on}$ is varied and t_{on} is delayed relative to the measured -20 min (see text). (H) Simulated single-nucleus traces of total Ftz for varying values of t_{on} with initial conditions held constant at $r_{-20} = 1.8$ a.u., $R_{-20} = 1.1$ a.u., $P_{-20} = 2.6$ a.u. The simulations show that the commitment window can cause the transient dynamics to vary quite dramatically. Dashed lines denote the t_{off} (gold) and t_{on} (gray) values used for each plot. The parameters are as in Table 3.1. See also Figure 3.12, Figure 3.13 and Figure 3.14.

of P_{-20} , even at the highest concentrations we measured, are partitioned between ON and OFF fates by the values of r_{-20} and R_{-20} (Figure 3.7C). Thus, the autoregulatory element does not simply perform an instantaneous readout of Ftz concentration at the time of onset to determine cell fate. Rather, a combination of high initial transcription r_{-20} and high initial mRNA R_{-20} —which together lead to a bump in Ftz concentration *after* the onset of autoregulation—seems to be essential for adopting the high fate.

Given this observation, we asked how long it takes for the autoregulatory element to establish developmental memory of Ftz levels. We posit that this timespan corresponds to the classical notion of a commitment window, defined as the period of time during which a cell integrates information from external factors to decide its fate [42, 122]. It can be difficult to access temporal features of development such as the commitment window *in vivo*, in part due to the technical challenge of measuring and systematically manipulating input signals while simultaneously monitoring the resulting gene expression programs in individual cells within intact tissues or organisms [10, 91].

Our mathematical model provided us with a unique opportunity to examine the commitment window by altering the timing of developmental events *in silico*. Because the low Ftz fate at steady state is the default expression state—the autoregulatory module will always produce zero protein in the absence of a transient signal from the early module—the commitment window primarily determines whether the cell has enough time to detect if transient Ftz concentrations are high and, if so, to adopt a trajectory destined for a high steady-state expression fate. Thus, we analyzed only the subset of the best predicted nuclei that were predicted to adopt the high fate according to the switching separatrix analysis ($N = 21$; Figure 3.7C, green surface).

We define the commitment window as $t_{off} - t_{on}$, where t_{on} indicates the start of autoregulatory responsiveness (Figure 3.7D) and t_{off} is the time when the early element ceases transcription (Figure 3.7E). The commitment window represents the total amount of time during which both the early and autoregulatory modules dictate *ftz* expression, and serves as an estimate for how long the autoregulatory module has to establish a memory of transient

Ftz state.

To determine how long of a commitment window allows cells to convert transiently high Ftz concentrations into permanently high Ftz fates, we asked whether our analyzed subset of nuclei still reached the high fate as we systematically varied the commitment window. We solved the dynamical system from Equation 3.2 and Equation 3.1 with commitment windows of increasing length and recorded which nuclei reached the high Ftz fate. Our results are reported in Figure 3.7F in terms of the fraction of nuclei within the subset that adopted the high Ftz fate for each timing condition. For example, assuming $t_{on} \leq -13$ min, a commitment window of 34 min results in 95% of nuclei achieving high steady-state Ftz levels (Figure 3.7F). Our results reveal that there is a gradual dropoff in the fraction of cells that do not commit to the high fate as the commitment window is shortened or t_{on} is delayed. This suggests that slight temporal perturbations during development are unlikely to cause catastrophic patterning failures in which all cells suddenly adopt the low Ftz fate. Rather, we expect small changes in timing to affect the fates of only a small percentage of cells.

Our analysis of the commitment window suggests that about half an hour suffices for the vast majority of high-fated cells to stably commit to that fate (see also Materials and Methods “Modeling - Convergence rate estimates and basin of attraction”). Proper fate specification, however, does not guarantee similarity in the temporal trajectories of Ftz concentration, as evidenced by the wide range of dynamics observed in simulated traces for varying t_{on} and t_{off} with the same initial conditions for the early module (Figure 3.7G; Figure 3.12). Thus, if the transient Ftz trajectory, not just its ultimate fate, is instructive for downstream genes, then the need for proper regulation of these genes may place stricter constraints on the relative timing of the early and autoregulatory elements than those that are imposed by the specification of Ftz fate alone.

3.3 Discussion

For decades, developmental biologists have used the concept of Waddington’s landscape to conceptualize cellular decision-making. Under this framework, cells roll down valleys in a predetermined landscape to adopt their ultimate fates. This framework has been repeatedly mathematicized using dynamical systems theory [173, 61, 87, 35, 141]. Many of these studies have hypothesized that autoactivation [38] helps establish and maintain binary cell fates through bistability [183, 153], which can be thought of as introducing forks in Waddington’s landscape. Though experiments in cell culture and fixed tissue have provided evidence for the bistability of various autoregulatory modules found within gene regulatory networks, until now, these results have not been confirmed by direct examination of dynamics in intact, living embryos.

In this work, we utilized live imaging to quantitatively characterize the dynamics of the fruit fly *ftz* regulatory system *in vivo*. We elucidated tight temporal coordination between the two enhancer elements that regulate *ftz* expression (Figure 3.3) and combined dynamical

systems modeling with biophysical measurements to show that the bistability of the autoregulatory module can maintain otherwise transient expression levels driven by upstream factors (Figure 3.7). Based on the prevalence of autoregulatory motifs in nature [3, 134], we speculate that the approach employed by the Ftz system to decide cell fate is not limited to fruit flies, but might also be widely adopted during development in other organisms.

One of our central discoveries is that *ftz* autoregulation is triggered at a specific developmental time rather than being triggered when Ftz reaches a certain threshold concentration. Recent work has suggested candidates for “timer genes” that are expressed at distinct developmental time points and appear to facilitate the expression of other genes [115, 29, 31, 154, 30]. We speculate that timer genes might also bind the *ftz* autoregulatory element to trigger its responsiveness to Ftz.

We relied on quantitative modeling with no free parameters to provide strong evidence demonstrating that the Ftz autoregulatory module is bistable. Dynamical systems models such as the one employed here are advantageous for this approach since the parameters are biophysically interpretable, which is not true in, for example, more coarse-grained Boolean models of genetic networks [8]. Although the model in the main text is deterministic, in preliminary work we also considered a stochastic chemical reaction network simulated by the Gillespie algorithm, as well as a stochastic differential equation model (Materials and Methods “Simulations - Stochastic simulations”). Aside from introducing noise to Ftz expression levels, and in contrast to other patterning networks where noise appears crucially to drive fate determination [184, 128], these models did not exhibit any unique behaviors that better accounted for our experimental observations compared to the deterministic model.

The results presented in this work were derived from experiments paired with mathematical analysis of our theoretical model. The conclusions could be further supported—or challenged—by experiments where Ftz initial conditions are altered through, for example, heat shock or optogenetic approaches. Future research on Ftz autoregulation would also benefit from technological advancements that extend quantitative live imaging capabilities beyond gastrulation in order to monitor Ftz concentrations over extended timescales, from the onset of autoregulation to the final adoption of cellular fate.

A basic assumption of the approach employed here is that the behavior of the whole network can be predicted from the behavior of the parts (modules) in isolation [78, 15, 46]. In our model, we divided the early and late populations of Ftz into two separate modules. The early module produces the early protein (P_{early}), which acts as an input to the autoregulatory module responsible for producing the late protein (P_{late}). Positive feedback arises because P_{late} also activates its own expression (Figure 3.5 and Figure 3.11). This is not the only way to define the module; for example, if we knew which regulatory factors rendered the autoregulatory element responsive, then we could include those as inputs instead of describing their activity implicitly through the parameter t_{on} . Similarly, if we knew the dynamics of the upstream regulatory factors for the early element, we could replace β , the phenomenological temporal decay in the early transcription rate, with a mathematical expression relating the concentrations of these regulatory factor inputs to the early Ftz transcription rate.

Our representation of the autoregulatory module can predict the fate of *ftz* expression

from arbitrary trajectories of early Ftz. As a result, we can predict the effect of modifications to upstream signaling on the resulting gene expression patterns. This allows us to ask what forms of input are appropriate to achieve particular patterning outcomes. Such ability to reverse engineer the process of cellular decision-making could facilitate designing perturbations to manipulate the system, identifying constraints placed on upstream modules by the needs of downstream modules, and analyzing whether biologically evolved signals match those that are mathematically “optimal” for such needs as patterning speed [136] or information transmission [163]. Different methods of generating predictions may be appropriate depending on the types of inputs under consideration. In this paper, the fact that increasing any one of the parameters that define the early Ftz input (r_{-20} , R_{-20} , P_{-20}) increases total Ftz concentration at all points in time (a property known as monotonicity; [5, 155]) made it possible to analyze our model using a switching separatrix. However, this may not be true for other regulatory systems, as in the case where a gene within a module represses its own production.

Throughout developmental biology, the concept of a commitment window has been repeatedly utilized to describe the amount of time cells need to be exposed to upstream signals in order to decide their developmental fates [42, 122]. Our quantitative dynamical systems model enabled us to conduct a detailed examination of this commitment window, and to identify what fraction of cells adopt certain fates as developmental timing is varied. From an engineering perspective, we may consider a gene expression pattern as an objective that must be achieved with a prescribed level of precision (i.e., as a design specification), and work backwards to see what inputs satisfy this requirement. Our approach complements existing work on precision that emphasizes how tightly protein concentrations are controlled [72] and how accurately cells can locate their position by reading out concentrations of upstream factors [48, 135]. In particular, the latter approaches indicate what level of precision is actually achieved by a patterning network, while our framing focuses rather on what range of parameters allow a system to attain a predefined level of precision. A combination of the two perspectives could help elucidate what biophysical and evolutionary factors influence stochastic variation in phenotypes, including how precise expression patterns must actually be to produce functional, healthy organisms.

In summary, by turning widespread schematic models of autoactivation modules into precise mathematical statements and experimentally testing the resulting predictions, we have provided support for a widely held hypothesis about how developmental fates are established in embryos. In the future, combining quantitative measurements with precise spatiotemporal perturbations [68] and synthetic reconstitution methods [121] promises to enable yet another iteration of the dialogue between theory and experiment that constitutes the basis of our work, ultimately leading to a predictive understanding of function in developmental networks and the myriad forms and fates to which they give rise.

3.4 Materials and Methods

3.4.1 Fly strains/Genotypes

3.4.1.1 Creation of tagged *fushi tarazu* (*ftz*) gene using CRISPR-Cas9

Ftz-EGFP-LlamaTag fusion design is based on previously published transgenic line [18]. To tag endogenous *ftz* locus with EGFP-LlamaTag, we used CRISPR-mediated homology-directed repair with donor plasmid synthesized by Genscript. gRNA was designed using the target finder tool from flyCRISPR (<https://flycrispr.org>), and cloned based on the protocol from [70]. *yw*;nos-Cas9(II-attP40) transgenic line was used as the genomic source for Cas9 and the embryos were injected and screened by BestGene Inc.

3.4.1.2 Creation of *ftz* early and autoregulatory element reporter

The *ftz* autoregulatory element sequence was based on the 4.4kb DNA segment described in [83]. The *ftz* early and autoregulatory element reporters driven by the *ftz* promoter were constructed by combining the respective enhancer sequence with an array of 24 MS2 stem loops inserted into the *ftz* coding sequence *D. melanogaster yellow* gene [18]. The *ftz* autoregulatory element reporter driven by the *eve* promoter was constructed by combining the enhancer sequence with an array of 24 MS2 stem loops inserted into the *D. melanogaster yellow* gene [18]. The constructs were synthesized by Genscript and injected by BestGene Inc into *D. melanogaster* embryos with a Φ C31 insertion site in chromosome 3 (Bloomington stock #9750; landing site VK00033; cytological location 65B2).

3.4.1.3 Transgenes expressing EGFP and MCP-mCherry

The fly line maternally expressing MCP-mCherry (chromosome 3) was constructed as described in [18]. The fly line maternally expressing vasa-EGFP (chromosome 2) was constructed as described in [97]. To simultaneously image protein dynamics using LlamaTags and transcription using MCP-MS2 system, we combined the vasa-EGFP transgene with MCP-mCherry to construct a new line (*yw*;vasa-EGFP;MCP-mCherry) that maternally expresses both proteins.

3.4.2 Fly lines

To measure Ftz transcription and protein levels simultaneously, we performed crosses to generate virgins carrying transgenes that drive maternal EGFP, MCP-mCherry, the LlamaTagged Ftz locus along with *ftz* autoregulatory element reporter (*yw*; $\frac{\text{vasa-EGFP}}{\text{CyO}}$; $\frac{\text{MCP-mCherry}}{\text{FtzAuto-MS2-Yellow, Ftz-LlamaTag}}$). These flies were then crossed with males having both the *ftz* autoregulatory element reporter and the LlamaTagged Ftz locus (*yw*; +; FtzAuto-MS2-Yellow, Ftz-LlamaTag). This resulted in the embryo carrying mater-

nally deposited EGFP, MCP-mCherry, and two copies of the LlamaTagged Ftz locus and *ftz* autoregulatory element reporter.

3.4.3 Method Details

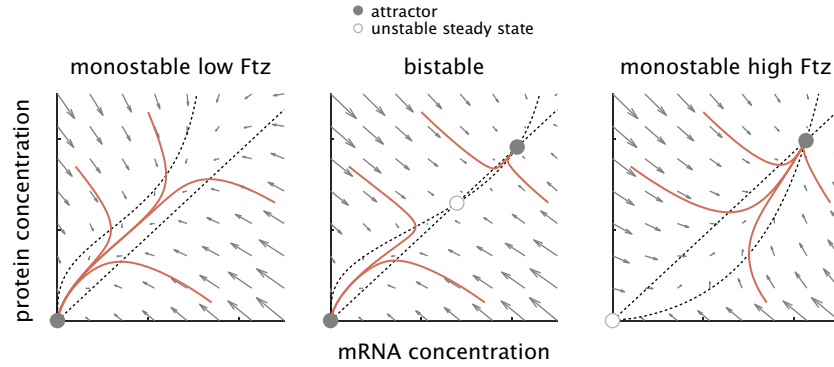
3.4.3.1 A primer on bistability

We model *ftz* regulation as a dynamical system represented by ordinary differential equations. Typically, in such models, the rate of change in mRNA and protein concentrations is given by a combination of their synthesis and degradation rates. For example, a model of the *ftz* autoregulatory module in isolation (ignoring regulation by the early element) describes the rate of change in *ftz* mRNA concentration R and protein concentration P as

$$\begin{cases} \frac{dR}{dt} = f(P) - \gamma_R R \\ \frac{dP}{dt} = \alpha R - \gamma_P P \end{cases} \quad (3.3)$$

In the first equation, $f(P)$ is the gene regulatory (or input-output) function that describes how input Ftz concentration controls the rate of *ftz* transcription, and γ_R is the *ftz* mRNA degradation rate. Since Ftz promotes its own production, $f(P)$ increases with P . In the second equation, α is the translation rate and γ_P corresponds to the Ftz protein degradation rate.

We can represent this dynamical system using a phase portrait, which is a two-dimensional plot that shows how the trajectory of mRNA and protein concentrations evolves over time. Specifically, the phase portrait has the mRNA and protein concentrations on its axes. At each point in the phase portrait we plug the corresponding mRNA and protein concentrations into Equation 3.3 in order to calculate $\frac{dR}{dt}$ and $\frac{dP}{dt}$. The values of these derivatives determine how quickly and in which direction the mRNA and protein concentrations will change over a small time period. This direction is encoded in the phase portrait by drawing an arrow. By following arrows throughout the phase portrait, we can qualitatively predict where the trajectory will travel. Hence, given initial concentrations of mRNA and protein, we can predict an entire trajectory by following the arrows.



Phase portraits for the isolated autoregulatory module described by Equation 3.3 show how mRNA and protein concentrations change over time. Sample trajectories (red) starting at different points follow the gray arrows encoding $\frac{dR}{dt}$ and $\frac{dP}{dt}$ to converge toward attractors (solid gray circles) and away from unstable steady states (open gray circle). The numerical values of the model parameters determine whether the module is monostable (left and right) or bistable (center), as visualized by the change in the number of intersections of the nullclines (dotted black lines). Parameters for all plots are given in Table 3.1 except for $c = 0.3$ on the left and $c = 0.65$, $n = 0.7$ on the right, which were chosen for the purpose of illustrating monostability.

We can draw a line on the phase portrait to indicate all points where the mRNA does not change concentration during a small time period (i.e., the arrows are vertical at these points). This line is defined by $\frac{dR}{dt} = 0$ and is called a nullcline. We can similarly identify a nullcline for protein concentration, defined by $\frac{dP}{dt} = 0$ and along which all arrows point horizontally. Trajectories become “stuck” where both $\frac{dR}{dt} = 0$ and $\frac{dP}{dt} = 0$, since concentrations do not change for either molecular species. Such points are called steady states and are found at the intersections of the nullclines. Crucially, the shapes of the nullclines depend on the shape of f as well as the quantitative values of the translation and decay rates, implying that the number of steady states also depends on these parameters.

All trajectories eventually converge to one or more steady states called attractors, so named because they “attract” trajectories toward them. Attractors are stable, meaning that if concentrations are perturbed slightly away from the attractor, they will return to the attractor. Alternatively, a steady state can be unstable, such that, after a small perturbation, the chemical concentrations evolve away from that steady state; in this case the state is not an attractor because trajectories in the vicinity tend to leave rather than approach it. The stability and number of steady states also depend on the system’s quantitative parameter values.

For models of the kind we consider here, there are either 1 or 3 steady states, such that the autoregulatory module is either monostable, meaning it possesses one attractor and possibly one unstable steady state, or bistable, meaning it possesses two attractors and one unstable steady state between these attractors. In the monostable case, all trajectories eventually reach the same attractor, whereas in the bistable case, the attractor reached by

a given trajectory depends on the initial concentrations of mRNA and protein.

In Section 3.2.3 of the main text, we discuss how to model *ftz* regulation by a combination of the early and autoregulatory elements, as it occurs in the wild-type embryo. The early element eventually stops driving *ftz* transcription, at which point we can use Equation 3.3 to describe the evolution of the Ftz mRNA and protein concentrations starting from the values they assumed at the time the early element shut down. Thus, if the autoregulatory module is bistable, then contributions from the early element will determine whether the steady-state Ftz concentration in a given nucleus (trajectory) is high or low, while if the autoregulatory module is monostable, then the steady-state Ftz concentration will always converge toward the same point regardless of the transient activity owing to the early element. Throughout the text, we will refer to the steady-state Ftz concentration that is reached by a trajectory as the Ftz *fate*, to distinguish it from the instantaneous *state* of its concentration at any given point in time.

3.4.3.2 Embryo preparation and data collection

The embryos were prepared following procedures described in [63, 18, 102]. Embryos were collected and mounted in halocarbon oil 27 between a semipermeable membrane (Lumox film, Starstedt, Germany) and a coverslip. Confocal imaging on a Zeiss LSM 780 microscope was performed using a Plan-Apochromat 40x /1.4NA oil immersion objective. EGFP and MCP-mCherry were excited with laser wavelengths of 488 nm (25.0 μ W laser power) and 594 nm (15.0 μ W laser power), respectively. Fluorescence was detected using the Zeiss QUASAR detection unit. Image resolution was 512×512 pixels, with pixel size of 0.231 μ m. Sequential z-stacks separated by 0.5 μ m were acquired. Specimens were imaged from mid nuclear cycle 14 until the start of gastrulation.

