UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Understanding patterning in the early fly embryo by transcriptional activators through the lens of theoretical models, live quantitative microscopy, and synthetic biology

Permalink

https://escholarship.org/uc/item/42m0n9zx

Author

Reimer, Armando

Publication Date

2021

Peer reviewed|Thesis/dissertation

Understanding patterning in the early fly embryo by transcriptional activators through the lens of theoretical models, live quantitative microscopy, and synthetic biology

by

Armando Reimer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Hernan Garcia, Chair
Professor Xavier Darzacq
Professor Michael Eisen
Professor Nipam Patel

Summer 2021

Understanding patterning in the early fly embryo by transcriptional activators through
the lens of theoretical models, live quantitative microscopy, and synthetic biology

Abstract

Understanding patterning in the early fly embryo by transcriptional activators through the lens of theoretical models, live quantitative microscopy, and synthetic biology

by

Armando Reimer

in Biophysics

University of California, Berkeley

Assistant Professor Hernan Garcia, Chair

Drosophila melanogaster has been a key model organism in the study of animal development for many decades. Since the revolutionary Nobel Prize winning work of Christiane Nüsslein-Volhard and Eric Wieschaus (Nusslein-Volhard and Wieschaus 1980) in the 1970's, we have understood that developmental patterning proceeds in stages wherein morphogen gradients provide the information to specify segments that give rise to distinct physiological compartments in the organism. However, knowing the genes involved and the qualitative features of developmental processes is only the first step on the road to truly understanding animal development. Detailed quantitative studies enabled by recent technological advances may allow us to go a step further and develop a predictive understanding of development and perhaps even engineer new developmental processes. As the Nobel Prize winning physicist Richard Feynman said, "What I cannot create, I do not understand." This simple idea was the driving force behind most of the work presented in this thesis. Standing on the shoulders of the giants in developmental biology, I have sought to create simple synthetic gene regulatory regions in the early fly embryo that are amenable to quantitative theoretical dissection. In addition, I have pursued experiments to challenge the theoretical assumptions underlying these models, namely how binding of activators leads to transcriptional activation and also developed new tools to enable these theoretical studies.

We have developed a minimal synthetic enhancer containing a single Dorsal binding site for the Dorsal activator as a tool for theoretical dissection of transcriptional regulation in the early *Drosophila melanogaster* embryo (Chapter 2). We found that a simple, theoretical model of transcriptional dynamics is sufficient to explain the fraction of loci in the embryo that engage in transcription and the timing of their transcription.

I also investigated transcription by Bicoid driven minimal synthetic enhancers (Chapter 3). In contrast to the Dorsal activator, a single Bicoid binding site was found to be

less capable of specifying positional information in the embryo, perhaps due to its extremely rapid on rates, which we measured during our binding studies on the lattice light sheet. Nonetheless, we made headway in developing a synthetic platform for studying the Bicoid activator, chiefly by creating transcription factor 'neutral' reporter sequences which do not bind other early embryonic transcription factors. Further study will be required to push the Bicoid synthetic platform to investigate similar questions possible using the Dorsal activator.

In addition to studying transcription driven by the Bicoid activator, we studied its binding kinetics in living embryos (Chapter 4) in an effort to shine light on assumptions about the binding events that precede transcriptional activation. We have pushed the envelope of the *in vivo* imaging of developing organisms by using the lattice light sheet microscope to probe the binding kinetics of single Bicoid molecules to the fly genome. We discovered that, contrary to expectations, Bicoid tends to bind in spatially localized clusters in the nucleus, and that these clusters facilitate binding to low-affinity clusters in the posterior embryo. Furthermore, the transcriptional pioneering factor Zelda is necessary for this clustering behavior and potentiates binding along the full length of the anteroposterior axis.

Finally, we developed tools for quantitative studies in the early fly embryo (Chapter 5). During the course of many other projects, including the ones outlined above, we have found that precise levels of maternally deposited proteins in the early embryo are often critical to quantitative studies, but there is a dearth of 'tunable' maternal promoter sequences available with well-characterized behaviors. We thus sought to develop such promoters, leading to the creation and characterization of promoters capable of driving the expression of maternally deposited proteins spanning several orders of magnitude in concentration. We additionally explored the effects of genomic location on the expression levels of our modular maternal promoters expression levels, which should expand their utility in the future. Lastly, we developed methods for calibrating measured fluorescence values from microscopy images into absolute numbers of molecules, an invaluable tool in quantitative studies.

To summarize, the experiments presented in this thesis lay the foundations for studying transcription in development with high spatiotemporal resolution and quantitative precision, contributing to a thoroughly predictive understanding of developmental processes.

Contents

Acknowledgments

Chapter 1

Introduction

1.1   Introduction

In this chapter, I will preview and summarize the later chapters of this thesis and provide some additional context for these experiments, starting with Chapter 2, an exposition of a manuscript currently in bioRxiv (Reimer et al. 2021) and under review at *Cell Systems* in which we develop a minimal synthetic enhancer containing only a single binding site for the Dorsal activator and subject it quantitative confocal microscopy and to theoretical dissection. In Chapter 3, I discuss an attempt to build similar synthetic enhancers driven by the Bicoid activator, which eventually evolved into the Dorsal synthetics project discussed in Chapter 2. In addition to studying transcription driven by Bicoid, I also investigated Bicoid's DNA binding kinetics, which play a pivotal role in determining Bicoid's effects on embryonic patterning. Chapter 4 includes a manuscript published in *Genes and Development* (Mir et al. 2017) in which we study the kinetic properties of an activator transcription factor binding in a live fly embryo using novel imaging modalities. Finally, Chapter 5 includes a summary of an ongoing project focused on creating better tools for the control of protein concentration in early fly embryo experiments.

1.2   Preview of Chapter 2: Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient

Although much progress has been made in recent decades towards decoding how gene regulatory architecture dictates developmental patterns in the early *Drosophila melanogaster* embryo, much work is left to be done in determining quantitatively how input transcription factors are read out by enhancers and promoters to give rise to precise levels of downstream output transcription factors. Inspired by the physicists' method of studying the most simple toy models to reach predictive understanding, we sought to investigate

transcription in development using a toy model of an enhancer, like a hydrogen atom for transcription in development Garcia2016.

Transcription factors in the early *Drosophila melanogaster* embryo form spatial gradients along the anterior-posterior and dorsoventral axes. The concentrations of these transcription factors are read out by enhancers, stretches of DNA with high densities of transcription factor binding sites, which then facilitate binding of general transcription factors to promoters that go on to produce as output some quantity of messenger RNA (mRNA) (Fig. 1.1 A, B). Spatially varying concentrations of this output mRNA are then translated into new, usually finer, gradients of protein transcription factors, leading to distinct cellular fates in different parts of the embryo.

Transcription factor binding to enhancer sequences can modulate different quantitative properties of output transcriptional dynamics such as the fraction of loci that transcribe at all (Fig. 1.1 C), the timing of transcriptional onset (Fig. 1.1 D), and the transcription rate (determined by the rate of RNAP loading onto the promoter; Fig. 1.1E). We sought to develop quantitative, predictive mathematical models of how transcription factor binding tunes each of these control 'knobs' that could be experimentally challenged. Similar models that could predict timing and rates of RNAP loading were investigated by Eck et al. 2020 and Dufourt et al. 2018. Thus, inspired by by Eck et al. 2020 and Dufourt et al. 2018, we developed a model of transcriptional control by the Dorsal activator in which readout of Dorsal concentration occurs in two stages: first, Dorsal binding 'primes' the promoter and prepares it for transcription (Fig. 1.1 F). This control over priming affects measurable dynamic quantities such as the transcriptional onset time during the nuclear cycle and the decision for each gene to engage in transcription. Second, after the promoter has been primed, Dorsal's equilibrium binding determines the rate of RNA Polymerase II (RNAP) loading on the promoter (Fig. 1.1 G). Together, these two stages and their respective measurable quantities encompass all of Dorsal's control over the transcriptional dynamics.

Figure 1.1: **Transcriptional regulatory strategies of enhancers in response to transcription factor concentration gradients; Integrated kinetic and thermodynamic model of simple activation by Dorsal. (A)** A *Drosophila* embryo with a transcription factor gradient along its dorsoventral axis. Caption continues on next page.
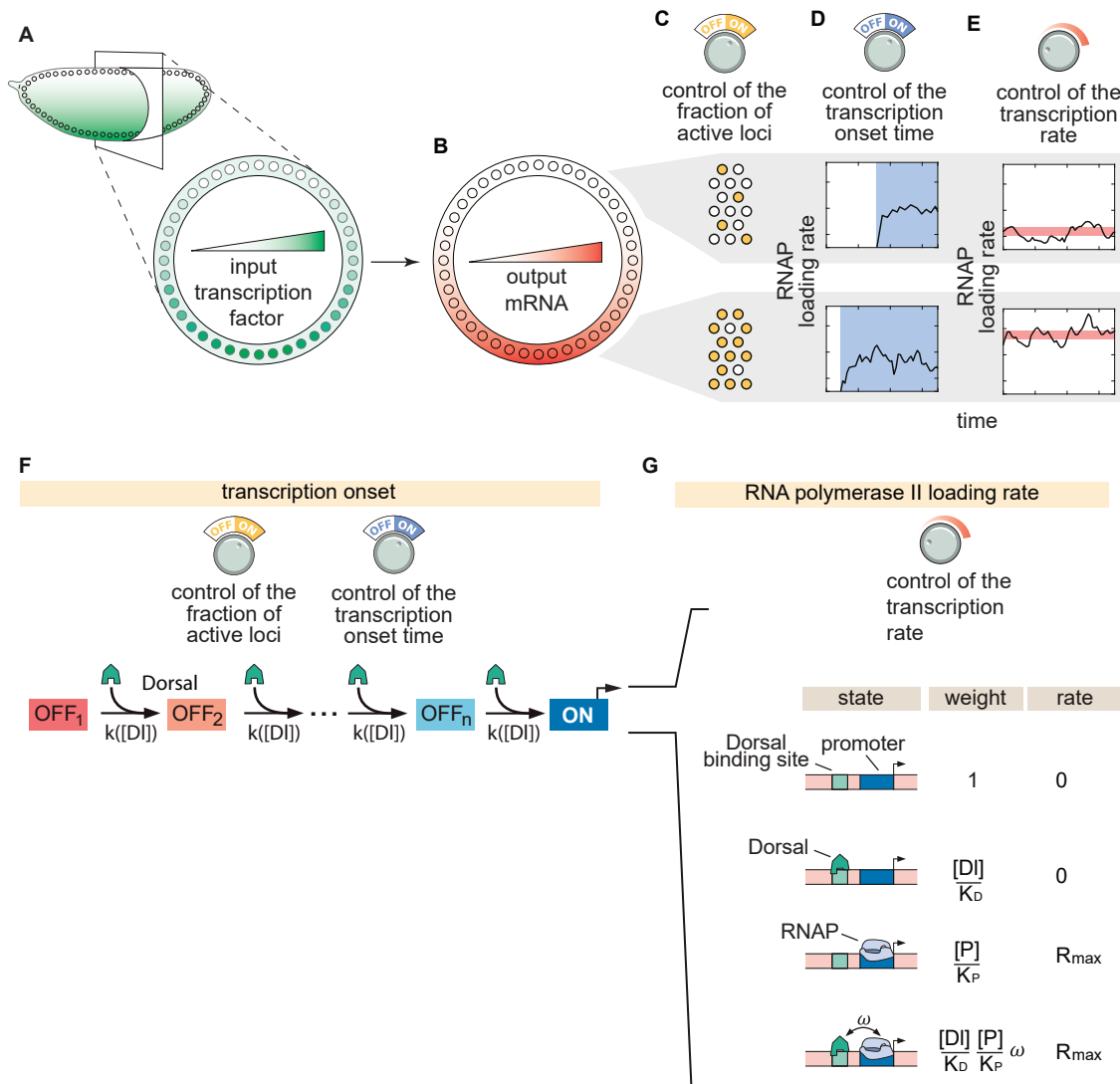
Figure 1.1: **Continued from previous page: Transcriptional regulatory strategies of enhancers in response to transcription factor concentration gradients; Integrated kinetic and thermodynamic model of simple activation by Dorsal. (B)** This input transcription factor dictates the emergence of output gene-expression patterns by controlling a combination of three enhancer regulatory 'knobs': **(C)** the probability of loci becoming transcriptionally active, **(D)** the transcriptional onset time, and **(E)** the mean transcription rate of active loci. (RNAP, RNA polymerase II). **(F)** The promoter undergoes kinetic transitions from transcriptionally inactive states (OFF_1 to OFF_n) to an active state (ON) with Dorsal accelerating the transition rate, k, by a factor proportional to the Dorsal occupancy at the promoter. **(G)** Thermodynamic states and weights for the simple activator model. The probability of finding RNAP bound to the promoter can be calculated from the statistical weights associated with all possible occupancy states of the enhancer-promoter system.

We then sought to challenge this model through experimentation. We created flies carrying a Dorsal-mVenus transgene Reeves et al. 2012a to report on Dorsal concentration (Fig. **??** A), along with the MCP-mCherry protein to report on transcriptional dynamics. We sought to measure transcriptional output from a reporter we constructed carrying MS2 loops driven by a minimal synthetic enhancer consisting of a single Dorsal binding site and an *even-skipped* core promoter (Fig. **??** B). Using a laser scanning confocal microscope, we imaged live, developing embryos during nuclear cycle 12, about an hour into development (Fig. 1.1 A, C). We successfully observed detectable signal from our single Dorsal binding site enhancers, proving viability of this single binding site enhancer as a toy model of enhancers in the embryo (Fig. 1.1 C, D).

As shown in Figure 1.1 D, we observed a sizable fraction of loci that never transcribe at all during nuclear cycle 12 (rows with all dark blue elements). It was important to determine whether this fraction of inactive loci were truly a different population or were simply active loci beneath our detection threshold (i.e., is there a binary switch at the level of transcriptional control that decides whether a promoter becomes active or not?). To answer this question, we employed the use of the ParB-ParS system *in vivo* DNA labelling system (Fig. **??** E), which enabled us to fluorescently label all reporter loci, regardless of whether they exhibited active transcription or not. We reasoned that if we observed transcription from both active and inactive loci, we would be able to differentiate two distinctively different populations of loci with different mean expression levels rather than just a continuum of transcriptional output (Fig. **??** F, G). After performing these measurements, we confirmed that signal at loci with detected spots was significantly different from loci without detected spots, thus demonstrating that there truly is a qualitatively separate population of nuclei which never exhibit transcription from our reporter(Fig. **??** H, I, J).

A

dorsal

10 µm

lateral

ventral

cross-section

dorsal

ventral

[Dl]

B

RNAP

MS2 stem loop

MCP-mCherry

14bp

Dorsal binding site

even-skipped minimal promoter

minimal enhancer

24xMS2 loops

lacZ

tubulin 3'UTR

C

+ Dorsal protein strong binding site

10 µm

D

spot fluorescence (a.u.)

nucleus

time into nuclear cycle 12 (min)

active loci (56%)

inactive loci (44%)

E

eGFP

ParB

MCP-mCherry

MS2 stem loop

parS site

spacer

enhancer + promoter

24xMS2

F

reporter locus position

ParB (DNA)

MS2 (RNA)

mCherry measurement location

'undetected' spots

detected spots

G

ParB (DNA)

MS2 (RNA)

5 µm

detected

undetected

H

detected

undetected

mCherry spot Intensity (a.u.)

time after nc12 mitosis (min)

I

detected spots

undetected spots

dorsal null

J

dorsal null

undetected spots

detected spots

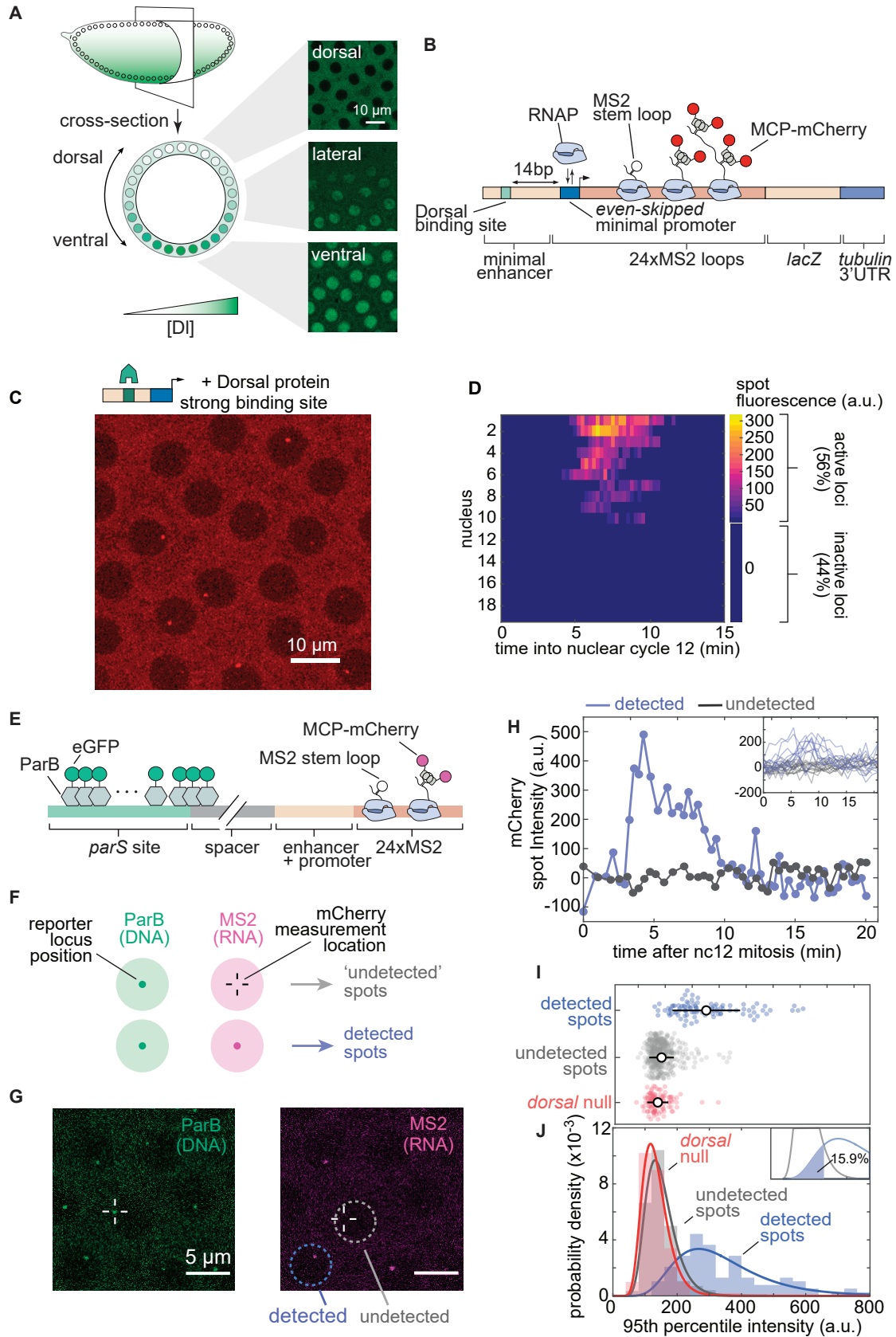probability density (×10⁻³)

95th percentile intensity (a.u.)

15.9%

Figure 1.3: **Simultaneously measuring transcription factor protein input and transcriptional output. (A)** Schematic of the Dorsal protein gradient in early *Drosophila* embryos. Dorsal protein accumulates in ventral nuclei and is progressively excluded from more dorsal nuclei. Example snapshots show Dorsal-mVenus in various positions along the dorsoventral axis. **(B)** Schematic of minimal synthetic enhancer system containing a single binding site for Dorsal that drives transcription of a reporter tagged with MS2 loops, which are visualized through the binding of MCP-mCherry. The Dorsal binding site is placed 14 bp upstream of the *even-skipped* minimal promoter. **(C)** Snapshots from embryos containing an optimal binding-site reporter in the presence of Dorsal protein. **(D)** Fluorescence of all transcription spots in individual nuclei in the field of view of one embryo as a function of time. If a transcription spot was detected within a nucleus at any point during the interphase of nuclear cycle 12, then the locus was considered active; otherwise, the locus was classified as inactive. **(E)** Schematic of ParB-eGFP construct. ParB-eGFP molecules bind and polymerize out from *parS* sequences, which are placed $\sim$400 bp upstream of the enhancer. The enhancer and promoter together drive transcription of MS2 loops that subsequently bind MCP-mCherry. **(F)** Schematic of the experiment. Loci are located by detecting a signal in the ParB-eGFP channel; these locations were used to fit a 2D Gaussian to the same area in the MS2-mCherry channel to estimate fluorescence intensity regardless of whether an MS2-mCherry signal was detected. **(G)** Example images of ParB-eGFP (left) and MCP-mCherry (right) channels. Detected and undetected transcriptionally active loci solely based on the MCP-mCherry signal alone are shown. **(H)** Example time traces of MCP-mCherry fluorescence over time at the ParB-eGFP loci in nuclei with (blue) and without (grey) detected MS2-mCherry spots of the DBS_6.23 enhancer showing clear qualitative differences between the two populations. Inset, all detected and undetected fluorescence traces obtained in the same embryo. Negative intensity values are due to spot intensities very close to the background fluorescence. **(I)** Swarm plots of 95th percentile MCP-mCherry fluorescence at loci with detected (blue; N = 125) and undetected MS2-mCherry transcription (gray; N = 425) driven by the DBS_6.23 enhancer in wild-type Dorsal embryos. Red (N = 96), maximum fluorescence of all loci in Dorsal null embryos, defined as the 95th percentile of intensity over time (black circles, mean; bars, standard deviation). Detected spots are significantly different from both null (ANOVA, $p < 0.01$) and undetected spots (ANOVA, $p < 0.01$) **(J)** Histograms of the data shown in (E). Solid lines correspond to log-normal fits performed for ease of visualization. Inset, undetected and detected distribution fits and the area used to estimate the false-negative detection rate of 15.9%.).

Having concluded that the fraction of active loci is not an experimental artifact and is a knob under the control of the Dorsal activator, we turned to challenging the theoretical models posed above using our experimental apparatus. To accomplish this, we developed a series of single Dorsal binding site minimal synthetic enhancers with different

binding affinities. Modulating the binding affinity was critical to testing our theoretical model since both stages of our integrated model depend heavily on Dorsal's binding affinity for its binding motif (Fig. 1.1 F, I). To that end, we developed a series of seven binding sites mutated away from the consensus Dorsal binding site (GGGAAAACCC), and used the Patser bioinformatics algorithm (Stormo, Schneider, and Gold 1986) to predict their relative binding affinities (Fig. 1.4 A, B). As can be seen in (Figure 1.4 C), the transcriptional response to these different binding affinities behaved qualitatively as expected, with the highest affinity binding sites leading to greater total transcriptional output.

Next, we directly compared the fraction of active loci, transcriptional onset times and RNAP loading rates (here, measured as maximum fluorescence over time) to the values predicted by our integrated model (Fig. 1.4 D, E). We found that the 'kinetic barrier' model's prediction of the fraction of active loci and the transcription onset times was fairly quantitatively consistent with our measurements (Fig. 1.4 D). However, with the throughput available with our experimental technique, along with the very high biological noise and the low dynamic range of the transcriptional response to our reporter, we were unable to adequately test whether the equilibrium model of transcription could predict the RNAP loading rates from our minimal synthetic system. Challenging this equilibrium model remains a high priority for future experiments.
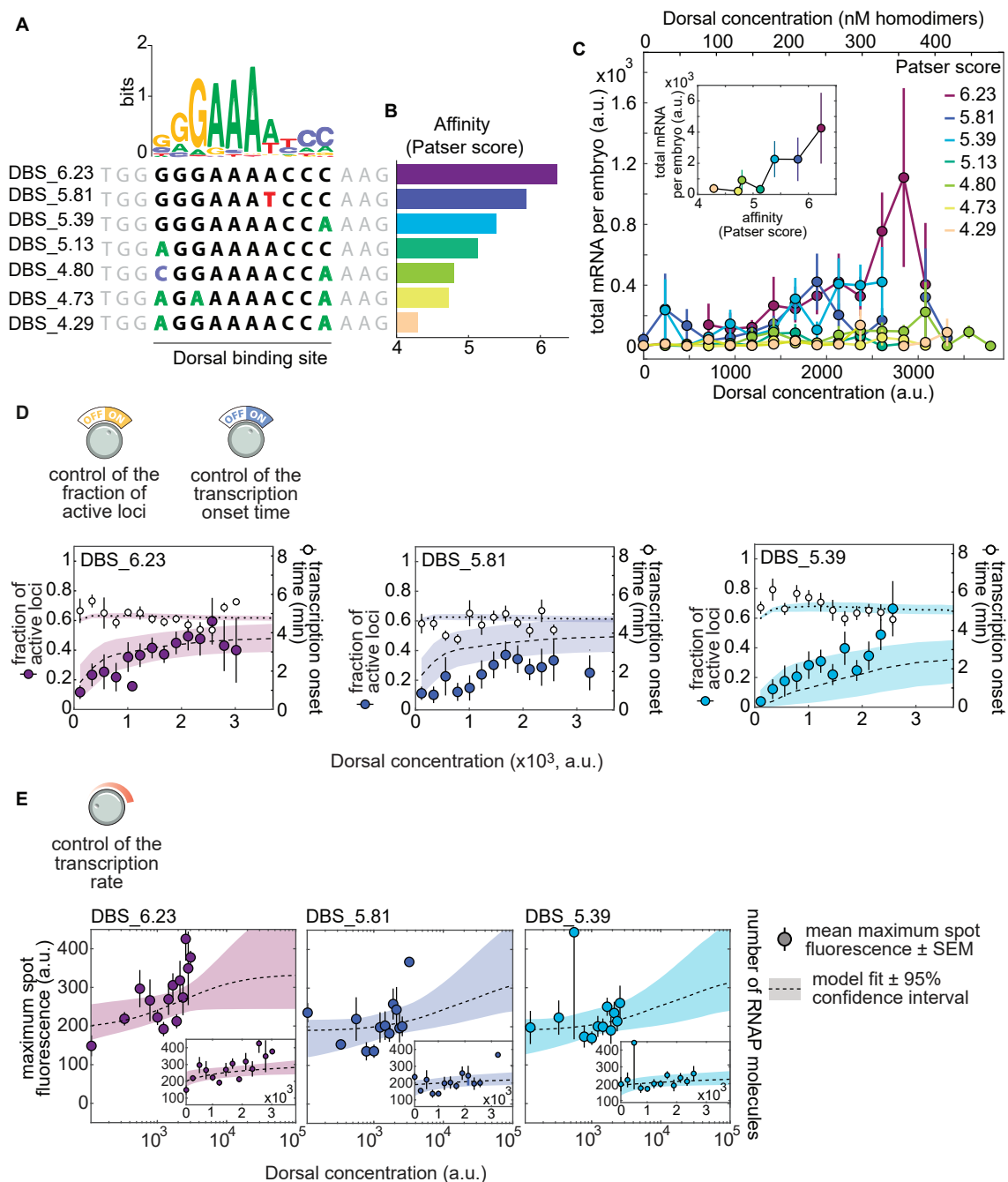
Figure 1.4: **A multi-step kinetic barrier model predicts the Dorsal-dependent fraction of active loci with constant mean transcriptional onset times; Testing RNAP loading rate predictions of the thermodynamic model. (A)** Top: Dorsal positional weight matrix logo from Ivan, Halfon, and Sinha 2008. Bottom: Sequence of the Dorsal binding sites engineered into minimal synthetic enhancers. Bold letters, 10 bp Dorsal motif. Black letters, consensus bases; colored letters, mutated bases; gray letters, sequence context. Caption continues on next page.

Figure 1.3: **Continued from previous page: Simultaneously measuring transcription factor protein input and transcriptional output. (B)** Relative affinities of Dorsal binding sites estimated from the Patser algorithm using the Dorsal position weight matrix. **(C)** Overall transcriptional activity driven by the enhancers containing the binding sites in (A) measured as the total produced mRNA (fluorescence integrated over nuclear cycle 12) as a function of Dorsal concentration. Inset, mean total mRNA produced per embryo integrated across all Dorsal concentrations. Error bars, SEM over N > 3 embryos containing 3 or more nuclei belonging to that fluorescence bin. The top x-axis shows the estimated nuclear Dorsal concentration according to the calibration described in Figure S8. **(D)** Data and model fits for the fraction of active loci (left y-axis) and mean transcription onset time (right y-axis) for each enhancer. Empty black circles, experimentally observed mean transcription onset time; filled circles, experimentally observed mean fraction of active loci. Fitted curves are represented as dashed lines (fraction of active loci) and dotted lines (mean onset times), corresponding to predictions using median parameter values from the joint posterior distribution. Shaded areas, 95% credible interval (see Table S1 for inferred parameter values). Error bars, SEM over N > 3 embryos containing 3 or more nuclei belonging to that fluorescence bin. **(E)** Mean maximum spot fluorescence as a function of Dorsal concentration for minimal synthetic enhancers with different affinities for Dorsal (filled circles). The right y-axis denotes the calibrated number of actively transcribing RNAP molecules (for details of calibration, see Section S1.3 and Fig. S2). Dashed curves correspond to a simultaneous Markov Chain Monte Carlo curve fit to all data using Equation 2. Fits share all parameters except K_D. Shaded areas, 95% prediction intervals. Insets, same data and fits plotted on a linear scale with axis ranges zoomed in on the data. See Table S2 for inferred parameter values. Error bars, SEM across N > 3 embryos containing 3 or more nuclei in a given fluorescence bin.

1.3   Preview of Chapter 3: Development of a Bicoid-driven minimal synthetic enhancer

Before investigating the Dorsal synthetic enhancers featured in Chapter 2, I attempted to develop a minimal synthetic enhancer for the Bicoid activator. Ultimately, perhaps due to Bicoid being an intrinsically weak activator relative to other early embryonic factors such as Dorsal, it was challenging to create synthetic enhancers with a single binding site for Bicoid that could drive detectable levels of gene expression above background signal with our imaging techniques. Much of the early results from this project were important for the development of the Dorsal synthetic enhancers, but many Bicoid-specific experiments have been left unpublished. Therefore, I will discuss here the difficulties associated with studying a Bicoid-dependent minimal synthetic enhancer, what I ex-

plored during the early years of my thesis project, and what may be done in the future to further develop this as a tool for studying early embryonic transcription driven by Bicoid.

One of the challenges working with a minimal synthetic enhancer is that because it carries a single activator binding site it necessarily has poor signal and dynamic range relative to other enhancers. This makes it technically challenging to observe the signal with a reasonable signal-to-noise ratio, and also makes it hard to generate enough useful data for statistics. In Figure 1.4 A, I show what an ideal experimental outcome would look like: gene expression is anterior-posterior dependent, requires the Bicoid protein, and requires the Bicoid binding site in the minimal synthetic enhancer.
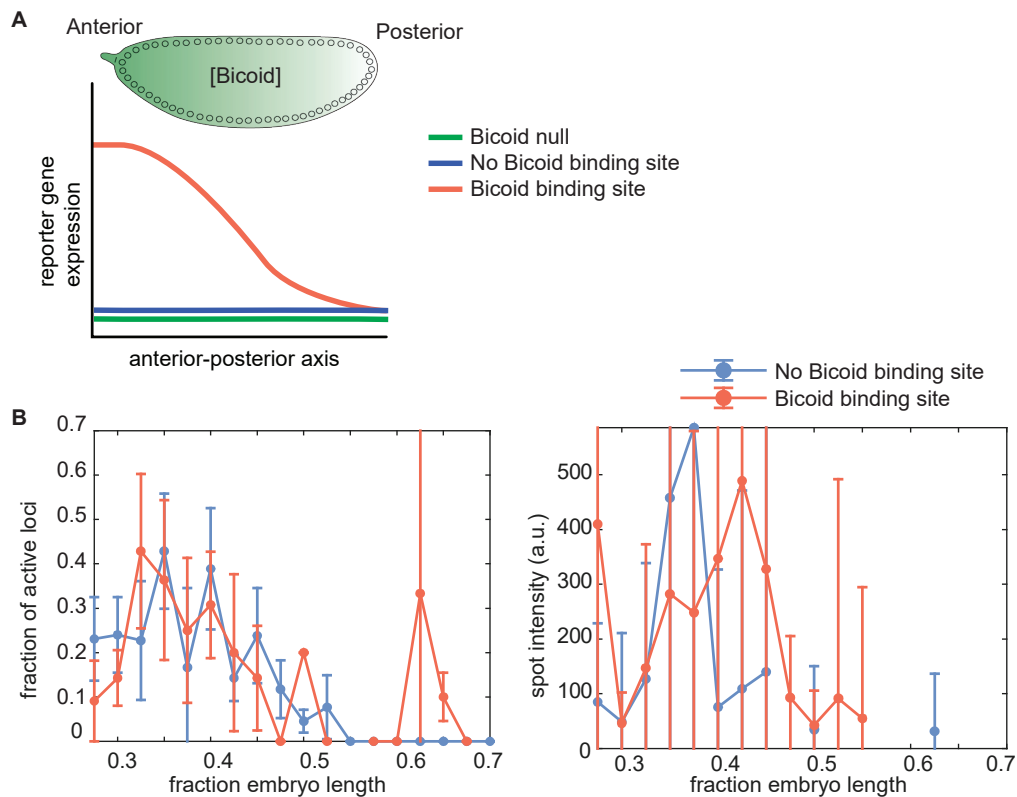
Figure 1.4: **Preview of Chapter 3: Development of a Bicoid-driven minimal synthetic enhancer (A)** Ideal experimental outcome where gene expression is anterior-posterior dependent, requires Bicoid, and requires the Bicoid binding site. **(B)** Experimental outcomes from live embryo confocal microscopy experiment imaging the *A3-evePr-MS2v7x24-lacZ* (A3: a strong Bicoid binding site, evePr: *even-skipped* core promoter) reporter and *evePr-MS2v7x24-lacZ* control. Although gene expression is anterior-poster dependent, it seems to be independent of the intended Bicoid binding site in the minimal synthetic enhancer.

In Figure 1.4 C, I have shown the actual experimental outcomes from a confocal microscopy experiment using the MS2-MCP transcriptional reporter system. The two constructs shown are the *A3-evePr-MS2v7x24-lacZ* (A3: a strong Bicoid binding site, evePr: *even-skipped* core promoter) reporter and a corresponding negative control, *evePr-MS2v7x24-lacZ*, which lacks a Bicoid binding site. Although gene expression here is anterior-poster dependent, it is clearly independent of the Bicoid binding site located in the minimal synthetic enhancer.