3.4.3.3 Modeling

Autoregulatory element We use $R_{early}(t)$ to describe the concentration of *ftz* mRNA transcribed from the early element, which is translated into protein $P_{early}(t)$. We define $R_{late}(t)$ as the *ftz* mRNA transcribed from the autoregulatory element and translated into protein $P_{late}(t)$. The total Ftz in the cell at time t is given by $P_{total}(t) = P_{early}(t) + P_{late}(t)$. The dynamical equations describing the temporal evolution of mRNA and protein are

$$\text{autoregulatory element: } \begin{cases} \frac{dR_{late}}{dt} = c\tilde{f}(t, P_{total}(t)) - \gamma_R R_{late}(t) \\ \frac{dP_{late}}{dt} = \alpha R_{late}(t) - \gamma_P P_{late}(t) \end{cases}, \quad (3.4)$$

where γ_R and γ_P are the decay rates of mRNA and protein respectively, α is the translation rate, and c is a scaling factor equivalent to the ratio of maximum production rate from the endogenous locus vs. the transgene (where $\tilde{f}$ is measured). Note that $\tilde{f}(t, \cdot)$, the gene regulatory function for the autoregulatory element, is time dependent. Unless otherwise

stated, we will assume

$$\tilde{f}(t, P) = \begin{cases} 0, & t < t_{on}, \\ f(P), & t \geq t_{on}, \end{cases} \quad (3.5)$$

where

$$f(P) = \frac{aP^n}{K^n + P^n} + l \quad (3.6)$$

is the sigmoid describing the autoregulatory relationship at maximum amplitude. Using the definition of $P_{total}(t)$ we will then write Equation 3.4 as

$$\text{autoregulatory element: } \begin{cases} \frac{dR_{late}}{dt} = cf(P_{early}(t) + P_{late}(t)) - \gamma_R R_{late}(t) \\ \frac{dP_{late}}{dt} = \alpha R_{late}(t) - \gamma_P P_{late}(t) \end{cases}, t \geq t_{on}. \quad (3.7)$$

Thus, $P_{early}(t)$ acts as the sole time-varying input to the autoregulatory element. Since autoregulation does not begin until time t_{on} , the initial conditions for Equation 3.7 are fixed at $R_{late}(t_{on}) = 0$ and $P_{late}(t_{on}) = 0$.

Early element From our empirical measurements, we observed that the production rate $r(t)$ of early *ftz* mRNA is well approximated by an exponential decay (Figure 3.5I and J), allowing us to model $R_{early}(t)$ and $P_{early}(t)$ through the dynamical system

$$\begin{cases} \frac{dr}{dt} = -\beta r(t) \\ \frac{dR_{early}}{dt} = r(t) - \gamma_R R_{early}(t) \\ \frac{dP_{early}}{dt} = \alpha R_{early}(t) - \gamma_P P_{early}(t) \end{cases}. \quad (3.8)$$

As before, α is the translation rate and γ_R and γ_P are the decay rates of mRNA and protein.

Because Equation 3.8 is linear, it can be equivalently written as

$$\frac{d\vec{x}(t)}{dt} = \begin{bmatrix} -\beta & 0 & 0 \\ 1 & -\gamma_R & 0 \\ 0 & \alpha & -\gamma_P \end{bmatrix} \vec{x}(t) =: \mathbf{A}\vec{x}(t), \quad (3.9)$$

where

$$\vec{x}(t) = \begin{bmatrix} r(t) \\ R_{early}(t) \\ P_{early}(t) \end{bmatrix}. \quad (3.10)$$

The analytical solution is then given by

$$\vec{x}(t) = c_1 e^{\lambda_1 t} \vec{v}_1 + c_2 e^{\lambda_2 t} \vec{v}_2 + c_3 e^{\lambda_3 t} \vec{v}_3, \quad (3.11)$$

where $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of A corresponding to eigenvectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ respectively. In particular,

$$\vec{x}(t) = \begin{bmatrix} 1 \\ \frac{1}{\gamma_R - \beta} \\ \frac{\alpha}{(\gamma_R - \beta)(\gamma_P - \beta)} \end{bmatrix} c_1 e^{-\beta t} + \begin{bmatrix} 0 \\ 1 \\ \frac{\alpha}{\gamma_P - \gamma_R} \end{bmatrix} c_2 e^{-\gamma_R t} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} c_3 e^{-\gamma_P t} \quad (3.12)$$

where

$$c_1 = r(t_{on}), \quad c_2 = \left(R_{early}(t_{on}) - \frac{r(t_{on})}{\gamma_R - \beta} \right), \quad c_3 = P_{early}(t_{on}) - \frac{\alpha}{\gamma_P - \gamma_R} c_2 - \frac{\alpha}{(\gamma_R - \beta)(\gamma_P - \beta)} c_1. \quad (3.13)$$

Thus, the input $P_{early}(t)$ to the autoregulatory element is completely characterized by three parameters ($r(t_{on}), R_{early}(t_{on}), P_{early}(t_{on})$) corresponding to the initial conditions for Equation 3.8. Since every term in the solution is multiplied by an exponential that decays in time, all state variables will tend to 0 as t goes to infinity.

Intersection test for bistability Here, we describe how to use an intersection test to identify steady states such as those introduced in Materials and Methods: “A primer on bistability” for our model of the autoregulatory module. To reiterate, our dynamical system Equation 3.7 is described by

$$\text{autoregulatory element: } \begin{cases} \frac{dR_{late}}{dt} = cf(P_{early}(t) + P_{late}(t)) - \gamma_R R_{late}(t) \\ \frac{dP_{late}}{dt} = \alpha R_{late}(t) - \gamma_P P_{late}(t) \end{cases}, \quad t \geq t_{on}. \quad (3.14)$$

Steady states are system states at infinite time. Since we know that $P_{early}(t)$ goes to 0 at long times, the steady states for Equation 3.7 are equivalent to the steady states of

$$\text{autoregulatory element: } \begin{cases} \frac{dR_{late}}{dt} = cf(P_{late}(t)) - \gamma_R R_{late}(t) \\ \frac{dP_{late}}{dt} = \alpha R_{late}(t) - \gamma_P P_{late}(t) \end{cases}, \quad (3.15)$$

which has no time-varying parameters and can therefore be analyzed for steady states by standard methods.

By definition, at steady state the derivative of the different molecules species with respect to time equals zero, allowing us to write

$$\text{autoregulatory element: } \begin{cases} 0 = cf(P_{late}^*) - \gamma_R R_{late}^* \\ 0 = \alpha R_{late}^* - \gamma_P P_{late}^* \end{cases} \quad (3.16)$$

where (R_{late}^*, P_{late}^*) is a steady state. Then, we rearrange the bottom equation to get

$$R_{late}^* = \frac{\gamma_P}{\alpha} P_{late}^*, \quad (3.17)$$

which we plug into the top equation to get

$$0 = cf(P_{late}^*) - \frac{\gamma_R \gamma_P}{\alpha} P_{late}^*. \quad (3.18)$$

From here, we rearrange terms to recover

$$\gamma_P P_{late}^* = \frac{\alpha c}{\gamma_R} f(P_{late}^*). \quad (3.19)$$

Hence, the intersections of a line of slope γ_P with the right-hand side give the steady-state late protein concentrations P_{late}^* , from which we can recover R_{late}^* through Equation 3.17. These intersections can be estimated graphically or through computational methods.

Convergence rate estimates and basin of attraction In this section, we show our calculation for the local convergence rate to the high Ftz steady state. The convergence rate is an estimate for how quickly a cell approaches the high fate, expressed as the time constant for an exponential decay of the distance between a trajectory and the steady state in the phase space. This section assumes familiarity with linear stability analysis, which is introduced in most nonlinear systems textbooks such as the classic [160].

Assume we combine Equation 3.7 and Equation 3.8 to obtain the full model

$$\begin{cases} \frac{dr}{dt} = -\beta r(t) \\ \frac{dR_{early}}{dt} = r(t) - \gamma_R R_{early}(t) \\ \frac{dR_{late}}{dt} = cf_0(P_{tot}(t)) - \gamma_R R_{late}(t) \\ \frac{dP_{tot}}{dt} = \alpha(r(t) + R(t)) - \gamma_P P_{tot}(t) \end{cases} \quad (3.20)$$

where $P_{tot}(t) = P_{early}(t) + P_{late}(t)$ and we have used the assumption that

$$f(t, P) = f_0(P) \quad (3.21)$$

for $t \geq 0$.

The Jacobian of this system about a steady state $(r^*, R_{early}^*, R_{late}^*, P_{tot}^*)$ is

$$J = \begin{bmatrix} -\beta & 0 & 0 & 0 \\ 1 & -\gamma_R & 0 & 0 \\ 0 & 0 & -\gamma_R & cf'(P^*) \\ 0 & \alpha & \alpha & -\gamma_P \end{bmatrix}. \quad (3.22)$$

Since J is a lower triangular block matrix, its eigenvalues are the eigenvalues of the diagonal blocks. The upper left block has eigenvalues

$$\lambda_1 = -\beta, \lambda_2 = -\gamma_R \quad (3.23)$$

and the lower block has eigenvalues

$$\lambda_+ = \frac{-(\gamma_P + \gamma_R) + \sqrt{(\gamma_P + \gamma_R)^2 - 4(\gamma_R\gamma_P - \alpha cf'(P^*))}}{2}, \quad (3.24)$$

$$\lambda_- = \frac{-(\gamma_P + \gamma_R) - \sqrt{(\gamma_P + \gamma_R)^2 - 4(\gamma_R\gamma_P - \alpha cf'(P^*))}}{2}. \quad (3.25)$$

If $\alpha cf'(P^*) > \gamma_R\gamma_P$, the eigenvalue λ_+ is positive, hence the system is bistable and the identified steady state is the unstable middle point. Otherwise the real parts of all eigenvalues are negative and the steady state is (marginally) stable.

Locally to a stable steady state (for which $\alpha cf'(P^*) < \gamma_R\gamma_P$), the convergence rate ξ is determined by the slowest of the decay rates along the eigenvectors, i.e.,

$$\xi = \min \left\{ b, \gamma_R, \left| \Re[\lambda_+] \right| \right\} \quad (3.26)$$

where

$$\left| \Re[\lambda_+] \right| < \frac{\gamma_R + \gamma_P}{2}. \quad (3.27)$$

The bound is determined by the point at which the imaginary components of λ_+, λ_- become nonzero. Thus, if $\gamma_P < \gamma_R$, then $\left| \Re[\lambda_+] \right| < \gamma_R$, which means that the local convergence rate will be slower than the mRNA decay rate.

The (nonnormalized) eigenvectors associated with each eigenvalue are

$$\begin{aligned} \vec{v}_1 &= \begin{bmatrix} -\frac{1}{\alpha}(\alpha cf'(P^*) - (\beta - \gamma_P)(\beta - \gamma_R)) \\ \frac{1}{\alpha(\beta - \gamma_R)}(\alpha cf'(P^*) - (\beta - \gamma_P)(\beta - \gamma_R)) \\ -\frac{(cf'(P^*))}{\beta - \gamma_R} \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \\ \vec{v}_+ &= \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha} \left(\gamma_P + \frac{-(\gamma_P + \gamma_R) + \sqrt{(\gamma_P + \gamma_R)^2 - 4(\gamma_R\gamma_P - \alpha cf'(P^*))}}{2} \right) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha}(\gamma_P + \lambda_+) \\ 1 \end{bmatrix}, \\ \vec{v}_- &= \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha} \left(\gamma_P + \frac{-(\gamma_P + \gamma_R) - \sqrt{(\gamma_P + \gamma_R)^2 - 4(\gamma_R\gamma_P - \alpha cf'(P^*))}}{2} \right) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha}(\gamma_P + \lambda_-) \\ 1 \end{bmatrix}. \end{aligned} \quad (3.28)$$

As expected, locally the convergence rate associated with $\lambda_2 = -\gamma_R$ (decay rate of mRNA) has no component in the “protein direction” (i.e., $[0 \ 0 \ 0 \ 1] \cdot v_2 = 0$), so if we are only interested in the local convergence rate of protein concentration we may ignore λ_2 and instead take

$$\xi_P = \min \left\{ b, \left| \Re[\lambda_+] \right| \right\}. \quad (3.29)$$

For the parameters measured in our system, we estimate convergence to the low steady state at the same rate as protein decay and to the high steady state at a rate of approximately 33.6 min, on par with the timescale over which the early element must remain active in order for fate to be appropriately specified.

3.4.3.4 Parameter estimation

Regulatory function $f(P)$ of the *ftz* autoregulatory element We calculated the regulatory function ($f(P)$) of the *ftz* autoregulatory element (shown in Figure 3.4D) for the anterior boundary of stripe 4. To make this possible, we identified the boundary in a manually selected image frame prior to gastrulation by extracting two adjacent columns of cells, each corresponding to high or low Ftz concentration. For each cell, we obtained the MS2 signal, which is a proxy for the instantaneous rate of transcription [63, 17, 102], and the Ftz fluorescence for each time point. Next, we binned data points within a specific temporal window into ten quantiles. We averaged the MS2 and Ftz signals belonging to the same quantile, then fit a Hill function to the resulting values to obtain the regulatory function of the *ftz* autoregulatory element within that time window. We repeated the process in four temporal windows (-20 to -15, -15 to -10, -10 to -5, and -5 to 0 min) to obtain the trend shown in Figure 3.4E).

Translation rate α To calculate the translation rate, we simultaneously imaged the *ftz* transcription rate and the resulting Ftz concentration in a Ftz-MS2-LlamaTag construct (Figure 3.5D). We focused on nuclei lacking initial Ftz transcription, meaning that for these nuclei, no MS2 spots were detected prior to the experiment’s initiation. For these nuclei, we measured the MS2 signal (see Figure 3.5E for a sample trace), which is an approximation of the *ftz* mRNA production rate [63, 102, 18], and integrated this signal in order to obtain the total amount of mRNA produced (see Figure 3.5F; [63]). This integration was done by solving the differential equation for the mRNA $R(t)$ given by

$$\frac{dR}{dt} = r(t) - \gamma_R R(t), \quad (3.30)$$

where $r(t)$ is the transcription rate (i.e., the mRNA production rate reported by MS2 fluorescence), and $\gamma_R=0.099$ 1/min (see Table 3.1). An example of a resulting prediction for the amount of mRNA as a function of time is shown in Figure 3.5F.

Next, we performed a parameter sweep for the translation rate α and, for each value of α , we integrated $R(t)$ to predict the protein dynamics $P(t)$ using the following equation.

$$\frac{dP}{dt} = \alpha R(t) - \gamma_P P(t). \quad (3.31)$$

Examples of these predicted protein dynamics traces for different values of α are shown as black lines in Figure 3.5G.

The translation rate that results in the best fit (Figure 3.5G, yellow line) is recorded for each nucleus. We then calculated α values for each embryo by averaging best-fitted α for each single cell (Figure 3.5H). These values are averaged across $N = 129$ cells (embryo 1) and $N = 119$ cells (embryo 2), respectively. Then we averaged the resulting value of α between two embryos, giving us $\alpha = 0.082 \pm 0.004$ protein AU / (mRNA AU min) (see Table 3.1), which is used in our dynamical systems model in the main text.

It is important to note that the correlation between these mRNA dynamics and the protein dynamics, as determined by the translation rate, is independent of which enhancer element drove mRNA expression initially. Furthermore, to calculate α , we utilized a construct that includes both the early and autoregulatory elements rather than focusing solely on the autoregulatory element.

Early element transcriptional activity decay rate β To determine the decay rate β for mRNA production from the early element, we utilized fluorescence measurements of the early element MS2 reporter construct (Figure 3.5I). For each embryo, an exponential fit was applied to its average trajectory (Figure 3.5J). Subsequently, we calculated the mean decay rate across all embryos (see Table 3.1, $N = 2$ embryos) for incorporation into the dynamical systems model.

Scaling factor c We estimated c , the ratio of maximum production rate from the endogenous locus vs. the transgene, by approximating the solution to the dynamical system in Equation 3.7 between two time points for which we have empirical data. In particular, we began from the experimental observation that $\tilde{f}(t, P)$ plateaus shortly before gastrulation for nuclei with total Ftz levels above about $P_\infty^{thresh} = 1.5 \times 10^6$ a.u. Therefore, for nuclei that satisfy $P_{early}(t) \geq P_\infty^{thresh}$ during this time (Figure 3.10A), we can approximate the nonlinear system for the autoregulatory element by the following *linear* system

$$\begin{cases} \frac{dR}{dt} = cf(P_\infty^{thresh}) - \gamma_R R(t) \\ \frac{dP}{dt} = \alpha R(t) - \gamma_P P(t) \end{cases}, \quad (3.32)$$

where $f(P_\infty^{thresh})$ is now a constant. This system of equations has an analytical solution given by

$$\begin{cases} R(t) = cR(0)e^{-\gamma_R t} + \frac{cf(P_\infty^{thresh})}{\gamma_R} (1 - e^{-\gamma_R t}) \\ P(t) = P(0)e^{-\gamma_P t} + \frac{\alpha}{\gamma_P - \gamma_R} \left(R(t) - R(0)e^{-\gamma_P t} - \frac{cf(P_\infty^{thresh})}{\gamma_P} (1 - e^{-\gamma_P t}) \right) \end{cases}. \quad (3.33)$$

Rearranging these expressions gives

$$c = \frac{\gamma_P}{f(P_{\infty}^{thresh})(1 - e^{-\gamma_P t})} \left((R(t) - R(0)e^{-\gamma_P t}) - \frac{\gamma_P - \gamma_R}{\alpha} (P(t) - P(0)e^{-\gamma_P t}) \right). \quad (3.34)$$

Note that here $t = 0$ is assigned to the beginning of the time window over which the simulation is performed.

We calculated the solution to Equation 3.34 for the individual boundary nuclei in the same embryos as used to fit the gene regulatory function (Figure 3.4D). We restricted our estimations to the 3 min before gastrulation based on personal observations that the prediction accuracy of the simulation of this linearized system tended to fall after ~ 3 min. We derived the initial conditions $(R(0), P(0))$ from estimates of the late protein obtained by subtracting simulated early protein from the total protein trace (where the autoregulatory element was assumed to begin contributing at -20 min before gastrulation). We estimated c in a windowed approach whereby, for each nucleus, we simulated only over one empirical sample interval (10 s) for all intervals from 3 min before gastrulation (Figure 3.10B). We pooled all samples for individual time windows across all nuclei (Figure 3.10C and D) and averaged them to give a final estimate of $c \cong 0.45 \pm 0.02$ across the 3 min before gastrulation (where the error range is the standard error).

3.4.3.5 Simulations

Extracting initial conditions Initial conditions for all nuclei were drawn from measurements of the endogenous Ftz protein tagged with LlamaTag. In particular, we assumed the autoregulatory element was unresponsive before t_{on} (corroborated by simultaneous measurement of MS2 for the late element transgene in these constructs), such that all protein and mRNA up until 20 min before gastrulation was contributed by the early element. Therefore, our measurements until gastrulation give us the trace $P_{early}(t)$ for $t < t_{on}$.

Protein traces were smoothed before use and provided P_{-20} directly. The early mRNA concentration $R_{early}(t)$ was calculated from the third equation in Equation 3.1 by plugging in measurements from the smoothed trace to

$$R_{early}(t) = \alpha^{-1} \left(\frac{dP_{early}}{dt} + \gamma_P P_{early}(t) \right), \quad (3.35)$$

from which R_{-20} could be extracted. The early transcription rate $r(t)$ was then estimated from the derivative of the estimated mRNA trace $R_{early}(t)$ evaluated at $t = -20$.

Best predicted nuclei In order to be assured of the accuracy of our conclusions concerning the dynamics of the commitment process, in Figure 3.7 we decided to restrict our analysis to sets of nuclei whose simulated trajectories well matched the empirical traces. From our dataset, we identified such “best predicted” nuclei based on the cumulative error between a measured trajectory $\{(R_t, P_t)\}$ at discrete time points t and a simulated trajectory $\{(\hat{R}_t, \hat{P}_t)\}$ at the same time points as

$$\epsilon = \sum_t \sqrt{(\hat{R}_t - R_t)^2 + (\hat{P}_t - P_t)^2}. \quad (3.36)$$

From a histogram of the errors (Figure 3.14A), which was roughly bimodal, we identified a threshold of 7×10^7 to identify the 79 best predicted nuclei out of 118 nuclei total. Some sample traces from these best predicted nuclei are shown in Figure 3.14B while traces for nuclei with high cumulative error are shown in Figure 3.14C.

We also calculated the cumulative error when simulations were conducted for a gradual increase in autoregulatory responsiveness described by

$$\hat{f}(t, P) = (1 - e^{-\xi^*(t-t_{on})} f(P)), \quad t > t_{on}, \quad (3.37)$$

where $f(P)$ is as defined in Equation 3.6 and $\xi = 0.14$ 1/min corresponds to a half-life of 5 min (from the observation that the autoregulatory element transitions from unresponsive to fully responsive over 10 min; see Figure 3.4E). The resulting sample traces are shown in Figure 3.14B–C. Compared to the case where we assume that the autoregulatory element becomes active instantaneously at t_{on} (Figure 3.14A), there was a slight shift in the distribution toward lower error (88 best predicted with the same cutoff as before in Figure 3.14A), but the binary classification accuracy at gastrulation was the same. Therefore we opted to use $f(P)$ rather than $\hat{f}(t, P)$ for the main analysis.

Whole-embryo simulations We were curious about how accurately our model, which is based on measurements at the anterior boundary of stripe 4, would predict Ftz state for all nuclei spanning stripes 3, 4 and 5 measured during our experiments. Assuming all nuclei follow the same dynamics as given in Equation 3.1 and Equation 3.7, we repeated the analysis from Figure 3.6 to predict Ftz concentration at gastrulation. We achieved a binary classification accuracy of (74.6%, or 763 of 1036 nuclei). Interestingly, this is worse than the accuracy achieved from thresholding early protein alone (84.3%, or 873 of 1036 nuclei), which is itself comparably accurate to the predictions for the anterior boundary of stripe 4 (with or without the autoregulatory contribution). The bulk of classification errors for the whole-embryo simulations, whether from thresholding full simulations or thresholding early protein alone, were false positives at the posterior boundaries of stripes, as in the example plotted in Figure 3.12B. This indicates that something differs in the regulation of Ftz at the anterior boundaries of stripes as compared to the posterior boundaries. For example, it has been noted that the posterior, but not the anterior, boundaries of Ftz stripes are repressed by *sloppy paired* (*slp*) [28].

Delaying the onset of responsiveness of the autoregulatory element From a mathematical standpoint, we can treat a delay in t_{on} as a change in the starting time of the simulation, which introduces a corresponding change to the initial conditions ($r(t_{on})$, $R_{early}(t_{on})$, $P_{early}(t_{on})$) of the early module. In this way, the time at which the trajectory of the early module ($r(t)$, $R_{early}(t)$, $P_{early}(t)$) crosses the switching separatrix is the latest time at which

the autoregulatory element can become responsive and still commit a cell to the appropriate (high) fate (Figure 3.12C). This follows from three conditions: (1) the early element is time invariant, (2) the early element is independent of the autoregulatory module (the same is not true for the autoregulatory module, which takes the output of the early module as its input), and (3) Ftz fate corresponds to the Ftz concentration state at infinite time. (1) and (2) ensure that, even if we delay the autoregulatory element, we can continue to use Equation 3.1 to simulate early protein by just changing the initial conditions, while (3) ensures that the delayed start of the autoregulatory element will not change the location of the switching surface (which is relative to steady state, not to a transient state of the trajectory at a fixed point in time). Intriguingly, the trajectories of early protein run parallel to the separatrix before converging to cross it in a restricted region of parameter space (Figure 3.12C).

In Figure 3.7 we analyze the commitment window by varying t_{off} at the same time as t_{on} . In Figure 3.12D we report full results for simultaneous variation in t_{on} and t_{off} . The strictness with which cell fate must be specified determines the variation in timing that can be tolerated. For example, if the early element ceases production at or after gastrulation ($t_{off} > 0$), the autoregulatory element can delay responsiveness until -15 min and still guide at least 75% of cells to the appropriate fate. If the early element does not cease production until 20 min, then the autoregulatory element may turn on just after -6 min and still direct 75% of cells to the correct fate. From these results, we see that almost no cells commit to the high fate when $t_{on} > t_{off}$. For this reason, in the main text we always set $t_{off} > t_{on}$.

Stochastic simulations Stochasticity in gene expression during embryonic development can compromise or improve system function depending on the context [184, 128]. We sought to investigate whether stochasticity in gene expression (1) was sufficient to explain the prediction error rates of our deterministic models, and (2) could drive stochastic switching at appreciable rates. We examined these questions using stochastic differential equations (SDEs), assuming that the noise in our data arises solely from stochastic dynamics within the cells rather than from measurement noise. Generally, we expect this method to overestimate the error.

In a stochastic differential equation model, changes in the amount of late RNA dR_{late} and total protein dP_{total} over a time interval dt are given by

$$\begin{cases} dR_{late}(t) = (cf(P_{total}(t)) - \gamma_R R_{late}(t)) dt + \sigma_R(R_{late}(t)) dW \\ dP_{total}(t) = (\alpha R_{tot}(t) - \gamma_P P_{total}(t)) dt + \sigma_P(P_{total}(t)) dW, \end{cases} \quad (3.38)$$

where dW is Gaussian with mean 0 and variance dt [67]. The terms $\sigma_R(R_{late})$ and $\sigma_P(P_{total})$ scale the variance of the noise from one time increment to the next.

For this analysis, we used our experimental setup featuring endogenous Ftz-LlamaTag driving an autoregulatory element transgene tagged with MS2 as introduced in Figure 3.4. Because, over a small time interval where mRNA degradation is negligible, the MS2 signal reports on the rate of mRNA production, this signal gave us direct access to $dR_{late}(t_i)$ at

discrete time points t_i . As a result, we can use this measure of $dR_{late}(t)$ to estimate for the late mRNA $R_{late}(t)$ by integrating the MS2 signal following

$$R_{late}(t_i) = R_{late}(t_{i-1}) + dR_{late}(t_i)(t_i - t_{i-1}) \quad (3.39)$$

under the assumption that $R_{late}(t_{on}) = 0$ (meaning that the autoregulatory element only becomes responsive at time t_{on}). Since we also have direct measurements of $P_{total}(t)$, this allows us to rearrange Equation 3.38 so as to estimate the noise contribution $\sigma_R(R_{late}(t))dW$ from

$$\sigma_R(R_{late}(t))dW = dR_{late}(t) - (cf(P_{total}(t)) - \gamma_R R_{late}(t)) dt. \quad (3.40)$$

If we further assume that $R_{early}(t)$ is noiseless and therefore given by the deterministic solution in Equation 3.12, we can estimate the total mRNA as $R_{total}(t) = R_{early}(t) + R_{late}(t)$. Then, since we have simultaneous measurements of total Ftz P_{total} , we can also rearrange the lower equation in Equation 3.38 to estimate $\sigma_P(P_{total}(t))dW$ from

$$\sigma_P(P_{total}(t))dW = dP_{total}(t) - (\alpha R_{total}(t) - \gamma_P P_{total}(t)) dt. \quad (3.41)$$

We can perform the above analysis on individual measured traces to produce a large number of sample points of $\sigma_R(R_{late})dW$ and $\sigma_P(P_{total})dW$. With these data we will aim to estimate $\sigma_R(R)$ and $\sigma_P(P)$. We assume the noise characteristics are time invariant, which allows us to pool all samples at all time points and bin them by the corresponding R_{late} or P_{total} . For example, for protein, we treat each protein concentration bin k as a population of samples of $\sigma_P(\bar{P}_{total}^k)dW$ where $\bar{P}_{total}^k$ is the mean of the samples in bin k (Figure 3.13A). Since we assume dW is normally distributed with variance dt , we divide all sample values by $\sqrt{dt}$ and fit a normal distribution to the resulting distribution within each protein concentration bin (Figure 3.13B). We found that the variances $\sigma_P(P)^2$ are quite well approximated by a linear relation $\sigma_P(P)^2 = aP + b$ (Figure 3.13C, right). The mRNA variance was estimated similarly and also found to fit a linear relation (Figure 3.13C, left). Noise was estimated from all available trajectories, regardless of whether they were part of the stripe 4 anterior boundary.

Having estimated the noise, we investigated whether stochasticity could explain the error rate in our predictions of Ftz expression state at gastrulation. We simulated $N_{sim} = 100$ experiments, each consisting of $N_{nuc} = 118$ nuclei evenly split between those deterministically predicted to be on and those deterministically predicted to be off. Specifically, we found the convex hull defined by the experimentally measured initial conditions for nuclei at the stripe 4 anterior boundary, and drew random initial conditions for the stochastically simulated nuclei from a uniform distribution within this hull. We assigned to each simulated experiment $\frac{N_{nuc}}{2}$ points below the blue surface in Figure 3.7C and $\frac{N_{nuc}}{2}$ above the surface without replacement (i.e., every nucleus in every simulated experiment has a unique set of initial conditions). Individual stochastic trajectories were generated using the Euler-Maruyama method, with the modification that protein and mRNA concentrations were forcibly lower bounded at 0 (i.e., random fluctuations that would bring concentrations to negative values were capped

to instead bring the concentration to zero). Trajectories were simulated for 20 min until gastrulation and thresholded with the same value as for our deterministic simulations (Figure 3.13D).

From our $N_{sim} = 100$ simulated experiments, we calculated a distribution of error rates given by

$$\frac{\text{total false positives} + \text{total false negatives}}{\text{total number nuclei}}, \quad (3.42)$$

where we compare the predicted outcomes from deterministic simulations to the “ground truth” of the stochastic simulations.

In Figure 3.13E we plot the cumulative distribution of error rates for our simulations (black), broken down into false negatives (red) and false positives (green). The dashed vertical lines indicate the experimentally measured error rates with the same color code. Where the vertical lines intersect the corresponding cumulative distributions indicates the probability of measuring an error rate up to that rate. If the system is really described by the stochastic dynamics we have inferred, then the most likely error rates are those that intersect the curves where their slope is highest.

We found that the empirical error rate across 3 embryos was roughly twice that of the most likely error rates from our stochastic simulations corresponding to the middle of calculated cumulative distribution functions (Figure 3.13E, top), with the majority of errors being false negatives. We knew from observation that one embryo had a large number of false negative predictions, and, interestingly, if we exclude this embryo from analysis, then the empirical error rate aligns well with what the stochastic model predicts to be most likely (Figure 3.13E, bottom). This result suggests that many of our prediction errors can likely be attributed to stochastic fluctuations.

Having determined that the error of the model in predicting Ftz expression state at gastrulation is comparable to the error expected when considering gene expression stochasticity, we next turned to the question of whether gene expression stochasticity is expected to play a large role in the long-term Ftz fate of cells in which the early element is no longer active. We ran $N = 100$ simulations beginning from the high steady state and calculated the distribution of first-passage times to particular protein values (Figure 3.13F, left) or to within some Euclidean distance of the opposite steady state (Figure 3.13F, right). These measures give an approximation of the switching rate depending upon how stringently one defines a threshold for switching.

Both trends indicate that switching from high to low occurs at a much faster rate than low to high, with conservative rates of stochastic switching between the high and low Ftz fates of around 2 hr and between low and high fates of approximately 3 hr. For comparison of timescales, Ftz stripes are no longer experimentally detected before the end of germband extension [74, 23], which occurs approximately 1.5 hours after gastrulation [149]. Thus, we find no strong evidence that stochastic switching should contribute significantly to Ftz stripe patterning.

3.4.4 Quantification and statistical analysis

3.4.4.1 Image processing

Image analysis of live embryo videos was performed based on the protocol in [63, 1], which included nuclear segmentation, spot segmentation, and tracking. In addition, the nuclear fluorescence of Ftz was calculated based on a nuclear mask generated from the MCP-mCherry channel. Ftz concentration for individual nuclei was extracted based on the integrated amount from maximum projection along the z-stack. The GFP background was calculated based on a control experiment and subsequently subtracted from the data.

3.4.4.2 Numerical analysis and simulations

Numerical analysis and simulations were carried out using custom scripts in MATLAB (2017b). The switching separatrix was estimated using a modification of the algorithm in [155], which employs a combination of bisection and random sampling to estimate the upper and lower bounds of the separatrix surface. Stochastic differential equations were simulated using the Euler-Maruyama method.

3.5 Supplementary Figures

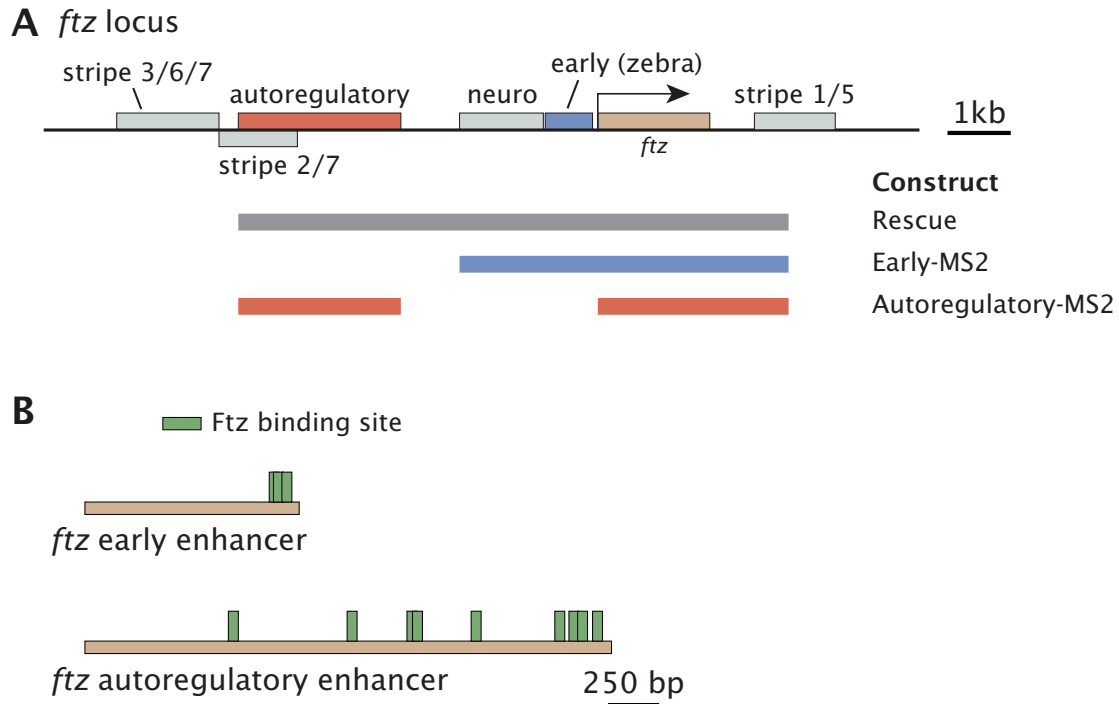


Figure 3.8: ***Ftz* regulatory region and binding sites.** (A) Schematic of the *ftz* locus with known regulatory regions. The regions included in both the early element and autoregulatory element reporters used in this study are highlighted, in addition to the region that can rescue the *ftz*-null mutant [83]. (B) Predicted Ftz binding sites in the *ftz* early and autoregulatory elements. The *ftz* early and autoregulatory elements are 1046 bp and 2571 bp segments from the endogenous *ftz* locus, respectively. Multiple Ftz binding sites are predicted to be found within the *ftz* early and autoregulatory enhancer by PATSER [80] using multiple Ftz position weight matrices (PWMs) from [52]. However, the Ftz binding sites in the early element appear not to be functional [83, 45, 138]. Only binding motifs with PATSER scores higher than 9.0 are shown.

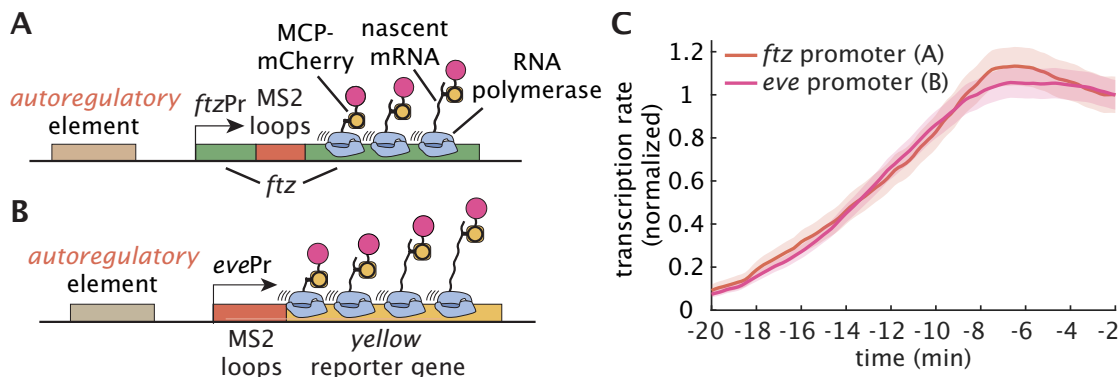


Figure 3.9: **Comparison of temporal dynamics between autoregulatory element constructs driven by the *ftz* and *eve* promoters.** (A-B) Schematics of constructs with *ftz* driven by the autoregulatory element and the (A) *ftz* or (B) *eve* promoters. (C) Both *ftz* promoter ($N = 6$ embryos) and *eve* promoter ($N = 7$ embryos) constructs show the same transcriptional initiation dynamics prior to gastrulation. Data are normalized based on the transcription rate at $t = -2$ min. The transcription rate is calculated from the measured MS2 signal, which is an approximation of the mRNA production rate [63, 102, 18].

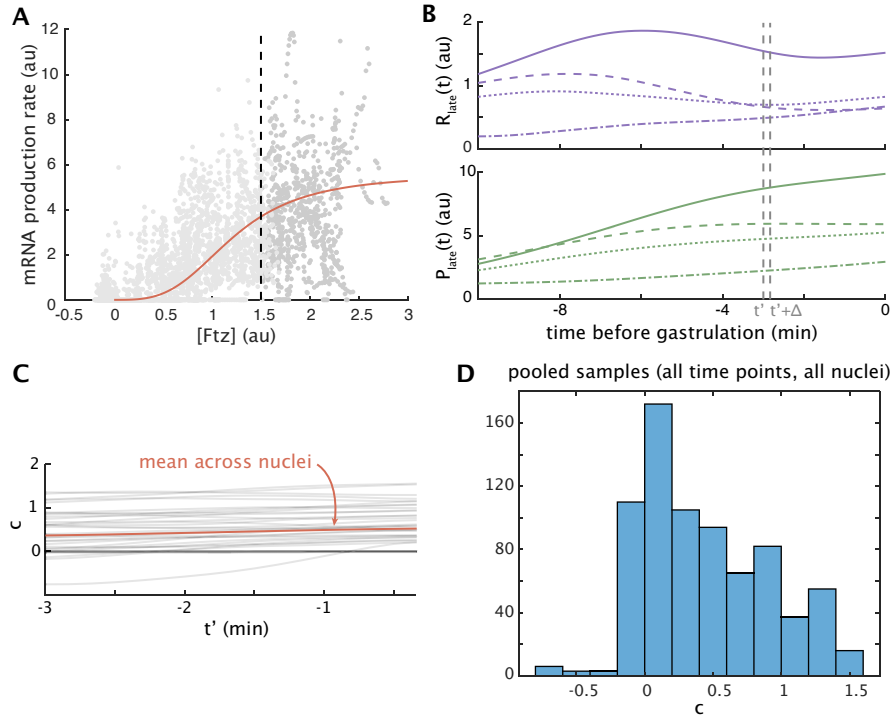


Figure 3.10: **Procedure for estimating c , the ratio of the maximum mRNA production rate from the endogenous to the transgene.** (A) We use a subset of samples for which the gene regulatory function is approximately saturated across the entire 3 min before gastrulation. Red, the gene regulatory function. Gray dots are individual time points from individual nuclei during the 3 min before gastrulation. Dark gray dots belong to nuclei for which $P_{total}(t) > P_{\infty}^{thresh}$ for all times between -3 and 0 min relative to gastrulation. Light gray dots correspond to nuclei that did not match this criterion. (B) Top, estimated late mRNA traces and bottom, estimated late protein traces for four sample nuclei drawn from the subset identified in (A). Gray dashed lines indicate one sample interval from t' to $t' + \Delta$ used to estimate c (see Materials and Methods). (C) Gray, estimates of c for each interval pictured in (B) and each nucleus identified in (A). Red, the mean at each time point across all nuclei. Note that most variation is between nuclei rather than across time points. (D) Histogram of c for all time intervals across all nuclei, mean 0.45, standard error 0.02.