We reasoned that in order to make expression from the minimal synthetic's Bicoid binding site detectable, we should try to remove background signal from the reporter that contributed to the negative control's expression. To accomplish this, we used a semi-rational approach where we screened various parts of the construct for transcription factor binding sites which could bind to transcription factors known to be present in the early embryo. After identifying sequences within the reporter rich for extraneous binding sites, we swapped these sequences out for similar sequences with those binding sites removed. Although we identified several constructs with minimal background expression (Fig. 1.5; elaborated on in Chapter 3), many of these had such minimal signal left over that it was difficult to collect statistics for. This may simply be due to the weakness of the Bicoid activator when acting without other Bicoid molecules or without Zelda proteins to facilitate binding or chromatin opening. In Figure 11, I have also highlighted some of the components of the reporter that were found to be most helpful in reducing background signal and may be a good starting point for future explorations. The rationale for these choices is discussed further in Chapter 3.

Modelling and visualizing the transcriptional output from a Bicoid driven reporter naturally leads to questions about how Bicoid binding actually leads to transcriptional output. However, our experiments are blind to Bicoid binding events, and the fact that Bicoid is binding to our enhancer is only an inference we make from seeing transcriptional output. Thus, we sought to develop ways to visualize Bicoid binding in the early embryo as discussed in the following section.
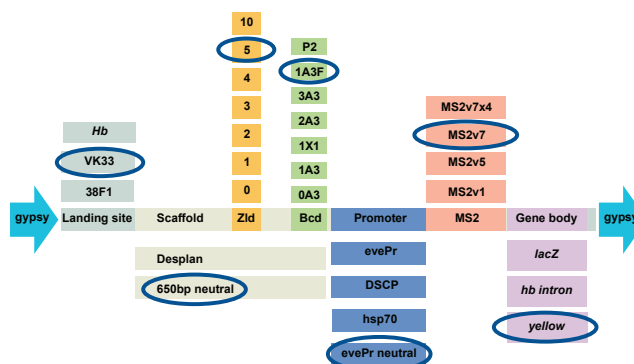
Figure 1.5: **Future directions for Bicoid minimal synthetic enhancer construction** Construction options for a Bicoid-dependent minimal synthetic enhancer. *gypsy*: the *gypsy* insulator sequence. VK33: Plasmid construct in pbϕ backbone and inserted at the VK33 landing site. 38F1: Plasmid construct in pIB backbone and inserted at 38F1 landing site. *Hb*: Construct placed near the endogenous hunchback gene in the genome (not tested in this study). Desplan: Original sequence upstream of the promoter into which the Bicoid A3 site(s) was inserted. 650 bp neutral: 650 bp sequence constructed with SiteOut into which Bicoid binding sites can be inserted. Zld: Number of Zelda sites additionally added to the enhancer. 0A3: No Bicoid binding site. 1A3: single A3 binding site originally taken from the hunchback P2 promoter. 2A3, 3A3: two and three A3 binding sites in the enhancer, respectively. 1A3F: single A3 Bicoid binding site with additional flanking base pairs. P2: the full or modified *hunchback* P2 promoter. evePr: *even-skipped* core promoter. DSCP: *Drosophila* Super Core Promoter. *hsp70*: *hsp70* core promoter. evePr neutral: *even-skipped* core promoter with extraneous binding sites removed via Site-Out. MS2v1, v5, v7 are defined above in previous sections. MS2v7x4: MS2v7 sequence repeated four times for a total of 96 loops. *hb intron*: endogenous *hb* intron sequence.

1.4 Preview of Chapter 4: Dense Bicoid hubs accentuate binding along the morphogen gradient

One of the holy grails of studying transcription in development is to simultaneously visualize both output transcription in a live embryo along with the binding events of single activator molecules that give rise to that transcription. Although we have some ways to go as a field before these simultaneous measurements become practical, in this chapter, I will discuss the first visualizations of single binding events of transcription factors in a developing embryo (Mir et al. 2017).

In order to achieve our goal of visualizing transcription factor binding events in a living embryo, for which the laser scanning confocal microscope would be inadequate due to limitations in speed and spatial resolution, we turned to using a state of the art lattice light sheet microscope (Chen et al. 2014a) built in the Tjian-Darzacq lab at UC Berkeley.

Using this microscope enabled us to achieve unprecedented single molecule resolution in a live embryo. We used the fluorescently labelled Bicoid-eGFP fusion developed by Gregor et al. 2007 and imaged embryos under the lattice light sheet microscope at very high laser powers and temporal resolution. High laser power enabled us to collect sufficient signal from single Bicoid molecules in a short time span so that they could be detectable above background noise. Meanwhile, high temporal resolution enabled us to measure residence times of single Bicoid molecules bound to the genome. Frame rate was an important experimental parameter because very fast-moving molecules diffusing in 3D would be blurred out at our frame rate, but molecules temporarily bound to DNA could be visualized as stationary puncta (Fig. 1.6 A, B, C). This frame rate, coupled with high laser intensity would be unachievable on a laser-scanning confocal microscope due to fundamental limitations arising from raster scanning and single-pixel detector technology (e.g. photomultiplier tubes (PMTs) or Leica's hybrid sequential avalanche photodiode-PMT detectors (HyDs)). The lattice light sheet, on the other hand, could achieve high frame rates by illuminating an entire, potentially large field of view simultaneously (rather than pixel-by-pixel), and also detecting it simultaneously (with a charge-coupled device (CCD) rather than a PMT).
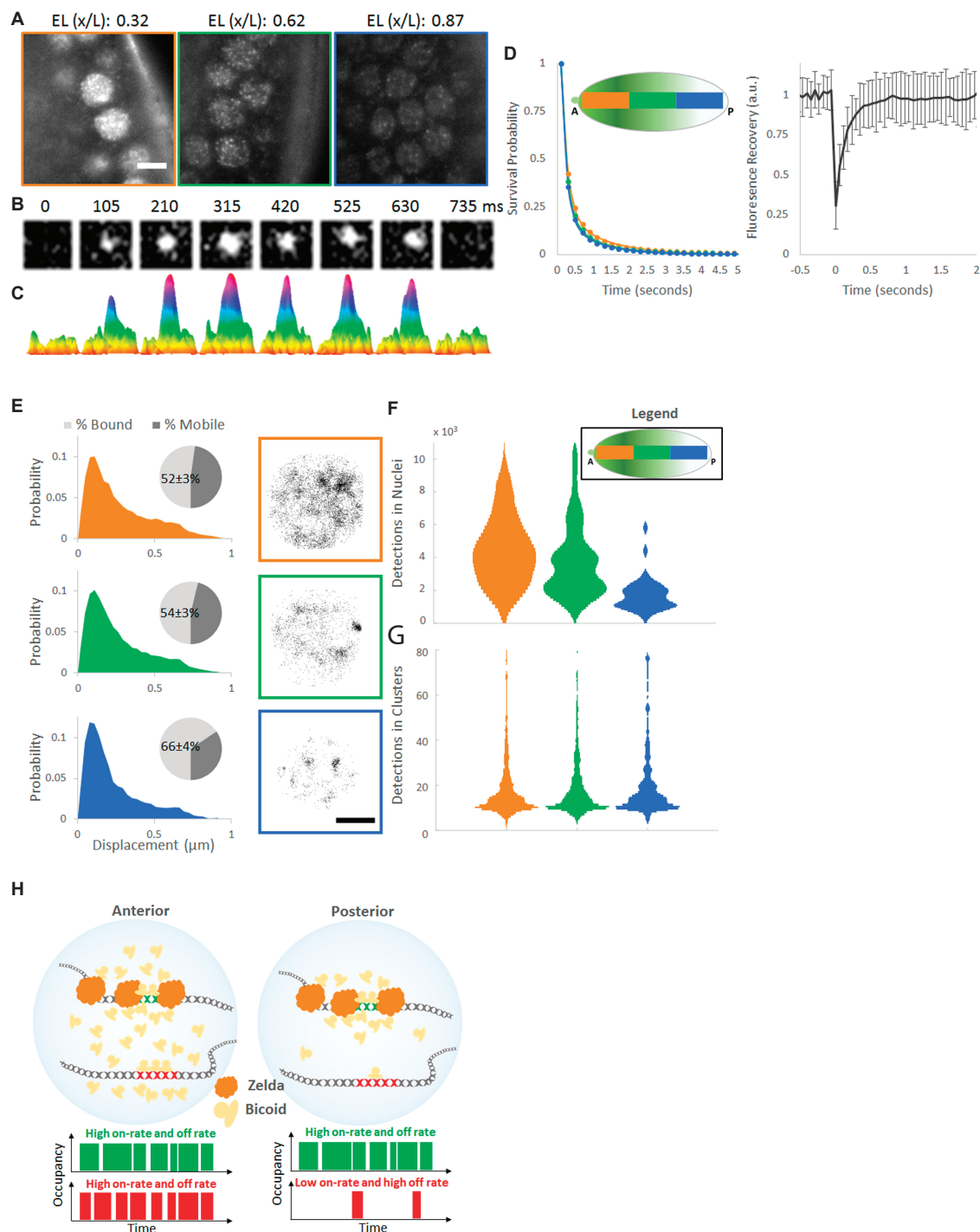
Figure 1.6: **Bicoid binding preview figure.** Single molecule kinetics of Bicoid in living Drosophila Embryos. **(A)** Raw images of BCD-eGFP molecules in a living *Drosophila* embryo acquired with a 100 millisecond exposure time. Scale bar is 5 μm. Positions along the A-P axis are shown as a fraction of the Embryonic Length (EL, x/L). Caption continues on next page.

Figure 1.5:   **Continued from previous page: Bicoid binding preview figure.**   Single molecule kinetics of Bicoid in living *Drosophila* Embryos. **(B)** Example of a single molecule binding event. Top row shows raw images from a 1.2 x 1.2 μm area, bottom row shows corresponding surface plot representations to illustrate the signal-to-noise. **(C)** Uncorrected survival probability curves for Bicoid binding (markers) in the Anterior (34 nuclei), Middle (70 nuclei) and Posterior (83 nuclei) segments of the embryo and corresponding fits to a two-exponent model (solid lines) show no significant differences. **D** FRAP curve for Bicoid shows a recovery time on the order of hundreds of milliseconds, error bars show standard deviation over 21 nuclei. Local modulation of Bicoid Concentration **(E)** Normalized probability distributions of measured displacements in the Anterior (30 nuclei), Middle (67 nuclei), and Posterior (66 nuclei) positions of the embryos, pie charts show the estimated mobile and bound fractions from fits to a two population distribution with the bound population percent labelled with the standard error of the fit parameter. **(F)** Examples of the spatial distribution of all detections in nuclei along the A-P axis, scale bar is 2.5 μm. **(G)** Distribution of the number of detections in all nuclei. (D) Distributions of the number of detections within all clusters. **(H)** Model of Zelda dependent modulation of the Bicoid on-rate at specific loci in the posterior embryo. At high concentrations in the anterior of the embryo, all target sites are highly occupied. At low concentrations, loci with Zelda occupancy have an increased time averaged occupancy through the formation of spatiotemporal hubs that enrich local concentrations and increase the on-rate.

A naïve expectation would be that the duration of individual Bicoid binding events should be unchanged along the length of the embryo (while, of course, the number of Bicoid binding events would be expected to decrease posteriorly), and this experiment permitted us to directly test this assumption. We imaged Bicoid binding events, measured residence times, and performed single particle tracking at three different positions along the entire anteroposterior (A-P) axis (Fig. 1.6 A). This allowed us to test hypotheses about how Bicoid binds to DNA as a function of its concentration. Puncta survival probabilities along the AP axis, how long Bicoid molecules remain bound after being detected, were virtually identical, which conformed to this expectation (Fig. 1.6 D, left). Fluorescence recovery after photobleaching (FRAP) validated this finding that Bicoid binding off-rates are constant along the anteroposterior axis (Fig. 1.6 D, right).

To get a better understanding of Bicoid's mobility, single particle tracking was performed to estimate the proportion of Bicoid that was bound (i.e., small frame to frame displacement) and mobile (i.e., large frame to frame displacement) across the length of the embryo (Fig. 1.6 E, left). Curiously, it was found that the bound fraction increased in the posterior of the embryo. Since we knew from prior experiments that Bicoid's off-rates were unchanged across the A-P axis, this experiment suggested Bicoid's on rates must increase in the posterior. However, on rates for transcription factor-DNA binding are typically diffusion limited and are thus simply concentration dependent. If Bicoid's

concentration was decreasing in the posterior, why would the on rates be increasing and leading to higher bound fractions?

To answer this question, we followed up on an observation that Bicoid may be binding in tightly localized spatial hubs, shown in Figure 1.6 E, right. Subsequent quantification showed that this clustering was indeed a real phenomenon and not due to random chance (Fig. 1.6 F). Furthermore, this observation explains how Bicoid's on rates could be relatively boosted in the posterior despite vanishingly low concentrations—Bicoid gets localized to hubs where local concentrations greatly exceed the global concentration, thereby increasing the DNA binding on rate. Despite the fact that global Bicoid concentration drops exponentially from anterior to posterior, many hubs can still be found in the posterior (Fig. 1.6 F), potentially providing Bicoid with the chance to activate binding sites within gene regulatory regions important for posterior segment specification. Further study additionally showed that hub formation was dependent on the early embryonic pioneer factor, Zelda, (Fig. 1.6 H) allowing activation of Bicoid targets in the posterior.

Having thoroughly discussed the binding and transcriptional output of transcriptional activators in the early *Drosophila* embryo, I will take a step back in the following section and discuss a project geared towards developing tools to enable better quantitative studies in the early embryo.

1.5   Preview of Chapter 5: Development of modular maternal promoters as a tool for generating tunable maternally deposited protein levels in the early *Drosophila melanogaster* embryo

In this chapter, I summarize a project focused on developing new tools for quantitative experiments in the early fly embryo. For experiments in the early fly embryo, we often need to drive the expression of maternally deposited proteins at precise levels and stoichiometries necessary for quantitative measurement. For example, in the Garcia lab, when using the MS2-MCP system for studying transcription, we have often found it necessary to test that the concentration of MCP in the embryo is sufficient to fully saturate transcribed loops without having such high concentrations that the signal is overwhelmed by background MCP. Achieving this balance can only be achieved by driving the expression of MCP at precise levels with a maternal promoter. However, there is a noticeable dearth of promoters available for driving gene expression maternally with precise, predictable, and tunable levels. Inspired by the work of Sano, Nakamura, and Kobayashi 2002 dissecting the maternal *vasa* enhancer, we sought to develop small enhancers that could be multimerized to drive tunable expression levels of maternally deposited proteins in the early embryo. A major motivation for this project was to be

able to drive a wide range of concentrations using maternal promoters By developing a range of expression levels with a single construct, we could create promoters with reproducible and precise levels of gene expression that could be integrated at specific genomic locations using phiC31 integration.

The minimal $\sim$40 bp *vasa* enhancer, necessary and sufficient for driving rescuing levels of Vasa protein, discovered by Sano, Nakamura, and Kobayashi 2002, which we call here *eVasa*, was our starting point. We concatenated 1 to 20 copies of *eVasa* upstream of the maternal transposase promoter to drive the expression of a fluorescent reporter (initially tandem-MCP-eGFP, which was chosen based on our initial motivations for pursuing this project (Fig. 1.6 A). By multimerizing up to ten copies, we can vastly increase expression levels, filling in a crucial gap in concentrations not easy to achieve with the endogenous drives in common use such as the full *vasa* and *nanos* enhancer-promoters. Curiously, we found that expression dropped substantially after ten copies, setting an upper limit on the concentrations that *eVasa* enhancers were capable of achieving(Fig. 1.6 C).

Since we were unable to increase expression after ten copies, but had not yet reached expression levels sufficient for many of our purposes (e.g., for saturating MS2 loops in a transcriptional reporter by MCP-mCherry, for which very high concentrations are required) we created the *eNos* series of multimerizable, modular maternal enhancers (Fig. 1.6 B). Through a trial and error approach, we identified the smallest part, $\sim$80 bp, of the full *nanos* enhancer yet reported in the literature that drives near wild-type levels. We multimerized this *eNos* minimal enhancer to up to 4 copies, but the expression levels plateaued at 3 copies (Fig. 1.6 D). This multimerizable *eNos* enhancer is incredibly useful for pushing past the limits of expression that the *nanos* enhancer can achieve with only a single copy. While it is possible to push to higher levels of maternally deposited proteins by integrating the transgene of interest at multiple places in the genome, such as on different chromosomes, this is a costly and time-consuming process that leads to more complex flies. Achieving higher levels with only a single promoter is much more convenient for many purposes.

To expand the utility of our modular maternal enhancers, we also sought to characterize their expression at different locations in the fly genome in order to accommodate a wider variety of fly crossing and recombination schemes. We injected one particular reporter transgene (MCP-mCherry driven by *eNosx2*), into ten different VK landing sites (Venken et al. 2009) across all chromosome arms in the fly genome (Fig. 1.6 E). We found that expression levels varied dramatically ($\geq$ 8-fold difference in mean expression between the highest and lowest expressing promoters) depending on genomic location (Fig. 1.6 E), and so this must be taken into account when trying to drive maternally deposited proteins. However, our characterization can be used as a guide for decision making when precise concentrations and genomic locations are desired.

A full *vasa* enhancer-promoter (289bp)

-96 to -57 minimal *vasa* enhancer (40bp)
CTAAATCACATTTGATGTGTTAGTGGAAAACGGCTATATA

eGFP

minimal *vasa* enhancer copies

transposase promoter (98bp)

B full *nanos* enhancer-promoter (730bp)

-108 to -29 minimal *nanos* enhancer (80bp)
TTACTAAAAATATCGATATATTTATTATCGCTGGAAAACTACATTATTCCACCTCTAAGCAAGAACCGTTAGTTGGCGCG

eGFP

minimal *nanos* enhancer copies

transposase promoter (98bp)

C

Normalized concentration

Copies of *eVasa* sequence

D

Concentration (a.u.)

Copies of *eNos* sequence

E

Chr X
centromere

Chr 2
L — R

Chr 3
L — R

F

Total fluorescence intensity (a.u.)

- Chr X
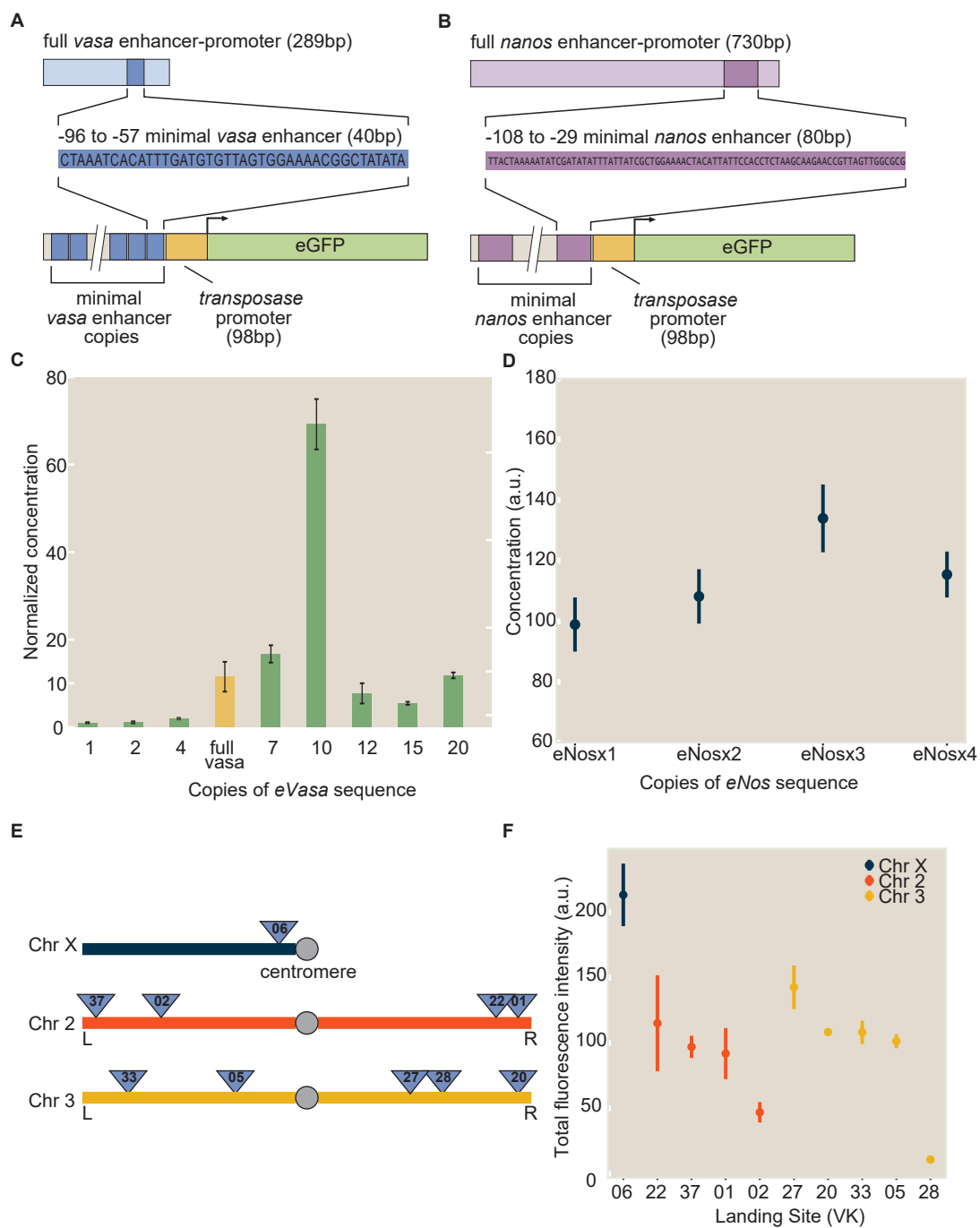- Chr 2
- Chr 3

Landing Site (VK)

Figure 1.6: **Modular maternal enhancers preview figure.** Caption continues on next page.

Figure 1.6: **Continued from previous page: Modular enhancers preview figure. (A)** Minimal *eVasa* design. From the 289 bp *vasa* enhancer-promoter in common use (e.g. for driving Cas9 used in germline genome editing), a 39 bp portion previously identified by Sano, Nakamura, and Kobayashi 2002 was multimerized and placed upstream of the maternal *transposase* core promoter sequence to drive the expression of eGFP. **(B)** Minimal *eNos* design. From the 730 bp annotated enhancer-promoter sequence in common use (e.g. for driving MCP-eGFP in MS2 studies in the early embryo (Garcia et al. 2013), we identified a minimal 89 bp sequence sufficient to drive expression levels comparable to the full sequence. This 89 bp minimal enhancer was then multimerized and placed upstream of the maternal *transposase* promoter to drive the expression of an eGFP reporter. **(C)** Mean nuclear concentration of minimal *eVasa* multimer driven tandem-MCP-eGFP (normalized to the level of *eVasax1*) and **(D)** mean nuclear concentration of minimal *eNos* multimer driven MCP-mCherry during nuclear cycle 12 including all nuclei in the field of view. Error bars are SEM over ≥ 3 embryos. **(E)** Schematic showing positions of genomic landing sites used for reporter integration. **(F)** Mean nuclear concentration of MCP-mCherry driven by *eNosx2-pTrans* at different genomic locations in (E) colored according to chromosome.

Finally, we sought to develop a method for reporting the protein levels driven by our tunable enhancers in absolute numbers of molecules instead of fluorescence arbitrary units so that the concentrations we report could be used independently of particular fluorophore fusions. We turned to the *in vivo* standard candle system developed by Hsia et al. 2016; King et al. 2014; Akamatsu et al. 2020 (Fig. 1.7 A; adapted with permission from Alamos et al. 2020). Here, protein subunits self-assemble *in vivo* into dodecahedra (a trimer of three polypeptides assembles at each vertex, leading to 60 subunits in total) with precise stoichiometries. By fusing fluorophores to these subunits, we can determine a conversion factor from arbitrary units into numbers of molecules from the measured fluorescence intensity of a single dodecahedron that we observe on a microscope.

In order to partially immobilize the standard candles within the cell for easier imaging, we chose to tether our standard candles to the nuclear lamina using a Lamin protein fusion (Fig. 1.7 C, D). It would be undesirable for not *all* polyhedron subunits to be tethered to the nuclear lamina in order to ensure that untethered subunits could self-assemble onto the tethered subunits. Thus, it was necessary to express a certain proportion of subunits fused to tethers, with the rest untethered, at $\sim$1:50 - 1:100 ratio. To accomplish this, we followed the path of Akamatsu et al. 2020 and used the F2A protein self-cleaving system. The F2A peptide is a self-cleaving peptide with only partial efficiency, $6\% \pm 4\%$. By placing the F2A self-cleaving peptide in between our standard candle and a tether, we could create two distinct polypeptides— those with or without the tether. Out of the protein tethers we tested, we found the Lamin fusion to be the most effective 1.7 E).

Additionally, it was necessary to determine the optimal expression levels of the standard candles for imaging. To determine the optimal level, we tested several of the *eVasa* enhancers to find which concentrations led to standard candles that were at low enough concentration to distinguish single molecules but high enough concentrations to fully self-assemble. From this exploration, we found that *eVasa*x5 to drove the optimal levels. It is worth noting that without the *eVasa* enhancers we developed, this would have been more challenging because the endogenous enhancers available would have been either too low (e.g., full length *vasa*) or too high (e.g., full length *nanos*). Going forward, more work needs to be done to complete the calibration that we initially set out to do, as we only measured intensities from a single 60mer polyhedron, and additional data points are needed using 120mer and 180mer dodecahedra to construct a calibration curve. Additionally, it will be useful to conduct more searches to find high expressing maternal enhancers than *eNos* that can be subjected to the sorts of dissections we performed in this study.
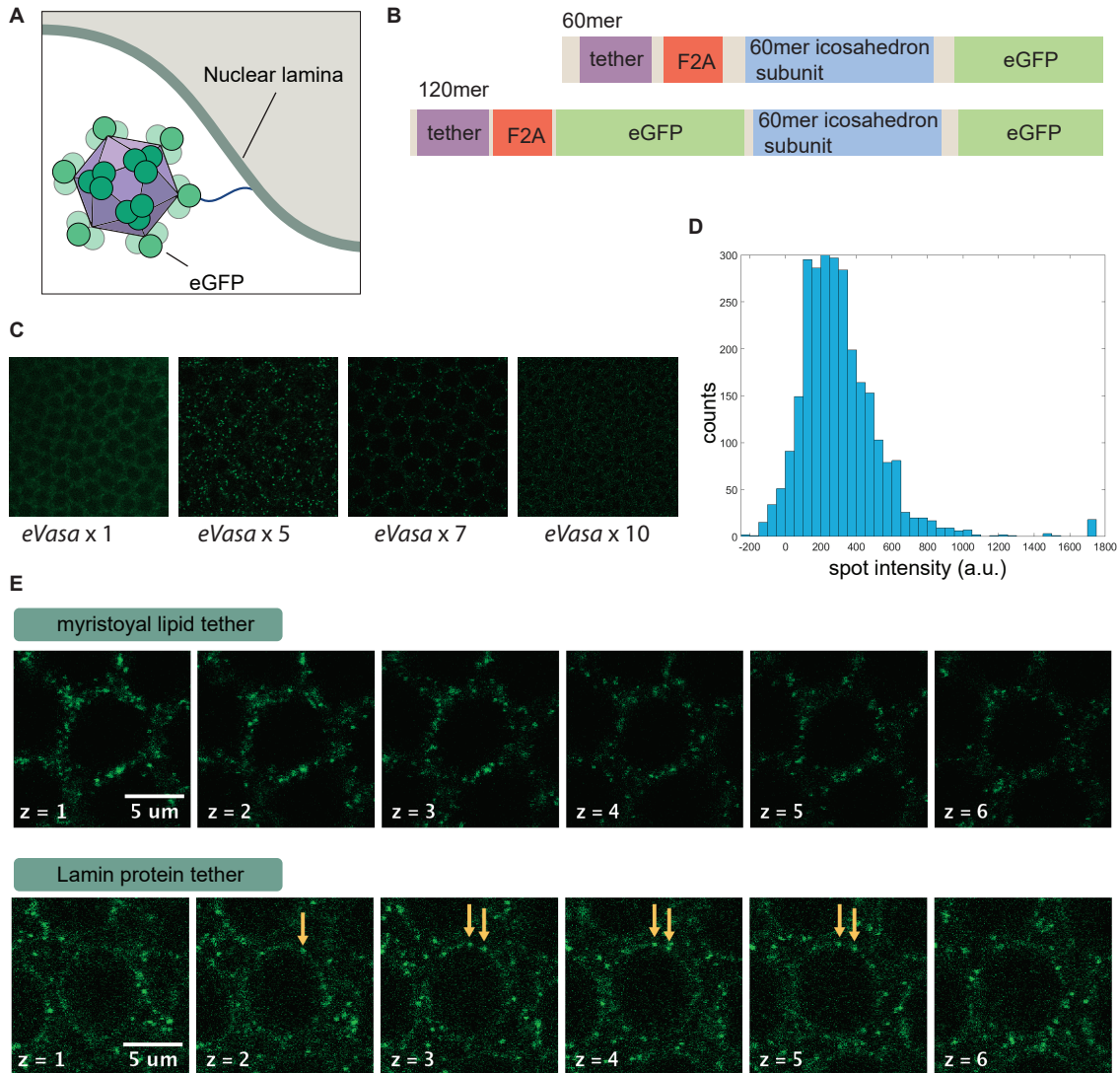
Figure 1.7: **Development of standard candles in *Drosophila* embryos (A)** Schematic of standard candle localization by membrane tether (adapted with permission from Alamos et al. 2020). Caption continues on next page.

Figure 1.7: **Continued from previous page: Development of standard candles in *Drosophila* embryos. (B) (B)** Schematic of standard candle genetic construct designs (tether: either myristoyal binding protein or Lamin protein, F2A: self-cleaving peptide). **(C)** Standard candles driven at different expression levels using multimer *eVasa-pTrans-tdMCP-eGFP* reporters. **(D)** Histogram showing distribution of *eVasax5-pTrans-6omer-mCherry-F2A-lamin-tub3'UTR* standard candle fluorescence. Intensities are the sum of fluorescence from 3 z-slices centered at the brightest z-slice. Intensities only drawn from spots with 3 to 8 z-slices. **(E)** Confocal microscopy images of standard candle localization with two different types of tethers: (top) myristoyal lipid tether for membranes, *eVasax10-pTrans-myr-F2A-eGFP-6omer-eGFP-Tub3'UTR*, and (bottom) Lamin protein tether for localization to the nuclear lamina, *eVasax5-pTrans-6omer-mCherry-F2A-lamin-tub3'UTR*.

Chapter 2

Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient

2.1 Abstract

How enhancers interpret morphogen gradients to generate spatial patterns of gene expression is a central question in developmental biology. Although recent studies have begun to elucidate that enhancers can dictate whether, when, and at what rate a promoter will engage in transcription, the complexity of endogenous enhancers calls for theoretical models with too many free parameters to quantitatively dissect these regulatory strategies. To overcome this limitation, we established a minimal synthetic enhancer system in embryos of the fruit fly *Drosophila melanogaster*. Here, a gradient of the Dorsal activator is read by a single Dorsal binding site. By quantifying transcriptional activity using live imaging, our experiments revealed that this single Dorsal binding site is capable of regulating whether promoters engage in transcription in a Dorsal concentration-specific manner. By modulating binding-site affinity, we determined that a gene's decision to engage in transcription and its transcriptional onset time can be explained by a simple theoretical model where the promoter has to traverse multiple kinetic barriers before transcription can ensue. The experimental platform developed here pushes the boundaries of live-imaging in studying gene regulation in the early embryo by enabling the quantification of the transcriptional activity driven by a single transcription factor binding site, and making it possible to build more complex enhancers from the ground up in the context of a dialogue between theory and experiment.

2.2   Introduction

The adoption of distinct cellular identities in multicellular organisms relies on the formation of spatial gene expression domains driven, in large part, by transcriptional regulatory programs. The positional information giving rise to these mRNA patterns is typically provided by transcription factor gradients (Fig. 2.1A) whose concentrations are interpreted by enhancer DNA sequences that, in turn, regulate transcription of developmental genes (Wolpert 1969; Briscoe and Small 2015). A long-standing goal in quantitative developmental biology is to precisely predict gene expression from knowledge of the DNA regulatory sequence and morphogen concentration (Garcia et al. 2020; Vincent, Estrada, and DePace 2016). Achieving this predictive understanding requires theoretical models that calculate how DNA sequence dictates the functional relation between input morphogen concentration and output transcriptional activity, and calls for testing these predictions by measuring input-output functions (Garcia et al. 2020). Precise genetic manipulations (Venken and Bellen 2005; Bier et al. 2018) and powerful imaging technologies (Gregor et al. 2005; Garcia et al. 2013; Mir et al. 2017) have rendered the early embryo of the fruit fly *Drosophila melanogaster* (*Drosophila*) a prime model system for quantitatively dissecting these input-output functions in development.

In recent years, several studies have reported that *Drosophila* enhancers can control various, potentially independent aspects of transcriptional dynamics in early embryonic development (Fig. 2.1; Lucas et al. 2013; Garcia et al. 2013; Fukaya, Lim, and Levine 2016a; Lammers et al. 2020; Fuqua et al. 2020; Eck et al. 2020; Berrocal et al. 2020; Fukaya 2021; Harden, Vincent, and DePace 2021). First, for a given gene, a fraction of loci remain transcriptionally inactive throughout entire mitotic cycles in development, even when exposed to the same activator concentration as active loci (Fig. 2.1B)—a behavior usually quantified through the fraction of active nuclei or loci. This stochastic decision for a locus to become active is a ubiquitous and potentially important regulatory feature for shaping gene-expression patterns in the embryo (Garcia et al. 2013; Dufourt et al. 2018; Lammers et al. 2020; Harden, Vincent, and DePace 2021). However, it remains unclear whether this feature constitutes a regulatory 'knob' or whether inactive loci are artifacts of experimental detection thresholds. Second, the timing of transcription onset (and cessation, which is not addressed in the present investigation) can also be controlled by input transcription-factor dynamics (Fig. 2.1C; Desponds et al. 2016; Tran et al. 2018; Dufourt et al. 2018; Eck et al. 2020; Lammers et al. 2020; Desponds, Vergassola, and Walczak 2020; Harden, Vincent, and DePace 2021). Finally, the rate of transcriptional initiation in active loci is under regulatory control (Fig. 2.1D) and has been the focus of most studies to date (e.g., Garcia et al. 2013; Fukaya, Lim, and Levine 2016b; Park et al. 2019; Lammers et al. 2020; Berrocal et al. 2020; Fukaya 2021). Thus, multiple regulatory strategies together realize gene-expression patterns in space and time (Fig. 2.1E).