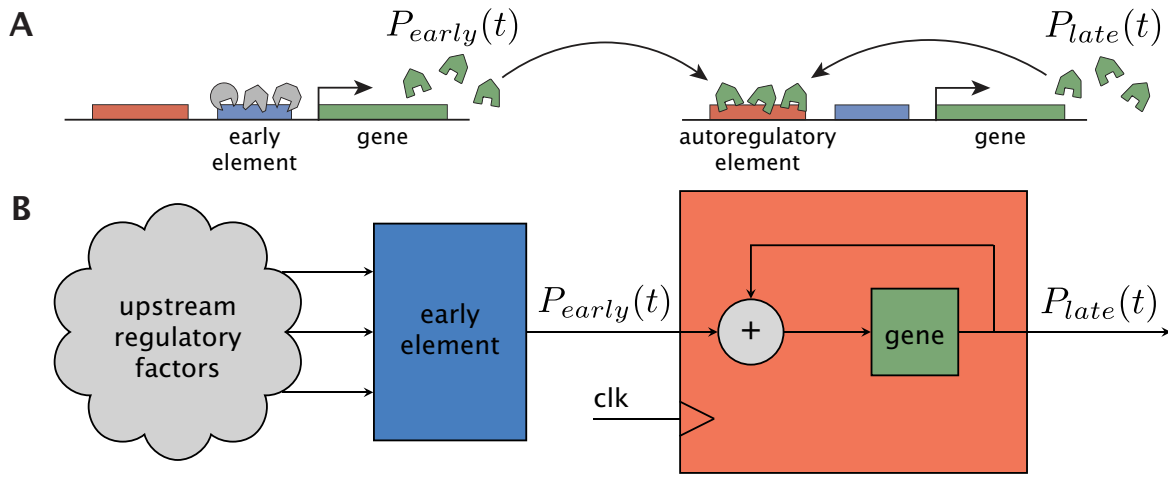


Figure 3.11: **Two ways to visualize a genetic network as an interconnection of modules.** (A) Schematic of Ftz regulation in which populations of Ftz protein are differentially labeled based on whether they are generated by regulation from the early element ($P_{early}(t)$) or from the autoregulatory element ($P_{late}(t)$). The two populations are summed to generate the total Ftz concentration that binds the autoregulatory element. (B) The Ftz regulatory system can be visualized as a block diagram to highlight how a modular network decomposition might be inspired by similar breakdowns of electrical circuits and other engineering control systems [36]. Here, unspecified upstream factors regulate the early element module, which produces output $P_{early}(t)$. This, in turn, acts as the input to the autoregulatory module, which has been schematized as a latch with a clock signal (clk) to show that the element only becomes responsive at a particular time. The output $P_{late}(t)$ can be summed with $P_{early}(t)$ to recover total Ftz concentration. Relative to the more traditional schematic in (A), this representation includes dynamic signals ($P_{early}(t)$ and $P_{late}(t)$), a timing element that triggers autoregulatory responsiveness, and a clear separation of inputs and outputs from each module.

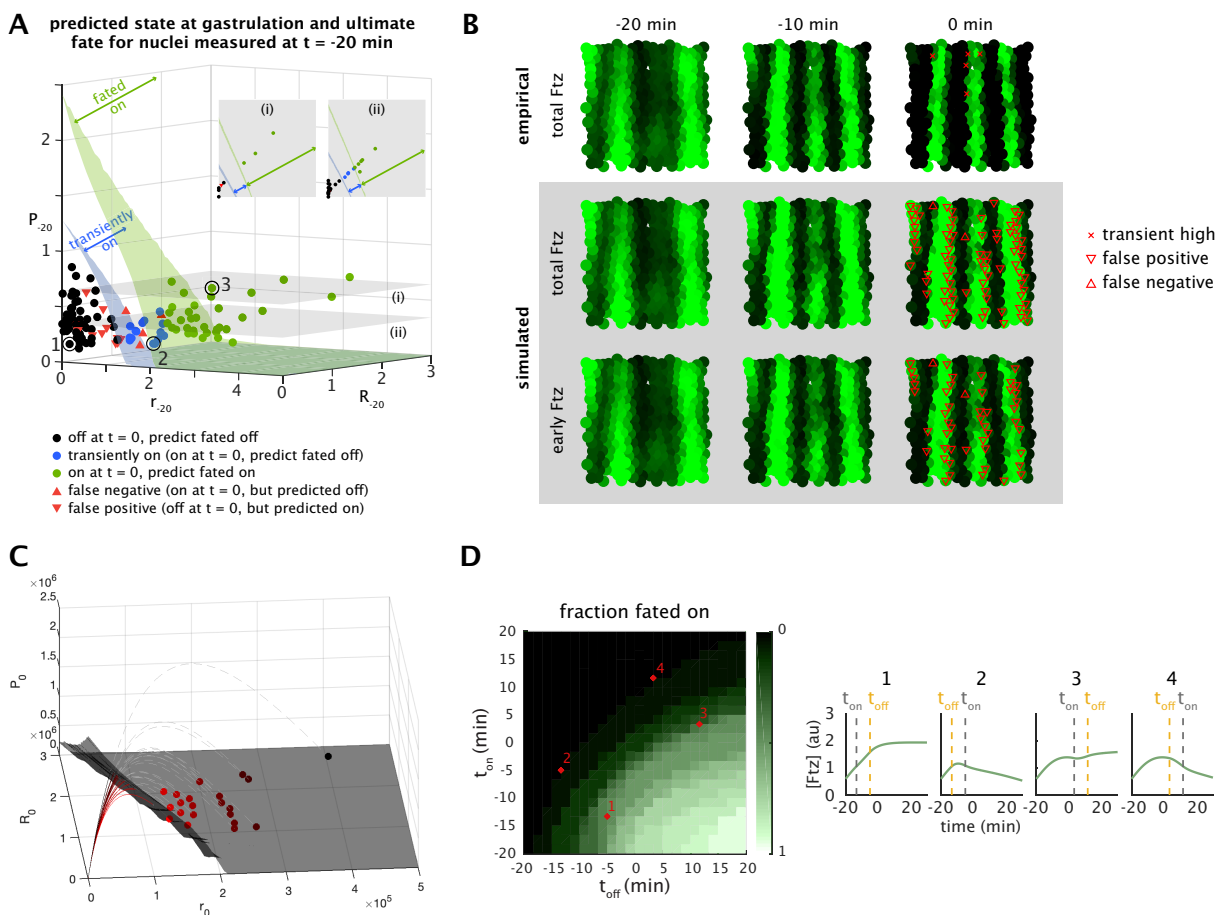


Figure 3.12: **Detailed deterministic simulation results.** (A) The switching separatrix (green surface; repeated from Figure 7) compared to an analogous surface for binary Ftz state at gastrulation (blue surface), which separates nuclei that are high at gastrulation (above surface) from those that are low at gastrulation (below surface). All analyzed nuclei whose binary state at gastrulation was accurately predicted are plotted as colored circles. Remaining nuclei are plotted as red triangles, upward-facing if predicted OFF at gastrulation but measured ON, vice versa for downward-facing. Insets show two thin-width slices through the surfaces at fixed initial protein levels P_{-20} between (i) 0.3–0.4 a.u. and (ii) 0.6–0.7 a.u. (B) Whole-embryo simulation for the same embryo as pictured in Figure 6D. Pixel intensities are normalized to the steady-state high Ftz level. *(continued on the next page)*

Figure 3.12: **(continued) Detailed deterministic simulation results.** (C) Dashed gray lines denote trajectories $(r(t), R_{early}(t), P_{early}(t))$ corresponding to individual high-fated nuclei ($N = 21$, same as analyzed in Figure 7). Circles denote initial conditions at time -20 min colored based on the amount of time until the corresponding trajectory crosses the switching separatrix, with darker colors indicating longer times. This crossing time corresponds to the longest acceptable delay in the onset of autoregulatory responsiveness without changing the fate of the cell. (D) Left, heat map showing the fraction of cells predicted to adopt the high Ftz fate for different combinations of t_{on} and t_{off} . Right, examples of simulated Ftz traces for a single nucleus with the timing conditions corresponding to the red marks on the heatmap.

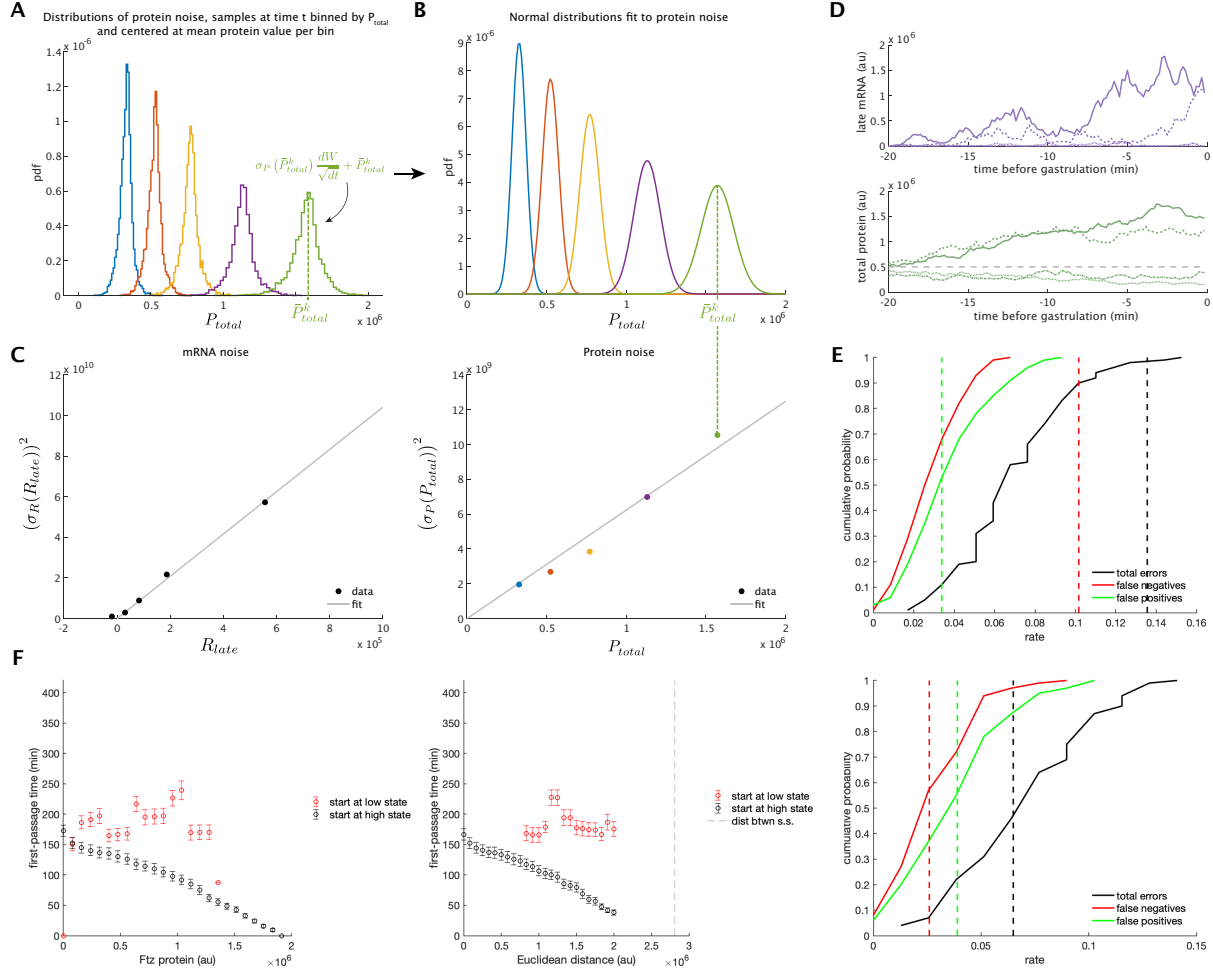


Figure 3.13: Detailed stochastic simulation results. (A–C) Procedure for estimating mRNA and protein expression noise following the method described in Materials and Methods “Stochastic simulations”. We centered the samples from each bin at the corresponding $\bar{P}_{total}^k$ to generate the pictured histograms. (D) Top, late mRNA and bottom, total protein traces simulated for individual nuclei obeying the stochastic dynamics in Equation 38. Gray dashed line in the bottom plot is the threshold to be classified as high Ftz at gastrulation. (E) Solid lines, cumulative distribution functions for error rates in the prediction of results for $N = 100$ simulated stochastic experiments (using the stochastic differential equation in Equation 38), when using a deterministic simulation to make the predictions. Dotted vertical lines correspond to empirical error rates for three embryos with $N_{nuc} = 118$ nuclei (left) across 3 embryos, or excluding one embryo that had a high number of deterministic false negatives for $N_{nuc} = 78$ nuclei (right) across 2 embryos. *(continued on the next page)*

Figure 3.13: *(continued)* **Detailed stochastic simulation results.** (F) Left, time to first passage of a stochastic trace starting from the high (black) or low (red) fate to a given protein value. Right, time to first passage of a stochastic trajectory with a given Euclidean distance from the opposite steady state. In both plots, red denotes simulations beginning in the low Ftz state and black denotes simulations beginning in the high Ftz state.

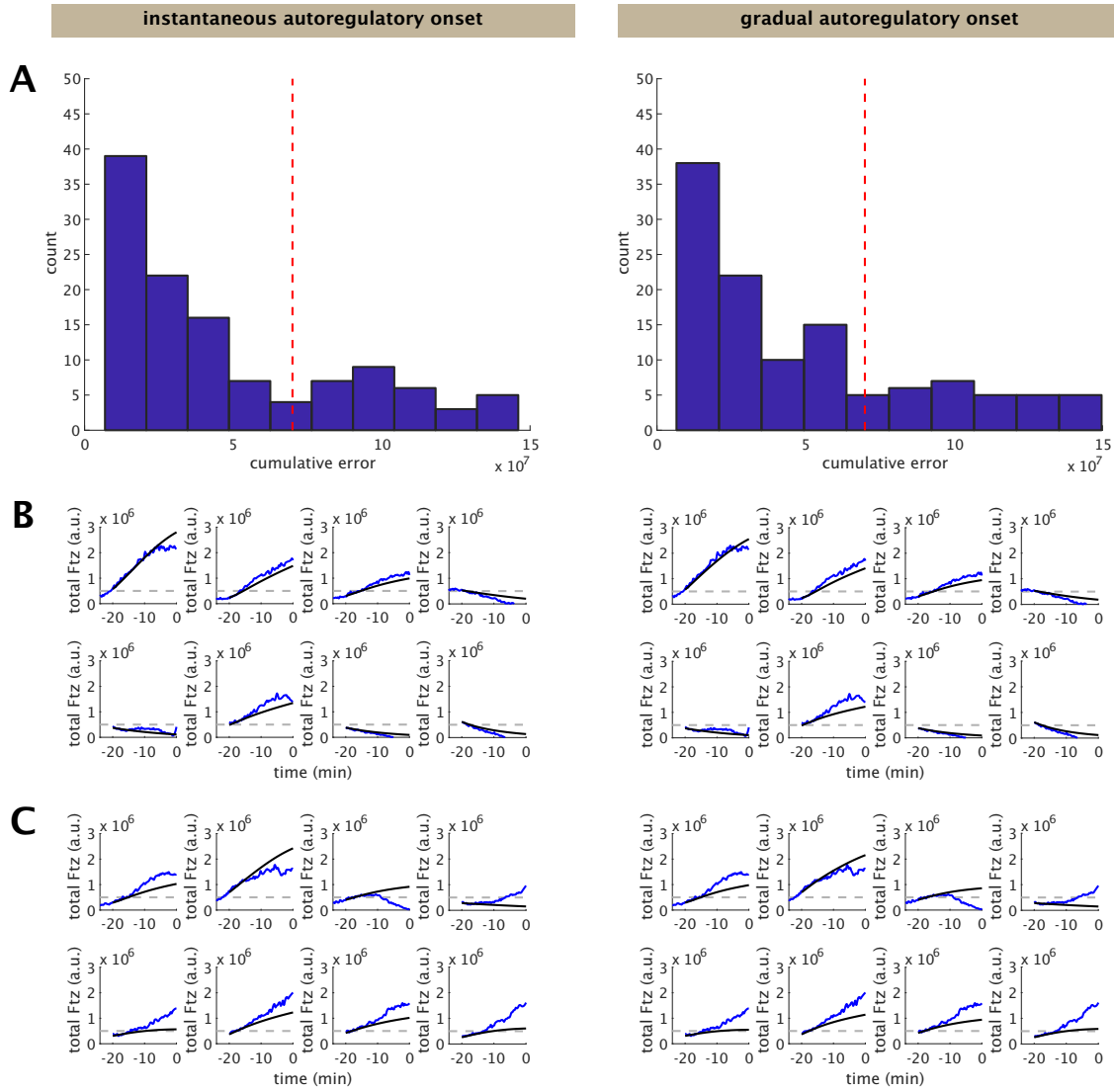


Figure 3.14: **Classifying nuclei by the cumulative error in model prediction.** Left column, simulations conducted assuming that the autoregulatory element becomes instantaneously responsive, as in Equation 6 and the main text. Right, simulations conducted assuming a gradual increase in responsiveness of the autoregulatory element, following the dynamics in Equation 37. **(A)** Histogram of cumulative errors for $N = 118$ nuclei. **(B)** Sample traces of best predicted nuclei. In each plot, the blue curve is the measured trace and the black line is the corresponding predicted trace. The dashed gray line is the threshold for classifying a nucleus as “on” at gastrulation. **(C)** As for (B), but for nuclei that did not qualify to be best predicted.

3.6 Supplemental Table

Parameter	Description	Source	Value	Units
α	translation rate	this work	0.082 ± 0.004	protein AU / (mRNA AU · min)
τ_R	mRNA half-life	[51]	7	min
γ_R	mRNA decay rate	$\frac{\ln 2}{\tau_R}$	0.0990	1/min
τ_P	protein half-life	[18]	7.9	min
γ_P	protein decay rate	$\frac{\ln 2}{\tau_P}$	0.0877	1/min
$f(P)$	autoactivation gene regulatory function	this work	$\frac{aP^n}{K^n + P^n}$	mRNA AU / (protein AU)
a	maximum transcription rate	this work	$5.555e5 \times 10^5$	mRNA AU / (protein AU)
K	-	this work	1.216×10^6	protein AU
n	Hill coefficient	this work	3.264	-
c	ratio of transcription rate endogenous to transgene	this work	0.45 ± 0.02	-
β	decay rate of early mRNA production	this work	0.0479 ± 0.0021	1/min

Table 3.1: **Quantitative parameters used throughout this work.**

Chapter 4

Enhancer design principles revealed by binding-site reorganization and modular dissection

4.1 Introduction

Enhancers are stretches of non-coding regulatory DNA sequence that transcription factors, such as activators and repressors, can bind to regulate when, and where a gene is expressed. These regulatory elements are essential during embryonic development [60], and their misregulation has been implicated in various diseases [90]. Yet, despite the ability to find enhancers and identify most of their transcription factor binding sites with relative ease, we are still incapable of predicting—let alone synthesizing—enhancer function from knowledge of the number, placement, and affinity of binding sites within them.

A significant challenge in synthesizing functional enhancers from scratch lies in determining the essential and complete sites that are essential for the enhancer function [169, 90]. Given that past attempts to create enhancers based solely on known binding sites have proven unsuccessful [169], it is clear that our understanding of the information within enhancers is still incomplete. Recent high-throughput mutation studies indicate that almost all alterations on a developmental enhancer have resulted in changes in gene expression patterns [59]. This suggests that these enhancers carry dense regulatory information, complicating our understanding [59]. To successfully construct a functional enhancer from scratch, it is crucial that we identify all the essential elements that can reproduce the activities of endogenous enhancers.

In addition to identifying necessary binding sites, understanding the organizational flexibility of enhancers is also crucial for successful enhancer synthesis. This pertains to how binding sites must be arranged to create a functional enhancer and to what extent their arrangement can be modified without affecting function. Two models, the “Enhanceosome model” and the “Billboard model”, attempt to answer this [6, 156]. The enhanceosome

model, drawn from interferon-beta enhanceosome studies, suggests that an enhancer only operates when all factors are in a specific DNA arrangement [127]. Conversely, the billboard model posits that enhancer function does not necessitate a particular DNA arrangement—the presence of all essential factors is sufficient. If the enhanceosome model proves accurate, additional mapping between binding site arrangement and functionality will be required. However, if the billboard model is correct, the synthesis process could be significantly simplified as it allows more flexible configurations.

A study involving evolutionary comparisons of enhancers from multiple closely related species reveals that various enhancer sequences can yield very similar expression patterns [76] and gives credit to the billboard model. Yet, there remains controversy over whether these sequences maintain a conserved global organization of transcriptional information [76, 77, 39]. Thus, to fully explore this concept and gain a deeper understanding of the constraints of enhancer organization, a systematic reorganization of enhancer sequences—particularly those not represented in evolutionary trajectories—is indispensable.