Intense theoretical scrutiny (Desponds et al. 2016; Fakhouri et al. 2010; Sayal et al. 2016; Estrada et al. 2016; Scholes and Depace 2017; Dufourt et al. 2018; Park et al. 2019; Eck et al. 2020; Cheng et al. 2021) has generated a compelling hypothesis: that the reg-
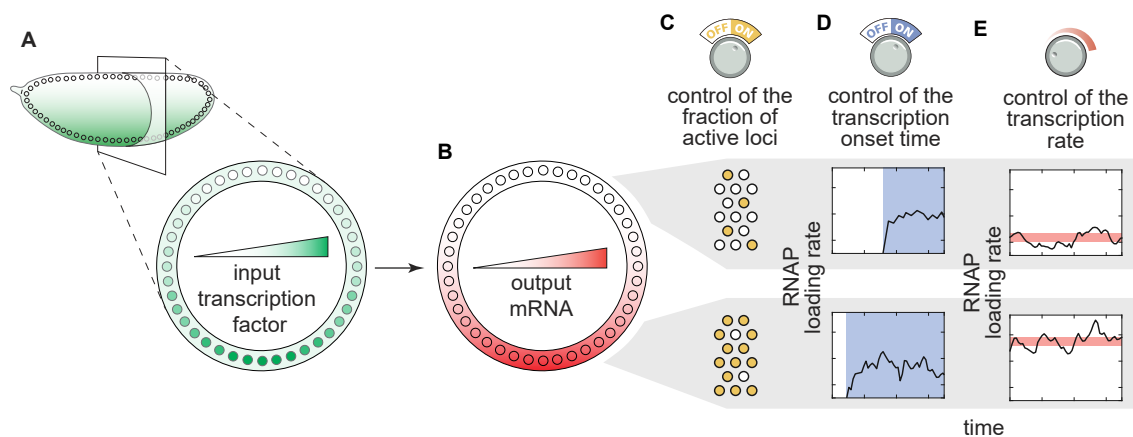
Figure 2.1: **Transcriptional regulatory strategies of enhancers in response to transcription factor concentration gradients. (A)** A *Drosophila* embryo with a transcription factor gradient along its dorsoventral axis. **(B)** This input transcription factor dictates the emergence of output gene-expression patterns by controlling a combination of three enhancer regulatory 'knobs': **(C)** the probability of loci becoming transcriptionally active, **(D)** the transcription onset time, and **(E)** the mean transcription rate of active loci. (RNAP, RNA polymerase II).

ulation of transcriptional dynamics can be separated into two stages. First, a promoter must pass through a series of kinetic barriers consisting of reactions catalyzed by transcription factors in order for for loci to engage in transcription. Previous analyses of the mean and distribution in transcriptional onset times have suggested that the number of inactive promoter states can range from one to three (Dufourt et al. 2018; Eck et al. 2020; Harden, Vincent, and DePace 2021). These reactions could be associated with, for example, the stepwise unwrapping of DNA from nucleosomes (Desponds et al. 2016; Dufourt et al. 2018; Eck et al. 2020) and/or the sequential recruitment of general transcriptional cofactors (Zhou et al. 1998). Second, after initial promoter activation, the rate of mRNA production is proportional to the probability of finding RNA polymerase II (RNAP) bound to the promoter. Statistical mechanical (also called thermodynamic) models have been used to calculate this probability of finding RNAP bound to the promoter, and have successfully use to predict mRNA production rates in bacteria (Razo-Mejia et al. 2018). However, whether they can be applied to the more complex context of eukaryotic transcriptional regulation—let alone to the dynamical processes of cellular decision-making in development—is still an open question (Polach and Widom 1995; Schulze and Wallrath 2006; Lam, Steger, and O'Shea 2008; Li et al. 2008; Kim and O'Shea 2008; Levine 2010; Fussner, Ching, and Bazett-Jones 2011; Bai, Ondracka, and Cross 2011; Li et al. 2014; Hansen and O???Shea 2015; Estrada et al. 2016; Li and Eisen 2018; Park et al. 2019; Eck et al. 2020).

One of the main challenges to systematically testing these models is the complexity

of endogenous regulatory regions (Fakhouri et al. 2010; Foo et al. 2014; Sayal et al. 2016; Dufourt et al. 2018; Park et al. 2019; Eck et al. 2020). Because endogenous enhancers contain multiple binding sites for different transcription factors, accounting for these sites and their interactions leads to a combinatorial explosion of model parameters (Garcia, Brewster, and Phillips 2016; Garcia et al. 2020); determining the values of these parameters from simple experiments constitutes a computational—and conceptual—challenge (Vincent, Estrada, and DePace 2016; Garcia, Brewster, and Phillips 2016; Garcia et al. 2020). To render complex transcriptional regulatory systems tractable to theory, minimal synthetic enhancers have been engineered in bacteria (Garcia and Phillips 2011; Brewster et al. 2014; Razo-Mejia et al. 2018; Phillips et al. 2019), eukaryotic cells (Popp et al. 2020), and developing organisms (Fakhouri et al. 2010; Sayal et al. 2016). In such experiments, a short, synthetic DNA sequence with only one to a few binding sites for a single transcription factor drives the expression of a reporter gene. Measuring the concentration of the transcription-factor input and reporter mRNA output makes it possible to test models of transcriptional regulation and to infer molecular parameters that can be used to predict the behavior of more complex regulatory architectures (Phillips et al. 2019).

Here we sought to use synthetic minimal enhancers to challenge our integrated model of transcriptional control using the dorsoventral patterning system in *Drosophila* embryos, in which a concentration gradient of the Dorsal transcription factor specifies spatial domains of transcription, as a case study. To test the integrated model of transcriptional dynamics (Fig. 2.2A,B), we performed simultaneous quantitative live-cell measurements of Dorsal concentration (input) and transcription (output) driven by minimal synthetic Dorsal-dependent enhancers in single nuclei. By repurposing the *parS*-ParB DNA labeling technology (Germier et al. 2017; Chen et al. 2018) to quantify transcriptional activity independent of RNA detection, we determined that the inactive loci described by our model constitute a distinct transcriptional state under regulatory control and are not the result of detection artifacts. Further, our theoretical model predicted how, through the Dorsal-mediated catalysis of reactions prior to transcriptional onset, regulatory architecture dictates both the transcriptional onset time and the fraction of active loci. Finally, once promoters turn on, we found that our measurements are compatible with an equilibrium model. Thus, the present investigation provides quantitative evidence supporting a unified model of transcriptional regulation in eukaryotes that accounts for whether loci become transcriptionally active, when this activity ensues, and, once transcription ensues, at what rate nascent RNA molecules are produced. More generally, our work demonstrates the feasibility of using minimal synthetic enhancers to engage in a dialogue between theory and experiment in the context of transcriptional control in development.

2.3 Results

2.3.1 An integrated model of transcriptional dynamics driven by a single activator binding site

To probe the transcriptional regulatory strategies (Fig. 2.1) of a minimal synthetic enhancer, we posit a theoretical model that predicts the fraction of loci that will become active, their transcriptional onset time, and RNAP loading dynamics once transcription ensues. Specifically, we consider a simplified case in which only one activator is present and can only bind to one site only a few base pairs away from the promoter (Fig. 2.2).

In order to explain the transcriptional onset dynamics of a locus and the probability of loci becoming active, we invoke recent experiments leading to a 'kinetic barrier' model (Desponds et al. 2016; Dufourt et al. 2018; Eck et al. 2020) proposing that, after exiting mitosis, all promoters are in an inactive state. In this state, labeled as 'OFF$_1$' in Figure 2.2A, transcription is not possible. Promoters must then traverse a series of distinct inactive states (labeled 'OFF$_2$' to 'OFF$_n$' in Fig. 2.2A) before reaching an active state in which transcription proceeds (labeled ON in Fig. 2.2A).

The temporal evolution of the enhancer-promoter system as it traverses the states shown in Figure 2.2A can be simulated by computing the probability that the promoter occupies each state. Here, the transition rates between states, k, determines how the states probability spreads from the initial condition where the promoter is in state OFF$_1$ to the active state as time passes (see Section S1.1 for details).

We propose that a transcriptional activator such as Dorsal can catalyze the transition between states in an affinity-dependent manner via binding to its cognate site in the enhancer. Because we assume that Dorsal binding and unbinding is faster than the transition rate k, we posit that k is a linear function of the equilibrium Dorsal occupancy at the enhancer such that

$$k(t) = c \cdot \frac{\frac{[Dl](t)}{K_D}}{1 + \frac{[Dl](t)}{K_D}}, \tag{2.1}$$

where c is a rate constant, $[Dl](t)$ is the Dorsal concentration at time t, and K_D is the Dorsal-DNA dissociation constant.

Because Dorsal is time-varying, the model cannot be solved analytically. As a result, we numerically calculated the probability of the promoter being in each state as a function of time using a particular set of model parameters (see details in Section S1.1). As seen in Figure 2.2C, because individual loci must traverse a sequence of intermediate states before reaching the ON state, this model introduces a delay in activation.

This kinetic barrier model accounts for loci that never transcribe during the nuclear cycle. Specifically, if nuclear cycles lasted indefinitely, all promoters would eventually reach the ON state as shown in Figure 2.2C. However, due to the rapid mitotic cycles that characterize early embryonic development in *Drosophila*, this duration is limited: transcription cannot initiate during mitosis and thus is only permissible during a time

window within interphase (Fig. 2.2C, vertical dashed line; Shermoen and O'Farrell 1991; Garcia et al. 2013; Eck et al. 2020). Consequently, if the time it takes a promoter to reach the ON state is longer than the duration of this window, then this hypothetical promoter will not initiate transcription at all during the nuclear cycle (Fig. 2.2C, horizontal dashed line).

The kinetic barrier model can be used to predict two of the three regulatory strategies, fraction of active loci and transcription onset times, that we aim to dissect quantitatively (Fig. 2.1). First, the model predicts how the fraction of active loci is determined by Dorsal nuclear concentration and binding affinity (Fig. 2.2D, left y-axis). Second, this same model calculates the mean transcriptional onset time of those loci that turn on as a function of these same Dorsal parameters (Fig. 2.2D, right y-axis).

To model a locus once it is active, we follow Eck et al. 2020 and propose a simple thermodynamic model (Bintu et al. 2005b; Bintu et al. 2005a) that assumes that the RNAP loading rate, R, is proportional to the probability of finding RNAP bound to the promoter p_{bound}, such that

$$R = R_{max} \cdot p_{bound}, \tag{2.2}$$

where R_{max} is a constant coefficient that dictates the maximum possible polymerase loading rate.

Thermodynamic models enable the calculation of p_{bound} by assigning a statistical weight to each possible state in which the regulatory system can be found. In the case of a minimal enhancer with one activator binding site, the enhancer-promoter DNA can be empty, occupied by Dorsal, occupied by RNAP, or simultaneously bound by Dorsal and RNAP (Fig. 2.2B). The statistical weight associated with each of these terms is shown in Figure 2.2B. Here, $[Dl]/K_D$ is the statistical weight associated with finding Dorsal (with concentration $[Dl]$ and binding dissociation constant K_D) bound to the promoter alone, while $[P]/K_P$ is the weight of finding RNAP (with concentration $[P]$ and binding dissociation constant K_P) bound to the promoter alone. Note that the weight of having both Dorsal and RNAP bound simultaneously includes an extra glue-like cooperativity coefficient, ω, that determines how strongly Dorsal recruits RNAP to the promoter. The value of ω is constrained to be > 1 so that higher Dorsal occupancy leads to higher RNAP occupancy.

To calculate p_{bound}, we divide the sum of the weights featuring a bound RNAP molecule by the sum of all possible weights. Substituting this calculation into Equation 1 yields

$$R = R_{max} \cdot p_{bound} = R_{max} \cdot \frac{\frac{[P]}{K_P} + \frac{[Dl]}{K_D}\frac{[P]}{K_P}\omega}{1 + \frac{[Dl]}{K_D} + \frac{[P]}{K_P} + \frac{[Dl]}{K_D}\frac{[P]}{K_P}\omega}, \tag{2.3}$$

which is plotted in Figure 2.2E. As shown in the figure, increasing K_D shifts the concentration at which the RNAP loading rate reaches half its maximum value toward higher Dorsal concentrations, but does not change the overall shape of the curve. We also note the presence of a non-zero baseline of RNAP loading rate due to the Dorsal-independent

$[P]/K_P$ term in the numerator of Equation 2. This baseline suggests that it could be possible for a promoter in the 'ON' state to produce low, basal-level transcription in the absence of bound Dorsal.

Together, the kinetic barrier model outlined in Figure 2.2A and the thermodynamic model's Equation 2 define a comprehensive quantitative framework that predicts how the fraction of active loci, the transcriptional onset time, and the RNAP loading rate as a function of Dorsal concentration vary as model parameters such as the Dorsal dissociation constant K_D are modulated (Fig. 2.2D,E). These predictions constitute hypotheses that we experimentally tested throughout the remainder of this work.
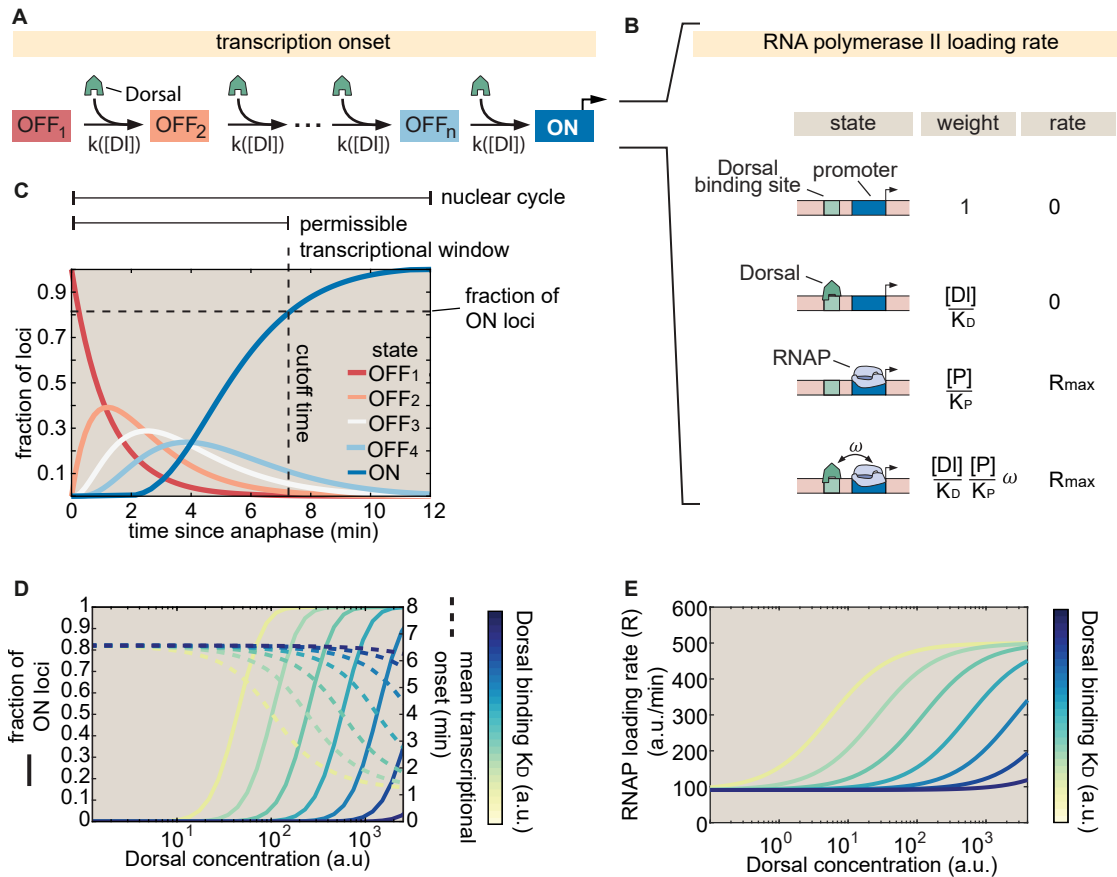
Figure 2.2: **Integrated kinetic and thermodynamic model of simple activation by Dorsal. (A)** The promoter undergoes kinetic transitions from transcriptionally inactive states (OFF_1 to OFF_n) to an active state (ON) with Dorsal accelerating the transition rate, k, by a factor proportional to the Dorsal occupancy at the promoter. **(B)** Thermodynamic states and weights for the simple activator model. The probability of finding RNAP bound to the promoter can be calculated from the statistical weights associated with all possible occupancy states of the enhancer-promoter system. **(C)** Visualization of a particular solution of the kinetic scheme from (A) showing the probability of finding a given locus in each of the states for an illustrative, representative set of parameters ($[Dl]$ = 1000 a.u., K_D = 1000 a.u., c = 10/min, n = 4 states, and 7 min nuclear cycle duration). The predicted fraction of active loci (dashed horizontal line) is calculated as the probability of being in the ON state by the end of the permissible time window (dashed vertical line) that is determined by mitotic repression. **(D)** Predictions for the fraction of active loci (solid lines plotted against the left y-axis) and mean transcriptional onset times (dashed lines plotted against the right y-axis) as a function of Dorsal concentration for different, illustrative values of the Dorsal dissociation constant K_D. **(E)** Rate of mRNA production across active loci as a function of Dorsal concentration for different values of K_D based on the model in (B) (R_{max} = 1000 a.u., Dorsal K_D ranging from 10 a.u. to 10^5 a.u., ω = 10, $[P]/K_P$ = 0.1).

2.3.2 Establishing a minimal synthetic enhancer system to test theoretical predictions

To test our model's predictions, we constructed single binding site enhancers driven by the Dorsal activator. Dorsal is one of the best characterized transcription factors in *Drosophila* and a classic example of a morphogen (Roth, Stein, and Nüsslein-Volhard 1989; Reeves et al. 2012b). Dorsal is provided maternally and forms a dorsoventral gradient of nuclear localization (Fig. 2.3A) (Gilbert 2010), acting as an activator by default (Thisse et al. 1991; Jiang et al. 1991) and as a repressor in the presence of nearby binding sites for corepressors (Kirov et al. 1993; Papagianni et al. 2018). Prior to activation of the zygotic genome (up to the 12th mitotic cycle), Dorsal is the only transcription factor with a nuclear protein gradient across the dorsoventral axis (Sandler and Stathopoulos 2016; Dufourt et al. 2020). Thus, the Dorsal nuclear concentration is the sole source of dorsoventral positional information for developmental enhancers at this stage in development. These features, combined, make Dorsal an ideal input transcription factor for activating a minimal synthetic reporter system.

In order to relate output transcriptional activity to the time-variant input Dorsal concentration throughout development, we measured the instantaneous Dorsal concentration in live embryos by creating a CRISPR knock-in Dorsal-mVenus fusion allele based on a previous Dorsal fusion (Reeves et al. 2012b) that rescues embryonic development (Kremers et al. 2006; Gratz et al. 2015; Materials and methods). Further, in order to increase the dynamic range of Dorsal concentration in our experiments, we further combined this CRISPR allele with a Dorsal-mVenus transgene (Reeves et al. 2012b), resulting in a line that will hereafter be referred to as 2x Dorsal flies. This fusion made it possible to quantify the concentration dynamics of the Dorsal protein input (Fig. 2.3A,B) in individual nuclei (Video S4, left; Materials and methods). Dorsal-mVenus nuclear fluorescence time traces quantified over nuclear cycle 12 confirmed the dynamic nature of Dorsal concentration and were quantitatively similar to previous measurements (Fig. 2.3B; Reeves et al. 2012b; details of Dorsal-mVenus quantification in Fig. S6A,B). Nuclear cycle 12 nuclei in 2x Dorsal flies experience a Dorsal concentration gradient spanning multiple orders of magnitude, from less than 1 nM to $\approx$ 400 nM (Fig. 2.3B; details of Dorsal-mVenus calibration in Fig. S8).

To visualize the dynamics of Dorsal-dependent transcription, we developed a reporter transgene containing a minimal synthetic enhancer consisting of a single high affinity, consensus binding site for the Dorsal transcription factor (Ip et al. 1992; Jiang et al. 1993; Szymanski and Levine 1995) (Fig. 2.3C). Hereafter we refer to this strong site enhancer as as DBS_6.23 for Dorsal Binding Site, followed by its binding affinity score according to the Patser algorithm (Stormo and Hartzell 1989; Materials and methods). To quantify the transcriptional activity of this enhancer, we used the MS2-MCP system to fluorescently label nascent RNA molecules in our reporter constructs, which appear as nuclear fluorescent puncta (hereafter "transcription spots") in laser-scanning confocal microscopy movies (Video S4, right; Bertrand et al. 1998; Garcia et al. 2013; Lucas

et al. 2013. We performed image analysis of the MS2 movies using a custom data analysis pipeline in Matlab and Fiji (Materials and methods; (Schindelin et al. 2012; Lammers et al. 2020).

To validate this minimal synthetic system, we determined that DBS_6.23-MS2 drives detectable and quantifiable levels of transcription, and that this transcriptional activity is mainly governed by Dorsal. We compared the transcriptional activity of DBS_6.23-MS2 in embryos laid by 2x Dorsal females with the activity in embryos laid by females homozygous for a *dorsal* null allele. While transcription spots were clearly present in the 2x Dorsal background (Fig. 2.3D, left), they were extremely rare in *dorsal* null embryos (Fig. 2.3D, middle): not a single transcription spot was detected during nuclear cycle 12 in any of 4 replicates containing > 60 nuclei in total. Dorsal is therefore necessary for transcriptional activity in our reporter constructs.

We next sought to determine whether the detected transcriptional activation is solely due to Dorsal interacting with the binding site explicitly engineered into the construct or whether there are cryptic Dorsal binding sites contributing to gene expression. We generated a second reporter, DBS_4.29-MS2 in which the Dorsal binding site was strongly perturbed using known point mutations (Ip et al. 1992). Transcription was rarely detectable in DBS_4.29-MS2 embryos (Fig. 2.3D, right), with the average transcriptional activity (mean instantaneous fluorescence) per detected spot being less than 10% of the optimal DBS_6.23 enhancer at any Dorsal concentration (Fig. S11). Thus, the Dorsal site placed within the synthetic enhancer is necessary for robust activation and is the main driver of this transcriptional activity.

Next, we identified which observable features in the MS2 signal could be used as metrics for quantifying Dorsal-dependent transcriptional activity. We collected DBS_6.23-MS2 time traces of MCP-mCherry fluorescence from transcription spots during nuclear cycle 12 along with four metrics of transcriptional activity (Fig. 2.3E,F). First, the maximum spot fluorescence corresponds to the 95th percentile of intensity over time, which is proportional to the transcription rate (Section S1.2). Second, the transcriptional onset time is defined as the time since the previous mitosis at which a transcription spot is first detected (Fig. S4). Third, the integrated spot fluorescence corresponds to the time integral of the spot fluorescence and is directly proportional to the amount of mRNA produced by the locus (Garcia et al. 2013) (Materials and methods). Finally, as previously observed in other genes in flies (Garcia et al. 2013; Dufourt et al. 2018; Lammers et al. 2020; Harden, Vincent, and DePace 2021), not all nuclei exposed to the same average nuclear Dorsal concentration exhibited detectable transcription (Fig. 2.3F). As a result, we quantified the fraction of active loci—regardless of their level of activity or temporal dynamics—by measuring the number of nuclei with observable transcription signal in at least one movie frame throughout nuclear cycle 12, divided by the total number of nuclei in the field of view. Thus, we have established quantitative metrics that enable us to engage in a dialogue between experiment and a theory of Dorsal-driven transcriptional dynamics.
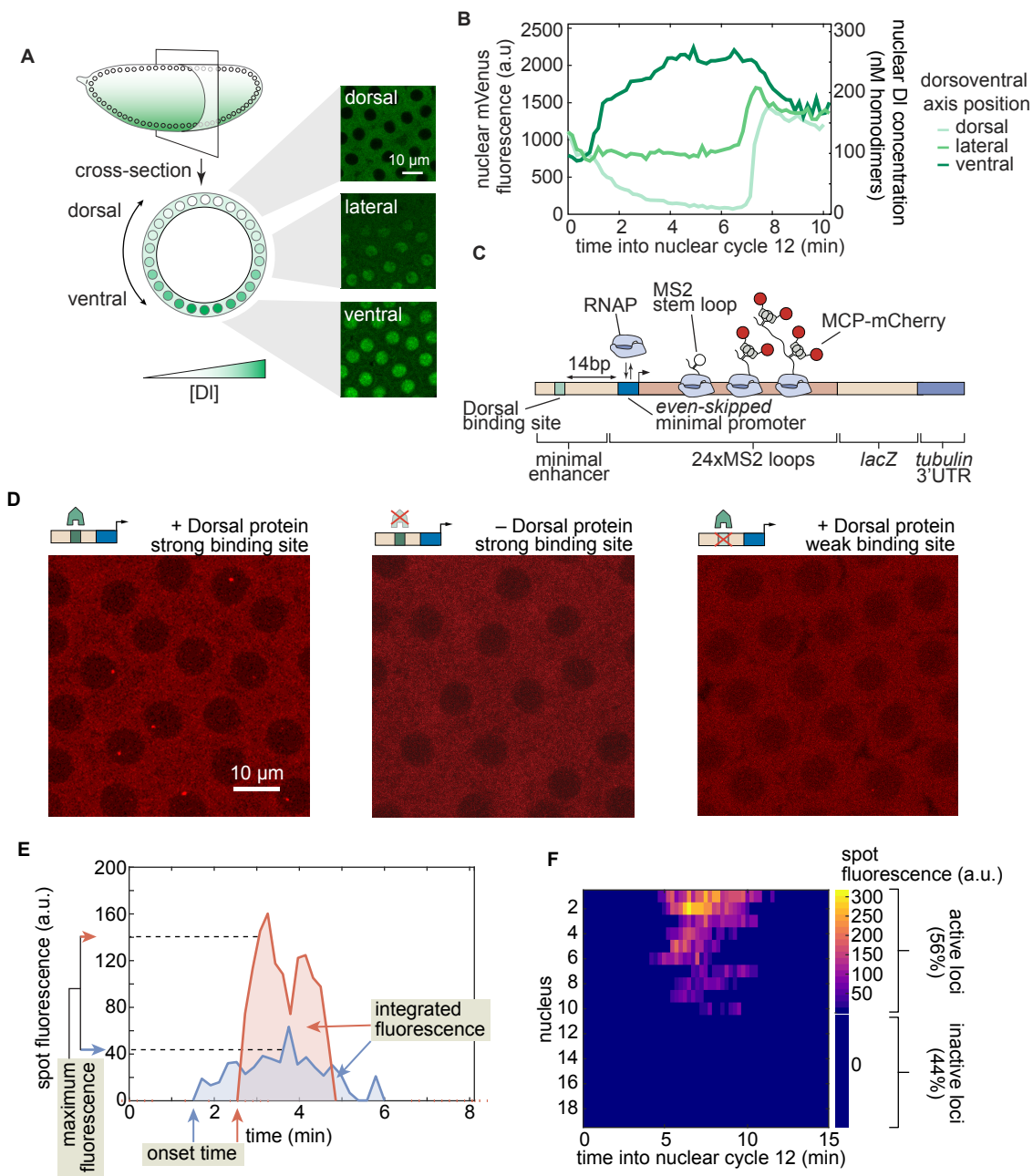
Figure 2.3: **Simultaneously measuring transcription factor protein input and transcriptional output. (A)** Schematic of the Dorsal protein gradient in early *Drosophila* embryos. Dorsal protein accumulates in ventral nuclei and is progressively excluded from more dorsal nuclei. Example snapshots show Dorsal-mVenus in various positions along the dorsoventral axis. **(B)** Representative time traces of nuclear Dorsal-mVenus fluorescence in various positions along the dorsoventral axis. The right y-axis shows the nuclear Dorsal concentration according to the calibration described in Figure S8. Caption continues on next page.

Figure 2.4: **Continued from previous page: Simultaneously measuring transcription factor protein input and transcriptional output. (C)** Schematic of minimal synthetic enhancer system containing a single binding site for Dorsal that drives transcription of a reporter tagged with MS2 loops, which are visualized through the binding of MCP-mCherry. The Dorsal binding site is placed 14 bp upstream of the *even-skipped* minimal promoter. **(D)** Snapshots from embryos containing an optimal binding-site reporter in the presence (left) or absence (middle) of Dorsal, or containing a strongly mutated Dorsal binding site (right). **(E)** Example fluorescence time traces and quantitative metrics of transcriptional activity. **(F)** Fluorescence of all transcription spots in individual nuclei in the field of view of one embryo as a function of time. If a transcription spot was detected within a nucleus at any point during the interphase of nuclear cycle 12, then the locus was considered active; otherwise, the locus was classified as inactive.

2.3.3 Transcriptionally active and inactive loci correspond to functionally distinct populations

Before attempting to predict Dorsal-driven transcriptional dynamics, it is important to ensure that the fact that only some loci engage in transcription is the result of Dorsal action and not of limitations of our experimental setup. Transcriptionally silent loci that remain inactive throughout interphase, such as those revealed by our experiment (Fig. 2.3F), have been observed using MS2 (and its sister mRNA labeling tool, PP7) in live-imaging experiments in flies (Garcia et al. 2013; Lammers et al. 2020; Berrocal et al. 2020), plants (Alamos et al. 2020), and mammalian cells (Hafner et al. 2020). However, it has not been possible to determine whether these inactive loci correspond to a separate transcriptional state from active loci, or whether they are an artifact of the fluorescence detection thresholds associated with various microscopy techniques.

To answer this question, it is necessary to quantify MS2 fluorescence at these inactive loci and determine whether they differ from loci not exposed to activators, which do not transcribe (Fig. 2.3F). However, to date this approach has not been feasible because most MS2 measurements have relied on the presence of an MS2 signal itself to segment and quantify the fluorescence of transcription spots. We hypothesized that, if undetected loci correspond to a distinct and weaker, Dorsal-independent state, then detected and undetected spots in embryos carrying wild-type Dorsal would appear as two distinct populations. In this scenario, the mCherry fluorescence of undetected spots corresponding to inactive loci in wild-type Dorsal embryos would be similar to that observed in Dorsal null embryos, and clearly distinct from the mCherry fluorescence of active loci in the presence of Dorsal.

To quantify MS2 fluorescence independently of whether a MS2 spot was detected, we implemented the *parS*-ParB DNA labeling system (Germier et al. 2017; Chen et al. 2018). Here, fluorescently labeled ParB proteins bind the parS DNA sequence resulting

in a fluorescence spot appearing at the locus independently of the transcriptional state of the locus (Fig. 2.5A). We created flies with and without functional Dorsal expressing ParB2-eGFP (subsequently referred to as ParB-eGFP) and MCP-mCherry to label our locus DNA and nascent RNA, respectively. We crossed flies containing parS-DBS_6.23-MS2 to flies carrying ParB-eGFP and MCP-mCherry to generate embryos that have our locus of interest labeled with ParB-eGFP colocalized with the transcriptional signal in the MCP-mCherry channel (Fig. 2.5A,B; Video S4).

Guided by the spatial positions reported by ParB-eGFP, we measured the MCP-mCherry signal at all DBS_6.23 reporter loci in embryos carrying wild-type Dorsal (Fig. 2.5C) or laid by mothers homozygous for the dl^1 null allele (Dorsal null embryos). We then classified loci from wild-type Dorsal embryos into two categories, detected and undetected, depending on whether they were identified as spots in the MCP-mCherry channel by our analysis pipeline (Fig. 2.5B,C; Section 2.5.5). As shown in the the examples presented in Figure 2.5D, there are clear qualitative differences between MCP-mCherry fluorescence time traces corresponding to detected or undetected transcriptional spots from wild-type embryos. Thus, our analysis made it possible to quantify MS2 fluorescence in three populations: all loci in Dorsal null embryos, undetected loci in wild-type Dorsal embryos, and detected loci in wild-type Dorsal embryos.

To compare these populations, we computed the 95th percentile value over each locus' MCP-mCherry fluorescence time trace (Fig. 2.5E). The distribution of mCherry fluorescence from undetected spots in wild-type Dorsal embryos largely overlapped with that of all spots in Dorsal-null embryos (Fig. 2.5F), consistent with these two populations corresponding to loci expressing Dorsal-independent levels of activity. Moreover, both distributions were clearly distinct from the distribution of detected spots in wild-type Dorsal embryos (Fig. 2.5E,F). Thus, our results provide strong evidence that inactive loci are not artifacts of the detection limit of our imaging techniques. Rather, loci can belong to one of two distinct populations: those that transcribe at a high, Dorsal-dependent level and those that are transcriptionally inactive (or active at a low, undetectable level that is comparable to that of embryos lacking Dorsal). We therefore conclude that the decision to transcribe made by each locus is an additional regulatory strategy controlled by Dorsal.

From the observations in Figure 2.5E and F, we estimated our error in classifying loci as inactive. This false-negative detection rate, corresponding to the area under the curve shaded in the inset of Figure 2.5F, is estimated as 15.9%. However, this false-negative rate is likely an underestimation. For example, this rate may depend on Dorsal concentration, which cannot be controlled for in this experiment. Additionally, the presence of ParB in the locus may itself affect transcriptional dynamics, impacting the false-negative rate. For these reasons, we do not attempt to correct our measurements of the fraction of active loci using this estimated false-negative rate.
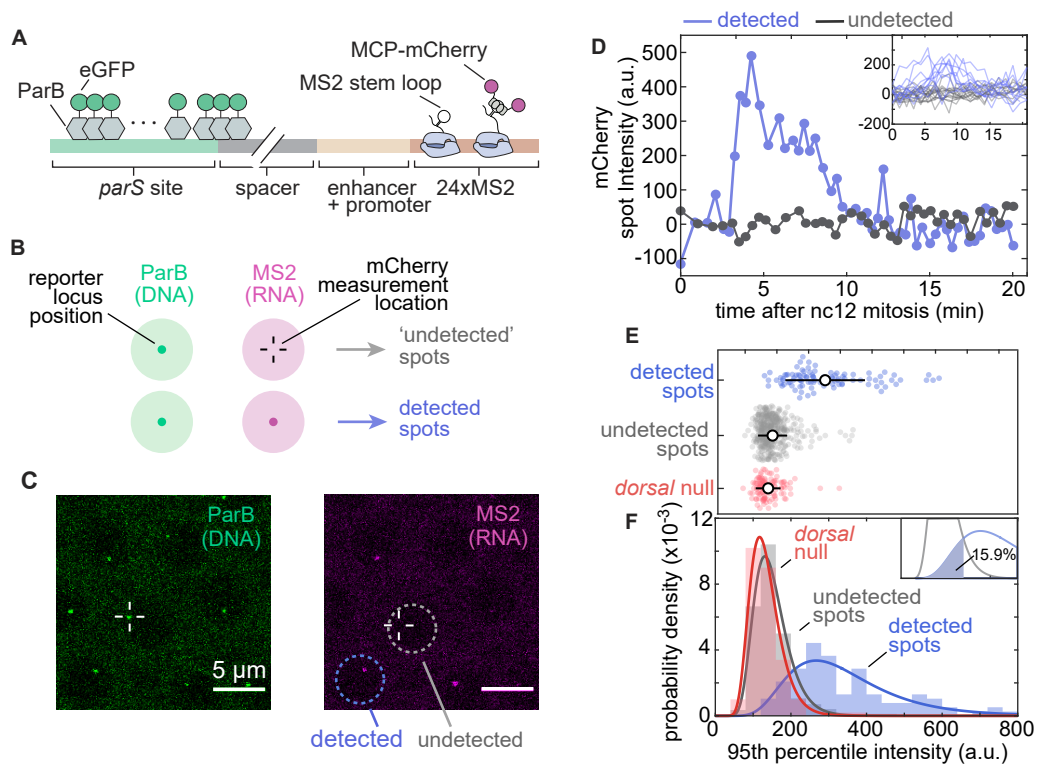
Figure 2.5: **Transcriptionally independent ParB labeling confirms that transcriptionally inactive loci are functionally distinct from active loci. (A)** Schematic of ParB-eGFP construct. ParB-eGFP molecules bind and polymerize out from *parS* sequences, which are placed ∼400 bp upstream of the enhancer. The enhancer and promoter together drive transcription of MS2 loops that subsequently bind MCP-mCherry. **(B)** Schematic of the experiment. Loci are located by detecting a signal in the ParB-eGFP channel; these locations were used to fit a 2D Gaussian to the same area in the MS2-mCherry channel to estimate fluorescence intensity regardless of whether an MS2-mCherry signal was detected (Materials and methods Sec. 2.5.4). **(C)** Example images of ParB-eGFP (left) and MCP-mCherry (right) channels. Detected and undetected transcriptionally active loci solely based on the MCP-mCherry signal alone are shown. **(D)** Example time traces of MCP-mCherry fluorescence over time at the ParB-eGFP loci in nuclei with (blue) and without (grey) detected MS2-mCherry spots of the DBS_6.23 enhancer showing clear qualitative differences between the two populations. Inset, all detected and undetected fluorescence traces obtained in the same embryo. Negative intensity values are due to spot intensities very close to the background fluorescence. Caption continues on next page.

Figure 2.5: **Continued from previous page: Transcriptionally independent ParB labeling confirms that transcriptionally inactive loci are functionally distinct from active loci. (E)** Swarm plots of 95th percentile MCP-mCherry fluorescence at loci with detected (blue; N = 125) and undetected MS2-mCherry transcription (gray; N = 425) driven by the DBS_6.23 enhancer in wild-type Dorsal embryos. Red (N = 96), maximum fluorescence of all loci in Dorsal null embryos, defined as the 95th percentile of intensity over time (black circles, mean; bars, standard deviation). Detected spots are significantly different from both null (ANOVA, $p < 0.01$) and undetected spots (ANOVA, $p < 0.01$) **(F)** Histograms of the data shown in (E). Solid lines correspond to log-normal fits performed for ease of visualization. Inset, undetected and detected distribution fits and the area used to estimate the false-negative detection rate of 15.9%.).

2.3.4 Dorsal-dependent kinetic barriers explain transcription onset dynamics and modulation of the fraction of active loci

Having established that transcriptionally inactive promoters mostly constitute a separate population from transcriptionally active promoters (Fig. 2.5), we sought to test whether our theoretical model (Fig. 2.2A) can quantitatively recapitulate the fraction of active loci and their transcription onset times. Tuning transcription factor-DNA binding affinity has been a powerful tool to test models of transcriptional regulation in the past (Meijsing et al. 2009; Phillips et al. 2019). Inspired by these previous works, we probed our model by adjusting the Dorsal-DNA interaction energy in our minimal synthetic enhancer.

We constructed a series of enhancers containing a single binding site with varying affinities for Dorsal. Building on the optimal DBS_6.23 and the mutated DBS_4.29 sites (Fig. 2.3D, left vs. right), we created five additional enhancers of varying intermediate strengths by introducing point mutations into the consensus Dorsal binding motif to obtain a range of predicted affinities (Fig. 2.6A,B; Materials and methods Section 2.5.1). As described above, we refer to these enhancers as DBS, followed by their corresponding Patser score.

For the purpose of quantifying output transcriptional activity as a function of Dorsal concentration, we assigned a single Dorsal concentration value to each nucleus corresponding to the mVenus fluorescence in the center of that nucleus at a fiducial time point halfway through each nucleus's lifetime, approximately in the middle of nuclear cycle 12 when Dorsal levels are relatively stable (Fig. S6A,B). We next grouped nuclei into 17 linearly spaced bins that span the dorsoventral axis based on their fiducial fluorescence (Fig. S6B).

We assessed whether these point mutations were sufficient to generate a graded response to Dorsal and to determine the dynamic range of gene expression afforded by these enhancers. To make this possible, we integrated the total mRNA output over nu-

clear cycle 12 of each enhancer as a function of Dorsal concentration across all nuclei exposed to a given Dorsal concentration. The integrated mRNA output of the four weakest enhancers changed little across the dorsoventral axis (Fig. 2.6C). However, an appreciable trend in integrated mRNA was observed for the three strongest affinities (Fig. 2.6C). Further, plotting the total mRNA integrated across the entire dorsoventral axis of the embryo as a function of Patser score revealed that binding-site affinity (as reported by Patser score) is strongly correlated with transcriptional output in our single binding site enhancers (Fig. 2.6C, inset). In the case of this measure, there was also a threshold affinity: enhancers containing binding sites with affinities below that of DBS_5.13 showed no substantial differences in transcriptional activity (inset, Fig. 2.6C).

We used these constructs to measure mean transcriptional onset time as a function of Dorsal concentration and binding affinity, one of the key magnitudes predicted by our model (Fig. 2.2D). The measured mean onset time was relatively constant at ≈ 5 minutes across all Dorsal concentrations and enhancer constructs (Fig. 2.6D, dotted lines). This value is consistent with the measured onset times of other early embryonic genes such as the minimal *hunchback* promoter P2P (Garcia et al. 2013; Lucas et al. 2013; Eck et al. 2020).

We also determined that the fraction of active loci is highly sensitive to Dorsal concentrations and Dorsal binding-site affinity (Fig. 2.6D, dashed lines). The strongest Dorsal binding sites showed a large modulation of the fraction of active loci across Dorsal concentrations, while the weakest drove a relatively constant and low fraction of active loci across all Dorsal concentrations (Fig. 2.6D).

Our kinetic barrier model assumes that loci which fail to become active during the permissible transcription time window will remain inactive during the rest of the nuclear cycle (Fig. 2.2C). As a result, to determine whether the kinetic barrier model recapitulates the observations in Figure 2.6D, it was necessary to assign a value to this time window. We reasoned that the end of this time window determines the time point at which new transcription spots can no longer appear, possibly due to the onset of the next round of mitosis. To estimate the time point when nearly all spots have turned on, we calculated the 95th percentile of the observed spot onset times across all affinities: ≈ 7.1 min after the previous anaphase (Fig. 2.6E).

Using the measured time window of permissible transcription, we performed a simultaneous fit to the fraction of active loci and mean transcription onset times across all enhancers using the kinetic barrier model from Section 2.3.1 (Fig. 2.6D). Consistent with our model, we forced all enhancers to share the same value for c, and only letting the Dorsal dissociation constant, K_D, vary for each enhancer separately. By systematically exploring models with different numbers of OFF states n (Fig. S12, Fig. S13), we determined that a biochemical cascade with at least 3 to 4 rate-limiting OFF states is capable of capturing the qualitative behavior of our observations: a Dorsal concentration- and binding affinity-dependent fraction of active loci (dashed lines in Fig. 2.6D) and a mean transcription onset time that is mostly constant across Dorsal concentrations and affinities (dotted lines in Fig. 2.6D). Interestingly, alternative functional forms for k, such as

modeling this transition rate as depending linearly on Dorsal concentration, instead of depending on Dorsal DNA occupancy, resulted in worse fits to the fraction of active loci at saturating concentrations of Dorsal (Section S1.5; Fig. S5). Thus, our observations can be explained by a model in which Dorsal, through DNA binding, accelerates the promoter's transition through a sequence of kinetic barriers to a state of active transcription.
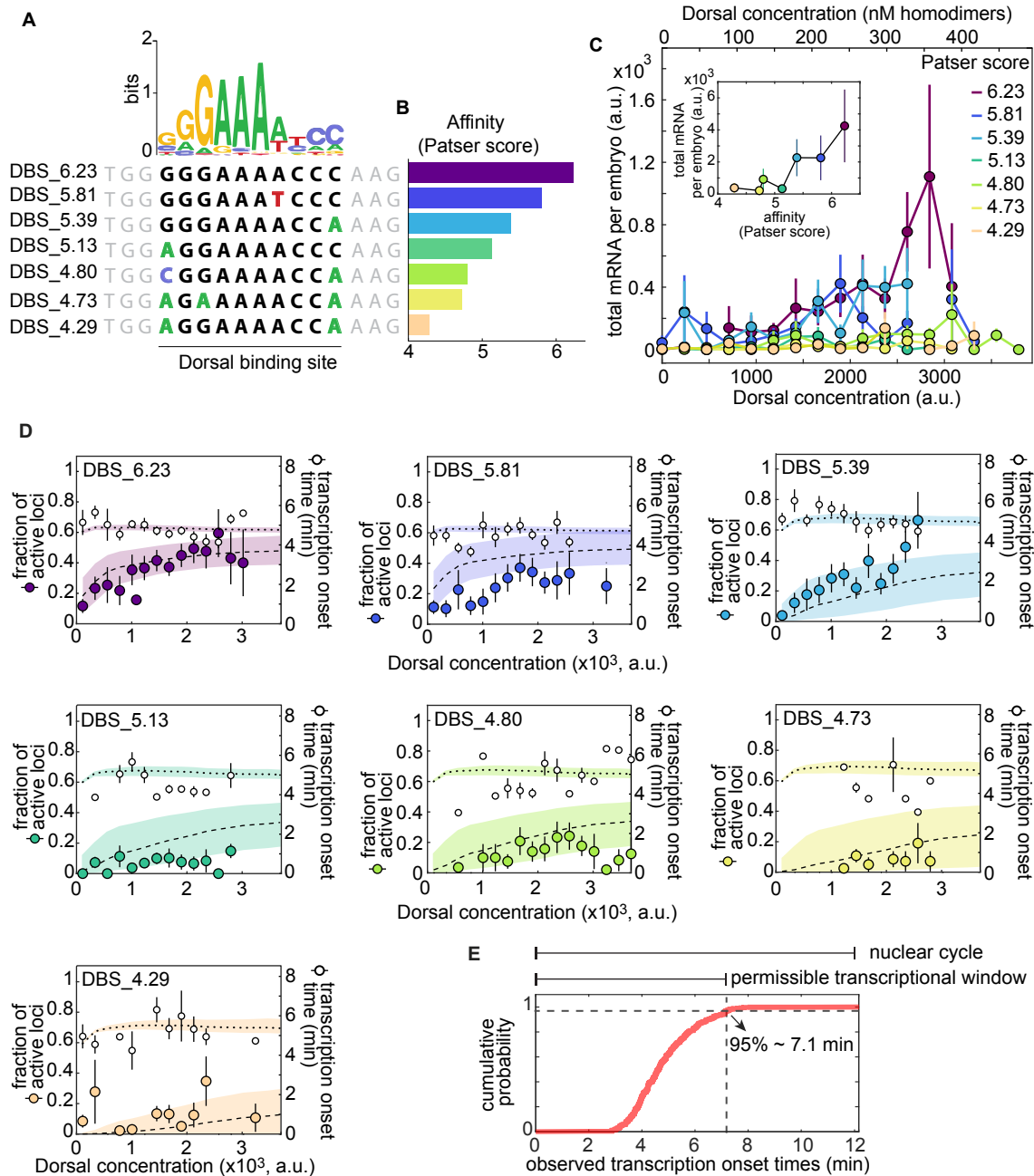
Figure 2.6: **A multi-step kinetic barrier model predicts the Dorsal-dependent fraction of active loci with constant mean transcriptional onset times. (A)** Top: Dorsal positional weight matrix logo from Ivan, Halfon, and Sinha 2008. Bottom: Sequence of the Dorsal binding sites engineered into minimal synthetic enhancers. Bold letters, 10 bp Dorsal motif. Black letters, consensus bases; colored letters, mutated bases; gray letters, sequence context. **(B)** Relative affinities of Dorsal binding sites estimated from the Patser algorithm using the Dorsal position weight matrix. Caption continues on next page.

Figure 2.5: **Continued from previous page: A multi-step kinetic barrier model explains the Dorsal-dependent fraction of active loci with constant mean transcriptional onset times. (C)** Overall transcriptional activity driven by the enhancers containing the binding sites in (A) measured as the total produced mRNA (fluorescence integrated over nuclear cycle 12) as a function of Dorsal concentration. Inset, mean total mRNA produced per embryo integrated across all Dorsal concentrations. Error bars, SEM over $N > 3$ embryos containing 3 or more nuclei belonging to that fluorescence bin. The top x-axis shows the estimated nuclear Dorsal concentration according to the calibration described in Figure S8. **(D)** Data and model fits for the fraction of active loci (left y-axis) and mean transcription onset time (right y-axis) for each enhancer. Empty black circles, experimentally observed mean transcription onset time; filled circles, experimentally observed mean fraction of active loci. Fitted curves are represented as dashed lines (fraction of active loci) and dotted lines (mean onset times), corresponding to predictions using median parameter values from the joint posterior distribution. Shaded areas, 95% credible interval (see Table S1 for inferred parameter values). Error bars, SEM over $N > 3$ embryos containing 3 or more nuclei belonging to that fluorescence bin. **(E)** Cumulative distribution of mean spot detection times per Dorsal fluorescence bin across all embryos and enhancers ($N = 344$ spots). Vertical dashed line, time at which 95% of spots have turned on (≈ 7.1 min) and end of the permissible transcription time window.

2.3.5 The experimentally measured RNAP loading rate are compatible with a thermodynamic binding model

As a next step in our theoretical dissection, we tested the performance of our theoretical model in explaining the rate of transcription after loci become active. Typically, in MS2 experiments, the loading rate is measured from the initial slope of spot fluorescence traces (Garcia et al. 2013; Eck et al. 2020; Liu et al. 2021). However, due to the weak expression driven by our enhancers, it was not possible to perform this analysis with confidence (Fig. S10). In lieu of directly measuring the transcription rate, we evaluated a related, more robust and readily observable quantity: the maximum trace fluorescence (Fig. 2.3E). We approximately relate the RNAP loading rate predicted by the simple activator model (Equation 2) to the maximum fluorescence by a constant factor (Appendix S1 .2), enabling direct comparison between theoretical predictions and experimental data.

Measurements of the maximum spot fluorescence over time as a function of Dorsal concentration for each of our seven minimal synthetic enhancers revealed that the maximum fluorescence is relatively constant across Dorsal concentration for most binding sites—particularly for the weakest of them, DBS_5.13, DBS_4.73, and DBS_4.23 (Fig. 2.6). However, the sparse and noisy nature of our data makes it challenging to draw confident conclusions from the fits, even for the stronger binding sites (i.e. DBS_6.23, DBS_5.81,

and DBS_5.39). In the case of the lower affinity binding sites, the constant maximum fluorescence suggests that the Dorsal concentration level in our embryos is far below the Dorsal dissociation constant K_D, even after increasing the Dorsal dosage by a factor of two as in our 2x Dorsal line. The effect of very low Dorsal concentrations relative to their respective K_D values can be clearly seen in Equation 2 and in Figure 2.2, where, for $[Dl]/K_D \ll 1$, the RNAP loading rate, R, adopts a basal level given by

$$R = R_{max} \frac{\frac{P}{K_P}}{1 + \frac{P}{K_P}} \tag{2.4}$$

that is independent of Dorsal concentration and binding affinity.

As shown on the right y-axes in Figure 2.6, this basal level corresponds to ≈ 20 RNAP molecules actively transcribing the gene ($\approx 15\%$ of the maximum number of RNAPs that can fit on the gene, as described in Section S1.3). For ease of visual comparison to the thermodynamic model predictions, we also plotted best-fit theoretical curves on top of the data using dashed curves (the insets in Fig. 2.6 show the same plots but zoomed into the measured data and plotted on a linear scale). These fits further underscore that our data do not explore a wide dynamic range with the precision necessary to determine the magnitude of K_D for each construct and to thoroughly test the thermodynamic model.
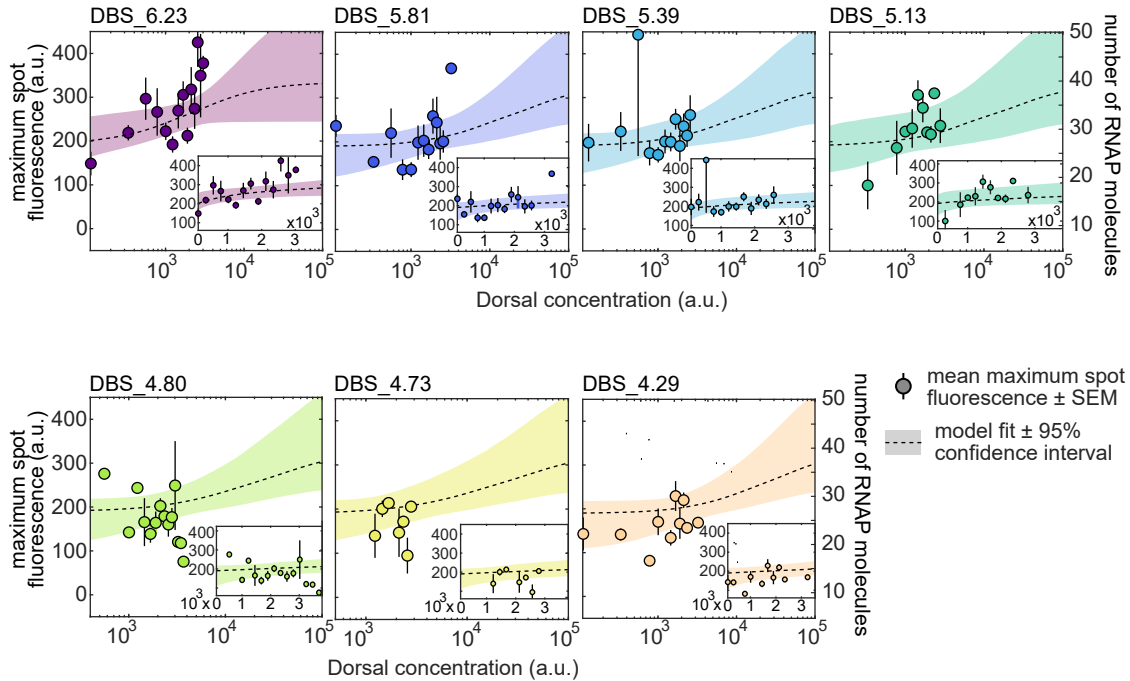
Figure 2.6: **Testing RNAP loading rate predictions of the thermodynamic model.** Mean maximum spot fluorescence as a function of Dorsal concentration for minimal synthetic enhancers with different affinities for Dorsal (filled circles). The right y-axis denotes the calibrated number of actively transcribing RNAP molecules (for details of calibration, see Section S1.3 and Fig. S2). Dashed curves correspond to a simultaneous Markov Chain Monte Carlo curve fit to all data using Equation 2. Fits share all parameters except K_D. Shaded areas, 95% prediction intervals. Insets, same data and fits plotted on a linear scale with axis ranges zoomed in on the data. See Table S2 for inferred parameter values. Error bars, SEM across N > 3 embryos containing 3 or more nuclei in a given fluorescence bin.

2.4 Discussion

A major obstacle to uncovering the mechanistic and quantitative underpinnings of enhancer action is the inherent complexity of endogenous regulatory sequences. Synthetic minimal enhancers are powerful alternatives to the complex experimental reality faced by modeling efforts in endogenous enhancers (Garcia, Brewster, and Phillips 2016; Garcia et al. 2020). Synthetic minimal enhancers contain binding sites for one or a handful of transcription factors, making them more amenable to theoretical dissection (Fakhouri et al. 2010; Sayal et al. 2016; Crocker, Tsai, and Stern 2017) and revealing the complex interplay among activators, repressors, and pioneer factors, as well as their contribution

to mRNA transcript accumulation (Fakhouri et al. 2010; Sayal et al. 2016; Crocker, Tsai, and Stern 2017). However, previous synthetic-based efforts to dissect enhancer function always involved fixed-embryo measurements, which cannot reveal the three inherently dynamical roles dictated by enhancer sequences (Fig. 2.1).

Here we augmented previous synthetic approaches by quantifying the real-time action of minimal enhancers with one binding site for the Dorsal activator in single cells of living, developing *Drosophila* embryos using the MS2 system. Contrary to theoretical speculations that single binding sites within eukaryotic genomes lack enough information to be recognized by transcription factors in the absence of other nearby binding sites (Wunderlich and Mirny 2009), we demonstrated that Dorsal can drive expression when bound to single binding sites (Fig. 2.3D). Additionally, we demonstrated that the fraction of active loci is a feature under regulatory control in our synthetic system (Fig. 2.3F; Fig. 2.5F), confirming the important role of this regulatory strategy in shaping the expression dynamics of endogenous enhancers (Garcia et al. 2013; Dufourt et al. 2018; Lammers et al. 2020; Harden, Vincent, and DePace 2021). Thus, while the signal driven by our minimal synthetic constructs is weak (Fig. 2.6), it can be quantified and recapitulates biologically relevant dynamic features of transcription that are also at play in endogenous enhancers.

It is important to note that the uncovering of a fraction of inactive loci in many reporter systems by us and others (Garcia et al. 2013; Dufourt et al. 2018; Lammers et al. 2020; Harden, Vincent, and DePace 2021) did not necessarily imply that this modulation of transcriptional engagement constitutes a biological control variable. Indeed, because live cell imaging techniques typically lack single-molecule resolution, it was unclear whether undetected loci in our study—and all previous studies—corresponded to a distinct population or were a detection artifact. By simultaneously labeling the locus with the transcription-independent reporter ParB-eGFP and nascent mRNA with MCP-mCherry (Fig. 2.5A), we demonstrated that a significant number of loci categorized as inactive do not constitute an experimental artifact and instead correspond to a distinct transcriptional state that is comparable to that measured in the absence of Dorsal protein (Fig. 2.5). In the future, conducting all live transcription measurements with DNA loci labeled by ParB could make it possible to confidently quantify the activity of all loci regardless of their activity.

Our minimal synthetic constructs and our validation of a distinct population of inactive loci enabled us to test an emerging theoretical model of enhancer action in development: a kinetic barrier model of transcriptional engagement (Fig. 2.2A; Fritzsch et al. 2018; Dufourt et al. 2018; Eck et al. 2020). Importantly, our model deviated from previous theoretical efforts that assumed that the transition rates between states preceding transcriptional engagement were either constant (Dufourt et al. 2018) or depended linearly on activator concentration (Eck et al. 2020). Instead, in order to account for the effects of Dorsal binding affinity on transcriptional dynamics, we assumed that this rate was proportional to Dorsal occupancy at its target DNA site. Thus, while the mechanisms underlying several aspects of this model, such as the molecular identity of the

various OFF states, remain unknown, this model can generate predictions for how the fraction of active loci and the transcriptional onset time are modulated by the Dorsal concentration and its binding affinity (Fig. 2.2C-E).

We systematically challenged this model by generating a small collection of minimal synthetic enhancers spanning a large range of affinities for Dorsal (Fig. 2.6A). Comparing the fraction of active loci and the transcription onset times of these enhancers revealed that the kinetic barrier model recapitulated our measurements (Fig. 2.6D). In past studies probing transcription dynamics in the *Drosophila* embryo (Dufourt et al. 2018; Eck et al. 2020), the pioneer factor Zelda was found to be largely responsible for ensuring constant transcription factor onset times and for determining the fraction of active loci. We cannot rule out the potential existence of distant or low-affinity Zelda binding sites (Rushlow and Shvartsman 2012) in our constructs. Alternatively, as it was recently demonstrated for the Bicoid activator Hannon, Blythe, and Wieschaus 2017, Dorsal could also have a pioneering activity. Indeed, the Dorsal homolog NF-κB has been recently shown to displace nucleosomes (Cheng et al. 2021). To further test the kinetic barrier model, it would be informative to directly perturb the temporal dynamics of nuclear Dorsal concentration to affect transcriptional engagement. For example, several optogenetics systems have been successfully deployed in the early fly embryo to inactivate transcription factors during discrete time widows (Huang et al. 2017; McDaniel et al. 2019; Irizarry et al. 2020). In the future, a version of one of these systems may dissect how the temporal dynamics of Dorsal concentration affect transcriptional activation.

Although the kinetic barrier model predicted the fraction of active loci and onset times (Fig. 2.6D) relatively well, we were unable to use our data to conclusively test the thermodynamic model's predictions of the rate of mRNA production (Fig. 2.6). Such limitation stemmed from the fact that only a fraction of loci display detectable transcription that can be used to quantify the mRNA production rate. Further, among these loci, the rate of transcription was found to be highly variable. As a result, our statistics were limited such that it was not possible to perform an unequivocal test of the thermodynamic model.

The apparent lack of substantial Dorsal concentration dependence observed in our measurements of RNAP loading rate could be explained in two possible ways. First, it is possible that there is a modulation of this rate in our measurements, but that this modulation is obscured by our experimental noise. Second, the Dorsal concentrations accessed by our experiment could be below the K_D of our binding sites. In this scenario, a modulation in the mRNA production rate would become apparent only at Dorsal concentrations higher than those attainable by our experimental system. While our embryos contained double the genetic dosage of Dorsal compared to wild type, perhaps 5-10 times the wild-type Dorsal concentration could be needed to exceed the K_D and modulate the rate of mRNA production. To express this high Dorsal concentration, which is certain to affect normal embryonic development, genetic approaches to increase Dorsal dosage in the embryos similar to those recently applied to flatten the Bicoid gradient might be necessary (Hannon, Blythe, and Wieschaus 2017).

It is important to note that, despite not seeing a modulation in the rate of mRNA production, we do see a significant change in the fraction of active loci as Dorsal concentration is varied (Fig. 2.6). This seeming contradiction could be explained through the presence of two dissociation constants in our model (Fig. 2.2): one dissociation constant for the first part of the model governing the onset of transcription, and a different dissociation constant for the second part of the model dictating the rate of RNAP loading once transcription has ensued. Interestingly, previous works quantifying transcriptional dynamics of a minimal Bicoid-activated *hunchback* P2 enhancers also hint at the existence of these two distinct dissociation constants (Garcia et al. 2013).

Further, this model is consistent with our surprising observation of a basal level of transcription in the presence of even extremely weak binding sites (Fig. 2.6) despite the lack of detected transcription in the absence of Dorsal protein (Fig. 2.3D, middle). This observation could be explained if Dorsal acted as both as a pioneer-like transcription factor triggering the onset of transcription, even at low concentrations relative to its K_D, and as an activator of the transcription rate at high concentrations.

Going forward, synthetic minimal enhancers could constitute the foundation for exploring the behavior of more complex regulatory regions. Independently inferring biophysical parameters such as Dorsal-DNA binding and dissociation constants could help constrain models of Dorsal participating in the activation of promoters with additional activators and repressors (Fakhouri et al. 2010; Sayal et al. 2016). Indeed, while Dorsal is the sole maternal nuclear-localized input specifying dorsoventral position in *Drosophila*, it rarely acts alone in endogenous enhancers (Hong, Hendrix, and Levine 2008). For example, the interaction of Dorsal with Twist is a classic example of positive cooperativity in development (Szymanski and Levine 1995). Dorsal can also act as a repressor depending on the presence of nearby Capicua binding sites (Shin and Hong 2014). The minimal synthetic enhancers presented here could be used as scaffolds for more complex minimal enhancers incorporating a second binding site for Twist or Capicua, for example.

In conclusion, we have developed a minimal synthetic enhancer system that has shed light on the fundamental assumptions about transcription in development. By engaging in a dialogue between theory and experiment, we have advanced our understanding of how kinetic processes give rise to important features of transcriptional dynamics in the embryo and made progress toward predictive understanding of how regulatory DNA sequence dictates the functional relation between input transcription factor dynamics and output transcriptional activity in development.

2.5   Methods and materials

2.5.1   Plasmids and reporter design

To design our minimal construct (Fig. 2.3), we placed the 10 bp consensus Dorsal binding site (Markstein et al. 2002) upstream of the *even-skipped* core promoter. This enhancer-

promoter construct drives the expression of the MS2v5 sequence containing 24 non-repetitive MS2 loops (Tutucci et al. 2018) followed by the *lacZ* coding sequence and the *tubulin* 3'UTR. (Garcia et al. 2013).

In addition to the consensus Dorsal binding site (DBS_6.23), we created six enhancers of varying strength by introducing point mutations to the consensus Dorsal binding motif. Some of these binding sites were taken from known validated Dorsal motifs (Markstein et al. 2002), while others were generated based on mutations known to decrease Dorsal binding (Ip et al. 1992; Jiang et al. 1991). To guide the design of these binding sites, we used an already existing position weight matrix computed with the MEME algorithm (Ivan, Halfon, and Sinha 2008; Bailey et al. 2006) using motifs generated by DNAse I footprinting assays (Bergman, Carlson, and Celniker 2005) and quantified the information content of each base pair using Patser (Hertz and Stormo 1999).

All plasmid sequences used in this study are shown in Table 2.1 and can be accessed from a public Benchling folder. Injections were carried out by Rainbow inc. or Bestgene inc.

2.5.2   Flies

Reporter plasmids were injected into BDSC fly line 27388 containing a landing site in position 38F1. Transgene orientation was confirmed by PCR using primers 18.8 (ggaac-gaaggcagttagttgt) and Ori-Seq-F1 (tagttccagtgaaatccaagcattttc) binding outside of the 5' 38F1 *attP* site and the *even-skipped* promoter, respectively. All reporter lines were confirmed to be in the same orientation. All flies used in this study can be found in Table 2.2.

To generate the embryos used in the experiments shown in all figures except for Figure 2.5, we crossed 2x Dorsal or 1x Dorsal virgins to males carrying synthetic enhancers. The genotype of 2x Dorsal flies is *yw;Dl-mVenus (CRISPR), MCP-mCherry; Dorsal-mVenus, MCP-mCherry, His2Av-iRFP*. The genotype of 1x Dorsal flies is *yw;dl[1], MCP-mCherry; Dorsal-mVenus, MCP-mCherry, His2Av-iRFP*. Because there does not seem to be a difference in transcriptional activity between the CRISPR knock-in and the transgene Dorsal-mVenus alleles (Fig. S9), we combined the 1x Dorsal and 2x Dorsal data for some enhancers.

MCP-mCherry and His-iRFP were described before by (Liu et al. 2021). The Dorsal-mVenus transgene was developed by Reeves et al. 2012b.

To generate the Dorsal-Venus knock-in allele we used the CRISPR/Cas9 protocol described by (Gratz et al. 2015). We generated a donor plasmid containing the mVenus sequence followed by a stop codon and a 3xP3-dsRed marker flanked by PiggyBac recombinase sites. This insert was flanked by two ≈1 kbp homology arms matching ≈2 kbp surrounding the Dorsal stop codon (plasmid Dl-mVenus-dsRed in Table 2.1). The Cas9 expressing BDSC line 51324 was injected with the donor plasmid in combination with a plasmid carrying a sgRNA targeting the sequence GTTGTGAAAAAGGTATTACG in the C-terminus of Dorsal (plasmid pU6-DlgRNA1 in in Table 2.1). Survivors were crossed to

yw and the progeny was screened for dsRed eye fluorescence. Several independent lines were established and tested for rescue. The insertion was confirmed by PCR using primers flanking the homology arms OutLHA (ccattaaaacggaaccaagaggtgag) and OutDl-RHA (tctaacaatggctcgatttttgcca). The dsRed eye marker cassette was flipped out of rescuing lines via crossing with a piggyBac recombinase line. The resulting Dorsal-mVenus locus was then resequenced using the same primers.

The data shown in Figure 2.5 were obtained from embryos laid by *yw;ParB2-eGFP, eNosx2-MCP-mCherry;+* (wild-type Dorsal mothers) or *yw;ParB2-eGFP, eNosx2-MCP-mCherry, dl[1];+* (Dorsal null mothers).

2.5.3 Microscopy

Fly cages were allowed to lay for 90 to 120 minutes prior to embryo collection. Embryos were then mounted on microscopy slides in Halocarbon 27 oil (Sigma-Aldrich, H8773) in between a coverslip and breathable membrane as described in (Garcia et al. 2013; Bothma et al. 2014; Garcia and Gregor 2018).

Confocal microscopy was performed on a Leica SP8 with HyD detectors and a White Light Laser. We used a 63x oil objective, and scanned bidirectionally with a scan rate of 420 Hz and a magnification of 3.4x zoom. We did not use line or frame accumulation. Time-lapse z-stacks were collected with $\sim$10 s frame rate and 106 nm x-y pixel dimensions and 0.5 μm separation between z-slices (7 μM range, 16 slices). x-y resolution was 512x512 pixels. Pinhole was set to 1.0 Airy units at 600 nm. mVenus was excited by a 510 nm laser line calibrated to 5 μW using the 10x objective and detected in a 520-567 nm spectral window. mCherry was excited by a 585 nm laser line calibrated to 25 μW and detected in a 597-660 nm spectral window. To image His2av-iRFP, the 700 nm laser line was set to 10% and detected in a 700-799 nm spectral window. In all channels, detection was performed using the counting mode of the HyD detectors.

All movies were taken at $\sim$50% along the anterior-posterior axis of the embryo.