To address these outstanding questions and challenges in the field of enhancer biology, we focused on the 497-bp minimal segment of the *eve* stripe 2 enhancer, known for generating a standard stripe pattern in the anterior region of the *Drosophila* embryo [7, 9]. Previous attempts to recreate this enhancer using only known binding sites were unsuccessful [169]. However, incorporating a few newly identified binding sites allowed us to reconstruct the enhancer successfully. The reconstructed enhancer yielded a stripe pattern similar to its wild-type counterpart, indicating that known binding sites alone can be sufficient to generate the *eve* stripe 2 pattern.

We then examined the flexibility of enhancer organization by randomly reorganizing the enhancer sequence. Past studies have attempted the reorganization of binding sites within developmental enhancers; however, the result can be difficult to interpret due to the possibility of introducing new variables [162]. Using our reconstituted enhancer avoids introducing these complication factors. Intriguingly, we found that even a completely random rearrangement often did not disrupt the enhancer’s stripe pattern. This finding strongly supports the “Billboard model”, which posits that alterations in the binding site arrangement do not disrupt the enhancer’s stripe pattern.

Finally, we attempted to further minimize the *eve* stripe 2 enhancer. We showed that the tandemization of merely a third of the enhancer sequence was adequate to generate a stripe pattern, showing that the complete stripe enhancer logic was contained within this local module. In other words, we succeeded in synthesizing a further “minimal” stripe enhancer.

Our findings have provided critical advancements in understanding and synthetically reproducing enhancer function, specifically within the context of the *eve* stripe 2 enhancer. By demonstrating that known binding sites alone can generate this enhancer pattern, validating the “billboard model”, and distilling the enhancer logic to a minimal sequence module, our study navigates significant hurdles in enhancer biology. This not only deepens our grasp of the organizing principles underlying enhancer function but also paves the way for the rational design of regulatory sequences.

4.2 Results

4.2.1 Successful reconstitution of *even-skipped* (*eve*) stripe 2 enhancer

To assert our understanding of a developmental enhancer and be able to synthetically produce one, it is imperative to establish confidence in its core regulatory logic, which includes the identification of the key binding sites within the enhancer, and of the identities of transcription factors that bind those sites and contribute to its expression pattern. The ultimate test of this understanding is our ability to reconstitute the enhancers' critical functional features based solely on known binding sites [169]. Interestingly, even for arguably one of the most understood enhancers in all of biology—the *even-skipped* (*eve*) stripe 2 enhancer [151, 9]—this feat of reconstituting a gene expression pattern from its known binding sites is yet to be accomplished [169].

To test this, we focused on one of the best-studied developmental enhancers in biology—the *even-skipped* (*eve*) stripe 2 enhancer [9]—and asked whether the known binding sites for the enhancer are sufficient for generating its characteristic stripe pattern. *Eve* stripe 2 enhancer is known to be activated by several activators (including Bicoid, Hunchback, Caudal, and Zelda), as well as repressors (including Giant, Krüppel, and Sloppy paired) (Figure 4.1A). Bicoid and Hunchback, transcription factors primarily expressed in the embryo's anterior region, are posited to be the primary activators for *eve* stripe 2, with Caudal and Zelda also playing a significant role [170]. Zelda, a uniformly expressed activator, is hypothesized to facilitate chromatin opening [169]. Contrastingly, repression from the anterior and posterior sides is driven by Giant and Krüppel, respectively, which results in the formation of a stripe pattern in the area in between.

To validate our understanding of the *eve* stripe 2 enhancer, we engineered a synthetic version, preserving the known binding sites of the key activators (Figure 4.1B). Previous research has indicated that flanking sequences directly affect the binding of Bicoid [131, 181]. Consequently, we included an additional eight base pairs of flanking sequences around the Bicoid and Caudal binding sites. We substituted the remaining sequence with neutral sequences created via SiteOut [52]. These computationally generated sequences are deprived of binding sites for transcription factors in the early embryo. Alongside the version replaced with neutral sequences, we also generated a shorter variant of the enhancer, removing all sequences except the binding sites (341 bp in total length).

To confirm that the reconstituted and shorter versions of the *eve* stripe 2 enhancer successfully recapitulate key endogenous features, we analyzed the temporal and spatial dynamics of stripe expression patterns. We created transgenic fly lines with an MS2 reporter [63, 114] directly linked to the enhancer of interest (Figure 4.1C). Upon transcription, the MS2 reporter forms a hairpin-like structure, leading to the binding of maternally deposited MCP-GFP proteins and creating a spot directly indicative of transcriptional activity (Figure 4.1C) [63, 114]. This process allows for the direct quantification of the essential temporal and spatial features of the enhancer elements.

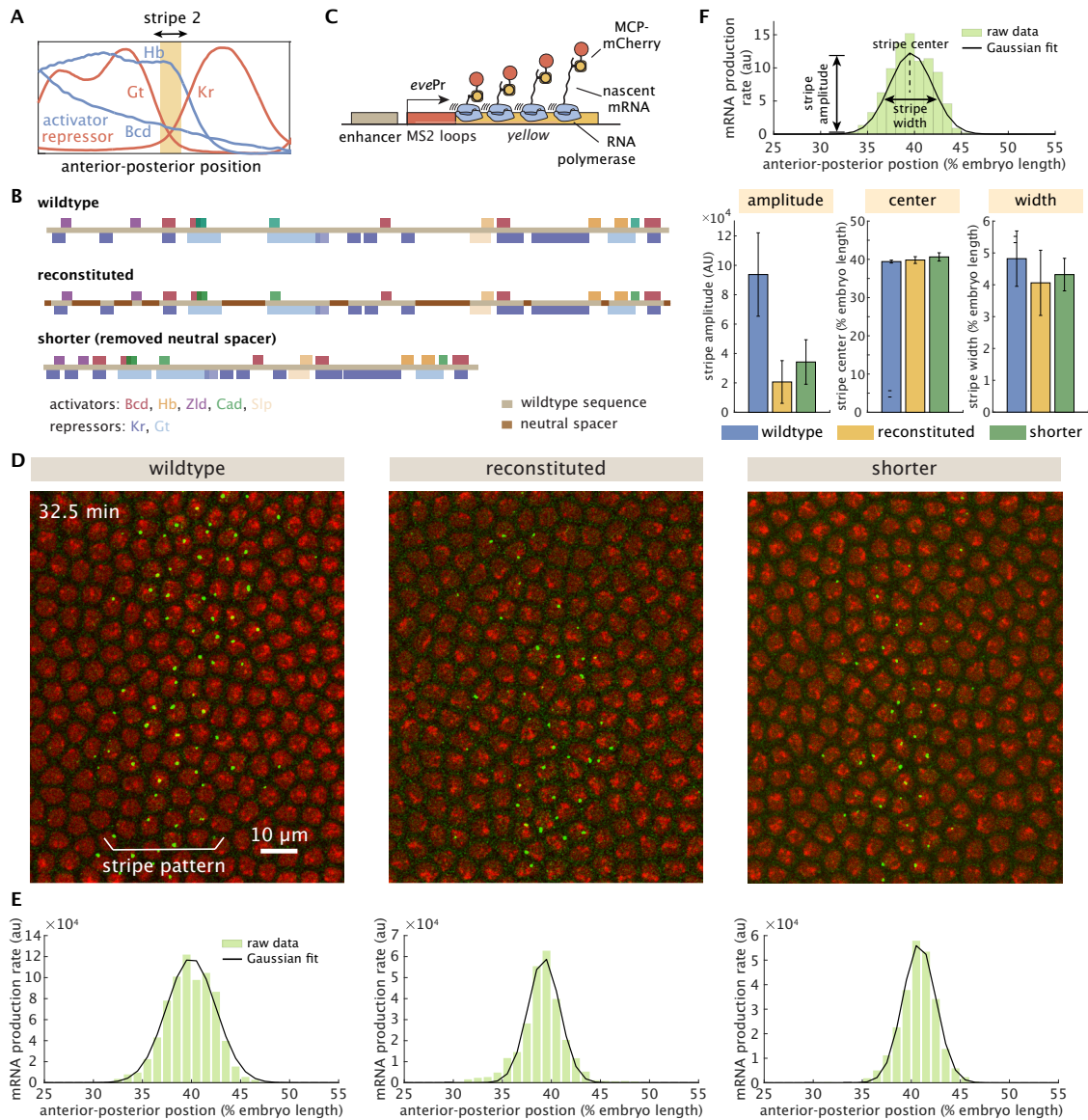


Figure 4.1: **Reconstituted *eve* stripe 2 enhancer recapitulates endogenous stripe features.** (A) Regulation of *even-skipped* stripe 2. The stripe is activated by transcriptional factors, including Bicoid (Bcd), Hunchback (Hb), Caudal (Cad, not shown) ubiquitous factor Zelda (Zld, not shown), while repressed by Giant (Gt), and Kruppel (Kr). (B) Sequence comparison between wildtype, reconstituted and shorter versions of the *eve* stripe 2 enhancer. (C) Design of the enhancer reporter construct. Enhancer is attached to basal *eve* promoter that is followed by 24x MS2 loops and a segment of the *yellow* gene. (*continued on the next page*)

Figure 4.1: *(continued)* **Reconstituted *eve* stripe 2 enhancer recapitulates endogenous stripe features.** (D) Example snapshots of wildtype, reconstituted and shorter versions of *eve* stripe 2 enhancers. Maternally deposited MS2 coat protein (MCP) binds to MS2 stem-loops in the nascent RNA formed by RNAP molecules elongating along the body of the reporter construct, leading to the accumulation of fluorescence at sites of nascent transcript formation. Nuclei are labeled in red using Histone-RFP fusion protein. (E) Raw distribution and corresponding Gaussian fits of snapshots. (F) Stripe amplitude, center and width were computed by fitting a Gaussian distribution to the stripe pattern and are compared between multiple constructs. (G) Quantitative comparison between wildtype ($n = 4$ embryos), reconstituted ($n = 5$ embryos), and shorter ($n = 4$ embryos) versions show that they give rise to almost identical stripe patterns but with distinct expression strength.

Intriguingly, both the reconstituted enhancer and its shortened version exhibited stripe patterns echoing the endogenous *eve* stripe 2 (Figure 4.1D and E). To quantify and compare these patterns, they can be fit to a Gaussian function to extract quantitative parameters, including the stripe amplitude, stripe center, and stripe width (Figure 4.1F). The stripe center and width were consistent with those of the endogenous pattern (Figure 4.1G), suggesting that we have successfully reconstituted the *eve* stripe 2 pattern. However, the amplitude presented a discrepancy, with both the reconstituted and shortened versions demonstrating a lesser degree of expression (Figure 4.1F). This is possibly due to the removal of some weak binding sites within the enhancer.

4.2.2 Systematic reorganization retains enhancer logic

Our successful reconstitution of the *eve* stripe 2 expression pattern led us to identify a minimal set of transcription factors capable of capturing its key regulatory logic. In other words, we have obtained the minimal set of binding sites that are required for the synthesis of a functional enhancer.

Utilizing the reconstituted enhancer, we then tackled a key question regarding developmental enhancers — their organizational flexibility and appropriate functional units. By comparing *eve* stripe 2 enhancer sequences from species with large evolutionary distances, past evolutionary studies observed that regulatory sequences whose binding sites have been rearranged could result in similar pattern [77]. Although the sequence is completely reorganized, there are still some conserved small modules of 10-50bp that can be discovered through careful sequence alignments [76, 39, 77]. While the existence of these conserved modules is by no means a definite proof of modularity, it nevertheless suggests a relatively flexible organization of the enhancers and posits these small modules as the basic units for enhancer function and evolution. Understanding the organizational flexibility of the enhancer with rational perturbation of the reconstituted enhancer, thus acts as complementary to the related evolutionary studies.

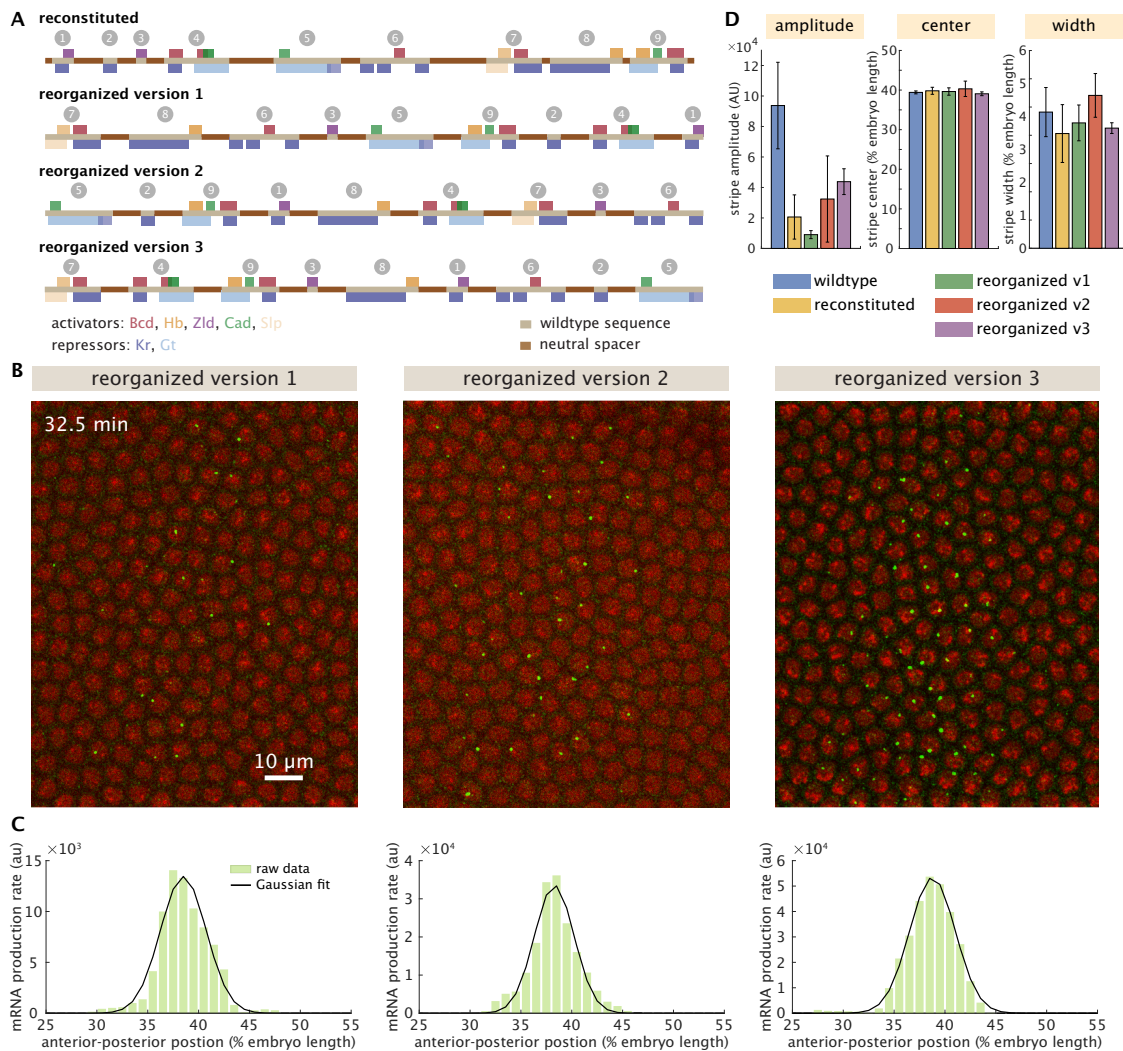


Figure 4.2: **Enhancer reorganization retains expression pattern.** (A) Sequence comparison between reconstituted and reorganized versions of the *eve* stripe 2 enhancer. (B) Example snapshots of reorganized versions of *eve* stripe 2 enhancers. (C) Raw distribution and corresponding Gaussian fits of snapshots. (D) Comparison of stripe amplitude, stripe center, and stripe center for wildtype ($n = 4$ embryos), reconstituted ($n = 5$ embryos), as well as reorganized versions 1 ($n = 3$ embryos), 2 ($n = 3$ embryos) and 3 ($n = 3$ embryos). After random permutation, stripe center and stripe width remain unchanged, while stripe amplitude changes significantly.

To study the organizational flexibility of the enhancer and its basic functional units, we asked whether the organization is constrained at the scale of small modules (10-50bp). Some of them are conserved across distant species, making them ideal candidates for the basic organization units [76, 39, 77]. To do so, we redistributed the binding sites inside the enhancer into 9 large functional modules based on their separation by neutral sequences (Figure 4.2A). Each module can be regarded as an independent functional unit. Then, we generated four randomly permuted versions of the enhancer to understand the influence of this reorganization on the enhancer function (Figure 4.2A). The enhanceosome model predicts this perturbation to be disruptive, while the billboard model allows for such a reorganization with minimal influence on the function. Surprisingly, such a drastic reorganization only resulted in a minor influence on its pattern (Figure 4.2B). The stripe pattern is retained, as the stripe position and width remain the same. However, the stripe amplitude was modulated, with two of them becoming stronger, while the other one became weaker. This is in direct contradiction to the enhanceosome model while in line with the billboard model if we consider these modules as basic organizational units.

This stunning result suggests that the interactions most important to the enhancers are local interactions rather than global ones. These smaller modules can be considered basic organizational units, and they do not follow the enhanceosome model, while their behavior is more similar to the billboard model except for its influence on amplitude. The consistent stripe pattern suggests that repression activity remains consistent throughout the reorganization. However, strong variations in the amplitude imply a larger influence on the activation activity.

4.2.3 Further minimization of *eve* stripe 2 enhancer

While we have demonstrated a flexible enhancer organization, it remains uncertain if every module is essential for the enhancer's function. In other words, not all modules might be required to ensure the correct expression pattern. Within the *eve* stripe 2 enhancer, certain modules house both activator and repressor binding sites. Should these modules independently produce a stripe pattern, this would suggest the potential for an even more minimized enhancer.

To further simplify the stripe enhancer, we segmented the *eve* stripe 2 enhancer into three approximately equal lengths: modules 1-4, 5-6, and 7-9. Notably, only the third segment (modules 7-9) encompasses all activators (e.g., Bicoid, Hunchback, and Caudal) and repressors (e.g., Giant, Kruppel, and Sloppy-paired), marking it as the most viable candidate for creating a minimal stripe enhancer.

Subsequently, we tandemized the third segment, assessing its expression pattern (see Figure 4.3A). We developed two versions with two or four tandem repeats. To ensure sufficient activation, we introduced three additional Zelda binding sites to the constructs. Given that Zelda is uniformly expressed throughout the embryo [94, 50], it's unlikely to alter the expression pattern.