2.5.4 ParB experiment fly crosses and microscopy

We created flies with and without functional Dorsal expressing ParB2-eGFP maternally driven by the *nanos* promoter and MCP-mCherry driven by two copies of a minimal *nanos* enhancer to label our locus DNA and nascent mRNA, respectively. In addition, we added a parS sequence followed by a 400 bp spacer (created with SiteOut, Estrada et al. 2016) to our DBS_6.23 enhancer. We then crossed male flies containing parS-DBS_6.23-MS2 to *yw; ParB2-eGFP; eNosx2-MCP-mCherry; +* females to create embryos that have our locus of interest labeled with eGFP colocalized with transcriptional loci in the MCP-mCherry channel (Fig. 2.5A and B). After mounting embryos using the protocol described above in Section 5.4, we used the sequential scanning mode on the Leica SP8 confocal microscope to eliminate bleedthrough from eGFP into the mCherry channel, and imaged at

approximately 20 s per stack, half the rate used in other imaging experiments in this study.

2.5.5   Image and time-series analysis

Image analysis was performed in Matlab using the custom pipeline described in Garcia et al. 2013 and Lammers et al. 2020 (this pipeline can be found in the mRNA Dynamics Github repository). Image segmentation was also aided by the Trainable Weka Segmentation plugin in FIJI (Witten et al. 2016; Arganda-Carreras et al. 2017). Further analysis of time-series and other data were likewise performed in Matlab. Movies for publication were made in FIJI (Schneider, Rasband, and Eliceiri 2012; Schindelin et al. 2012).

2.5.6   Measuring Dorsal-mVenus concentration

Dorsal-mVenus concentration was calculated as in (Fig. S6). As shown in the figure, we measured the average mVenus fluorescence intensity in a circle of 2 μm radius at the center of the nucleus in every z-slice of each nucleus. This results in a z-profile of fluorescence values covering the nucleus itself and the cytoplasm below and above it. The reported concentration corresponds to the value at the middle z-plane of each nucleus. To find this plane, we fit a parabola to the fluorescence z-profile. We use as the nuclear concentration the fluorescence value at the plane corresponding to the fitted parabola's vertex (Fig. S6B). We then plotted this value over time and selected a single time point for each trace corresponding to the middle of each nucleus's observed trajectory (Fig. S6B). To determine the background fluorescence in the mVenus channel we imaged flies with the same genotype as 2x Dorsal except for the Dorsal-Venus fusions. We calculated the average nuclear fluorescence in the mVenus channel across nuclear cycle 12 and subtracted this value from our Dorsal-Venus measurements.

2.5.7   Curve fitting and parameter inference

Curve fitting and parameter inference were performed in Matlab using the MCMCSTAT Matlab package using the DRAM Markov Chain Monte Carlo algorithm (Haario et al. 2006). For simplicity, uniform priors were assumed throughout.

2.6   Acknowledgments

2.7 Biological material

Plasmids	
Name (hyperlinked to Benchling)	**Function**
pIB-1Dg-evePr-MS2v5-LacZ-Tub3UTR	DBS_6.23-MS2 reporter
pIB-1DgS-MS2v5-LacZ-Tub3UTR	DBS_5.81-MS2 reporter
pIB-1DgW-MS2v5-LacZ-Tub3UTR	DBS_5.39-MS2 reporter
pIB-1DgAW-MS2v5-LacZ-Tub3UTR	DBS_5.13-MS2 reporter
pIB-1DgSVW-MS2v5-LacZ-Tub3UTR	DBS_4.8-MS2 reporter
pIB-1DgVVW-MS2v5-LacZ-Tub3UTR	DBS_4.73-MS2 reporter
pIB-1DgVW-MS2v5-LacZ-Tub3UTR	DBS_4.29-MS2 reporter
pIB-2xIntB2-Neutral400-1Dg-MS2v5-LacZ-Tub3UTR	DBS_6.23-MS2 reporter with two ParB2 binding sites (note that 2xIntB2 is termed a parS sequence in the main text)
Dl-mVenus-dsRed	Donor plasmid for Dorsal-mVenus CRISPR knock-in fusion
pU6-DlgRNA1	Synthetic guide RNA for Dorsal-mVenus CRISPR knock-in fusion
pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR	Maternally deposited MCP-mCherry
pCasper4-His2Av-iRFP	Histone2Av fusion to infrared RFP (His-iRFP)
pCasper4-Pnos-NoNLS-MCP-mCherry-TUB3'UTR	Maternally deposited MCP-mCherry
pCasper-pNos-NoNLS-ParB2-GFP-TUB3'UTR	ParB-eGFP

Table 2.1: List of plasmids used to create the transgenic fly lines used in this study.

Fly lines	
Genotype	**Usage**
yw; ParB2-eGFP; eNosx2-MCP-mCherry; +	Label reporter DNA and nascent RNA
yw; Dorsal-mVenus, pNos-MCP-mCherry; pNos-MCP-mCherry, His2Av-iRFP	Females to visualize Dorsal protein, label nascent RNA, label nuclei
yw; Dorsal-mVenus, pNos-MCP-mCherry; Dorsal-mVenus, pNos-MCP-mCherry, His2Av-iRFP	Females to visualize Dorsal protein, label nascent RNA, label nuclei
yw; dl^1, pNos-MCP-mCherry; pNos-MCP-mCherry, His2Av-iRFP	Females to label nascent mRNA and label nuclei in embryos lacking Dorsal protein
yw; 1Dg(11) ; +	Males carrying the DBS_6.23-MS2 reporter
yw; 1DS(2) ; +	Males carrying the DBS_5.81-MS2 reporter
yw; 1DgW(2) ; +	Males carrying the DBS_5.39-MS2 reporter
yw; 1DgAW(3) ; +	Males carrying the DBS_5.13-MS2 reporter
yw; 1DgSVW(2) ; +	Males carrying the DBS_4.8-MS2 reporter
yw; 1DgVVW(3) ; +	Males carrying the DBS_4.73-MS2 reporter
yw; 1DgVW) ; +	Males carrying the DBS_4.29-MS2 reporter
yw; 2xIntB2-1Dg(4)(5)(6) ; +	Males carrying the DBS_6.23-MS2 reporter with two ParB2 binding sites (note that 2xIntB2 is termed *parS* in the main text and in figures)

Table 2.2: List of fly lines used in this study and their experimental usage

Supplementary information

S1 Appendix

S1.1 Calculating the fraction of active loci and the transcriptional onset time by solving the kinetic barrier model

We describe here in detail the method we used to solve kinetic barrier model presented in Section 2.3.1 and Figure 2.2A. The problem posed in Figure 2.6A, namely the time evolution of the probability of nuclei occupying a discrete number of consecutive states, can be described by the following system of linear differential equations (also known as the 'master equation')

$$\frac{d\overrightarrow{P}}{dt} = \boldsymbol{K}(t)\overrightarrow{P}, \tag{S1}$$

where $\overrightarrow{P}$ is a column vector containing the probability as a function of time of each of the states that the system can be in. $\boldsymbol{K}$ corresponds to the transition rate matrix containing the rates that dictate the passage from each OFF state to the next and to the final ON state.

For n OFF states followed by a ON state connected by irreversible transitions with a rate of $k(t)$, Equation S1 can be written as

$$\begin{bmatrix} \frac{dP(OFF_1,t)}{dt} \\ \frac{dP(OFF_2,t)}{dt} \\ ... \\ \frac{dP(OFF_n,t))}{dt} \\ \frac{dP(ON,t)}{dt} \end{bmatrix} = \begin{bmatrix} -k(t) & 0 & ... & 0 & 0 \\ k(t) & -k(t) & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & -k(t) & 0 \\ 0 & 0 & ... & k(t) & 0 \end{bmatrix} \times \begin{bmatrix} P(OFF_1,t) \\ P(OFF_2,t) \\ ... \\ P(OFF_n,t) \\ P(ON,t) \end{bmatrix}, \tag{S2}$$

where $P(s,t)$ indicates the probability of the system being in state s at time t.

As described in Section 2.3.4, the transition rate matrix, $\boldsymbol{K}$, is a function of time as a consequence of the assumption that the transition rate between states, k, depends on the time-varying Dorsal concentration. In our model, k is given by

$$k(t) = c \cdot \frac{\frac{[Dl](t)}{K_D}}{1 + \frac{[Dl](t)}{K_D}}, \tag{S3}$$

where K_D is the Dorsal binding dissociation constant and c is a rate constant. If k were a constant, then the system of equations describing transcriptional dynamics could be solved analytically. However, because $k(t)$ depends on the empirical Dorsal-mVenus fluorescence dynamics, which does not have a concrete functional form, solving the system in Equation S2 becomes analytically intractable. Thus, in order to obtain the probability of each state as a function of time, $\overrightarrow{P}$, and calculate the fraction of active loci and the mean transcription onset times, we solve the system in Equation S2 numerically

for a given number of n OFF states. Specifically, at each time step dt, we calculated how the probability of each state changes with respect to the previous time step.

To calculate $P(s, t)$ we need to consider the previous time step $t - 1$ and take into account three possible scenarios:

1. Loci that were already in state s at time $t - 1$ and stay in this state at time t.

2. Loci that were in state $s - 1$ at $t - 1$ that transition into state s at time t.

3. Loci that were in state s at time $t - 1$ that leave this state by transitioning to the next state $s + 1$ at time t.

The likelihood of a locus jumping from one state to the next at time t during an arbitrarily small time window of dt is given by the transition rate $k(t) \times dt$. As a result, the probability of the promoter locus being in state s at time t can be calculated as

$$\underbrace{P(s,t)}_{\substack{\text{Probability of} \\ \text{state } s \text{ at time } t}} = \underbrace{P(s, t - 1)}_{\text{being in state } s \text{ at } t - 1} + \underbrace{k(t)dtP(s - 1, t - 1)}_{\text{enter from state } s - 1} - \underbrace{k(t)dtP(s, t - 1)}_{\text{leave for state } s + 1} .$$

(S4)

It is clear that, for $s = 1$, $P(s - 1, t - 1) = 0$, since there is not a previous state from which loci can enter the first OFF state. Similarly, since promoters cannot leave the final ON state once they have entered it, $P(n + 2, t - 1) = 0$ for n OFF states.

To obtain the fraction of active loci, we initialize the system to $P(s = 1, t = 0) = 1$ and calculate $P(s = n + 1, t = T/dt)$, where T is the duration of the transcriptional window such that

$$\text{Fraction of active loci} = P(n + 1, T/dt). \tag{S5}$$

To obtain the mean transcriptional onset time, we calculate the expected value $\mathbb{E}[onset]$ of the time to reach the final $n + 1$ state before the end of the transcriptional time window at $t = T$. From the definition of expected value,

$$\mathbb{E}[onset] = \sum_{i=1}^{T} t_i \times p_i, \tag{S6}$$

where t_i indicates a given onset time and p_i the probability of loci having that specific onset time. Note that the sum only runs until the end of the transcription time window T, as loci that will remain inactive for the duration of the nuclear cycle should not be considered in our calculation of the mean transcriptional onset time. This means that p_i is a normalized probability, calculated only amongst loci that turn on before time T such that

$$\sum_{i=1}^{T} p_i = 1. \tag{S7}$$

In terms of the system described in Equation S4, the probability p_i of loci reaching the ON state n+1 at time t_i is

$$\begin{array}{c}\text{Probability of loci}\\\text{to turn on at time } t_i\end{array} = P(n+1, t_i) - P(n+1, t_i - 1). \qquad (S8)$$

And the normalized probability p_i of loci reaching the ON state n+1 at time t_i among loci that reach it before T is

$$\begin{array}{c}\text{Probability of loci to turn on}\\\text{at time } t_i \text{ (normalized)}\end{array} = p_i = \frac{P(n+1, t_i) - P(n+1, t_i - 1)}{\sum_{i=1}^{T}\left[P(n+1, t_i) - P(n+1, t_i - 1)\right]}. \qquad (S9)$$

Replacing p_i in Equation S6 with its definition in Equation S9, we arrive at the formula for the mean transcriptional onset time

$$\begin{array}{c}\text{Mean transcriptional}\\\text{onset time}\end{array} = \mathbb{E}[onset] = \sum_{i=0}^{T} t_i \frac{P(n+1, t_i) - P(n+1, t_i - 1)}{\sum_{i=1}^{T}\left[P(n+1, t_i) - P(n+1, t_i - 1)\right]}. \qquad (S10)$$

Note that the solutions for the fraction of active loci (Eqn. S5) and their mean transcription onset time (Eqn. S9) ultimately depend on the Dorsal concentration over time $[Dl](t)$ as they determine $P(t, n)$. Hence, to generate predictions that can be directly compared to our live-imaging measurements, we need to solve these equations accounting for the Dorsal-mVenus fluorescence dynamics that determine $[Dl](t)$.

S1.2   Relating MS2 signal to the statistical mechanical model

In order to understand how the maximum MCP-mCherry fluorescence of a locus relates to the average RNAP loading rate, a model for the fluorescence trajectory during a nuclear cycle is required. We start by assuming that RNAP molecules begin loading at a time t_0 into the nuclear cycle and continue to load at a constant rate proportional to R, as shown in Equation 1 ($R = R_{max} \cdot p_{bound}$) and step (1) in Figure S1. The observed signal increases linearly until the first polymerase terminates transcription. At this point, the signal plateaus at the value f_{max} because polymerase molecules continue to be loaded onto the gene at a constant rate while simultaneously terminating at the same rate at the end of the gene (Fig. S1, step (2)). We note that, in this model, initiation halts at step (3), leading to a decrease in fluorescence as elongating polymerases finish transcribing. Note that this step is not accounted for in any analyses or models in this study.

Given this model, the maximum fluorescence observed in a trace is given by

$$f_{max} \approx \alpha \cdot R \cdot \Delta t_{elongation}, \qquad (S11)$$

, where R is the loading defined in Equation 1, and α is the instantaneous fluorescence per mRNA molecule that we estimate in Section S1.3. As a result, the maximum fluorescence is proportional to the loading rate, namely

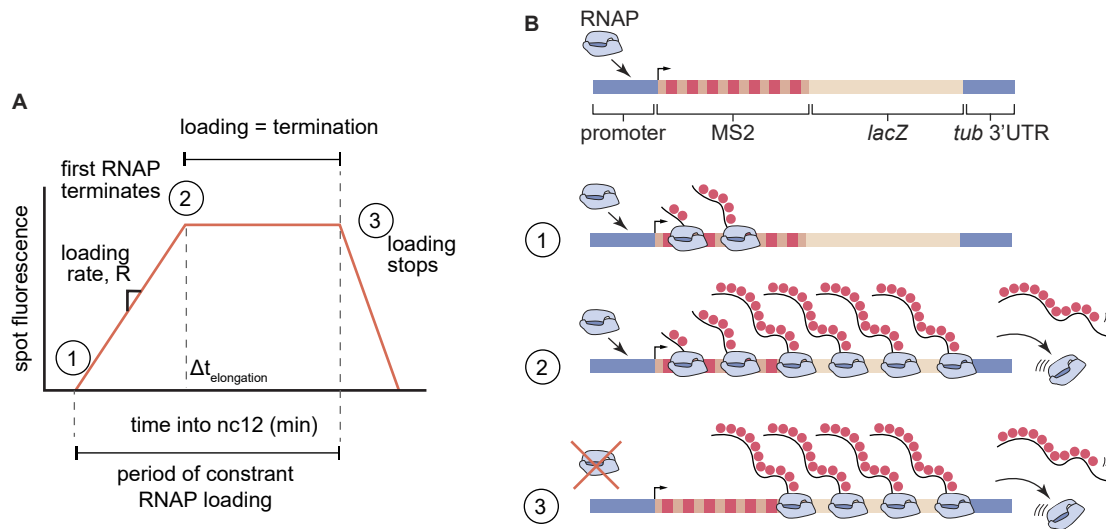$$f_{max} \propto \alpha \cdot R. \qquad (S12)$$

Figure S1: **Trapezoid model of transcription dynamics during early embryonic nuclear cycles in *Drosophila*. (A)** Depiction of a piece-wise linear approximation to average measured fluorescence of loci as a function of time during nuclear cycle 12. In step (1), RNAP molecules are loaded on to the gene at an average constant rate, R. After the first RNAP terminates transcription at time $\Delta t_{elongation}$, initiation and termination balance each other out, leading to a constant fluorescence value (step (2)). In step (3), initiation ends, causing the observed fluorescence to monotonically decrease. **(B)** Schematic of the RNAP loading behavior at each step in (A).

Thus, we now have an expression for f_{max} that enables us to relate our measurements to the thermodynamic model's prediction for R, the RNAP loading rate (Fig. 2.2E).

S1.3   MS2 Calibration

To estimate the fluorescence detection threshold in our system, we calibrated the MCP-mCherry signal to single molecule fluorescence *in situ* hybridization (smFISH) data from Garcia et al. 2013. This calibration is based on the fact that, to produce one mRNA molecule, RNAP has to spend a defined amount of time on the reporter thus contributing to the integrated spot fluorescence. We define α as the fluorescence of one RNAP molecule bearing a labeled nascent RNA and $\Delta t_{elongation}$ as the time RNAP spends on the reporter gene to synthesize one mRNA molecule (Fig. S2A). Then, the integrated spot fluorescence corresponding to the production of one mRNA molecule, β, is

$$\beta(a.u. \times min \times molecule^{-1}) = \alpha(a.u. \times molecule^{-1}) \times \Delta t_{elongation}(min). \qquad \text{(S13)}$$

From the definition of β above, it follows that the integrated fluorescence of a spot over time corresponds to the total number of mRNA molecules produced by that locus in that period (Fig. S2A). Using smFISH, Garcia et al. 2013 measured the mean number of mRNA molecules produced per nucleus by a P2P-MS2 reporter transgene during nuclear cycle 13 as a function of anterior-posterior position (Fig. S2B). To compare these data with the measurements obtained from our imaging setup, we imaged the same reporter using 2x Dorsal flies and calculated the mean integrated spot fluorescence across all nuclei as a function of position along the anterior-posterior axis (Fig. S2B). We plotted these two measurements against each other and fitted the data to a line going through the origin (Fig. S2C). The slope of this line indicates β, the integrated spot fluorescence corresponding to a single produced mRNA molecule.

With this fluorescence calibration factor in hand, we can now estimate α, the spot fluorescence corresponding to a single RNAP molecule attached to one nascent mRNA molecule with 24 MS2 loops. We can estimate $\Delta t_{elongation}$ by invoking the elongation rate of RNAP in the fly embryo, v_{elon}, and the length of our reporter, L, such that

$$\Delta t_{elongation} = \frac{L}{v_{elon}}. \qquad \text{(S14)}$$

Using this expression for $\Delta t_{elongation}$, we can solve for α in Equation S13 to obtain the fluorescence of a single RNAP molecule given by

$$\alpha = \frac{\beta \times v_{elon}}{L}. \qquad \text{(S15)}$$

We next replace L by the length of our reporter transgene, 5.2 kbp. In addition we replace v_{elon} by a previously experimentally measured value of 1.5 ± 0.14 kbp/min (Garcia et al. 2013), and β by the calibration factor shown in Figure S2C. We then arrive at

$$\alpha = \frac{30.3\frac{a.u.\cdot min}{RNAP} \times 1.5\ \frac{kbp}{min}}{5.2\ kbp} = 8.837 \text{ a.u. per molecule}. \qquad \text{(S16)}$$

Note that v_{elon} and β have an associated error that leads to uncertainty in the calculation of α. Propagating these errors results in an uncertainty of 0.046 a.u. per RNAP, or approximately 14%. This uncertainty should be viewed as an underestimate since, for example, we are not accounting for embryo-to-embryo variability in the accumulated mRNA measured by microscopy or smFISH.

Using this calibration factor, we can now determine the detection threshold of our experimental setup in terms of absolute number of RNAP molecules. One way of determining this threshold is by comparing the mean fluorescence of the dimmest spots with the magnitude of their corresponding background fluctuations. If these values overlap, then it is not possible to determine with certainty whether a spot correspond to actual signal or to background. This approach reveals a detection threshold of ≈ 80 a.u. or ≈ 9 RNAP molecules (Fig. S2D). A second strategy to determining the detection threshold is looking at the fluorescence of the dimmest detected spots. Their average fluorescence indicates the value under which no reliable detection is possible. This analysis reveals a detection limit of ≈ 54 a.u. or ≈ 6 RNAP molecules (Fig. S2E). These values for our detection limit using MCP-mCherry are on the order of twice the limit determined for similar experiments that used MCP-eGFP or PCP-eGFP (Garcia et al. 2013; Alamos et al. 2020), most likely due to mCherry being a dimmer fluorophore than eGFP (Lambert 2019).

Finally, in the main text (Section 2.3.5), we estimated the maximum fluorescence corresponding to the basal level of RNAP molecules on our reporter constructs (Section 2.3.5). We include here details of the calculation. Since the length of the coding region of our reporter constructs is 5.2 kbp, and the footprint of RNA Polymerase II is 40 bp (Selby et al. 1997), 130 RNAPs can fit on the gene at any given time. Since we estimate the maximum fluorescence corresponding to basal transcription to be ≈ 20 RNAP molecules (Section 2.3.5), the reporter is 20/130$\approx$ 15% saturated by RNAPs.
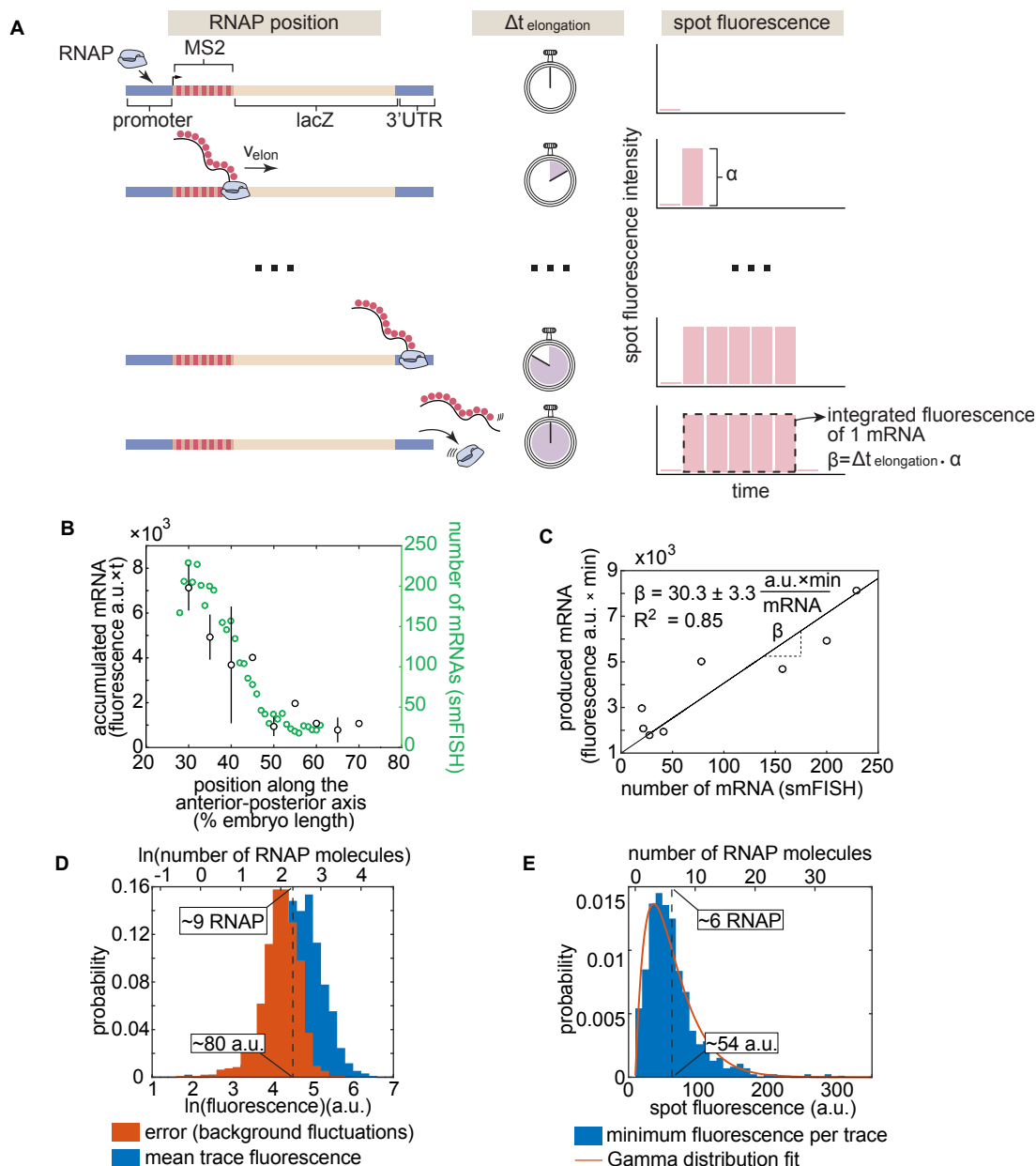
Figure S2: **Absolute calibration of MS2 using single molecule FISH. (A)** Schematic showing that the integrated spot fluorescence corresponding to the production of one mRNA, β, is equal to the fluorescence of a single RNAP molecule, α, multiplied by the time it spends on the gene, $\Delta t_{elongation}$. **(B)** Mean accumulated mRNA per nucleus (in nuclear cycle 13) based on the integrated MS2 fluorescence of P2P-MS2 employing the imaging conditions used for our reporter data (N = 6 embryos) compared to the number of mRNA molecules per nucleus produced in nuclear cycle 13 as reported by single molecule FISH by Garcia et al. 2013. Caption continues on next page.

Figure S3: **Continued from previous page: Absolute calibration of MS2 using single molecule FISH. (C)** Scatter plot showing data from (B) corresponding to the same anterior-posterior bin. The solid line shows the best linear fit to all data points. The slope error corresponds to the standard error of the fit. The error in the fluorescence per RNAP is the propagated standard error taking the errors in elongation rate and calibration slope into account as described in this section's text. **(D)** Histograms of mean trace fluorescence in all particles across all experiments and the error in the fluorescence of these particles as reported by fluctuations in the fluorescence background. Because the spot fluorescence was obtained by integrating over three slices, the corresponding error was propagated by multiplying the error from one slice (using the method described in (Garcia et al. 2013)) by $\sqrt{3}$. The dashed line indicates the center of where the two distributions overlap, suggesting a detection limit of approximately 9 RNAP molecules. **(E)** Histogram of the minimum spot fluorescence per trace across all experiments. The dashed line indicates the mean of the distribution, suggesting a detection limit of approximately 6 RNAP molecules. Note that in (D) and (E) the top x-axis is expressed in terms of absolute number of RNAP molecules using the calibration from (C). A best fit to a Gamma distribution is shown in red for ease of visualization.

S1.4 Measuring transcriptional onset times

We measured the time at which each locus turns on by determining the first time point where a spot was detected. To make this possible, we needed a reliable way to estimate t = 0 which corresponded to the beginning of the nuclear cycle.

Typically, fluorescently labeled histone is used to determine the timing of anaphase (Garcia et al. 2013). However, only a small fraction of our embryos had measurable levels of visible Histone-iRFP, most likely due to embryo-to-embryo variability and the low density of DNA in the nucleus in nuclear cycle 12 (compared to later nuclear cycles when His-iRFP is more visible). When the Histone-iRFP signal was insufficient to determine anaphase, we relied on the Dorsal-mVenus channel. As we describe below, just like Histone-iRFP, the nuclear Dorsal fluorescence also shows a characteristic pattern during mitosis.

To precisely determine which features of the Dorsal-mVenus channel to use for mitosis timing, we imaged Dorsal-mVenus and Histone-RFP—which, as opposed to Histone-iRFP, can be consistently detected—simultaneously (Fig.S4). This exercise showed that the edges of nuclei become less well defined as they enter mitosis and then elongate at the beginning of anaphase (Fig. S4). In this way, we could identify precise anaphase frames in movies with no visible Histone-iRFP. Despite using this method, we still estimate that there may be a 2-3 frame error (i.e. 20-30 s) in our determination of anaphase. Thus, this error is $< 20\%$ of the measured period of transcriptional activity within nuclear cycle 12 ($\sim$ 3 min).
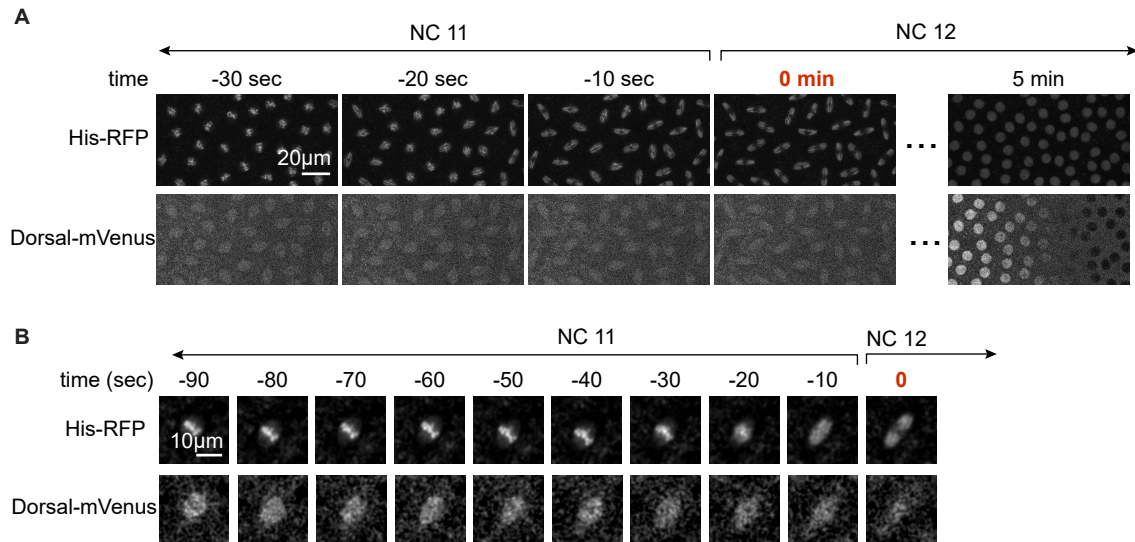
Figure S4: **Using the Dorsal-mVenus channel to determine the timing of mitosis**. **(A)** Visual comparison of nuclei in the field of view of Histone-RFP and Dorsal-mVenus channels during nuclear division. **(B)** Same as (A), but zoomed into a single nucleus. In (A) and (B), t = 0 min in red text corresponds to anaphase.

S1.5 Kinetic barrier fits with a different functional form of the transition rate k

In the main text, we hypothesize that the transition rate between OFF states and between the last OFF state and the ON state is proportional to Dorsal occupancy (Eqn. S3). Here, we show that another functional form for k in the kinetic barrier model can only partially recapitulate the fraction of active loci and transcriptional onset times for each of our enhancers. This functional form is motivated by the idea that Dorsal could catalyze a change in the promoter (e.g. opening of chromatin) in a manner dependent on the speed of its first occurrence of binding rather than its equilibrium occupancy. Specifically, inspired by (Eck et al. 2020), we posit that

$$k = c \cdot [Dl]. \tag{S17}$$

In this alternate model, we assume that the Dorsal binding site affinity dependence is wrapped up into the c parameter. Thus, we fit each enhancer using a distinct value of c. As can be seen in Figure S5, this alternate model cannot fit the data as well as when k is assumed to be proportional to the Dorsal occupancy as described in the main text and in Figure 2.6. Specifically, this functional form is less capable of recapitulating the saturation plateau of the fraction of active loci at high Dorsal concentrations.
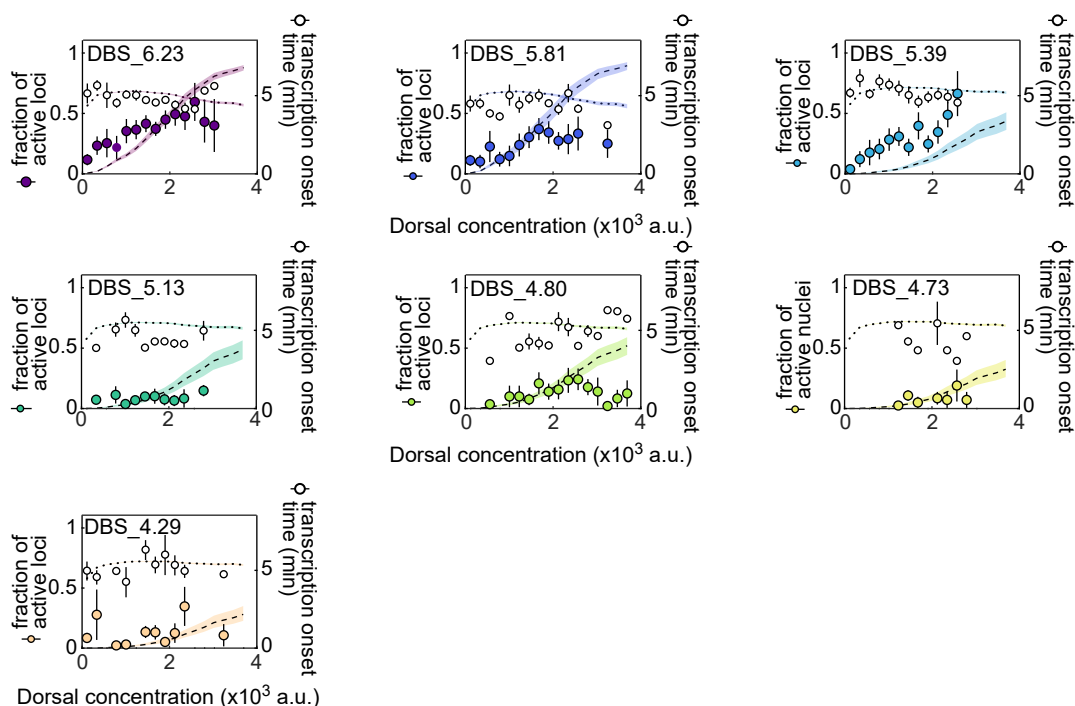
Figure S5: **Fits to kinetic barrier model using** $\mathrm{k} = \mathrm{c} \cdot [\mathrm{Dl}]$. Data and model fits for the fraction of active loci (left y-axis) and mean transcription onset time (right y-axis) for each enhancer. Empty black circles correspond to the experimentally observed mean transcription onset time. Filled colored circles correspond to experimentally observed mean fraction of active loci. Error bars on observations correspond to the standard error of the mean. Fitted curves are represented as black dashed lines (fraction of active loci) and black dotted lines (mean transcription onset times), which correspond to predictions using median parameter values from the joint posterior distribution. Colored shaded areas indicate the 25%-75% credible interval.