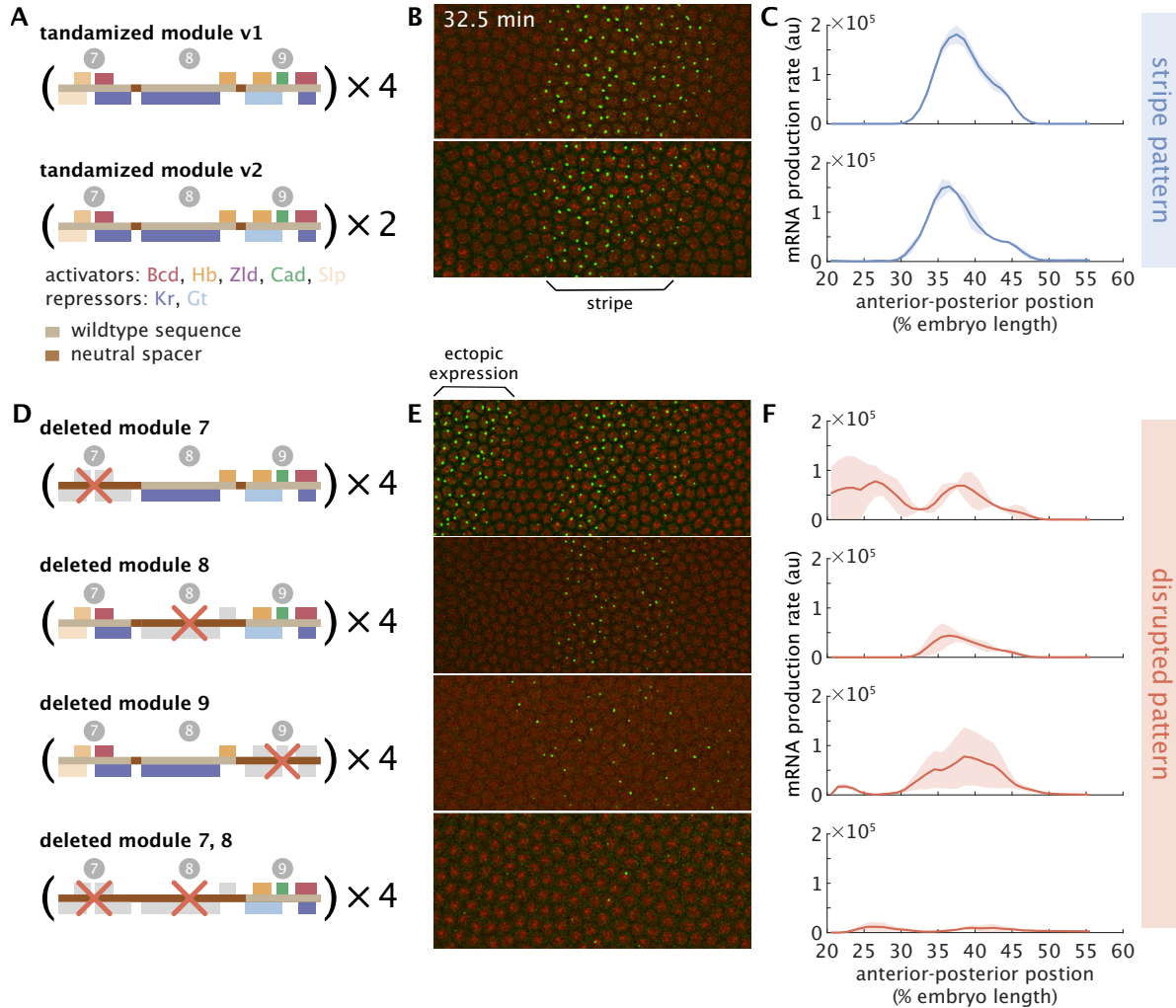


Figure 4.3: **Attempt to construct a further minimized stripe enhancer from a segment of *eve* stripe 2 enhancer.** (A-C) Designs of the tandemized version of a segment from the *eve* stripe 2 enhancer (A), snapshots of its expression (B) and quantified expression pattern (C). (D-F) Tandemized versions with submodules 7, 8, 9 or 7,8 mutated. Designs of the mutated versions (D), snapshots of its expression (E), and quantified expression pattern (F).

Remarkably, the tandemization of the third segment elicited a clear stripe pattern at the position similar to the endogenous stripe (Figure 4.3B and C) while broader in its expression range. This discovery implies the feasibility of crafting a further minimized enhancer that manifests in a stripe pattern using merely a third of the endogenous *eve* stripe 2 enhancer, which is 174 bp in length.

This result raises the subsequent question: can this enhancer be further minimized? To address this, we partitioned the segment into three submodules based on the reconstituted version, interspersed by neutral sequences. Then, we made versions in which each module was substituted with a neutral sequence (Figure 4.3D). We measured their respective expression patterns in live embryos. The results revealed that further sequence removal disrupts the stripe pattern significantly (Figure 4.3E and F).

Specifically, when module 7 was deleted, there was not only a decline in expression but also significant derepression at the embryo's anterior side. This suggests the elimination of a critical repressor binding site. This observation is consistent with the absence of the Sloppy-paired (Slp) binding site, believed to be instrumental in ensuring repression at the embryo's anterior end. Further, removing both modules 8 and 9 resulted in a marked decrease in expression, indicating a potential disruption in the activation mechanism.

Collectively, our findings underscore that each segment is pivotal in preserving the expression pattern's integrity. The disruption of even a single submodule introduces noticeable irregularities in the stripe pattern. This emphasizes the notion that the complexities of these modules and their interplay are finely tuned and balanced.

4.3 Discussion

Our work suggests that we now possess the capability to reconstitute the pattern of *eve* stripe 2 enhancer using solely binding sites. This breakthrough underscores our comprehensive grasp of its regulatory contributors and is a testament to the advancements in the field.

Historically, the endeavor to understand the organization of binding sites has been full of challenges. This is predominantly due to the inherent complexity of most developmental enhancers. While some studies have attempted to reveal them [162], their efforts have often been limited by several factors [90]. One of the primary constraints has been the intrinsic difficulty in reorganizing binding sites within endogenous contexts without inadvertently altering or eliminating other functional sites. Such a limitation, coupled with the absence of a comprehensive understanding of the full set of binding sites vital for a functional enhancer, has made the interpretation of results difficult. Furthermore, most of the experiments were conducted in the context of single cells, which amplifies the complexity, making it challenging to discern the broader implications in an actual developmental system.

In contrast, our research stands as a significant stride forward, representing the most profound reorganization effort undertaken on a complex developmental enhancer. The fact that we've been able to retain a majority of its regulatory logic, even after such an extensive reorganization, is truly remarkable.

In the realm of transcription factors, our findings are illuminating. The reconstructed enhancer we've presented mirrors the hallmark characteristics of the classical *eve* stripe 2 expression. This resemblance has profound implications, shedding light on the minimal set of factors essential for a functional enhancer. Our research pinpoints a unique combination of activators—Bicoid, Hunchback, Caudal, and Zelda—as foundational components for broad activation, which then sets the stage for further refinements by repressors. This synergy of transcriptional activators, in tandem with the pioneering factor Zelda, likely underpins the robust expression pattern sculpted across the embryo's anterior-posterior axis.

However, still, there are several problems inherent to the study. Although we have reconstituted the *eve* stripe 2 pattern, there are still differences in the amplitude between the reconstituted and endogenous enhancers. This is likely due to the omission of weaker activator binding sites while deciding on the sequence to keep. Previous research has underscored the significance of weak activator sites to the functionality of developmental enhancers [41, 59]. Another inherent challenge is the vast combinatorial landscape of possible permutations, making it virtually impossible to explore all conceivable module rearrangements, especially within the context of a developing embryo.

In future works, functional validation will be pivotal. Using CRISPR at the endogenous locus will determine whether it can rescue endogenous expression. Moreover, there's potential to extend the focus to other enhancers. Reconstitution with other intricate enhancers, particularly those rooted in evolutionary conservation, may lead to further refinement and minimization of regulatory sequences.

With the minimized version of the *eve* enhancer available, a more direct dissection and interpretation become feasible. This presents an opportunity for detailed sequence-level analysis. Future investigations could address how various factors converge to produce such robust expression. Understanding the significance of overlapping binding sites, the dynamics of binding site turnover, and the mechanisms behind synergistic activation will be paramount.

Incorporating a tandemized module based on the reconstituted enhancer will establish a minimal system that mirrors the intricacy of an endogenous enhancer. This system is optimally positioned to address fundamental questions about transcriptional regulation.

4.4 Methods and Materials

4.4.1 Cloning and Transgenesis

The newly constructed fly lines used in this study were generated by combining enhancer sequence with an array of 24 MS2 stem loops fused to the *D. melanogaster yellow* gene [18]. The enhancer segment was synthesized by Genscript and injected by BestGene Inc into *D. melanogaster* embryos with a Φ C31 insertion site in chromosome 3 (landing site VK00033).

4.4.2 Fly strains

To image the transcriptional dynamics of synthetic variants of *eve* stripe 2 enhancer, we crossed the male flies carrying transgenic construct with female flies, which were previously reported [63], that maternally express both MCP-GFP and His-RFP into the embryo.

4.4.3 Embryo preparation and data collection

The embryos were prepared following procedures described in [63, 18, 102]. Embryos were collected and mounted in halocarbon oil 27 between a semipermeable membrane (Lumox film, Starstedt, Germany) and a coverslip. Confocal imaging on a Zeiss LSM 780 microscope was performed using a Plan-Apochromat $40\times/1.4\text{NA}$ oil immersion objective. MCP-GFP and His-RFP were excited with laser wavelengths of 488 nm and 564 nm, respectively. Fluorescence was detected using the Zeiss QUASAR detection unit. Image resolution was 624×624 pixels, with a pixel size of 170 nm. Specimens were imaged for a minimum of 40 minutes into nuclear cleavage cycle 14.

Bibliography

- [1] Simon Alamos et al. “Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient”. In: *Cell Systems* 14 (Mar. 2023), pp. 1–17. ISSN: 24054720. DOI: [10.1016/j.cels.2022.12.008](https://doi.org/10.1016/j.cels.2022.12.008).
- [2] Bruce Alberts. *Molecular biology of the cell*. Garland science, 2017. ISBN: 1317563751.
- [3] Uri Alon. *An Introduction to Systems Biology*. Chapman and Hall/CRC, July 2006. ISBN: 9780429092794. DOI: [10.1201/9781420011432](https://doi.org/10.1201/9781420011432). URL: <https://www.taylorfrancis.com/books/9781420011432>.
- [4] Uri Alon. “Network motifs: Theory and experimental approaches”. In: *Nature Reviews Genetics* 8 (6 June 2007), pp. 450–461. DOI: [10.1038/nrg2102](https://doi.org/10.1038/nrg2102).
- [5] David Angeli, James E. Ferrell, and Eduardo D. Sontag. “Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems”. In: *Proceedings of the National Academy of Sciences* 101.7 (2004), pp. 1822–1827. DOI: [10.1073/pnas.0308265100](https://doi.org/10.1073/pnas.0308265100).
- [6] David N. Arnosti and Meghana M. Kulkarni. “Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?” In: *Journal of Cellular Biochemistry* 94.5 (2005), pp. 890–898. ISSN: 07302312. DOI: [10.1002/jcb.20352](https://doi.org/10.1002/jcb.20352).
- [7] David N. Arnosti et al. “The eve stripe 2 enhancer employs multiple modes of transcriptional synergy”. In: *Development* 122.1 (Jan. 1996), pp. 205–214. ISSN: 0950-1991. DOI: [10.1242/dev.122.1.205](https://doi.org/10.1242/dev.122.1.205). URL: <https://journals.biologists.com/dev/article/122/1/205/38869/The-eve-stripe-2-enhancer-employs-multiple-modes>.
- [8] Ahmet Ay and David N Arnosti. “Mathematical modeling of gene expression: A guide for the perplexed biologist.” In: *Critical Reviews in Biochemistry and Molecular Biology* 46.2 (Apr. 2011), pp. 137–51. ISSN: 1549-7798. DOI: [10.3109/10409238.2011.556597](https://doi.org/10.3109/10409238.2011.556597). URL: <http://www.ncbi.nlm.nih.gov/pubmed/21417596><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3086598>.
- [9] Michael Barresi J.F. and Scott Gilbert F. *Developmental Biology*. Thirteenth Edition. Oxford University Press, 2023.

- [10] David Bending et al. “A timer for analyzing temporally dynamic changes in transcription during differentiation *in vivo*”. In: *Journal of Cell Biology* 217.8 (2018), pp. 2931–2950. DOI: [10.1083/jcb.201711048](https://doi.org/10.1083/jcb.201711048).
- [11] Augusto Berrocal et al. “Kinetic sculpting of the seven stripes of the *Drosophila even-skipped* gene”. In: *eLife* 9 (2020), e61635. DOI: [10.1101/335901](https://doi.org/10.1101/335901).
- [12] 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\[prii\]](https://doi.org/S1097-2765(00)80143-4[prii]).
- [13] Lacramioara Bintu et al. “Transcriptional regulation by the numbers: models”. In: *Current Opinion in Genetics & Development* 15.2 (Apr. 2005), pp. 116–124. ISSN: 0959437X. DOI: [10.1016/j.gde.2005.02.007](https://doi.org/10.1016/j.gde.2005.02.007).
- [14] 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](https://doi.org/10.1126/science.aab2956).
- [15] Hamid Bolouri and Eric H. Davidson. “Modeling transcriptional regulatory networks”. In: *BioEssays* 24.12 (2002), pp. 1118–1129. DOI: [10.1002/bies.10189](https://doi.org/10.1002/bies.10189).
- [16] Jacques P. Bothma, Joe Magliocco, and Michael Levine. “The snail repressor inhibits release, not elongation, of paused Pol II in the drosophila embryo”. In: *Current Biology* 21.18 (2011), pp. 1571–1577. ISSN: 09609822. DOI: [10.1016/j.cub.2011.08.019](https://doi.org/10.1016/j.cub.2011.08.019). URL: <http://dx.doi.org/10.1016/j.cub.2011.08.019>.
- [17] 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 (July 2014), pp. 10598–10603. ISSN: 0027-8424. DOI: [10.1073/pnas.1410022111](https://doi.org/10.1073/pnas.1410022111). URL: <https://pnas.org/doi/full/10.1073/pnas.1410022111>.
- [18] Jacques P. Bothma et al. “LlamaTags: A versatile tool to image transcription factor dynamics in live embryos”. In: *Cell* 173.7 (2018), pp. 1810–1822. ISSN: 10974172. DOI: [10.1016/j.cell.2018.03.069](https://doi.org/10.1016/j.cell.2018.03.069). URL: <http://dx.doi.org/10.1016/j.cell.2018.03.069>.
- [19] Yassine X. Bouchoucha et al. “Dissection of a Krox20 positive feedback loop driving cell fate choices in hindbrain patterning”. In: *Molecular Systems Biology* 9 (2013), p. 690. ISSN: 17444292. DOI: [10.1038/msb.2013.46](https://doi.org/10.1038/msb.2013.46).
- [20] James E. Bradner, Denes Hnisz, and Richard A. Young. “Transcriptional addiction in cancer”. In: *Cell* 168.4 (2017), pp. 629–643. ISSN: 10974172. DOI: [10.1016/j.cell.2016.12.013](https://doi.org/10.1016/j.cell.2016.12.013). URL: <http://dx.doi.org/10.1016/j.cell.2016.12.013>.
- [21] Simon M G Braun et al. “Rapid and reversible epigenome editing by endogenous chromatin regulators”. In: *Nature Communications* 8.1 (2017), p. 560. ISSN: 2041-1723. DOI: [10.1038/s41467-017-00644-y](https://doi.org/10.1038/s41467-017-00644-y). URL: <https://doi.org/10.1038/s41467-017-00644-y>.

- [22] John H. Bushweller. “Targeting transcription factors in cancer — from undruggable to reality”. In: *Nature Reviews Cancer* 19.11 (2019), pp. 611–624. ISSN: 14741768. DOI: [10.1038/s41568-019-0196-7](https://doi.org/10.1038/s41568-019-0196-7). URL: <http://dx.doi.org/10.1038/s41568-019-0196-7>.
- [23] S Carroll. “Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis”. In: *Cell* 43.1 (1985), pp. 47–57. DOI: [10.1016/0092-8674\(85\)90011-x](https://doi.org/10.1016/0092-8674(85)90011-x).
- [24] Yick Bun Chan, Olga V. Alekseyenko, and Edward A. Kravitz. “Optogenetic control of gene expression in *Drosophila*”. In: *PLoS ONE* 10.9 (2015), e0138181. ISSN: 19326203. DOI: [10.1371/journal.pone.0138181](https://doi.org/10.1371/journal.pone.0138181).
- [25] 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](https://doi.org/10.1126/science.aar4199).
- [26] Shasha Chong et al. “Imaging dynamic and selective low-complexity domain interactions that control gene transcription.” In: *Science (New York, N.Y.)* 361.6400 (2018), eaar2555. ISSN: 1095-9203. DOI: [10.1126/science.aar2555](https://doi.org/10.1126/science.aar2555). URL: <http://www.ncbi.nlm.nih.gov/pubmed/29930090>.
- [27] Ke-Huan K. Chow et al. “Imaging cell lineage with a synthetic digital recording system”. In: *Science* 372.6538 (2021). ISSN: 0036-8075. DOI: [10.1126/science.abb3099](https://doi.org/10.1126/science.abb3099).
- [28] Erik Clark. “Dynamic patterning by the *Drosophila* pair-rule network reconciles long-germ and short-germ segmentation”. In: *PLoS biology* 15.9 (2017), e2002439. ISSN: 1544-9173.
- [29] Erik Clark and Michael Akam. “Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network”. In: *eLife* 5 (Aug. 2016), e18215. ISSN: 2050-084X. DOI: [10.7554/eLife.18215](https://doi.org/10.7554/eLife.18215). URL: <https://elifesciences.org/articles/18215>.
- [30] Erik Clark, Margherita Battistara, and Matthew A Benton. “A timer gene network is spatially regulated by the terminal system in the *Drosophila* embryo.” In: *eLife* 11 (Dec. 2022), e78902. ISSN: 2050-084X. DOI: [10.7554/eLife.78902](https://doi.org/10.7554/eLife.78902).
- [31] Erik Clark and Andrew D. Peel. “Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors”. In: *Development* 145.10 (May 2018). ISSN: 14779129. DOI: [10.1242/dev.155580](https://doi.org/10.1242/dev.155580).
- [32] Dorothy E. Clyde et al. “A self-organizing system of repressor gradients establishes segmental complexity in *Drosophila*”. In: *Nature* 426.6968 (2003), pp. 849–853. ISSN: 00280836. DOI: [10.1038/nature02189](https://doi.org/10.1038/nature02189).
- [33] Robert A. Coleman et al. “Imaging transcription: past, present, and future”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 80 (2016), pp. 1–8. ISSN: 19434456. DOI: [10.1101/sqb.2015.80.027201](https://doi.org/10.1101/sqb.2015.80.027201).

- [34] Adam M. Corrigan et al. “A continuum model of transcriptional bursting”. In: *eLife* 5.FEBRUARY2016 (2016), e13051. ISSN: 2050084X. DOI: [10.7554/eLife.13051](https://doi.org/10.7554/eLife.13051).
- [35] Francis Corson and Eric D Siggia. “Gene-free methodology for cell fate dynamics during development”. In: *eLife* 6 (Dec. 2017), e30743. ISSN: 2050-084X. DOI: [10.7554/eLife.30743](https://doi.org/10.7554/eLife.30743). URL: <https://elifesciences.org/articles/30743>.
- [36] Carlo Cosentino and Declan Bates. *Feedback Control in Systems Biology*. CRC Press, 2011. DOI: [10.1201/b11153](https://doi.org/10.1201/b11153).
- [37] James Cotterell and James Sharpe. “An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients”. In: *Molecular Systems Biology* 6 (Nov. 2010), p. 425. DOI: [10.1038/msb.2010.74](https://doi.org/10.1038/msb.2010.74).
- [38] Stephen T. Crews and Joseph C. Pearson. “Transcriptional autoregulation in development”. In: *Current Biology* 19.6 (Mar. 2009), R241–R246. ISSN: 09609822. DOI: [10.1016/j.cub.2009.01.015](https://doi.org/10.1016/j.cub.2009.01.015).
- [39] Justin Crocker and Albert Erives. “A closer look at the eve stripe 2 enhancers of *Drosophila* and *Themira*”. In: *PLoS Genetics* 4.11 (2008), pp. 1–7. ISSN: 15537390. DOI: [10.1371/journal.pgen.1000276](https://doi.org/10.1371/journal.pgen.1000276).
- [40] Justin Crocker, Garth R. Ilsley, and David L. Stern. “Quantitatively predictable control of *Drosophila* transcriptional enhancers *in vivo* with engineered transcription factors”. In: *Nature Genetics* 48.3 (2016), pp. 292–298. ISSN: 15461718. DOI: [10.1038/ng.3509](https://doi.org/10.1038/ng.3509). URL: <http://dx.doi.org/10.1038/ng.3509>.
- [41] Justin Crocker et al. “Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness”. In: *Cell* 160.0 (2015), p. 191. ISSN: 10974172. DOI: [10.1016/J.CELL.2014.11.041](https://doi.org/10.1016/J.CELL.2014.11.041). URL: [/pmc/articles/PMC4449256/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC4449256/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4449256/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC4449256/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4449256/).
- [42] Stephen Dalton. “Linking the Cell Cycle to Cell Fate Decisions”. In: *Trends in Cell Biology* 25.10 (2015), pp. 592–600. DOI: [10.1016/j.tcb.2015.07.007](https://doi.org/10.1016/j.tcb.2015.07.007).
- [43] Xavier Darzacq et al. “Imaging transcription in living cells”. In: *Annual Review of Biophysics* 38.1 (2009), pp. 173–196. ISSN: 1936122X. DOI: [10.1146/annurev.biophys.050708.133728](https://doi.org/10.1146/annurev.biophys.050708.133728).
- [44] George von Dassow et al. “The segment polarity network is a robust developmental module”. In: *Nature* 406.6792 (July 2000), pp. 188–192. ISSN: 0028-0836. DOI: [10.1038/35018085](https://doi.org/10.1038/35018085). URL: <http://www.nature.com/articles/35018085>.
- [45] C R Dearolf, J Topol, and C S Parker. “Transcriptional control of *Drosophila fushi tarazu* zebra stripe expression.” In: *Genes & development* 3.3 (1989), pp. 384–398. ISSN: 0890-9369.

- [46] Domitilla Del Vecchio et al. “What Have the Principles of Engineering Taught Us About Biological Systems?” In: *Cell Systems* 2 (1 Jan. 2016), pp. 5–7. DOI: [10.1016/j.cels.2016.01.008](https://doi.org/10.1016/j.cels.2016.01.008).
- [47] J. 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 (Electronic) 1553-734X (Linking). DOI: [10.1371/journal.pcbi.1005256](https://doi.org/10.1371/journal.pcbi.1005256).
- [48] Julien O. Dubuis et al. “Positional information, in bits”. In: *Proceedings of the National Academy of Sciences* 110.41 (Oct. 2013), pp. 16301–16308. ISSN: 0027-8424. DOI: [10.1073/pnas.1315642110](https://doi.org/10.1073/pnas.1315642110).
- [49] Mo R. Ebrahimkhani and Miki Ebisuya. “Synthetic developmental biology: build and control multicellular systems”. In: *Current Opinion in Chemical Biology* 52 (2019), pp. 9–15. ISSN: 18790402. DOI: [10.1016/j.cbpa.2019.04.006](https://doi.org/10.1016/j.cbpa.2019.04.006). URL: <https://doi.org/10.1016/j.cbpa.2019.04.006>.
- [50] Elizabeth Eck et al. “Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility”. In: *eLife* 9 (Oct. 2020), pp. 1–99. ISSN: 2050084X. DOI: [10.7554/eLife.56429](https://doi.org/10.7554/eLife.56429).
- [51] Bruce A. Edgar et al. “Repression and turnover pattern *fushi tarazu* RNA in the early *Drosophila* embryo”. In: *Cell* 47.5 (1986), pp. 747–754.
- [52] Javier Estrada et al. “SiteOut: An online tool to design binding site-free DNA sequences”. In: *PLoS ONE* 11.3 (2016), e0151740. ISSN: 19326203. DOI: [10.1371/journal.pone.0151740](https://doi.org/10.1371/journal.pone.0151740).
- [53] Walid D Fakhouri et al. “Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo”. In: *Molecular Systems Biology* 6.1 (2010), p. 341.
- [54] James E Ferrell. “Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability”. In: *Current Opinion in Cell Biology* 14.2 (2002), pp. 140–148. DOI: [10.1016/s0955-0674\(02\)00314-9](https://doi.org/10.1016/s0955-0674(02)00314-9).
- [55] Victoria E. Foe, Garrett M. Odell, and Bruce E. Edgar. “Mitosis and morphogenesis in the *Drosophila* embryo: Point and counterpoint”. In: *The Development of Drosophila melanogaster*. Ed. by Michael Bate and Alfonso Martinez Arias. Plainview, N.Y.: Cold Spring Harbor Laboratory Press, 1993. Chap. 3. ISBN: 0879694238.
- [56] M. Frasch et al. “Characterization and localization of the *even-skipped* protein of *Drosophila*.” In: *The EMBO Journal* 6.3 (1987), pp. 749–759. ISSN: 02614189. DOI: [10.1002/j.1460-2075.1987.tb04817.x](https://doi.org/10.1002/j.1460-2075.1987.tb04817.x).
- [57] 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* 126.11 (1999), pp. 2527–2538. ISSN: 09501991.

- [58] Takashi Fukaya, Bomyi Lim, and Michael Levine. “Enhancer control of transcriptional bursting”. In: *Cell* 166.2 (2016), pp. 358–368. ISSN: 10974172. DOI: [10.1016/j.cell.2016.05.025](https://doi.org/10.1016/j.cell.2016.05.025). URL: <http://dx.doi.org/10.1016/j.cell.2016.05.025>.
- [59] Timothy Fuqua et al. “Dense and pleiotropic regulatory information in a developmental enhancer”. In: *Nature* 587.7833 (2020), pp. 235–239. ISSN: 14764687. DOI: [10.1038/s41586-020-2816-5](https://doi.org/10.1038/s41586-020-2816-5). URL: <http://dx.doi.org/10.1038/s41586-020-2816-5>.
- [60] 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](https://doi.org/10.1126/science.aau0320).
- [61] Chikara Furusawa and Kunihiro Kaneko. “A dynamical-systems view of stem cell biology”. In: *Science* 338.6104 (Oct. 2012), pp. 215–217. ISSN: 0036-8075. DOI: [10.1126/science.1224311](https://doi.org/10.1126/science.1224311).
- [62] Hernan G. Garcia, Robert C. Brewster, and Rob Phillips. “Using synthetic biology to make cells tomorrow’s test tubes”. In: *Integrative Biology* 8.4 (2016), pp. 431–450. ISSN: 17579708. DOI: [10.1039/c6ib00006a](https://doi.org/10.1039/c6ib00006a). URL: <http://dx.doi.org/10.1039/C6IB00006A>.
- [63] 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](https://doi.org/10.1016/j.cub.2013.08.054). URL: <http://dx.doi.org/10.1016/j.cub.2013.08.054>.
- [64] Hernan G. Garcia et al. “Lighting up the central dogma for predictive developmental biology”. In: *Current Topics in Developmental Biology* 137 (Jan. 2020), pp. 1–35. ISSN: 00702153. DOI: [10.1016/bs.ctdb.2019.10.010](https://doi.org/10.1016/bs.ctdb.2019.10.010).
- [65] Charles J Geyer. “Practical Markov Chain Monte Carlo”. In: *Statistical Science* (1992), pp. 473–483.
- [66] Daniel T. Gillespie. “Exact stochastic simulation of coupled chemical reactions”. In: *Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361. ISSN: 00223654. DOI: [10.1021/j100540a008](https://doi.org/10.1021/j100540a008).
- [67] Daniel T. Gillespie. “Stochastic Simulation of Chemical Kinetics”. In: *Annual Review of Physical Chemistry* 58.1 (2007), pp. 35–55. DOI: [10.1146/annurev.physchem.58.032806.104637](https://doi.org/10.1146/annurev.physchem.58.032806.104637).
- [68] Alexander G Goglia and Jared E Toettcher. “A bright future: Optogenetics to dissect the spatiotemporal control of cell behavior”. In: *Current Opinion in Chemical Biology* 48 (2019), pp. 106–113. DOI: [10.1016/j.cbpa.2018.11.010](https://doi.org/10.1016/j.cbpa.2018.11.010).
- [69] Thomas G. W. Graham et al. “Modeling bistable cell-fate choices in the *Drosophila* eye: qualitative and quantitative perspectives”. In: *Development* 137.14 (2010), pp. 2265–2278. DOI: [10.1242/dev.044826](https://doi.org/10.1242/dev.044826).

- [70] Scott J. Gratz et al. “CRISPR-Cas9 genome editing in *Drosophila*”. In: *Current Protocols in Molecular Biology* 111.1 (2015), pp. 1–31. ISSN: 19343647. DOI: [10.1002/0471142727.mb3102s111](https://doi.org/10.1002/0471142727.mb3102s111).
- [71] Thomas Gregor et al. “Probing the limits to positional information”. In: *Cell* 130.1 (2007), pp. 153–164. ISSN: 0092-8674. DOI: <https://doi.org/10.1016/j.cell.2007.05.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0092867407006629>.
- [72] 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](https://doi.org/10.1016/j.cell.2007.05.025).
- [73] Ryan N Gutenkunst et al. “Universally Sloppy Parameter Sensitivities in Systems Biology Models”. In: *PLoS Computational Biology* 3.10 (2007). Ed. by Adam P Arkin, e189. DOI: [10.1371/journal.pcbi.0030189](https://doi.org/10.1371/journal.pcbi.0030189).
- [74] Ernst Hafen, Atsushi Kuroiwa, and Walter J Gehring. “Spatial distribution of transcripts from the segmentation gene *fushi tarazu* during *Drosophila* embryonic development”. In: *Cell* 37 (1984), pp. 833–841.
- [75] Saiyu Hang and J. Peter Gergen. “Different modes of enhancer-specific regulation by Runt and Even-skipped during *Drosophila* segmentation”. In: *Molecular Biology of the Cell* 28 (2017), pp. 681–691. ISSN: 1939-4586. DOI: [10.1091/MBE.E16-09-0630](https://doi.org/10.1091/MBE.E16-09-0630). URL: <https://pubmed.ncbi.nlm.nih.gov/28077616/>.
- [76] Emily E. Hare, Brant K. Peterson, and Michael B. Eisen. “A careful look at binding site reorganization in the even-skipped enhancers of *Drosophila* and sepsids”. In: *PLoS Genetics* 4.11 (2008), pp. 1–5. ISSN: 15537390. DOI: [10.1371/journal.pgen.1000268](https://doi.org/10.1371/journal.pgen.1000268).
- [77] Emily E. Hare et al. “Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation”. In: *PLoS Genetics* 4.6 (2008). ISSN: 15537390. DOI: [10.1371/journal.pgen.1000106](https://doi.org/10.1371/journal.pgen.1000106).
- [78] Leland H. Hartwell et al. “From molecular to modular cell biology”. In: *Nature* 402 (S6761 Dec. 1999), pp. C47–C52. DOI: [10.1038/35011540](https://doi.org/10.1038/35011540).
- [79] Nathaniel A. Hathaway et al. “Dynamics and memory of heterochromatin in living cells”. In: *Cell* 149.7 (2012), pp. 1447–1460. ISSN: 10974172. DOI: [10.1016/j.cell.2012.03.052](https://doi.org/10.1016/j.cell.2012.03.052). URL: <http://dx.doi.org/10.1016/j.cell.2012.03.052>.
- [80] Gerald Z. Hertz and Gary D. Stormo. “Identifying DNA and protein patterns with statistically significant alignments of multiple sequences”. In: *Bioinformatics* 15.7-8 (1999), pp. 563–577. ISSN: 13674803. DOI: [10.1093/bioinformatics/15.7.563](https://doi.org/10.1093/bioinformatics/15.7.563).
- [81] Hans Martin Herz, Deqing Hu, and Ali Shilatifard. *Enhancer malfunction in cancer*. Mar. 2014. DOI: [10.1016/j.molcel.2014.02.033](https://doi.org/10.1016/j.molcel.2014.02.033).
- [82] Yasushi Hiromi and Walter J. Gehring. “Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*”. In: *Cell* 50.6 (1987), pp. 963–974.

- [83] Yasushi Hiromi, Atsushi Kuroiwa, and Walter J. Gehring. “Control elements of the *Drosophila* segmentation gene *fushi tarazu*”. In: *Cell* 43.3 (Dec. 1985), pp. 603–613. ISSN: 00928674. DOI: [10.1016/0092-8674\(85\)90232-6](https://doi.org/10.1016/0092-8674(85)90232-6). URL: <https://linkinghub.elsevier.com/retrieve/pii/0092867485902326>.
- [84] Hiroyuki Hosokawa and Ellen V. Rothenberg. “How transcription factors drive choice of the T cell fate”. In: *Nature Reviews Immunology* 21.3 (Mar. 2021), pp. 162–176. ISSN: 14741741. DOI: [10.1038/s41577-020-00426-6](https://doi.org/10.1038/s41577-020-00426-6).
- [85] Anqi Huang et al. “Decoding temporal interpretation of the morphogen Bicoid in the early *Drosophila* embryo”. In: *eLife* 6 (2017), e26258. ISSN: 2050084X. DOI: [10.7554/eLife.26258](https://doi.org/10.7554/eLife.26258).
- [86] Emiliano Izquierdo, Theresa Quinkler, and Stefano De Renzis. “Guided morphogenesis through optogenetic activation of Rho signalling during early *Drosophila* embryogenesis”. In: *Nature Communications* 9.1 (2018), p. 2366. ISSN: 20411723. DOI: [10.1038/s41467-018-04754-z](https://doi.org/10.1038/s41467-018-04754-z). URL: <http://dx.doi.org/10.1038/s41467-018-04754-z>.
- [87] Johannes Jaeger and Nick Monk. “Bioattractors: Dynamical systems theory and the evolution of regulatory processes”. In: *The Journal of Physiology* 592.11 (June 2014), pp. 2267–2281. ISSN: 00223751. DOI: [10.1113/jphysiol.2014.272385](https://doi.org/10.1113/jphysiol.2014.272385).
- [88] Johannes Jaeger et al. “Dynamic control of positional information in the early *Drosophila* embryo”. In: *Nature* 430.6997 (2004), pp. 368–371. ISSN: 1476-4687. DOI: [10.1038/nature02678](https://doi.org/10.1038/nature02678). URL: <https://doi.org/10.1038/nature02678>.
- [89] Johannes Jaeger et al. “Dynamic control of positional information in the early *Drosophila* embryo”. In: *Nature* 430 (July 2004), pp. 368–371.
- [90] Granton A. Jindal and Emma K. Farley. “Enhancer grammar in development, evolution, and disease: dependencies and interplay”. In: *Developmental Cell* 56.5 (2021), pp. 575–587. ISSN: 18781551. DOI: [10.1016/j.devcel.2021.02.016](https://doi.org/10.1016/j.devcel.2021.02.016). URL: <https://doi.org/10.1016/j.devcel.2021.02.016>.
- [91] Heath E Johnson and Jared E Toettcher. “Illuminating developmental biology with cellular optogenetics”. In: *Current Opinion in Biotechnology* 52 (2018), pp. 42–48. DOI: [10.1016/j.copbio.2018.02.003](https://doi.org/10.1016/j.copbio.2018.02.003).
- [92] Heath E. Johnson et al. “The spatiotemporal limits of developmental Erk signaling”. In: *Developmental Cell* 40.2 (2017), pp. 185–192. ISSN: 18781551. DOI: [10.1016/j.devcel.2016.12.002](https://doi.org/10.1016/j.devcel.2016.12.002).
- [93] Heath E. Johnson et al. “Optogenetic rescue of a patterning mutant”. In: *Current Biology* 30.17 (2020), pp. 3414–3424. ISSN: 18790445. DOI: [10.1016/j.cub.2020.06.059](https://doi.org/10.1016/j.cub.2020.06.059). URL: <https://doi.org/10.1016/j.cub.2020.06.059>.
- [94] J. S. Kanodia et al. “Pattern formation by graded and uniform signals in the early *Drosophila* embryo”. In: *Biophys J* 102.3 (2012), pp. 427–33.

- [95] Shannon E. Keenan et al. “Rapid dynamics of signal-dependent transcriptional repression by Capicua”. In: *Developmental Cell* 52.6 (2020), pp. 794–801. ISSN: 18781551. DOI: [10.1016/j.devcel.2020.02.004](https://doi.org/10.1016/j.devcel.2020.02.004). URL: <https://doi.org/10.1016/j.devcel.2020.02.004>.
- [96] Yang Joon Kim et al. “Predictive modeling reveals that higher-order cooperativity drives transcriptional repression in a synthetic developmental enhancer”. In: *bioRxiv* (2021), p. 2021.07.28.454075. DOI: <https://doi.org/10.1101/2021.07.28.454075>. URL: <http://biorxiv.org/content/early/2021/07/29/2021.07.28.454075.abstract>.
- [97] Yang Joon Kim et al. “Predictive modeling reveals that higher-order cooperativity drives transcriptional repression in a synthetic developmental enhancer”. In: *eLife* 11 (Dec. 2022). ISSN: 2050084X. DOI: [10.7554/ELIFE.73395](https://doi.org/10.7554/ELIFE.73395).
- [98] Anna C. Kögler et al. “Extremely rapid and reversible optogenetic perturbation of nuclear proteins in living embryos”. In: *Developmental Cell* 56.16 (2021), pp. 2348–2363. ISSN: 15345807. DOI: [10.1016/j.devcel.2021.07.011](https://doi.org/10.1016/j.devcel.2021.07.011).
- [99] Hao Yuan Kueh et al. “Positive Feedback Between PU.1 and the Cell Cycle Controls Myeloid Differentiation”. In: *Science* 341.6146 (2013), pp. 670–673. DOI: [10.1126/science.1240831](https://doi.org/10.1126/science.1240831).
- [100] Karen Lai, Matthew J. Robertson, and David V. Schaffer. “The Sonic Hedgehog signaling system as a bistable genetic switch”. In: *Biophysical Journal* 86.5 (2004), pp. 2748–2757. DOI: [10.1016/s0006-3495\(04\)74328-3](https://doi.org/10.1016/s0006-3495(04)74328-3).
- [101] Nicholas C. Lammers et al. “A matter of time: using dynamics and theory to uncover mechanisms of transcriptional bursting”. In: *Current Opinion in Cell Biology* 67 (2020), pp. 147–157. ISSN: 18790410. DOI: [10.1016/j.ceb.2020.08.001](https://doi.org/10.1016/j.ceb.2020.08.001). URL: <https://doi.org/10.1016/j.ceb.2020.08.001>.
- [102] 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](https://doi.org/10.1073/pnas.1912500117).
- [103] 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/>.
- [104] P. Laslo et al. “Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system”. In: *Seminars in Immunology* 20.4 (2008), pp. 228–35. ISSN: 1044-5323 (Print) 1044-5323 (Linking). DOI: [10.1016/j.smim.2008.08.003](https://doi.org/10.1016/j.smim.2008.08.003).
- [105] Peter Laslo et al. “Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates”. In: *Cell* 126.4 (2006), pp. 755–766. DOI: [10.1016/j.cell.2006.06.052](https://doi.org/10.1016/j.cell.2006.06.052).

- [106] Michel Laurent and Nicolas Kellershohn. “Multistability: a major means of differentiation and evolution in biological systems”. In: *Trends in Biochemical Sciences* 24.11 (Nov. 1999), pp. 418–422. ISSN: 09680004. DOI: [10.1016/S0968-0004\(99\)01473-5](https://doi.org/10.1016/S0968-0004(99)01473-5).
- [107] 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](https://doi.org/10.1146/annurev-biophys-062215-010838).
- [108] Congxin Li et al. “Frequency modulation of transcriptional bursting enables sensitive and rapid gene regulation”. In: *Cell Systems* 6.4 (2018), pp. 409–423. ISSN: 24054720. DOI: [10.1016/j.cels.2018.01.012](https://doi.org/10.1016/j.cels.2018.01.012).
- [109] Li M. Li and David N. Arnosti. “Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes”. In: *Current Biology* 21.5 (2011), pp. 406–412. ISSN: 09609822. DOI: [10.1016/j.cub.2011.01.054](https://doi.org/10.1016/j.cub.2011.01.054).
- [110] Bomyi Lim et al. “Temporal dynamics of pair-rule stripes in living *Drosophila* embryos”. In: *PNAS* 115.33 (2018), pp. 8376–8381. ISSN: 10916490. DOI: [10.1073/pnas.1810430115](https://doi.org/10.1073/pnas.1810430115).
- [111] Shawn C. Little, Mikhail Tikhonov, and Thomas Gregor. “Precise developmental gene expression arises from globally stochastic transcriptional activity”. In: *Cell* 154.4 (2013), pp. 789–800. ISSN: 10974172. DOI: [10.1016/j.cell.2013.07.025](https://doi.org/10.1016/j.cell.2013.07.025). URL: <http://dx.doi.org/10.1016/j.cell.2013.07.025>.
- [112] Zhe Liu and Robert Tjian. “Visualizing transcription factor dynamics in living cells”. In: *Journal of Cell Biology* 217.4 (2018), pp. 1181–1191. ISSN: 15408140. DOI: [10.1083/jcb.201710038](https://doi.org/10.1083/jcb.201710038).
- [113] Francisco J. P. Lopes et al. “Spatial bistability generates *hunchback* expression sharpness in the *Drosophila* embryo”. In: *PLoS Computational Biology* 4.9 (2008). Ed. by Uwe Ohler, e1000184. DOI: [10.1371/journal.pcbi.1000184](https://doi.org/10.1371/journal.pcbi.1000184).
- [114] Tanguy Lucas et al. “Live imaging of Bicoid-dependent transcription in *Drosophila* embryos.” In: *Current Biology* 23.21 (2013), pp. 2135–2139. ISSN: 18790445. DOI: [10.1016/j.cub.2013.08.053](https://doi.org/10.1016/j.cub.2013.08.053). URL: <http://dx.doi.org/10.1016/j.cub.2013.08.053>.
- [115] Stewart MacArthur et al. “Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions”. In: *Genome Biology* 10.7 (2009), R80. ISSN: 1465-6906. DOI: [10.1186/gb-2009-10-7-r80](https://doi.org/10.1186/gb-2009-10-7-r80).
- [116] Manu et al. “Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation”. In: *PLoS Biology* 7.3 (Mar. 2009), e1000049.
- [117] Leslie J. Mateo et al. “Visualizing DNA folding and RNA in embryos at single-cell resolution”. In: *Nature* 568.7750 (2019), pp. 49–54. ISSN: 0028-0836. DOI: [10.1038/s41586-019-1035-4](https://doi.org/10.1038/s41586-019-1035-4).