S2 Supplementary figures
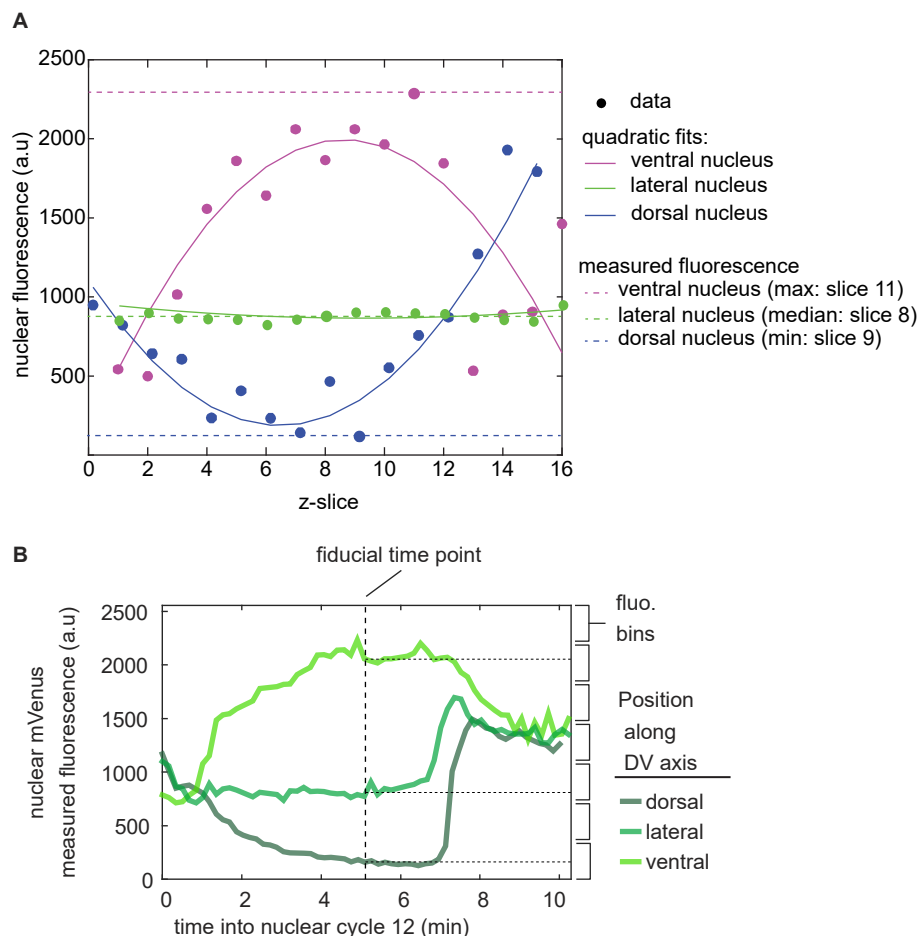
Figure S6: **Measuring Dorsal-mVenus nuclear fluorescence across the dorsoventral axis. (A)** In each frame, the Dorsal-mVenus fluorescence is measured in each z-slice across nuclei. This creates a series of fluorescence values as a function of z-slice (filled circles). z-slices at the top and the bottom correspond to cytoplasmic fluorescence. Thus, in ventral nuclei, the brightest slice is the z-slice corresponding to the best estimate of the true nuclear fluorescence (magenta circles). On the other hand, dorsal nuclei have a lower Dorsal concentration than the cytoplasm, so the darkest slice is a better estimate of the true Dorsal concentration (blue circles). In lateral nuclei, the nuclear fluorescence is similar to that of the cytoplasm (green circles). To identify which z-slice to use for nuclear fluorescence calculations, we fit the fluorescence, f, over z-slices, z, to a quadratic equation, $f = az^2 + bz$, where a and b are the coefficients of this quadratic equation. Then, we use the value of a to determine whether the nucleus is ventral ($a < -0.5$), lateral ($-0.5 < a < 0.5$), or dorsal ($a > 0.5$). Next, in ventral nuclei, we take the brightest z-slice as the Dorsal-mVenus fluorescence of that frame (dashed horizontal magenta line). In lateral nuclei, we take the median of fluorescence values over z-slices (dashed horizontal green line). In dorsal nuclei, we take the darkest z-slice as the respective frame's Dorsal-mVenus fluorescence (dashed horizontal blue line). Caption continues on next page.

Figure S7: **Continued from previous page: Measuring Dorsal-mVenus nuclear fluorescence across the dorsoventral axis. (B)** Representative time traces of nuclear Dorsal-mVenus fluorescence. To calculate transcriptional activity as a function of Dorsal protein, we sort nuclei into Dorsal concentration bins based on the the Dorsal-mVenus fluorescence at a single fiducial time point halfway through the respective lifetime of each nucleus.
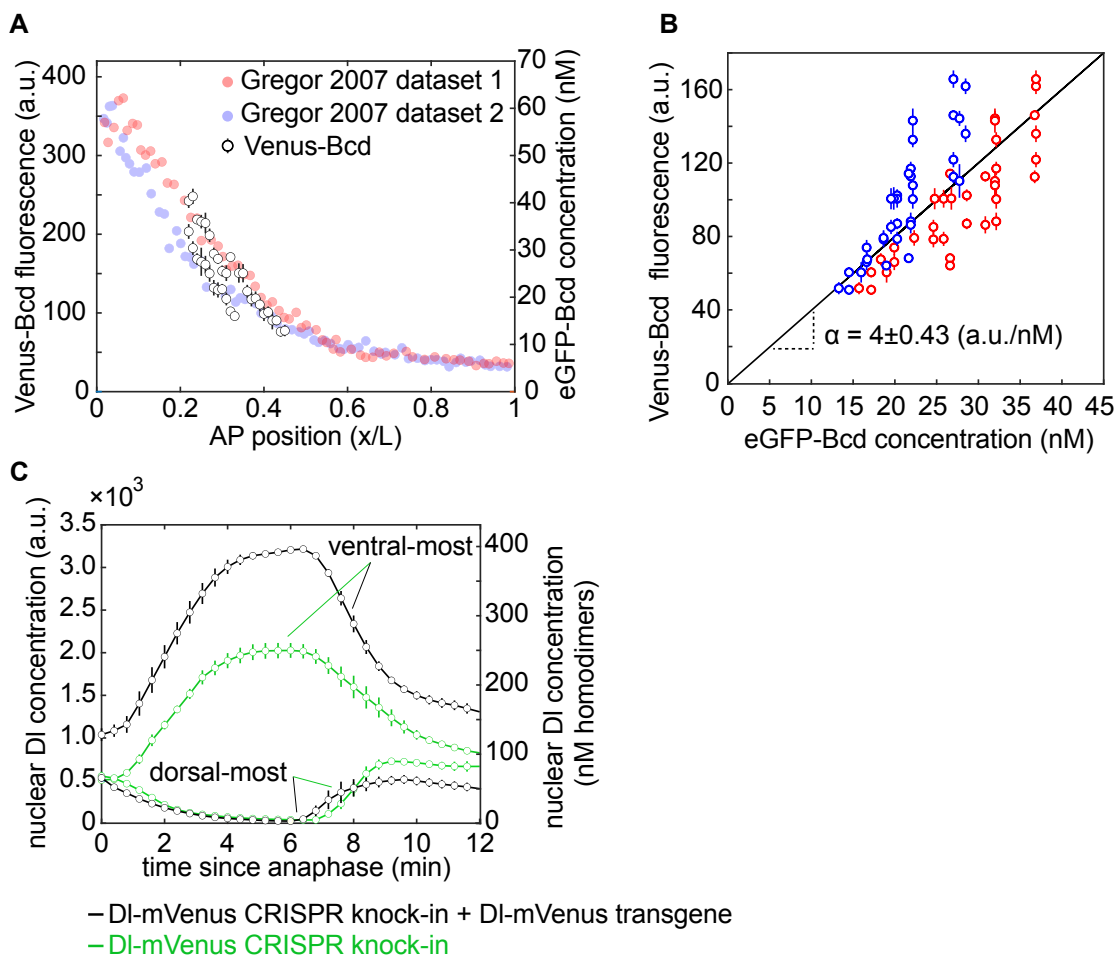
Figure S8: **Absolute calibration of Dorsal-mVenus fluorescence using Venus-Bicoid and previously measured eGFP-Bicoid concentration. (A)** Three embryos derived from *yw;Venus-Bicoid;BcdE1* homozygous mothers were imaged in nuclear cycle 14 using the imaging conditions of our MS2 experiments. The nuclear fluorescence was calculated 15 min into nuclear cycle 14 for cross-comparison with absolute eGFP-Bicoid concentration measurements from Figure 2B of Gregor et al. 2007. We compare the fluorescence values of Venus-Bicoid to the absolute concentration of eGFP-Bicoid along the anterior-posterior axis of the embryo. **(B)** Plot of Venus-Bicoid fluorescence as a function of eGFP-Bicoid fluorescence. Each data point corresponds to the mean $\pm$ standard deviation of the fluorescence of all nuclei belonging to the same 1% spatial window along the anterior-posterior axis. These data were compared to two different absolute measurements of eGFP-Bicoid, shown in red and blue. Linear fit was performed assuming no intercept term since we are estimating a proportionality constant. The slope's error (α) corresponds to the 95% confidence interval. **(C)** Mean and SEM of the Dorsal nuclear concentration in the ventral-most and dorsal-most nuclei across four embryos. 1x and 2x correspond to embryos from homozygous females containing one or two Dorsal-mVenus alleles, respectively. The right y-axis shows the concentration of Dorsal homodimers assuming 6 fluorescence a.u. per mVenus molecule based on (A) and (B).
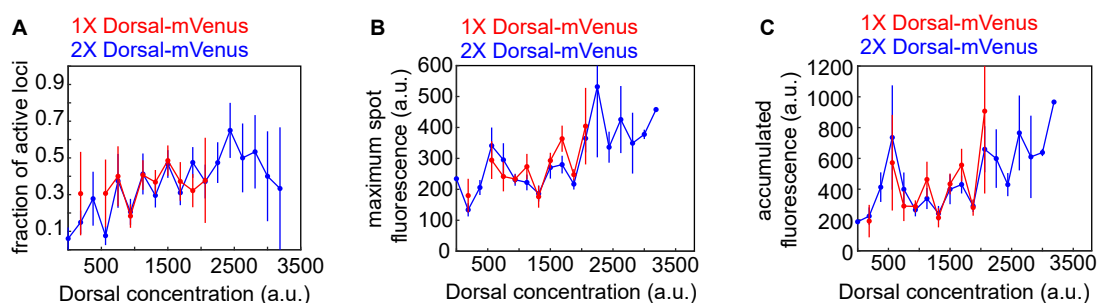
Figure S9: **Comparing the activity of the Dorsal-mVenus transgene to that of two copies of Dorsal-mVenus provided by a transgene plus a CRISPR knock-in**. For the DBS_6.23 reporter construct, we imaged embryos laid by two different mothers. 1x mothers (red) carry dl^1 (a null Dorsal allele) and a Dorsal-mVenus transgene created by Reeves et al. 2012b. 2x mothers (blue) carry a Dorsal-mVenus CRISPR knock-in and the aforementioned Dorsal-mVenus transgene. Nuclei from these different mothers were binned according to their mVenus fluorescence and different activity metrics were measured for each bin. The two Dorsal-mVenus populations are not different within error such that it is valid to treat embryos laid by these different mothers as equivalent. (Error bars correspond to the standard error across at least three embryos per Dorsal-mVenus fluorescence bin.)
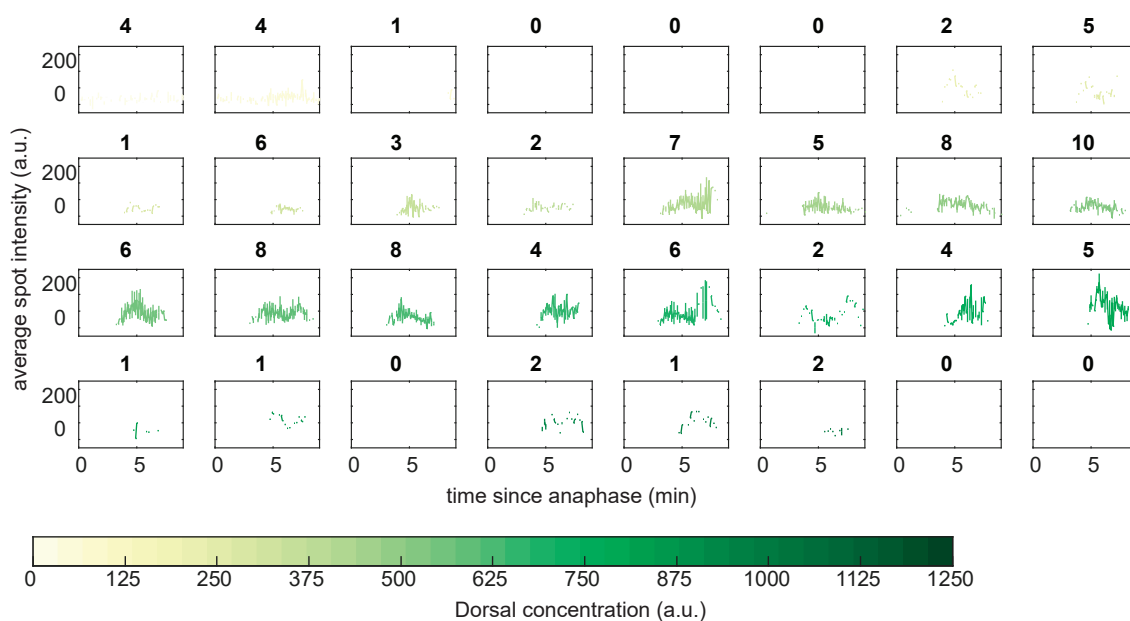
Figure S10: **Mean DBS_6.23 transcription spot intensity over time.** Mean spot intensity from DBS_6.23 transcription spots over time. Each plot corresponds to a different Dorsal-mVenus concentration as indicated by the legend. The nature of the data makes it challenging to estimate the RNAP loading rate by fitting a line to the initial rise of fluorescence. Bold letters above each plot indicate the number of particles included in each bin.
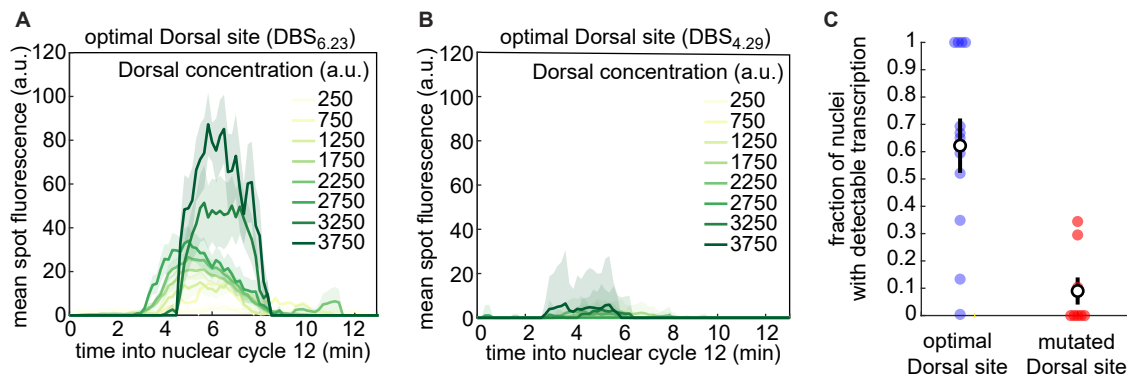
Figure S11: **Transcription driven by a minimal Dorsal synthetic enhancer with a mutated Dorsal binding site.** **(A,B)** Mean fluorescence over time across all loci in the field of view from an embryo carrying a minimal synthetic enhancer with a **(A)** single optimal and **(B)** a mutated Dorsal binding site. **(C)** Fraction of nuclei in which we detected a transcription spot at any time during the duration of nuclear cycle 12 in nuclei exposed to high Dorsal concentration (2600–3200 a.u.) within the field of view. Filled circles correspond to individual embryos. Black circles show the mean across all embryos. Shaded areas in (A) and (B) and error bars in (C) correspond to the standard error of the mean.
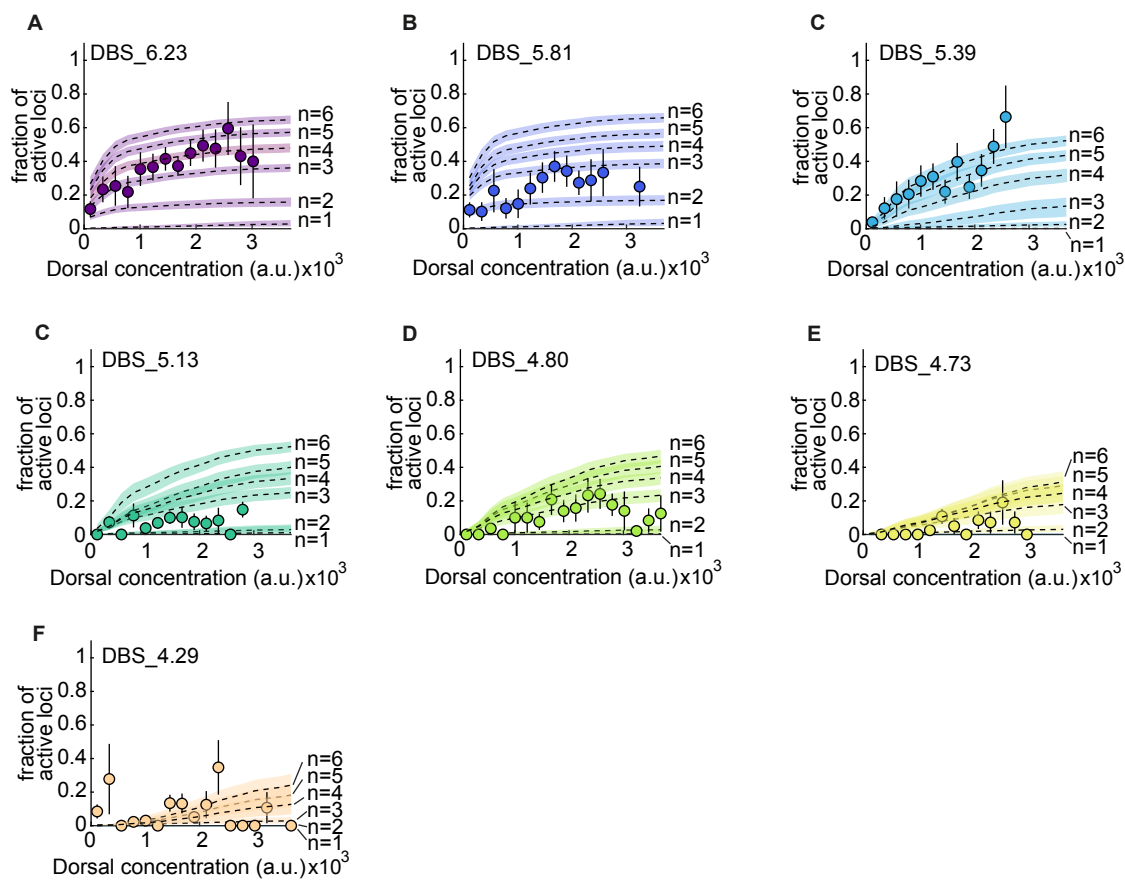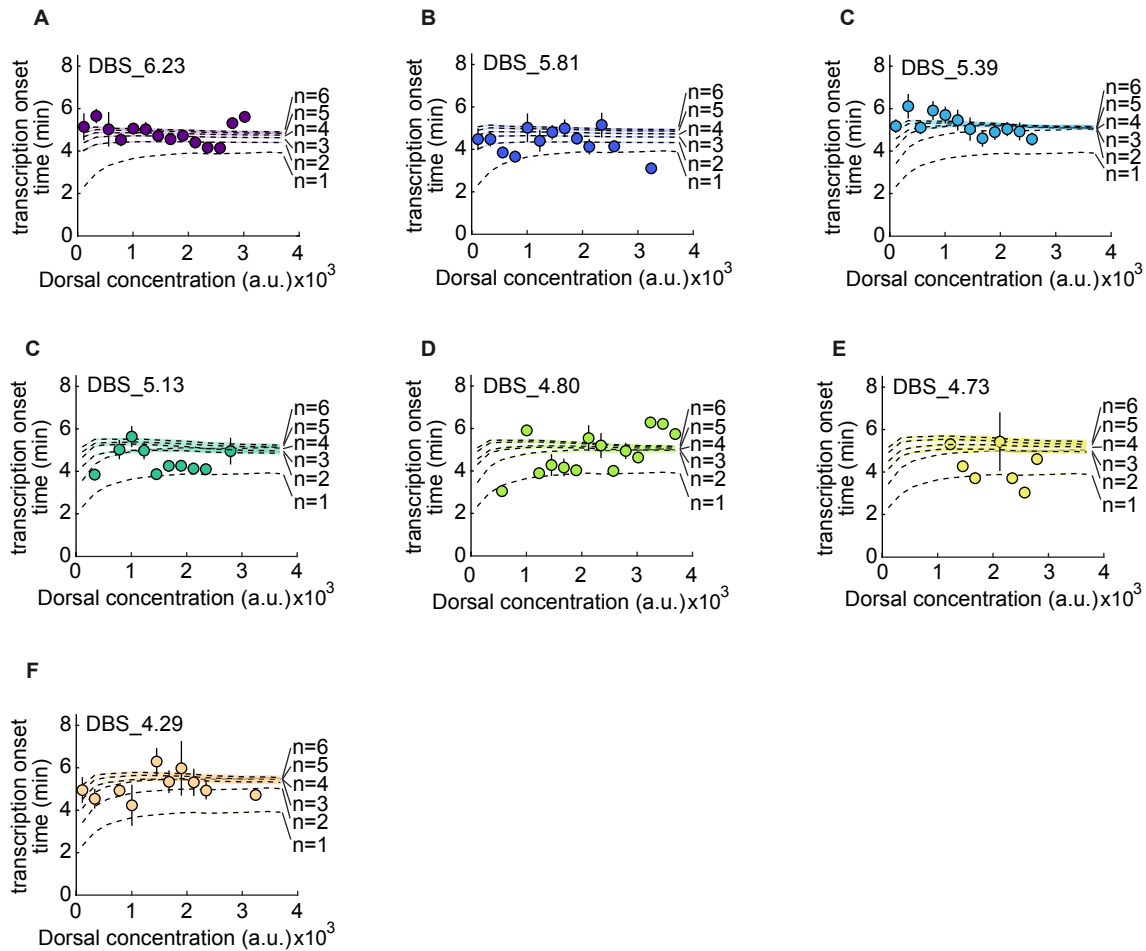
Figure S12: **Fits of the kinetic barrier model to the fraction of active nuclei using different numbers of transitions, n. (A)** Mean fraction of active loci as a function of Dorsal concentration in the DBS_6.23 enhancer. Dashed lines show model fits using different number of OFF states n = 1, 2, 3, 4, 5, and 6, corresponding to predictions using median parameter values from the joint posterior distribution. Fits are performed simultaneously across all enhancers with the value of c being shared and the value of K_D being allowed to vary across enhancers. The shaded areas indicate the 25%-75% credible interval. **(B-F)** Same as (A) for the rest of minimal synthetic enhancers. Error bars in (A)-(F) correspond to the SEM taken over N $>$ 3 embryos containing 3 or more nuclei in a given bin.

Figure S13: **Fits of the kinetic barrier model to the transcription onset times using different numbers of transitions, n. (A)** Mean transcription onset time as a function of Dorsal concentration in the DBS_6.23 enhancer. Dashed lines show model fits using different number of OFF states $n = 1, 2, 3, 4, 5,$ and 6, corresponding to predictions using median parameter values from the joint posterior distribution. Fits are performed simultaneously across all enhancers with the value of c being shared and the value of K_D being allowed to vary across enhancers. The shaded areas indicate the 25%-75% credible interval. **(B-F)** Same as (A) for the rest of minimal synthetic enhancers. Error bars in (A)-(F) correspond to the SEM taken over N > 3 embryos containing 3 or more nuclei in a given bin.

S3 Supplementary tables

Parameter	Mean	Std. Dev.
c (min^{-1})	0.55	0.037
K_D(DBS_6.23)	210	85
K_D(DBS_5.81)	150	52
K_D(DBS_5.39)	980	450
K_D(DBS_5.13)	870	360
K_D(DBS_4.80)	870	340
K_D(DBS_4.73)	1.5×10^3	680
K_D(DBS_4.29)	3.7×10^3	3.1×10^3

Table S1: Inferred parameters from kinetic barrier model fits in Figure 2.6. Each K_D has units of a.u..

Parameter	Mean	Std. Dev.
R_{max}	510	190
K_D(DBS_6.23)	6.3×10^3	5.5×10^3
K_D(DBS_5.81)	5.4×10^4	2.6×10^4
K_D(DBS_5.39)	4.3×10^4	2.7×10^4
K_D(DBS_5.13)	3.9×10^4	2.7×10^4
K_D(DBS_4.80)	6.7×10^4	2.2×10^4
K_D(DBS_4.73)	6.6×10^4	2.3×10^4
K_D(DBS_4.29)	6.8×10^4	2.2×10^4
ω	14	23
P/K_P	0.65	0.23

Table S2: Inferred parameters from fits of the thermodynamic model to the RNAP loading rates measured in Figure 2.6. R_{max} and K_D each have units of a.u., while the remaining parameters are unitless.

S4 Supplementary videos

For better quality of visualization, we recommend downloading these videos.

- Video S1. **DBS_6.23 confocal movie.** Confocal microscopy movie taken on the ventral side of a developing fly embryo (*yw; MCP-mCherry, Dl-mVenus(CRISPR) /*

DBS_6.23-MS2; MCP-mCherry, Dl-mVenus, His-iRFP / +) during nuclear cycle 12. Left: Dorsal-mVenus; Right: MCP-mCherry.

- Video S2. **ParB experiment confocal movie.** Confocal microscopy movie taken on the ventrolateral side of a developing fly embryo (*yw; ParB-eGFP, MCP-mCherry / intB2-DBS_6.23-MS2; +*) during nuclear cycle 12. Left: ParB-eGFP; Right: MCP-mCherry.

Chapter 3

Development of a Bicoid-driven minimal synthetic enhancer

1   Introduction

In the past few decades, developmental biologists have made tremendous progress identifying enhancers and binding motifs for transcription factors that drive the expression of genes responsible for cell fate decisions. Despite this progress, we are still incapable of determining mRNA transcript copy numbers given a regulatory DNA architecture and transcription factor concentrations. Recently, several genetic and microscopy tools have been developed that make it possible to engage in such dialogue between theory and experiment in the context of living embryos of the fruit fly *Drosophila melanogaster*. We use thermodynamic models to predict how the placement, number, and affinity of binding sites for transcription factors govern transcriptional dynamics in the embryo. To experimentally test the predictions stemming from these models, we designed a minimal gene cassette that serves as a scaffold for systematically modulating regulatory parameters. Our gene construct includes MS2 bacteriophage stem loop sequences to report on gene expression from an enhancer with one or more copies of the Bicoid activator and Zelda co-activator binding motifs.

In this chapter, I will mainly focus on the challenges associated with developing a minimal synthetic enhancer system in the early embryo. The particular minimal synthetic enhancers studied here carry a Bicoid activator binding site, but what we have learned also applies to other minimal synthetic systems such as the Dorsal synthetics featured in Chapter 2.

Experiments involving minimal synthetic enhancers come with several unique challenges not present in the study of high expressing enhancers using live-imaging like the *hunchback* P2 promoter (Garcia et al. 2013; Eck et al. 2020) or the *even-skipped* stripe enhancers (Lammers et al. 2020; Berrocal et al. 2020). Fundamentally, in minimal synthetic experiments, there is a trade-off between sensitivity and signal. An overly sensi-

tive enhancer might activate with low input concentrations, and thus wouldn't be able to detect changes across the gradient. However, this strongly sensitive system would have a reasonable signal. Meanwhile, an insensitive construct wouldn't respond to activator concentration at all— leading to too low a signal. There should exist a Goldilocks regime where the gene is sensitive enough to the activator to have reasonable signal at high concentrations, but low signal at low concentrations. Minimal synthetic enhancers have intrinsically weak signal relative to their stronger counterparts (e.g., the *hunchback* P2 promoter) by virtue of their few activator binding sites. Thus, the first challenge associated with this experiment is being able to detect this weak signal against background signal fluctuations —either biological or optical.

2 Results

2.1 Theory

To motivate the idea of minimal synthetic enhancers, I will provide here a brief discussion on theoretical predictions that form testable hypotheses that only synthetic enhancers can challenge. The equilibrium statistical mechanics here are the same as those found in Chapter 2, but I will briefly review the important equations.

The rate of RNAP loading onto our promoter of interest, R, is assumed to be proportional to the probability of RNAP being bound to the promoter such that,

$$R = R_{max} \cdot p_{bound}. \tag{1}$$

Furthermore, an equilibrium thermodynamic model can be used to predict the probability that RNAP is bound to the promoter given that it is recruited by an activator bound to its respective binding site in the enhancer,

$$R = R_{max} \cdot p_{bound} = R_{max} \cdot \frac{\frac{[P]}{K_P} + \frac{[Bcd]}{K_D} \frac{[P]}{K_P} \omega}{1 + \frac{[Bcd]}{K_D} + \frac{[P]}{K_P} + \frac{[Bcd]}{K_D} \frac{[P]}{K_P} \omega}. \tag{2}$$

There are some additional considerations here that we did not cover in Chapter 2. For instance, the dependence on of expression on the number of binding sites and binding site position (Fig. 1 A, B, C). In Figure 1 A, left, we show how an activator's binding site position relative to the promoter could potentially be incorporated into a thermodynamic model. Figure 1 C shows another example of how RNAP loading rate could vary sinusoidally as a function of binding site position, which could occur due to the periodicity of the DNA helix. Additionally, Figure 1 A, right shows how additional binding sites could be incorporated into a thermodynamic model and how two new parameters are introduced into the model: the cooperativity of activator molecules binding together, ω, and the distance between activator binding sites, D. Figure 1 B shows predictions this

model makes for one, two, and three Bicoid binding sites, respectively with a high level of cooperativity (ω).

Having explored thermodynamic models that could be used to dissect single or few activator binding site minimal synthetic enhancers, we now turn to the experimental apparatus we use to study the enhancers' gene expression and challenge these models.
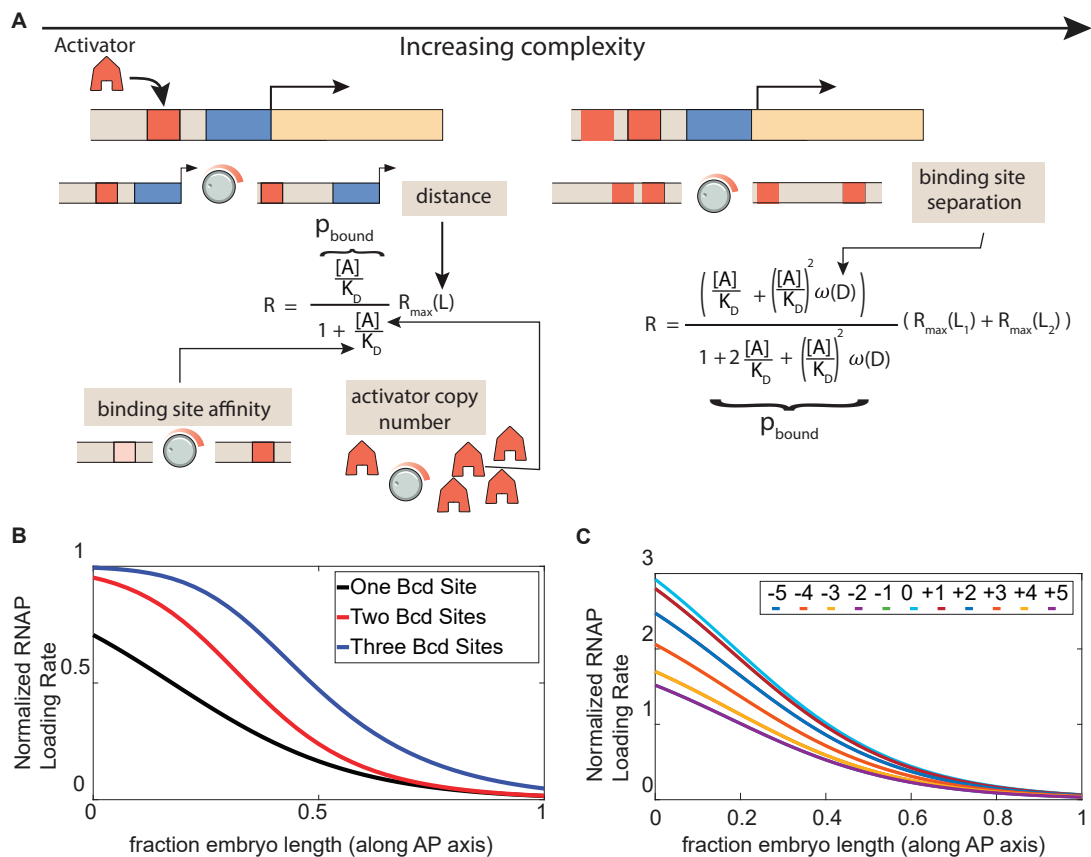
Figure 1: **Theoretical predictions for transcription driven by a minimal synthetic enhancer with one to a few activator binding sites. (A)** (left) A single activator binding site can have varying affinity, affecting the activator's binding site dissociation constant, K_D. Additionally, there is an interaction term between RNAP and the activator determining the rate of RNAP loading, R_{max}, which is a function of L, distance of the binding site from the promoter. Gene expression level, R, is also determined by activator concentration, $[A]$. (right) Adding additional activator binding sites adds two extra parameters to the model: ω, which characterizes the cooperativity of multiple transcription factors and D, which is the distance between activator binding sites. **(B)** Example of how RNAP loading rate predictions vary with activator binding site number with arbitrary parameter values from (A). The x-axis here is distance along the anterior-posterior axis, which corresponds to Bicoid concentration that decreases from anterior to posterior. **(C)** One possible way that RNAP loading rate may vary with binding site position (legend shows number of base pairs relative to an arbitrary initial position). In this example, the activator-RNAP interaction term varies sinusoidally with position with a period of 11 base pairs corresponding to a helical turn of DNA.

2.2 Experimental setup

As the experimental design here is very similar to the experimental design found in Chapter 2, this section will be brief. We placed a single Bicoid binding site upstream of the *eve* core promoter to drive the expression of *MS2-lacZ*. Male flies containing these minimal synthetic enhancers to female flies carrying *MCP-eGFP; His-RFP* (Fig. 2 A). We imaged early stage embryos on a laser scanning confocal microscope during nuclear cycle 12, 13 and 14 as they developed and observed the transcription from our reporter via the MS2-bound MCP-eGFP signal (Fig. 2 B). We then quantified the signal from individual loci and determined that this was different from background noise in the absence of a reporter construct (Fig. 2 D), and measured quantities such as the rate of mRNA production, the transcriptional onset times, and the fraction of active loci (or nuclei) which could be detected transcribing in a given nuclear cycle within the field of view (Fig. 2 E).

Although these initial results were promising as a proof of principle of being able to quantify the activity of single binding site synthetic enhancers, additional controls were necessary to demonstrate that our reporter was actually responsive to Bicoid binding to the binding site in the enhancer and not to other extraneous inputs.
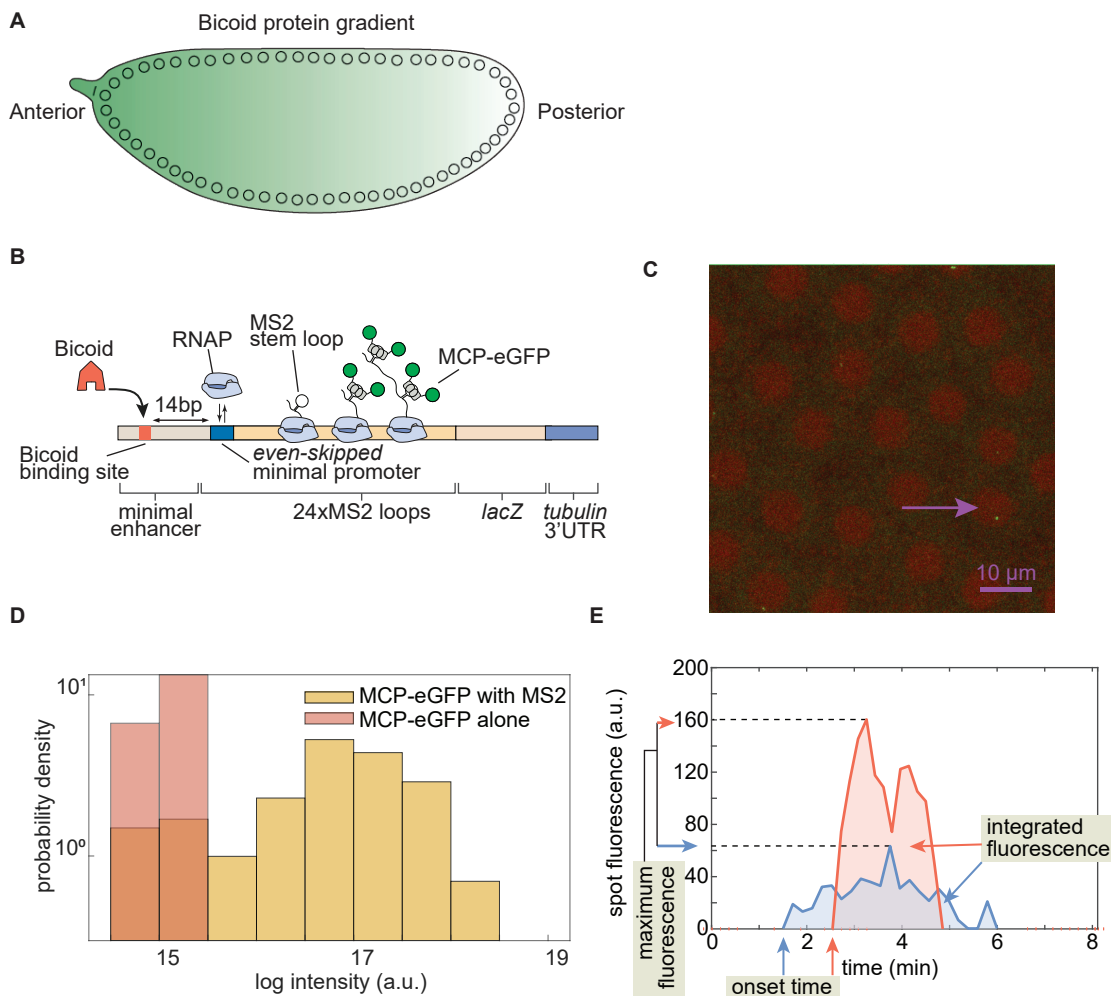
Figure 2: **Synthetic enhancers drive detectable and quantifiable levels of gene expression. A** Schematic of the Bicoid protein gradient in the *Drosophila melanogaster* embryo during nuclear cycle 14. **(B)** Schematic of single Bicoid binding site transcriptional reporter. **(C)** Snapshot from embryo in nuclear cycle 12 with transcription being driven by single Bicoid binding site reporter (green channel: MCP-eGFP; red channel: Histone-RFP nuclear marker). Arrow points to example nucleus with active transcription spot. **(D)** Histograms of trace intensities of embryos with or without an MS2 reporter. Intensities are a sum over 3 z-planes of difference-of-Gaussian filtered images in a 13 pixel radius around the detected spots. **(E)** Cartoon showing quantified intensity over time traces with measurable quantities such as integrated fluorescence, maximum fluorescence and transcriptional onset times emphasized.

2.3 Creating a transcriptionally neutral reporter

In order to ensure that our reporter was only reporting on the binding of Bicoid to the Bicoid binding motif we placed upstream of the promoter, it was important to determine whether our entire reporter sequence and its adjacent DNA sequences were free of binding motifs for other early embryonic transcription factors. Indeed, early results comparing the expression from *A3-evePr-MS2v5-lacZ* (A3: a strong Bicoid binding site, evePr: *even-skipped* core promoter) to a negative control without the Bicoid binding site, *evePr-MS2v5-lacZ*, suggested that there may have activation coming from elsewhere in our cassette (Fig 3 A, B).
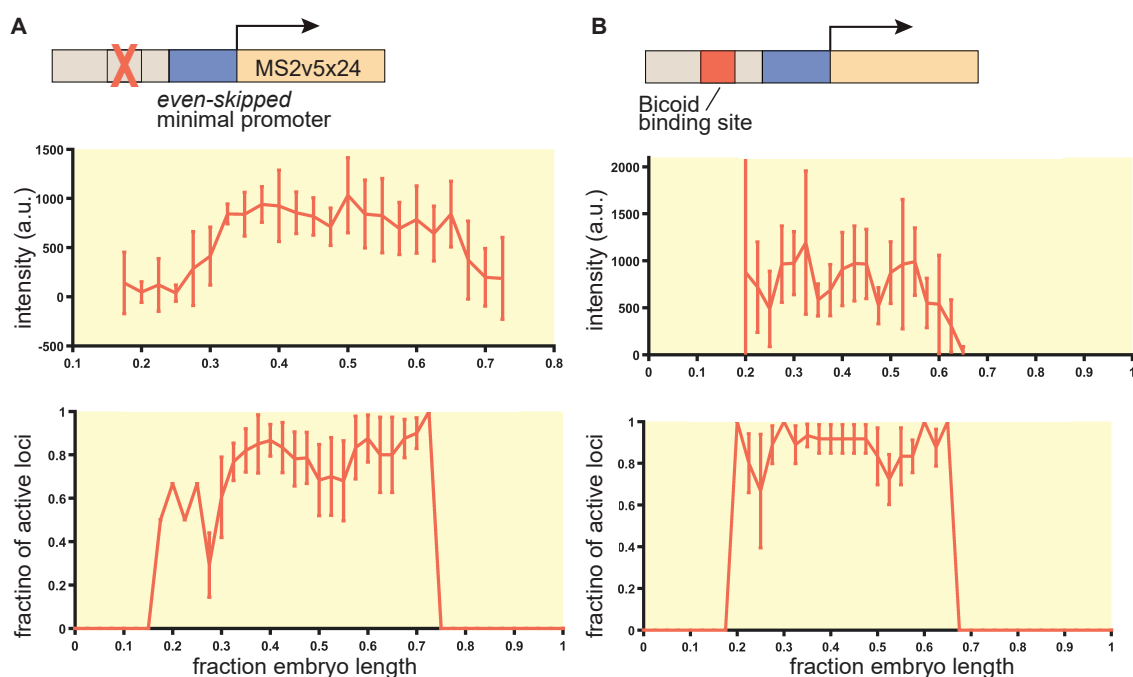
Figure 3: **Creating a neutral minimal synthetic enhancer reporter construct. (A)** Two measures of gene expression from *evePr-MS2v5-lacZ*: time-integrated intensity (top) and fraction of active loci (bottom). **(B)** Two measures of gene expression from *A3-evePr-MS2v5-lacZ*: time-integrated intensity (top) and fraction of active loci (bottom). Data for (A) and (B) were collected during nuclear cycle 12 and used ≥10 embryos in each case. Error bars are standard error of the mean (SEM) over embryos.

To control for this background activity, we sought to ensure that all sequences within our reporter, with the exception of the Bicoid site, were "neutral," or, free of transcription factor binding sites. From our experience developing synthetic enhancers, it is clear that many factors influence how transcription factor neutral a DNA sequence is in the early embryo. Those include: genomic position, presence of Zelda sites, presence of

other transcription factors, and the choice of core promoter (Fig. 4) B. In this section, we explore these different factors in detail.
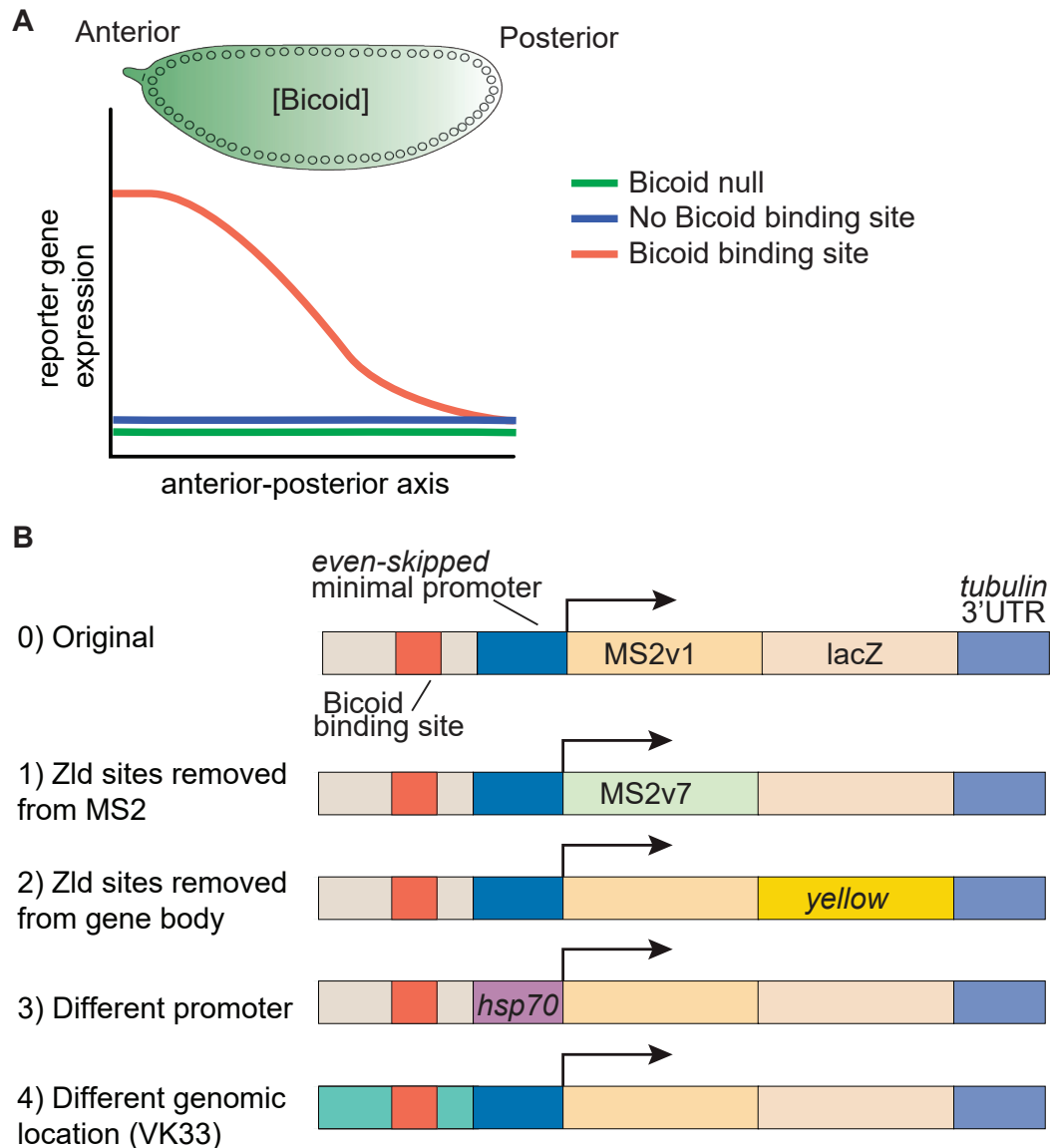
Figure 4: **Creating a neutral minimal synthetic enhancer reporter construct.** **(A)** Ideal experimental outcome where gene expression is anterior-posterior dependent, requires Bicoid, and requires the Bicoid binding site. **(B)** Schematic showing proposed changes to constructs imaged in Figure 2 A, B to reduce background expression levels in the absence of a Bicoid binding site.

In order to predict the neutrality of a DNA sequence, the reporter's DNA sequence can be run through a bioinformatic algorithm such as Patser (Stormo, Schneider, and

Gold 1986) that detects the presence of binding sites for transcription factors known to be present in the early embryo. Subsequently, those binding sites can be removed and a transcriptional reporter assay (e.g., MS2-MCP) can be performed to determine whether the resulting reporter has expression in the absence of Bicoid protein or in the absence of a Bicoid binding site.

We hypothesized that much of the observed basal expression was due to the pioneering factor Zelda. We initially suspected this because of Zelda's ubiquitous presence in early embryonic genes, where it serves to potentiate their expression (Foo et al. 2014). To confirm that Zelda was actually responsible for the background activation we observed, we performed our single Bicoid binding site MS2 experiments in a Zelda null background using the Zelda germline clone protocol also used in studies such as Mir et al. 2017, Eck et al. 2020, Liang et al. 2008, and Nien et al. 2011. The dramatic reduction in basal activity in the absence of maternally deposited Zelda protein (Fig. 5 red curves) confirmed our hypothesis that Zelda played a large role in Bicoid-independent reporter transcription. Indeed, in nuclear cycle 12 we observed no activity at all in the absence of Zelda protein, but transcription was slightly increased in later nuclear cycles. This suggests that there may be other factors present in later cycles capable of activating our promoter, perhaps including Bicoid. However, most of our imaging studies were performed in nuclear cycle 12, so we proceeded to focus our efforts to reduce basal expression by removing Zelda sites present in the reporter constructs.

Since we confirmed that the Zelda pioneering factor contributed much of the background signal in our experiments, we needed to be able to account for the presence of Zelda binding sites in our cassette. We also looked for the presence of early embryonic transcription factors, but Zelda was the predominant focus here. We hypothesized that there could be activating binding sites for transcription factors in the area immediately upstream of the enhancer, in the core promoter, in the MS2 sequence, or in the gene body following the MS2 sequence.

In our early studies, we initially employed the MS2v1 sequence used in Garcia et al. 2013, but we soon switched to the MS2v5 sequence developed by Tutucci et al. 2018 (MBSV6 in their nomenclature) because of the tendency for the more repetitive MS2v1 loops to be lost during the cloning steps or during integration into the fly genome, which often resulted in flies with fewer than half the desired number of loops. These less repetitive MS2v5 loops were also used in Chapter 2 for studying Dorsal synthetics. However, even these MS2v5 loops have a significant drawback that is more apparent in expression driven by Bicoid synthetics than the Dorsal synthetics because Bicoid synthetics have low expression to begin with, making the background signal predominant in the measurement. To find the location of transcription factor binding sites that might be contributing to background expression in our cassette, we used the Patser algorithm developed by (Hertz and Stormo 1999), which uses position weight matrices (PWMs) of transcription factor binding motifs to locate and score binding affinities on any given DNA sequence. We used preexisting position weight matrices computed with the MEME algorithm (Ivan, Halfon, and Sinha 2008; Bailey et al. 2006) using motifs generated by
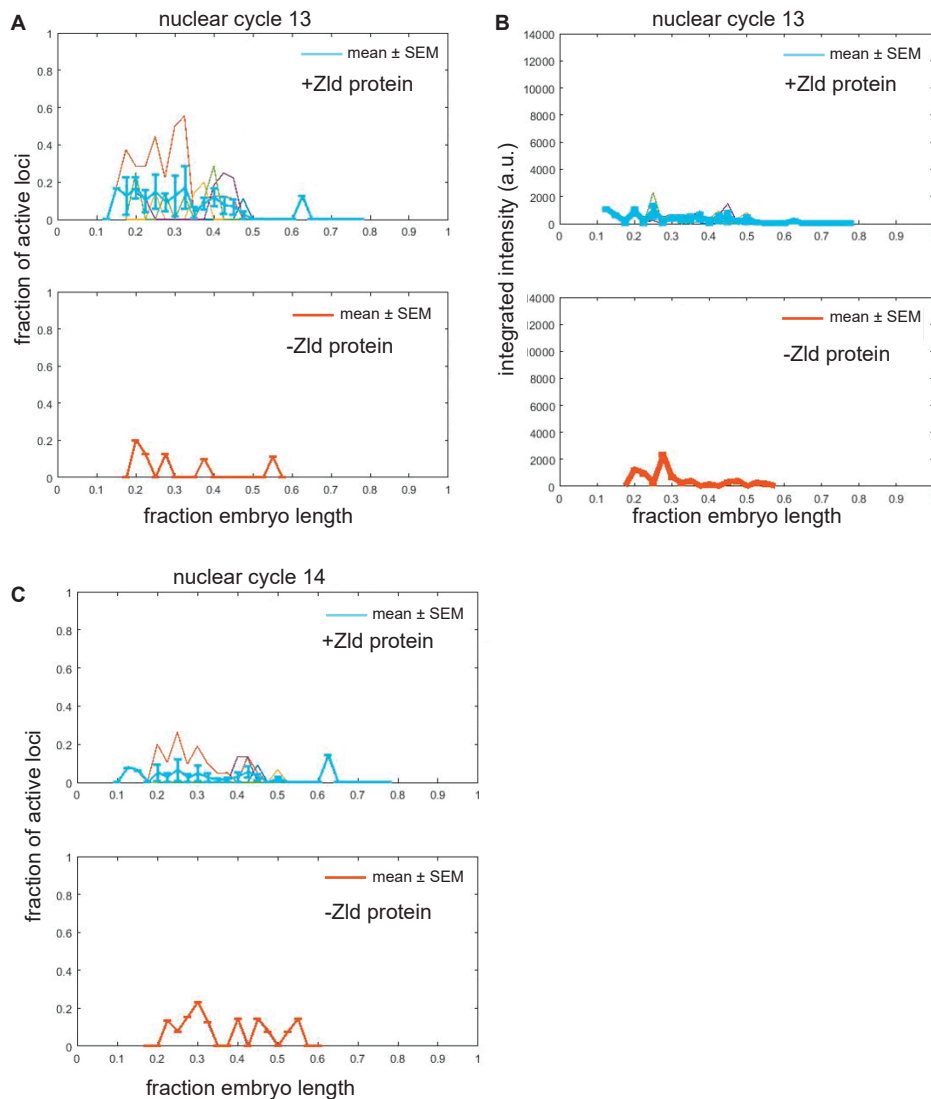
Figure 5: **Comparison of gene expression from single Bicoid binding site enhancers with or without Zelda protein. (A)** *A3-evePr-MS2v7-lacZ* fraction of active loci during nuclear cycle 13. (top) with Zelda protein and (bottom) in Zelda germline clone. **(B)** *A3-evePr-MS2v7-lacZ* time-integrated MS2 spot fluorescence intensity during nuclear cycle 13 (top) with Zelda protein and (bottom) in Zelda germline clone. **(C)** *A3-evePr-MS2v7-lacZ* fraction of active loci during nuclear cycle 14. (top) with Zelda protein and (bottom) in Zelda germline clone. For wild-type Zelda plots, the individual thin curves correspond to different data sets, while the thick curves correspond to standard error of the mean across five embryos. For Zelda null plots, curve corresponds to one embryo.

DNAse I footprinting assays (Bergman, Carlson, and Celniker 2005) and quantified the information content of each base pair using Patser (Hertz and Stormo 1999).As shown in Figure 6, the original MS2 loops contain a large number of high Patser-scoring Zelda sites. Because we were concerned about background transcriptional activity originating from Zelda binding, we swapped the MS2v5 loops for our so-called MS2v7 loops engineered by Elizabeth Eck in the Garcia lab to have all Zelda sites removed that had Patser scores above 3.0 (Fig. 7).
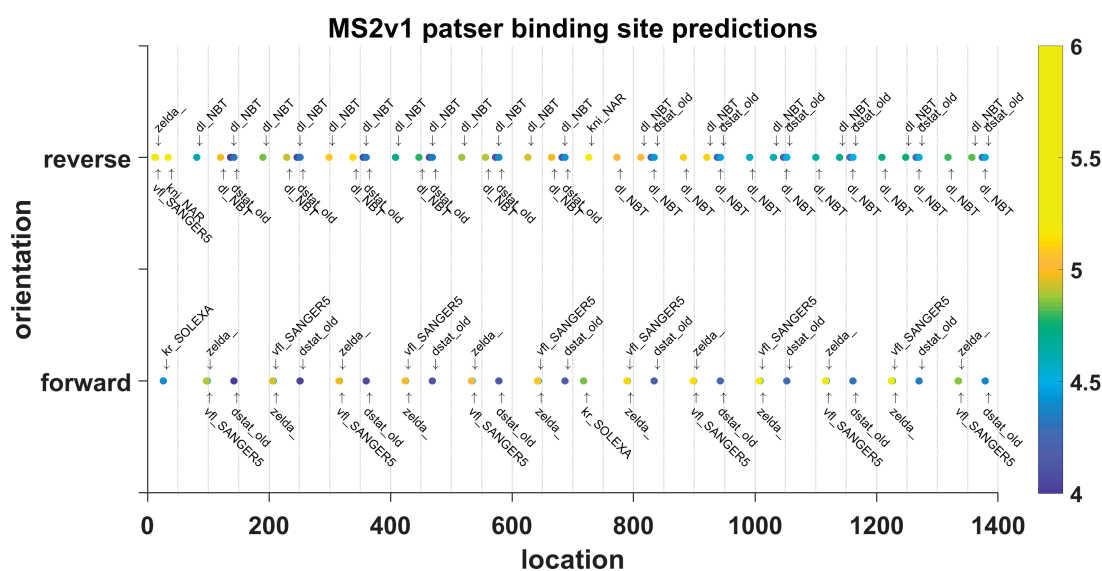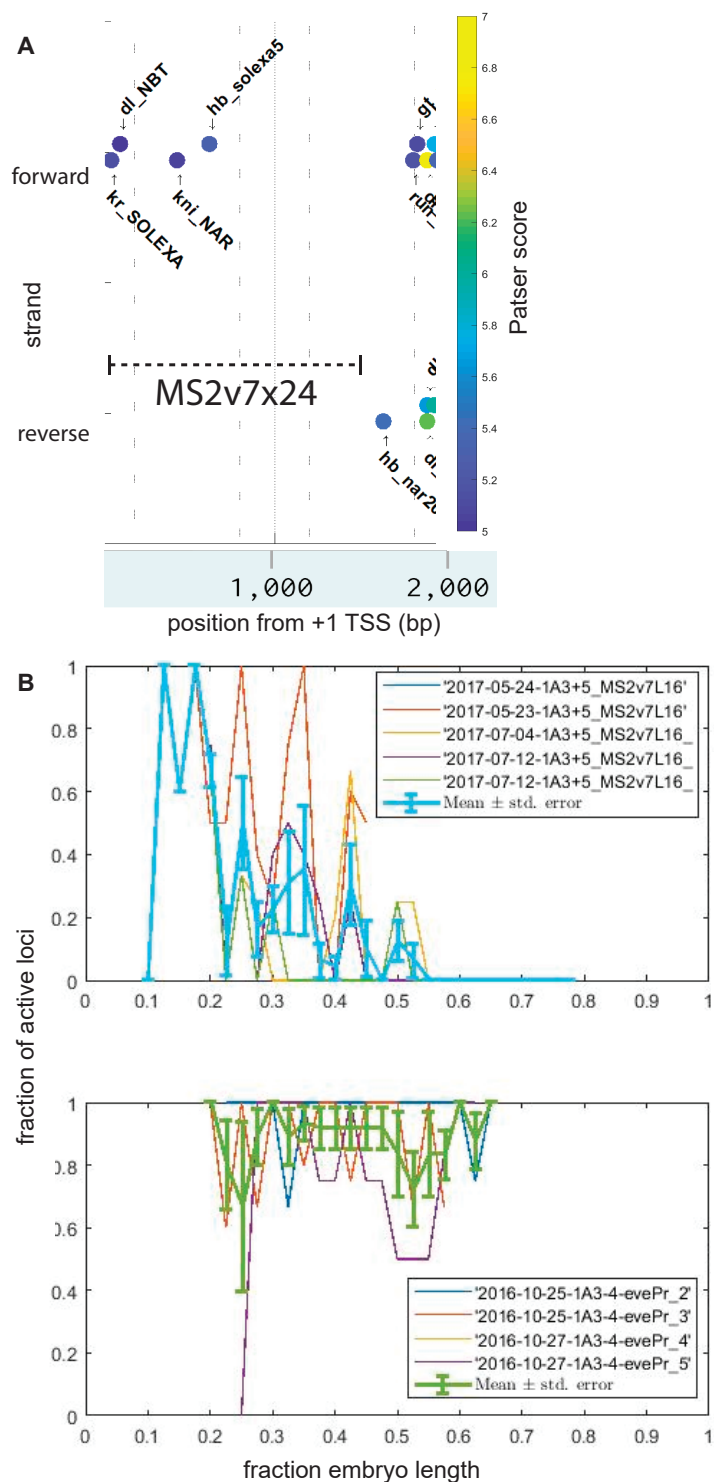

Figure 6: **MS2v1 binding sites**. Schematic showing the bioinformatically predicted binding scores for early embryonic transcription factors binding to the MS2v1x24 sequence.

Finally, we explored different genes that were part of the cassette downstream of the sequence of MS2 loops. We observed that the *lacZ* gene present in our constructs had many binding sites for Zelda, in particular (Fig. 8 A, B). Thus, we searched for a different gene of similar length that lacked strong Patser score > 3 Zelda binding motifs. In particular, we tested the *Drosophila melanogaster yellow* gene. We found that swapping *lacZ* for *yellow* successfully removed nearly all of the background signal present in our constructs (Fig. 9 B, C), most likely due to the lack of Zelda binding sites present in the yellow gene (Fig. 9 A). Note that we're reporting the data here in the form of signal histograms because there was too little activity to plot spot intensity as a function of anterior-posterior length as we could with stronger constructs.

Additionally, we sought to determine the effect of genomic position of the inserted transgene reporter. Thus, we injected our reporter in an additional location. The original location of our reporter was 38F1 on chromosome 2L. For our new location, we chose the position of the VK33 landing site on chromosome 3L. To facilitate screening of transformants using a red eye marker, we used a ϕC31 integration (Groth et al. 2004) with an

attP landing site generated by Venken et al. 2009 instead of the RMCE method Bateman, Lee, and Wu 2006 used for most of our original constructs. With RMCE, plasmids can be inserted randomly into one of two orientations in the genome, such that one must check orientations of inserted transgenes before experimentation to ensure consistency across different reporters. In addition to allowing easy screening, with the ϕC31 system, no such orientation checks are necessary. While we admit that our experiment requires an additional control to make sure that the effect is due to the genomic location and not the plasmid used, this remains a demonstration of how sensitive the reporter is to its environment (whether that comes from the injected plasmid or genomic DNA). We show the gene expression level, the fraction of active loci, of the minimal synthetic reporter (*A3-evePr-MS2v5-lacZ*) located at the VK33 landing site across the anteroposterior axis in Figure 10 A. It is clear that much of the basal expression greatly decreased relative to the constructs located at 38F1 (Fig. 10 A, B), but there is still some expression remaining. We conclude that the genomic context of the reporter's integration affects expression, and that there is an overabundance of ectopic activation at the 38F1 *attP* landing site.

Figure 7: **MS2v7 reporter (A)** Schematic showing the bioinformatically predicted binding scores for early embryonic transcription factors binding to the MS2v7 reporter. **(B)** Fraction of active loci during nuclear cycle 12: (top) *A3-evePr-MS2v5-lacZ* and (bottom) *A3-evePr-MS2v7-lacZ*.
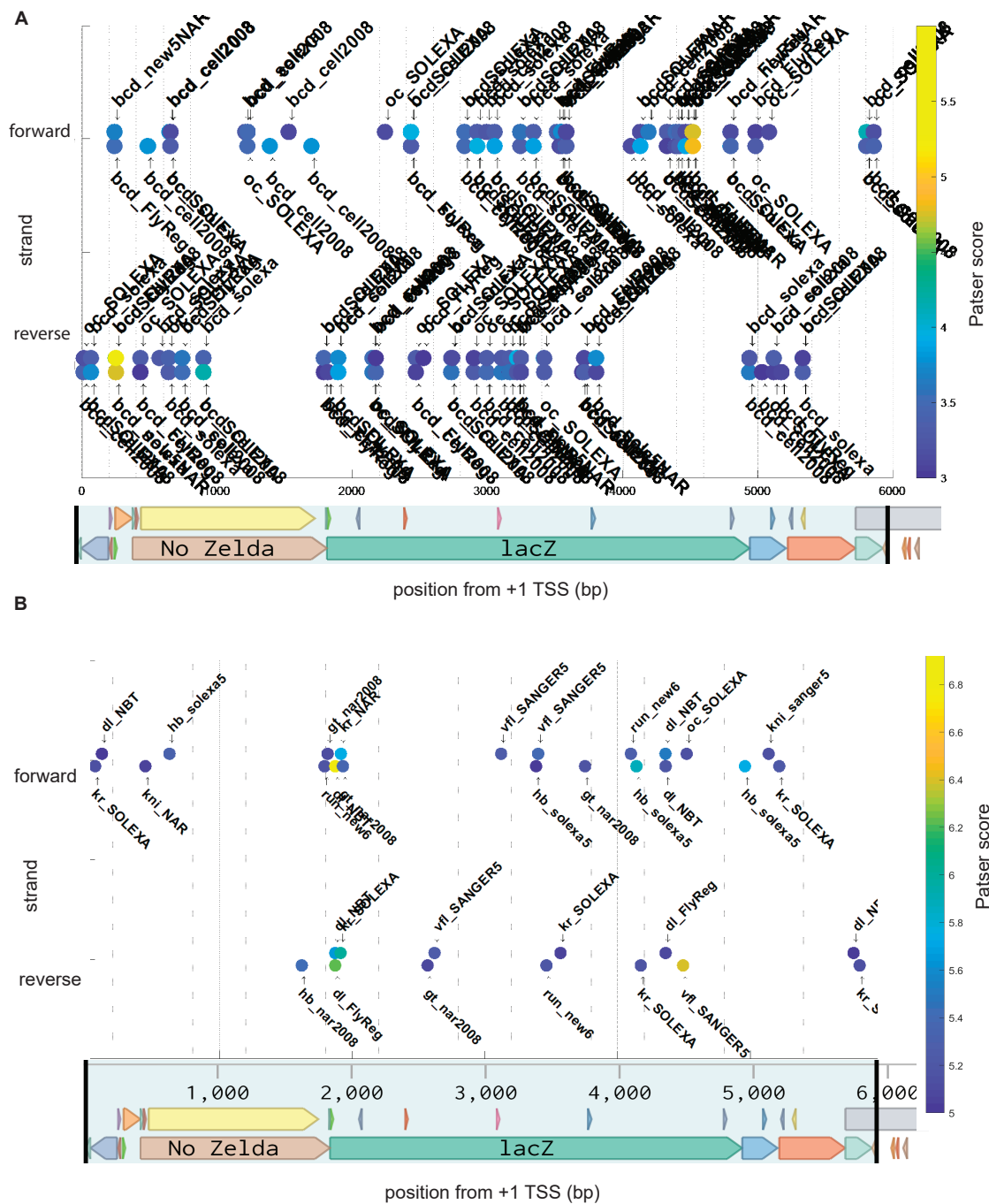
Figure 8: **Zelda and Bicoid DNA binding motifs present in reporter construct's *lacZ* gene.** **(A)** Bicoid and **(B)** Zelda DNA binding motifs present in the original 1A3 reporter construct's *lacZ* gene as identified by Patser with Patser scores above 3.0.
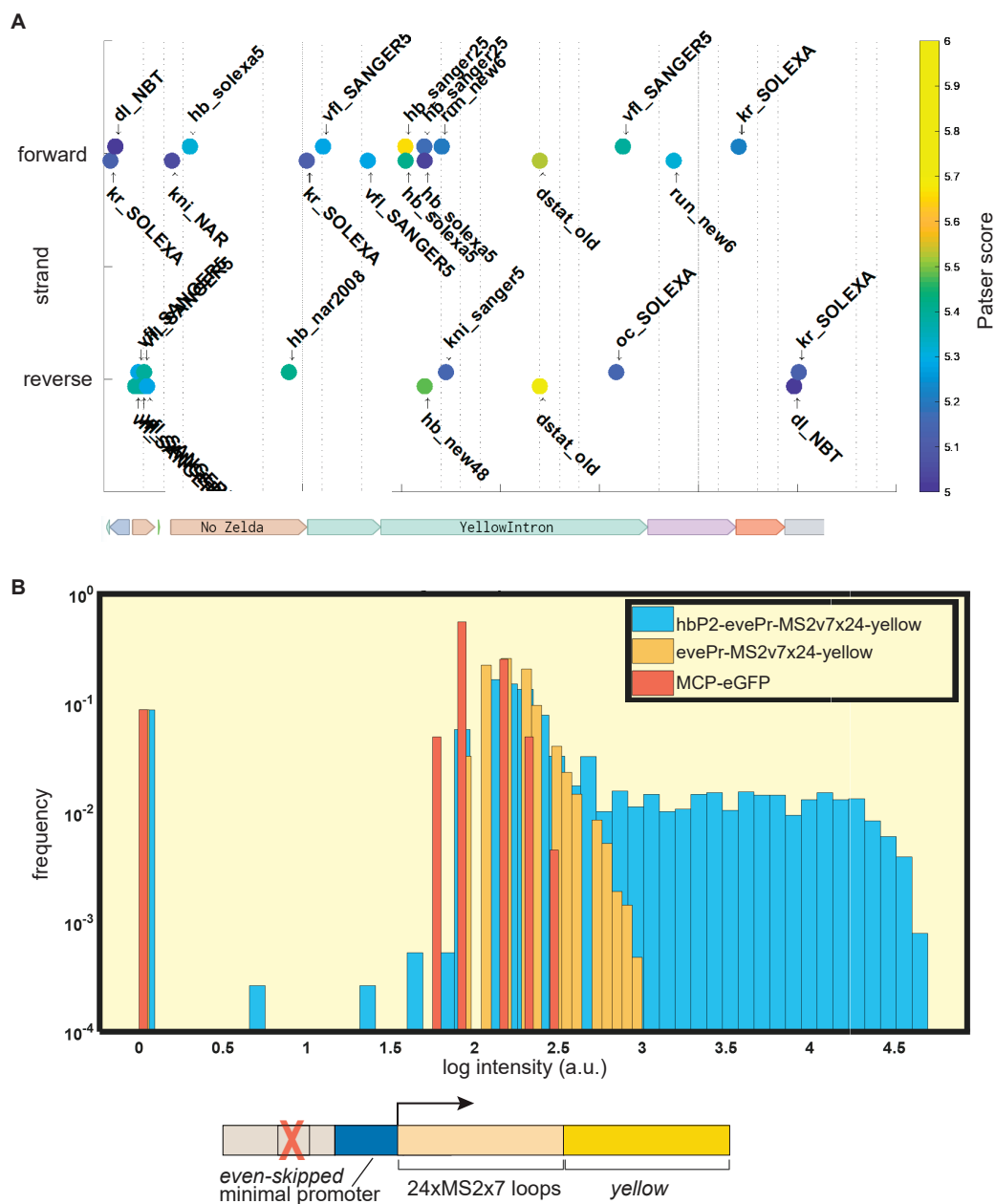
Figure 9: **Reporter construct with the *yellow* gene has no basal expression. A** Bioinformatically predicted binding sites present in the *evePr-MS2v5x24-yellow* construct. **B** Histograms of spot intensity for *evePr-MS2v5x24-yellow* compared to controls.
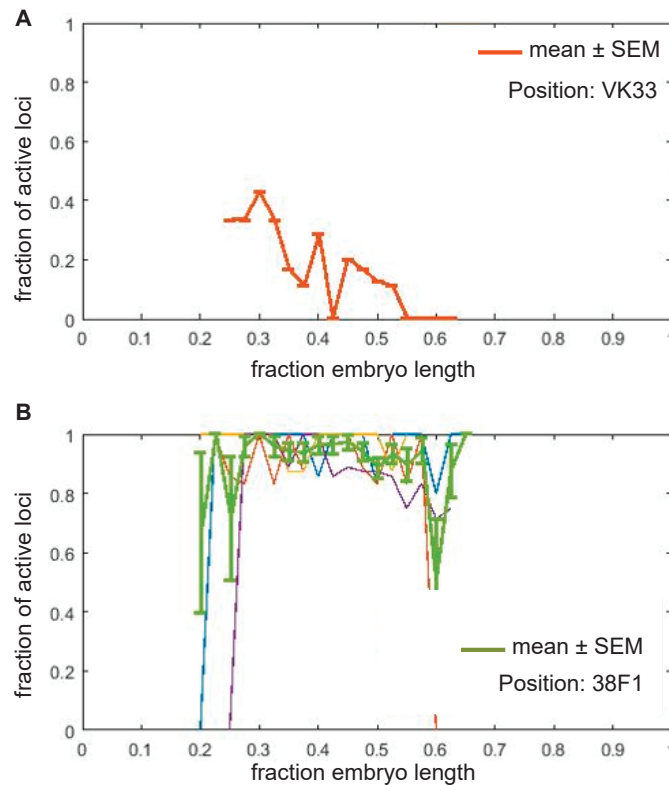
Figure 10: **Minimal synthetic Bicoid enhancer injected at VK33 landing site.** Fraction of active loci during nuclear cycle 13: **(A)** *A3-evePr-MS2v5-lacZ* at VK33 in the pbϕ plasmid (one embryo) and **(B)** *A3-evePr-MS2v7-lacZ* at 38F1 injected with the pIB RMCE plasmid in the forward genomic orientation. Thin individual curves correspond to different embryos, while the thick green curve corresponds to the mean. Error bars for the mean curve are standard error of the mean across 4 data sets.