- [118] Manos Mavrakakis, Richa Rikhy, and Jennifer Lippincott-Schwartz. “Plasma membrane polarity and compartmentalization are established before cellularization in the fly embryo”. In: *Developmental Cell* 16.1 (Jan. 2009), pp. 93–104. ISSN: 15345807. DOI: [10.1016/j.devcel.2008.11.003](https://doi.org/10.1016/j.devcel.2008.11.003).
- [119] Stephen L. McDaniel et al. “Continued activity of the pioneer factor Zelda is required to drive zygotic genome activation”. In: *Molecular Cell* 74.1 (2019), pp. 185–195. ISSN: 10974164. DOI: [10.1016/j.molcel.2019.01.014](https://doi.org/10.1016/j.molcel.2019.01.014). URL: <https://doi.org/10.1016/j.molcel.2019.01.014>.
- [120] Sarah McFann et al. “Temporal integration of inductive cues on the way to gastrulation”. In: *PNAS* 118.23 (2021), e2102691118. ISSN: 10916490. DOI: [10.1073/pnas.2102691118](https://doi.org/10.1073/pnas.2102691118).
- [121] Harold M. McNamara, Beatrice Ramm, and Jared E. Toettcher. “Synthetic developmental biology: New tools to deconstruct and rebuild developmental systems”. In: *Seminars in Cell & Developmental Biology* (2022). DOI: [10.1016/j.semcd.2022.04.013](https://doi.org/10.1016/j.semcd.2022.04.013).
- [122] Katrina C. McNeely and Noelle D. Dwyer. “Cytokinetic Abscission Regulation in Neural Stem Cells and Tissue Development”. In: *Current Stem Cell Reports* 7.4 (2021), pp. 161–173. DOI: [10.1007/s40778-021-00193-7](https://doi.org/10.1007/s40778-021-00193-7).
- [123] Nicholas Metropolis et al. “Equation of state calculations by fast computing machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. ISSN: 00219606. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114).
- [124] Kirstin Meyer et al. “Decoding of YAP levels and dynamics by pluripotency factors”. In: *bioRxiv* (2022), p. 2022.10.17.512504. DOI: [10.1101/2022.10.17.512504](https://doi.org/10.1101/2022.10.17.512504). URL: <https://www.biorxiv.org/content/10.1101/2022.10.17.512504v2><https://www.biorxiv.org/content/10.1101/2022.10.17.512504v2.abstract>.
- [125] Dominik Niopek et al. “Optogenetic control of nuclear protein export”. In: *Nature Communications* 7 (2016), pp. 1–9. ISSN: 20411723. DOI: [10.1038/ncomms10624](https://doi.org/10.1038/ncomms10624). URL: <http://dx.doi.org/10.1038/ncomms10624>.
- [126] Christiane Nüsslein-Volhard and Eric Wieschaus. “Mutations affecting segment number and polarity in *Drosophila*”. In: *Nature* 287.5785 (1980), pp. 795–801. ISSN: 1476-4687.
- [127] Daniel Panne. *The enhanceosome*. Apr. 2008. DOI: [10.1016/j.sbi.2007.12.002](https://doi.org/10.1016/j.sbi.2007.12.002).
- [128] Dimitrios K. Papadopoulos et al. “Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch”. In: *Development* 146.12 (June 2019). ISSN: 14779129. DOI: [10.1242/dev.168179](https://doi.org/10.1242/dev.168179).
- [129] Dmitri Papatsenko and Michael Levine. “The *Drosophila* gap gene network is composed of two parallel toggle switches”. In: *PLoS ONE* 6.7 (2011). Ed. by Vladimir N. Uversky, e21145. DOI: [10.1371/journal.pone.0021145](https://doi.org/10.1371/journal.pone.0021145).

- [130] A. Pare et al. “Visualization of individual Scr mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting”. In: *Current Biology* 19.23 (2009), pp. 2037–2042. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: [S0960-9822\(09\)01848-X\[pii\]10.1016/j.cub.2009.10.028](https://doi.org/10.1016/j.cub.2009.10.028).
- [131] Jeehae Park et al. “Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity”. In: *eLife* 8 (2019), pp. 1–25. ISSN: 2050084X. DOI: [10.7554/eLife.41266.001](https://doi.org/10.7554/eLife.41266.001).
- [132] David S Parker et al. “The cis-regulatory logic of Hedgehog gradient responses: key roles for Gli binding affinity, competition, and cooperativity”. In: *Science Signaling* 4.176 (2011), ra38. DOI: [10.1126/scisignal.2002077](https://doi.org/10.1126/scisignal.2002077). URL: <https://doi.org/10.1126/scisignal.2002077>.
- [133] Aleena L. Patel et al. “Capicua is a fast-acting transcriptional brake”. In: *Current Biology* 31.16 (2021), pp. 3639–3647. ISSN: 09609822. DOI: [10.1016/j.cub.2021.05.061](https://doi.org/10.1016/j.cub.2021.05.061).
- [134] Isabelle S Peter and Eric H Davidson. *Genomic control process: Development and evolution*. Academic Press, 2015. ISBN: 0124047467.
- [135] Mariela D. Petkova et al. “Optimal Decoding of Cellular Identities in a Genetic Network”. In: *Cell* 176.4 (2019), 844–855.e15. DOI: [10.1016/j.cell.2019.01.007](https://doi.org/10.1016/j.cell.2019.01.007).
- [136] A. Pezzotta and J. Briscoe. “Optimal control of gene regulatory networks for morphogen-driven tissue patterning”. In: *bioRxiv* (2022). DOI: [10.1101/2022.07.26.501519](https://doi.org/10.1101/2022.07.26.501519).
- [137] Rob Phillips et al. “Figure 1 theory meets Figure 2 experiments in the study of gene expression”. In: *Annual Review of Biophysics* 48 (2019), pp. 121–163. DOI: [10.1146/annurev-biophys-052118](https://doi.org/10.1146/annurev-biophys-052118). URL: <https://doi.org/10.1146/annurev-biophys-052118>.
- [138] Leslie Pick et al. “Analysis of the *ftz* upstream element: Germ layer-specific enhancers are independently autoregulated”. In: *Genes and Development* 4.7 (1990), pp. 1224–1239. ISSN: 08909369. DOI: [10.1101/gad.4.7.1224](https://doi.org/10.1101/gad.4.7.1224).
- [139] Armando Reimer et al. “Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient”. In: *bioRxiv* (2021), p. 2021.07.10.451524.
- [140] Christian P Robert and George Casella. “The Metropolis-Hastings algorithm”. In: *Monte Carlo statistical methods*. Springer, 2004, pp. 267–320.
- [141] Meritxell Sáez et al. “Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions”. In: *Cell Systems* 13.1 (Jan. 2022), pp. 12–28. ISSN: 24054712. DOI: [10.1016/j.cels.2021.08.013](https://doi.org/10.1016/j.cels.2021.08.013).

- [142] Keisuke Sako et al. “Optogenetic control of Nodal signaling reveals a temporal pattern of Nodal signaling regulating cell fate specification during gastrulation”. In: *Cell Reports* 16.3 (2016), pp. 866–877. ISSN: 22111247. DOI: [10.1016/j.celrep.2016.06.036](https://doi.org/10.1016/j.celrep.2016.06.036).
- [143] Hanae Sato et al. “Imaging of DNA and RNA in living eukaryotic cells to reveal spatio-temporal dynamics of gene expression”. In: *Annual Review of Biochemistry* 89 (2020), pp. 159–187.
- [144] Rupinder Sayal et al. “Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early *Drosophila* embryo”. In: *eLife* 5 (2016), e08445. ISSN: 2050-084X. DOI: [10.7554/ELIFE.08445](https://doi.org/10.7554/ELIFE.08445). URL: <https://pubmed.ncbi.nlm.nih.gov/27152947/>.
- [145] A. F. Schier and W. J. Gehring. “Analysis of a *fushi tarazu* autoregulatory element: Multiple sequence elements contribute to enhancer activity”. In: *EMBO Journal* 12.3 (1993), pp. 1111–1119. ISSN: 02614189. DOI: [10.1002/j.1460-2075.1993.tb05752.x](https://doi.org/10.1002/j.1460-2075.1993.tb05752.x).
- [146] Alexander F Schier and Walter J Gehring. “Direct homeodomain–DNA interaction in the autoregulation of the *fushi tarazu* gene”. In: *Nature* 356.6372 (1992), pp. 804–807. ISSN: 1476-4687.
- [147] Gavin Schlissel and Pulin Li. “Synthetic developmental biology: Understanding through reconstitution”. In: *Annual Review of Cell and Developmental Biology* 36 (2020), pp. 339–357. ISSN: 15308995. DOI: [10.1146/annurev-cellbio-020620-090650](https://doi.org/10.1146/annurev-cellbio-020620-090650).
- [148] Mark D Schroeder, Christina Greer, and Ulrike Gaul. “How to make stripes: Deciphering the transition from non-periodic to periodic patterns in *Drosophila* segmentation”. In: *Development* 138.14 (2011), pp. 3067–3078. ISSN: 1477-9129.
- [149] Sara Morais da Silva and Jean-Paul Vincent. “Oriented cell divisions in the extending germband of *Drosophila*”. In: *Development* 134.17 (2007), pp. 3049–3054. DOI: [10.1242/dev.004911](https://doi.org/10.1242/dev.004911).
- [150] Anand P. Singh et al. “Optogenetic control of the Bicoid morphogen reveals fast and slow modes of gap gene regulation”. In: *Cell Reports* 38.12 (2022), p. 110543. ISSN: 22111247. DOI: [10.1016/j.celrep.2022.110543](https://doi.org/10.1016/j.celrep.2022.110543).
- [151] 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.
- [152] Ruslan Soldatov et al. “Spatiotemporal structure of cell fate decisions in murine neural crest”. In: *Science* 364.6444 (2019). ISSN: 0036-8075. DOI: [10.1126/science.aas9536](https://doi.org/10.1126/science.aas9536).
- [153] Ruslan Soldatov et al. “Spatiotemporal structure of cell fate decisions in murine neural crest”. In: *Science* 364.6444 (2019). DOI: [10.1126/science.aas9536](https://doi.org/10.1126/science.aas9536).

- [154] Isabella V Soluri et al. “Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the *Drosophila* segmentation network”. In: *eLife* 9 (Apr. 2020). ISSN: 2050-084X. DOI: [10.7554/eLife.53916](https://doi.org/10.7554/eLife.53916). URL: <https://elifesciences.org/articles/53916>.
- [155] Aivar Sootla et al. “Shaping pulses to control bistable systems: Analysis, computation and counterexamples”. In: *Automatica* 63 (2016), pp. 254–264. DOI: [10.1016/j.automatica.2015.10.037](https://doi.org/10.1016/j.automatica.2015.10.037).
- [156] François Spitz and Eileen E.M. Furlong. “Transcription factors: from enhancer binding to developmental control”. In: *Nature Reviews Genetics* 13.9 (2012), pp. 613–626. ISSN: 14710056. DOI: [10.1038/nrg3207](https://doi.org/10.1038/nrg3207).
- [157] David Sprinzak et al. “Cis-interactions between Notch and Delta generate mutually exclusive signalling states”. In: *Nature* 465 (May 2010), pp. 86–90.
- [158] Shyam Srinivasan et al. “A BMP-FGF Morphogen Toggle Switch Drives the Ultrasensitive Expression of Multiple Genes in the Developing Forebrain”. In: *PLoS Computational Biology* 10.2 (2014). Ed. by Stanislav Shvartsman, e1003463. DOI: [10.1371/journal.pcbi.1003463](https://doi.org/10.1371/journal.pcbi.1003463).
- [159] Ralph Stadhouders, Guillaume J. Filion, and Thomas Graf. “Transcription factors and 3D genome conformation in cell-fate decisions”. In: 569.7756 (May 2019), pp. 345–354. ISSN: 14764687. DOI: [10.1038/s41586-019-1182-7](https://doi.org/10.1038/s41586-019-1182-7).
- [160] Steven H. Strogatz. *Nonlinear Dynamics and Chaos*. CRC Press, 2018. DOI: [10.1201/9780429492563](https://doi.org/10.1201/9780429492563).
- [161] Paolo Struffi and David N. Arnosti. “Functional interaction between the *Drosophila knirps* short range transcriptional repressor and RPD3 histone deacetylase”. In: *Journal of Biological Chemistry* 280.49 (2005), pp. 40757–40765. ISSN: 00219258. DOI: [10.1074/jbc.M506819200](https://doi.org/10.1074/jbc.M506819200). URL: <http://dx.doi.org/10.1074/jbc.M506819200>.
- [162] Christina I. Swanson, Nicole C. Evans, and Scott Barolo. “Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer”. In: *Developmental Cell* 18.3 (Mar. 2010), pp. 359–370. ISSN: 15345807. DOI: [10.1016/j.devcel.2009.12.026](https://doi.org/10.1016/j.devcel.2009.12.026). URL: <https://linkinghub.elsevier.com/retrieve/pii/S153458071000095X>.
- [163] Gašper Tkačik and Thomas Gregor. “The many bits of positional information”. In: *Development* 148.2 (2021), dev176065. DOI: [10.1242/dev.176065](https://doi.org/10.1242/dev.176065).
- [164] Pearlyn J Y Toh et al. “Optogenetic control of YAP cellular localisation and function”. In: *EMBO Reports* (2022), e54401. ISSN: 1469-221X.
- [165] Edward Tunnaciffe and Jonathan R. Chubb. “What Is a Transcriptional Burst?” In: *Trends in Genetics* 36.4 (2020), pp. 288–297. ISSN: 13624555. DOI: [10.1016/j.tig.2020.01.003](https://doi.org/10.1016/j.tig.2020.01.003). URL: <https://doi.org/10.1016/j.tig.2020.01.003>.

- [166] Ludovica Vanzan et al. “Mechanisms of histone modifications”. In: *Handbook of epigenetics*. Elsevier, 2023, pp. 27–54.
- [167] Berta Verd, Anton Crombach, and Johannes Jaeger. “Dynamic maternal gradients control timing and shift-rates for *Drosophila* gap gene expression”. In: *PLoS Computational Biology* 13 (2 2017), e1005285. DOI: [10.1371/journal.pcbi.1005285](https://doi.org/10.1371/journal.pcbi.1005285).
- [168] Alejandro F Villaverde et al. “BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology”. In: *BMC Systems Biology* 9.1 (2015). DOI: [10.1186/s12918-015-0144-4](https://doi.org/10.1186/s12918-015-0144-4).
- [169] Ben J. Vincent, Javier Estrada, and Angela H. DePace. “The appeasement of Doug: a synthetic approach to enhancer biology”. In: *Integrative Biology* 8.4 (2016), pp. 475–484. ISSN: 17579708. DOI: [10.1039/c5ib00321k](https://doi.org/10.1039/c5ib00321k). URL: <http://dx.doi.org/10.1039/C5IB00321K>.
- [170] 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). ISSN: 15537404. DOI: [10.1371/journal.pgen.1007644](https://doi.org/10.1371/journal.pgen.1007644).
- [171] C.H. Waddington. *The Strategy of the Genes*. Routledge, 1957. DOI: [10.4324/9781315765471](https://doi.org/10.4324/9781315765471).
- [172] B T Wakimoto, F R Turner, and T C Kaufman. “Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila melanogaster*”. In: *Developmental biology* 102.1 (1984), pp. 147–172. ISSN: 0012-1606.
- [173] Jin Wang et al. “Quantifying the Waddington landscape and biological paths for development and differentiation”. In: *Proceedings of the National Academy of Sciences* 108.20 (May 2011), pp. 8257–8262. ISSN: 0027-8424. DOI: [10.1073/pnas.1017017108](https://doi.org/10.1073/pnas.1017017108).
- [174] Xiao Wang et al. “Three-dimensional intact-tissue sequencing of single-cell transcriptional states”. In: *Science* 361.6400 (2018). ISSN: 0036-8075. DOI: [10.1126/science.aat5691](https://doi.org/10.1126/science.aat5691).
- [175] Xiaobo Wang et al. “Light-mediated activation reveals a key role for Rac in collective guidance of cell movement *in vivo*”. In: *Nature Cell Biology* 12.6 (2010), pp. 591–597. ISSN: 14657392. DOI: [10.1038/ncb2061](https://doi.org/10.1038/ncb2061).
- [176] Amy J Weiner, Matthew P Scott, and Thomas C Kaufman. “A molecular analysis of *fushi tarazu*, a gene in *Drosophila melanogaster* that encodes a product affecting embryonic segment number and cell fate”. In: *Cell* 37 (1984), pp. 843–851.
- [177] John C. Wheeler et al. “Distinct *in vivo* requirements for establishment versus maintenance of transcriptional repression”. In: *Nature Genetics* 32.1 (2002), pp. 206–210. ISSN: 10614036. DOI: [10.1038/ng942](https://doi.org/10.1038/ng942).
- [178] Franz-Georg Wieland et al. “On structural and practical identifiability”. In: *Current Opinion in Systems Biology* 25 (2021), pp. 60–69. DOI: [10.1016/j.coisb.2021.03.005](https://doi.org/10.1016/j.coisb.2021.03.005).

- [179] Wen Xiong and James E. Ferrell. “A positive-feedback-based bistable ‘memory module’ that governs a cell fate decision”. In: *Nature* 448.7157 (Aug. 2007), pp. 1076–1076. ISSN: 0028-0836. DOI: [10.1038/nature06125](https://doi.org/10.1038/nature06125). URL: <http://www.nature.com/articles/nature06125>.
- [180] Heng Xu et al. “Combining protein and mRNA quantification to decipher transcriptional regulation”. In: *Nature Methods* 12.8 (2015), pp. 739–742. ISSN: 15487105. DOI: [10.1038/nmeth.3446](https://doi.org/10.1038/nmeth.3446).
- [181] Venkata Rajesh Yella et al. “Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif”. In: *Nucleic Acids Research* 46.22 (Dec. 2018), pp. 11883–11897. ISSN: 13624962. DOI: [10.1093/nar/gky1057](https://doi.org/10.1093/nar/gky1057).
- [182] 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](https://doi.org/10.1074/jbc.M109.063784).
- [183] Magdalena Zernicka-Goetz, Samantha A. Morris, and Alexander W. Bruce. “Making a firm decision: Multifaceted regulation of cell fate in the early mouse embryo”. In: *Nature Reviews Genetics* 10.7 (2009), pp. 467–477. DOI: [10.1038/nrg2564](https://doi.org/10.1038/nrg2564).
- [184] Lei Zhang et al. “Noise drives sharpening of gene expression boundaries in the zebrafish hindbrain”. In: *Molecular Systems Biology* 8.1 (2012), p. 613. DOI: [10.1038/msb.2012.45](https://doi.org/10.1038/msb.2012.45).
- [185] Jiayi Zhao et al. “A bistable autoregulatory module in the developing embryo commits cells to binary expression fates”. In: *Current Biology* (July 2023). ISSN: 09609822. DOI: [10.1016/j.cub.2023.06.060](https://doi.org/10.1016/j.cub.2023.06.060). URL: <https://linkinghub.elsevier.com/retrieve/pii/S096098222300845X>.
- [186] Jiayi Zhao et al. “Optogenetic dissection of transcriptional repression in a multicellular organism”. In: *bioRxiv* (Jan. 2023), p. 2022.11.20.517211. DOI: [10.1101/2022.11.20.517211](https://doi.org/10.1101/2022.11.20.517211). URL: <http://biorxiv.org/content/early/2023/04/28/2022.11.20.517211.abstract>.
- [187] Benjamin Zoller, Shawn C. Little, and Thomas Gregor. “Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting”. In: *Cell* 175.3 (2018), pp. 835–847. ISSN: 10974172. DOI: [10.1016/j.cell.2018.09.056](https://doi.org/10.1016/j.cell.2018.09.056). URL: <https://doi.org/10.1016/j.cell.2018.09.056>.