3   Discussion and future directions

Although we explored a wide range of construct designs, we ultimately were unable to achieve a construct capable of driving Bicoid-dependent gene expression with sufficient detectable signal for our experimental design. Although some constructs such as those containing the *yellow* gene had reduced background activity, they also had too little activity even in the presence of an enhancer containing a Bicoid binding site for any statistics to be generated (data not shown). It is possible that Bicoid is simply too weak an activator to produce signal we can quantify with our experimental design and requires other Bicoid binding sites or Zelda binding sites in order for these experiments to be possible using MS2 and a confocal microscope.

In Figure 11, we show a schematic of most of the designs we tested, with the most promising components circled in blue. Firstly, VK33 coupled with a pbϕ plasmid backbone is suggested due to the lower background activity at that location and the fact that this construct can only integrate in a single orientation in the genome. 650 bp neutral sequence is chosen as the scaffold because it is designed to reduce extraneous background activity. MS2v7 and *yellow* are chosen for coding region since they have reduced Zelda sites and measurably lower background signal. Adding five Zelda sites to the enhancer to potentiate the activity of Bicoid activation could allow Bicoid to have stronger activity. With no added Zelda sites, it is unlikely that the enhancer would have activity that could be measured with sufficiently good statistics. the 1A3F Bicoid binding site is chosen since it is a longer site (9 bp rather than the 6 bp of the original 1A3 Bicoid binding site) that may have stronger Bicoid binding affinity, although this is uncertain at this point. There are also some additions present in the figure that were untested in this study, but would be promising prospects in the future. In particular, flanking the construct with the *gypsy* insulator (shown as arrows in Figure 11) may successfully keep background expression low by preventing transcription factors that bind outside of the construct from activating the promoter.

Regardless of future construct designs, going forward, some improved experimental protocols would greatly aid this project. Faster screening of viable constructs is a top priority. One promising method may be fluorescence *in situ* hybridization (FISH) or single molecule FISH (smFISH). By bypassing live imaging, this higher-throughput approach may enable faster screening of viable enhancers before more careful study is done in live embryos. Another possibility for faster testing is perform live imaging of cultured S2 cells. This would obviate the need for fly husbandry and for long imaging sessions of developing embryos. Again, this would need to be supplemented with live imaging of embryos after screening viable enhancer candidates, since it is not clear how well transcription within S2 cells would translate to full embryos.

Another fruitful approach may be simply abandoning the idea of a completely minimal single site enhancer and working with enhancers as minimal as possible but still detectable and Bicoid-dependent. Alternatives to a single minimal synthetic Bicoid site include having more sites. We know that three strong Bicoid binding sites can achieve
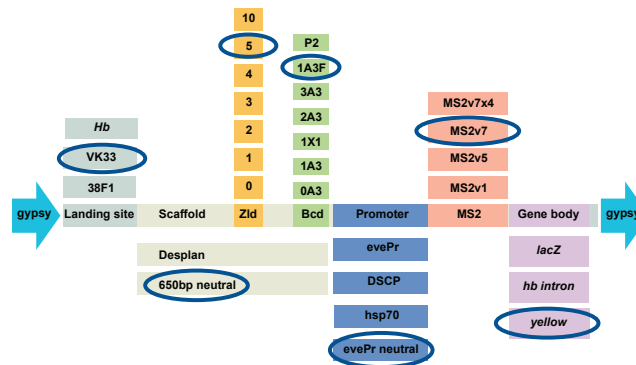
Figure 11: **Future directions for Bicoid minimal synthetic enhancer construction** Construction options for the synthetic enhancer. *gypsy*: the *gypsy* insulator sequence. VK33: Plasmid construct in pbϕ backbone and inserted at the VK33 landing site. 38F1: Plasmid construct in pIB backbone and inserted at 38F1 landing site. *Hb*: Construct placed near the endogenous hunchback gene in the genome (not tested in this study). Desplan: Original sequence upstream of the promoter into which the Bicoid A3 site(s) was inserted. 650 bp neutral: 650 bp sequence constructed with SiteOut into which Bicoid binding sites can be inserted. Zld: Number of Zelda sites additionally added to the enhancer. 0A3: No Bicoid binding site. 1A3: single A3 binding site originally taken from the hunchback P2 promoter. 2A3, 3A3: two and three A3 binding sites in the enhancer, respectively. 1A3F: single A3 Bicoid binding site with additional flanking base pairs (inspired by Park et al. 2019). P2: the full or modified *hunchback* P2 promoter. evePr: *even-skipped* core promoter. DSCP: *Drosophila* Super Core Promoter. *hsp70*: *hsp70* core promoter. evePr neutral: *even-skipped* core promoter with extraneous binding sites removed via SiteOut. MS2v1, v5, v7 are defined above in previous sections. MS2v7x4: MS2v7 sequence repeated four times for a total of 96 loops. *hb intron*: endogenous *hb* intron sequence.

Bicoid-dependent expression, so this may be a good starting point here. Alternatively, one could attempt to achieve a minimal enhancer that is not synthetic by removing binding sites from the endogenous *hunchback* P2 promoter, as was done in Park et al. 2019.

To summarize, we have developed here a platform for probing Bicoid dependent transcription in a way that is amenable to theoretical challenge. However, our theoretical models come loaded with assumptions about how Bicoid binding relates to transcriptional activity. Despite this, the methods here do not actually involve directly observing Bicoid binding and only infer its occurrence. In the following chapter, I will discuss a project that aims to observe Bicoid DNA binding events in living embryos to help shine light on the events that must take place prior to transcriptional activation.

4   Materials and methods

4.1   Plasmids and reporter design

To design our minimal constructs, (Fig. 2), we placed the 6 bp strong Bicoid binding site (Simpson-Brose, Treisman, and Desplan 1994) upstream of the *even-skipped* core promoter. This enhancer-promoter construct drives the expression of the MS2 sequence containing 24 nonrepetitive MS2 loops (Tutucci et al. 2018) followed by the *lacZ* coding sequence and the *tubulin* 3'UTR. (Garcia et al. 2013).

To guide the design of these synthetic constructs, we used preexisting position weight matrices computed with the MEME algorithm (Ivan, Halfon, and Sinha 2008; Bailey et al. 2006) using motifs generated by DNAse I footprinting assays (Bergman, Carlson, and Celniker 2005) and quantified the information content of each base pair using Patser (Hertz and Stormo 1999).

4.2   Flies

Reporter plasmids were injected into BDSC fly line 27388 containing a landing site in position 38F1. Transgene orientation was confirmed by PCR using primers 18.8 (ggaac-gaaggcagttagttgt) and Ori-Seq-F1 (tagttccagtgaaatccaagcattttc) binding outside of the 5' 38F1 *attP* site and the *even-skipped* promoter, respectively. All reporter lines were confirmed to be in the same orientation. All synthetic enhancer flies used in this study can be found in Table 1.

To generate the embryos used in the experiments shown in all, we crossed *MCP-eGFP;His-RFP* virgins to males carrying synthetic enhancers.

Zelda germline clones were produced using the protocol found in Liang et al. 2008.

4.3   Microscopy

As described in Chapter 2, fly cages were allowed to lay for 90 to 120 minutes prior to embryo collection. Embryos were then mounted on microscopy slides in Halocarbon 27 oil (Sigma-Aldrich, H8773) in between a coverslip and breathable membrane as described in (Garcia et al. 2013; Bothma et al. 2014; Garcia and Gregor 2018).

Confocal microscopy was performed on a Leica SP8 with HyD detectors and a White Light Laser. We used a 63x oil objective, and scanned bidirectionally with a scan rate of 420 Hz and a magnification of 3.4x zoom. We did not use line or frame accumulation. Time-lapse z-stacks were collected with $\sim$10 s frame rate and 106 nm x-y pixel dimensions and 0.5 μm separation between z-slices (7 μM range, 16 slices). x-y resolution was 512x512 pixels. Pinhole was set to 1.0 Airy units at 510 nm. eGFP was excited by a 488 nm laser line calibrated to 25 μW using the 10x objective. In all channels, detection was performed using the counting mode of the HyD detectors.

Images of the full embryo at the bottom surface and the mid-sagittal plane were taken after each imaging session in order to precisely locate the field of view along the anterior-posterior axis of the embryo.

4.4 Image and time-series analysis

Image analysis was performed in Matlab using the custom pipeline described in Garcia et al. 2013 and Lammers et al. 2020 (this pipeline can be found in the mRNA Dynamics Github repository). Image segmentation was also aided by the Trainable Weka Segmentation plugin in FIJI (Witten et al. 2016; Arganda-Carreras et al. 2017). Further analysis of time-series and other data were likewise performed in Matlab. Movies for publication were made in FIJI (Schneider, Rasband, and Eliceiri 2012; Schindelin et al. 2012).

5 Supplementary tables

Shorthand name	Scaffold	Enhancer	Promoter	Loop placement	Loops	Gene body	Method of integration	Genome Position	Status (12/19)
0A3	Desplan	None	evePr	Tub 5'UTR	MS2v1	LacZ	RMCE	38F1	Flies
1A3	Desplan	1A3	evePr	Tub 5'UTR	MS2v1	LacZ	RMCE	38F1	No full loop flies
2A3	Desplan	2A3	evePr	Tub 5'UTR	MS2v1	LacZ	RMCE	38F1	No full loop flies
3A3	Desplan	3A3	evePr	Tub 5'UTR	MS2v1	LacZ	RMCE	38F1	Flies
Phase shifted 1A3 (1A3-5,...,1A3+5)	Desplan	1A3	evePr	Tub 5'UTR	MS2v1	LacZ	RMCE	38F1	Some flies (See OneNote)
1A3v5	Desplan	1A3	evePr	Tub 5'UTR	MS2v5	LacZ	RMCE	38F1	Flies?
1A3+5v7	Desplan	1A3+5	evePr	Tub 5'UTR	MS2v7	LacZ	RMCE	38F1	Two Flies
hsp70	Desplan	None	hsp70	Tub 5'UTR	MS2v5	LacZ	RMCE	38F1	One fly
evePr-yellow	Desplan	None	evePr	Tub 5'UTR	MS2v5	yellow	RMCE	38F1	One fly
evePr-pbϕ	Desplan	None	evePr	Tub 5'UTR	MS2v5	LacZ	ϕ c31	VK33	Flies
evePr-hb	Desplan	None	evePr	hb intron	MS2v5	hb null	RMCE	38F1	5 lines
650N-10xZld	650 Neutral + 2x 5XZld (JC)	None	evePr	Tub 5'UTR	MS2v5	lacZ	RMCE	38F1	Two flies
650N-1A3	650 Neutral + 2x 5XZld (JC)	1A3	evePr	Tub 5'UTR	MS2v5	lacZ	RMCE	38F1	Many lines
650N-2A3	650 Neutral + 2x 5XZld (JC)	2A3	evePr	Tub 5'UTR	MS2v5	lacZ	RMCE	38F1	4 lines
650N-1X1	650 Neutral + 2x 5XZld (JC)	1X1	evePr	Tub 5'UTR	MS2v5	lacZ	RMCE	38F1	4 lines
650N-5xZ1A3	650 Neutral + 5xZld1A3 (JC)	5xZ1A3	evePr	Tub 5'UTR	MS2v5	lacZ	RMCE	38F1	?
650N-0A3-v7	650 Neutral No Zld	None	evePr	Tub 5'UTR	MS2v7	lacZ	RMCE	38F1	Many lines
650N-1A3-v7	650 Neutral No Zld	1A3	evePr	Tub 5'UTR	MS2v7	lacZ	RMCE	38F1	Many lines
650N-2A3-v7	650 Neutral No Zld	2A3	evePr	Tub 5'UTR	MS2v7	lacZ	RMCE	38F1	4 lines
650N-1X1-v7	650 Neutral No Zld	1X1	evePr	Tub 5'UTR	MS2v7	lacZ	RMCE	38F1	4 lines
1A3+5v7x96	Desplan	1A3+5	evePr	Tub 5'UTR	MS2v7x96	LacZ	RMCE	38F1	flies
p2-evePr-yellow	Desplan	p2	evePr	Tub 5'UTR	MS2v5	yellow	RMCE	38F1	flies
p2-evePr-hb	Desplan	p2	evePr	hb intron	MS2v5	hb null	RMCE	38F1	flies
650N-1A3FJS-v7x4	650 Neutral No Zld	1A3 Flanking JS	evePr	Tub 5'UTR	MS2v7x96	lacZ	RMCE	38F1	plasmid
650N-2A3FJS-v7x4	650 Neutral No Zld	2A3 Flanking JS	evePr	Tub 5'UTR	MS2v7x96	lacZ	RMCE	38F1	designed
650N-1A3FTD-v7x4	650 Neutral No Zld	1A3 Flanking TD	evePr	Tub 5'UTR	MS2v7x96	lacZ	RMCE	38F1	designed
650N-2A3FTD-v7x4	650 Neutral No Zld	2A3 Flanking TD	evePr	Tub 5'UTR	MS2v7x96	lacZ	RMCE	38F1	designed

Table 1: Summary of Bicoid minimal synthetic enhancer constructs

Chapter 4

Dense Bicoid hubs accentuate binding along the morphogen gradient

1 Abstract

Morphogen gradients direct the spatial patterning of developing embryos, however, the mechanisms by which these gradients are interpreted remain elusive. Here we use lattice light-sheet microscopy to perform *in vivo* single molecule imaging in early *Drosophila melanogaster* embryos of the transcription factor Bicoid that forms a gradient and initiates patterning along the anteroposterior axis. In contrast to canonical models, we observe that Bicoid binds to DNA with a rapid off-rate throughout the embryo, such that its average occupancy at target loci is on-rate dependent. We further observe Bicoid forming transient "hubs" of locally high density that facilitate binding as factor levels drop, including in the posterior where we observe Bicoid binding despite vanishingly low protein levels. We propose that localized modulation of transcription factor on-rates via clustering provides a general mechanism to facilitate binding to low-affinity targets, and that this may be a prevalent feature of other developmental transcription factors.

2 Introduction

Spatial patterning during embryonic development is orchestrated through concentration gradients of regulatory molecules known as morphogens (Turing 1952; Wolpert 1969). The maternally deposited transcription factor (TF) Bicoid (BCD) in *Drosophila melanogaster* was the first identified morphogen (Driever and Nusslein-Volhard 1988a), and remains an iconic and widely studied developmental regulator. Bicoid is distributed in an exponentially decaying concentration gradient along the anteroposterior (A-P) axis of embryos and predominantly regulates the activity of $\sim$100 genes in distinct spatial expression domains ranging from the anterior tip to the middle of the embryo

(Driever and Nusslein-Volhard 1988b; Driever and Nusslein-Volhard 1988a; Struhl, Struhl, and Macdonald 1989; Driever, Siegel, and Nusslein-Volhard 1990).

The ability of BCD and other morphogens to activate different target genes at locations along concentration gradients is classically thought to arise from variations in the number and strength of cognate DNA binding sites within different enhancers (Burz et al. 1998; Lebrecht et al. 2005; Xu et al. 2015), with sharp expression domain boundaries then set through cooperative binding (Ephrussi and St Johnston 2004; Lebrecht et al. 2005). Under this model, enhancers with lower affinity binding sites would only be activated at high concentrations while enhancers with higher affinity sites would also be activated at lower concentrations. This explains how particular enhancers differentially interpret the same gradient to establish spatial domains of gene expression. In recent years this classical model of concentration dependent activation has been challenged through experiments on mutant embryos with flattened BCD distributions which reveal that segment order and polarity can be maintained even without a concentration gradient (Ochoa-espinosa et al. 2009). It has been suggested that instead of a pure concentration dependence, the activation of BCD target genes and the resulting sharp expression domain boundaries is tightly regulated by spatially opposing gradients of repressors (Chen et al. 2012) and the combinatorial actions of other transcription factors (Combs and Eisen 2017).

The recent discovery of the ubiquitous factor Zelda and its role in the regulation of chromatin accessibility (Liang et al. 2008; Harrison et al. 2011; Foo et al. 2014; Li et al. 2014; Xu et al. 2014; Schulz et al. 2015; Sun et al. 2015; Blythe and Wieschaus 2016) and in modulating the timing and strength of BCD (Xu et al. 2014; Hannon, Blythe, and Wieschaus 2017) and Dorsal (Foo et al. 2014) controlled enhancer activation in a concentration dependent fashion has further strengthened the hypothesis that the interpretation of the BCD and other morphogen gradients is more complex than previously thought (Lucchetta et al. 2005).

Extant models cannot, for example, address how BCD is sufficient for activating its targets such as Knirps (Rivera-Pomar et al. 1995) and Hairy (La Rosee et al. 1997), in the posterior of the embryo, where BCD nuclear concentrations are < 2 nM (Morrison et al. 2012) in the short interphase times (5-10 min) of the early nuclear cycles. Since the question of how BCD molecules can find their targets in these short times requires dynamic measurements, genomic assays and biochemical approaches which provide static snapshots have proven inadequate to resolve outstanding mechanistic questions about morphogen activity.

In this study, we address gaps in our understanding of morphogen activity by performing direct measurements of BCD-DNA interactions *in vivo* by single molecule imaging. Single molecule imaging in living cells has been increasingly used in recent years to measure the dynamics of TF-DNA interactions (Liu, Lavis, and Betzig 2015). However, the techniques commonly used are not suitable for whole embryos and thick tissues. Total Internal Reflection (TIRF) and Highly-Inclined illumination (Hi-Lo) (Tokunaga, Imamoto, and Sakata-Sogawa 2008), which have enabled single molecule imaging in monolayer

cell cultures, use wide-field excitation geometries and restrict the illumination volume to a small distance above the microscope coverslip in order to limit the excitation of out-of-focus fluorophores. This confinement of the illumination volume is necessary to achieve the signal-to-background ratios (SBRs) required for single molecule detection. Consequentially, if the thickness of the illumination volume is extended to image further away from the coverslip, the SBR degrades as more and more out-of-focus fluorophore emission raises the background level and reduces contrast. This degradation is further exacerbated when imaging highly autofluorescent samples such as embryos or thick tissues.

Lattice Light-Sheet Microscopy (LLSM) was recently developed to overcome these technical barriers (Chen et al. 2014a). The principle of LLSM is to create an excitation light-sheet that matches the depth-of-field of the detection objective such that only fluorophores that are in focus are excited (Chen et al. 2014a). As in all light-sheet microscopes, in LLSM the excitation and detection objectives are independent and oriented orthogonally to each other. However, unlike conventional light-sheet modalities that use Gaussian beam illumination, in LLSM an array of Bessel beams is used. Light-sheets that are generated with Gaussian beams are generally useful for achieving cellular level resolution over large fields of views but a severe tradeoff exists between the thickness of the sheet and field-of-view (Planchon et al. 2011) and are thus not suitable for imaging with sub-cellular resolution. On the other hand Bessel beams which are optically non-diffracting allow for the generation of light-sheets with sub-micron thickness while maintaining a suitable field-of-view (Planchon et al. 2011). In LLSM, the spacing and phase of Bessel beams in an array is controlled such that their side-lobes destructively interfere in order to achieve maximal axial confinement of the light-sheet while also minimizing photo- toxicity and bleaching by spreading the excitation energy across the array of beams (Chen et al. 2014a). Unlike in wide-field excitation geometries the thickness of the excitation volume in LLSM is independent of the distance from the coverslip which is being probed.

Here, we apply LLSM to developing *Drosophila melanogaster* embryos in order to characterize, for the first time, the single molecule DNA-binding kinetics of BCD along its concentration gradient. We find that BCD, binds to chromatin in a highly transient manner, with specific binding events lasting on the order of a second, in all portions of the embryo. Examination of the spatial distribution of BCD binding events reveals spatiotemporal hubs of high local BCD concentration that increase its local DNA binding on-rate and facilitate specific binding even with such a high off-rate. This effect is most dramatic in posterior nuclei, where, given that there is minimal BCD, we were surprised to observe a significant number of binding events. Through genome wide analysis of BCD-DNA binding on dissected posterior segments of embryos, we show that the binding we observe via single molecule imaging in posterior nuclei occurs at specific regulatory regions, a result that cannot be explained by classical models of BCD activity.

We further find that the regions which are enriched for BCD in the posterior segments are highly correlated with binding of the maternal factor Zelda (ZLD), which has

previously been shown to affect the regulation of BCD targets, especially at lower concentrations. Through single molecule imaging of BCD in ZLD null embryos, we show that ZLD is required for the formation of BCD hubs in the posterior embryo. Together, these data advocate for a model in which ZLD mediates the formation of hubs of high local BCD concentration that facilitate BCD binding and enable the activation of BCD dependent targets at all position along the anteroposterior axis of the embryo.

3 Results

3.1 Single molecule imaging in living Drosophila embryos using Lattice Light-Sheet Microscopy

We constructed a lattice light-sheet microscope (Supplemental Fig. S1) (Chen et al. 2014a) and adapted it to image BCD-eGFP in living Drosophila embryos over a large field of view with high temporal resolution (Fig. 1 A-B, Supplemental Fig. S2 and Supplemental Movies 1-3). We utilize a *yw; His2av-mRFP1; BcdE1, eGFP-Bcd* fly line in which only the labelled BCD is expressed indicating proper functionality and expression levels (Gregor et al. 2007) and to ensure that all molecules we observe are functionally relevant. The single molecule nature of the data is reflected in the distribution of intensities (Supplemental Movie 2 and Supplemental Fig. S2 A-B) and discrete characteristics (Supplemental Movie 2 and Supplemental Fig. S2 C-D) of the observed binding events.

Bicoid nuclear concentrations, as measured by two-photon microscopy, on the same fly line we utilize in this work, range from $\sim$50 nM at the anterior most positions down to $<$ 2 nM (Morrison et al. 2012) in the posterior. This translates to a range on the order of 104-102 BCD molecules/nucleus. To gain some preliminary insight on what to expect when imaging with the LLSM, we assume an isotropic distribution of molecules and a 400 nm thick excitation sheet and estimate a range on the order of 103-101 BCD molecules per imaging plane in a single nucleus. This simple calculation provides an intuitive feeling of why it is possible to perform single molecule imaging using BCD-EGFP. This range of concentrations is reflected in the data shown in Supplemental Movie S1, where unambiguous single molecule detections can be seen in the middle and posterior positions from the start, whereas in the anterior positions they can only be detected when a sufficient amount of bleaching has occurred. This natural concentration range allows us to perform single molecule tracking at all positions in the embryo without utilizing sparse labelling strategies or photo-switchable fluorophores.

3.2 Bicoid binds chromatin in a highly transient manner across the concentration gradient

At high concentrations of Bicoid in the anterior, the on-rate and thus time average occupancy is high at both low- and high- affinity binding sites. Under the classical model,

the strength of binding sites within specific enhancers is set such that the time average occupancy varies depending on the affinity of binding sites and position along the concentration gradient. Thus, at vanishingly low concentrations in the posterior embryo, the time average occupancy at all enhancers is expected to be low, since even the highest affinity sites would not be occupied frequently enough to drive expression. To test this model at the single molecule level we therefore first performed single molecule imaging and tracking at long (100 msec) exposure times, effectively blurring out the fast moving (unbound) population (Supplemental Movies 1-2 and Supplemental Fig. S2) (Chen et al. 2014b) to estimate the residence times (RTs) of BCD binding in nuclei at all positions along the A-P axis.

Previous single molecule studies of transcription factors have consistently found two populations in the survival probability distributions, a short-lived population with RTs on the order of hundreds of milliseconds, and a longer-lived population with RTs on the order of tens of seconds to minutes (Chen et al. 2014b; Normanno et al. 2015; Hansen et al. 2017). These two populations have often been shown to be the non-specific and specific binding populations, respectively. The survival probability distributions of BCD similarly are fit better with a two-exponent model than with a single-exponent model indicating the presence of two sub-populations (Supplemental Fig. S3). Fits to the survival probability distributions of BCD binding events (Fig. 1C and Supplemental Table 1) in the anterior, middle, and posterior thirds of the embryo identified short-lived populations with average RTs (after photo-bleaching correction) on the order of 100s of milliseconds and longer lived populations with average RTs on the order of one second, and with no significant dependence on position along the A-P axis for either population (Supplemental Table 1). The validity of our RT estimation of specific binding on the order of 1 second is supported by additional measurements at 500 msec exposure times (Supplemental Fig. S4 and Supplemental Table1). We note that the values reported here are the genome wide averages for both the long and short-lived binding populations. It is also likely that there are even shorter-lived BCD-DNA interactions that we cannot access due to the practical trade-off between temporal resolution and signal-to-noise ratio in single molecule imaging.

To further validate our observation that the RT of the long-lived population of BCD is highly transient compared to the 10-60 sec typically observed for other sequence specific DNA binding TFs using single molecule tracking (Chen et al. 2014b; Hansen et al. 2017), we performed Fluorescence Recovery After Photobleaching (FRAP) experiments on BCD in the embryo (Fig. 1D) which revealed halftimes of BCD recovery on the order of hundreds of milliseconds (Supplemental Fig. S5). The rapid dynamics of BCD indicated by the FRAP data are consistent with previous measurements by others using Fluorescence Correlation Spectroscopy (FCS) (Porcher and Dostatni 2010). Although the FCS measurements on nuclear BCD dynamics (Porcher and Dostatni 2010) were previously only analyzed to estimate the diffusion coefficients of a fast and slow moving population, when this data is re-analyzed using a "stick-and-diffuse" model (Yeung, Shtrahman, and Wu 2007), which takes into account binding events between diffusion, the RT for a

short-lived binding population of 122 msec is found, consistent with our measurements (C. Fradin, Personal Communication and manuscript in preparation). These results confirm the transient nature of BCD binding, however, due to the limitations on the dynamic range of FCS measurements and the kinetic modeling utilized, the less prevalent, longer-lived binding population we quantify with single molecule imaging cannot be detected. The dominance of the short-lived interactions (Fig. 1C) highlights the preponderance of low-affinity BCD binding sites in the Drosophila melanogaster genome and the resulting large number of non-specific interactions, as previously suggested by genomic studies (Ochoa-Espinosa et al. 2005; Li et al. 2008). The observation of a significant number of binding events in posterior nuclei is surprising as we expected the majority of the few BCD molecules left to be diffusing and binding too infrequently to specific targets to be detected.

3.3 Spatiotemporal hubs of Bicoid binding enrich local concentrations in the posterior embryo

The observation of significant binding events in the posterior embryo where BCD has been reported to be at vanishingly low ($<$ 2nM posterior vs $\sim$ 50 nM anterior) concentrations (Morrison et al. 2012) motivated us to next measure the fraction of the BCD population that is diffusing vs. bound along the concentration gradient. Since longer exposure times only allow detection of molecules bound for at least the span of the exposure and do not provide any data on the mobile population that is blurred into the background, we performed single molecule tracking measurements at a decreased exposure time of 10 msec. The signal-to-noise ratios at these lower exposure times are adversely affected as expected, limiting the type of analysis that can performed on this data (Supplemental Movie 3). However, despite this reduced contrast, we were able to perform single particle tracking, and through analyses of displacement distributions (Supplemental Fig. S6), we estimated the fraction of BCD that is bound along the A-P axis (Fig. 2A). Surprisingly, a greater fraction of the BCD population is bound in more posterior positions of the embryo where BCD is present at the lowest concentrations.

This counter-intuitive result, which suggests that the on-rate of BCD-DNA binding is decoupled from its nuclear concentration, motivated us to re-examine the 100 msec exposure-time data set. One way to resolve this discrepancy would be for BCD to be restricted to a small volume within the nucleus, increasing its effective local concentration. Indeed, when we analyzed the spatial distribution of BCD binding events in the 100 msec data set we see a distinct spatio-temporal clustering (Supplemental Movie S4) of binding events that becomes more pronounced toward posterior positions (Fig. 2B and Supplemental Fig. S7) with a greater fraction of binding events occurring within clusters (Supplemental Fig. S8). Remarkably, although the number of binding events per nucleus follows the trend dictated by the global concentration gradient across the embryo (Fig. 2C), the distribution of BCD molecules detected per cluster is maintained

even in the posterior (Fig. 2D).

Together these data suggest that BCD forms hubs with high local concentration that lead to high time-averaged occupancy at specific sites in nuclei across the A-P axis. More hubs are formed at higher concentrations, but the characteristics of hubs are independent of position along the BCD concentration gradient. The surprising observation that BCD binding is concentrated in hubs led us to next ask if there is a mechanisms to enrich functional BCD target sites in these regions. To do this we concentrated on posterior positions where the majority of binding events are occurring within the hubs.

3.4 Bicoid binds specific regulatory regions in the posterior embryo in a Zelda dependent manner

To test whether BCD is binding with specificity in the posterior embryo we analyzed its binding profiles in a spatially segregated manner (Combs and Eisen 2013) by comparing ChIP-seq profiles derived from individually dissected posterior thirds of embryos to previously published data from whole embryos (Bradley et al. 2010). Our analysis reveals that BCD indeed binds to known targets in the posterior but with increased relative enrichment at specific enhancer elements (Fig. 3A). For example, in the *hunchback* locus, binding at the posterior stripe enhancer (Perry et al. 2012) is highly enriched over the background in nuclei from the dissected posterior third relative to the whole embryo. Intriguingly, genomic regions that exhibit a relative increase in BCD occupancy in the posterior are correlated with an enrichment of Zelda (ZLD) binding (Fig. 3A and Supplemental Fig. S9), a ubiquitous activator often described as a pioneer factor active during early embryonic development (Liang et al. 2008; Harrison et al. 2011; Foo et al. 2014; Xu et al. 2014).

Remarkably, enrichment of ZLD is more predictive of BCD binding in the posterior than previously determined positions of enhancer activity for the loci shown in Fig. 3A. Analysis of the correlation between BCD and ZLD enrichment at the *cis*-regulatory modules of 12 gene loci, and at ZLD and BCD peaks genome-wide reveal that binding of BCD in posterior nuclei is highly correlated with ZLD co-binding (Fig. 3A, Supplemental Figs. S9 and S10). The difference in the binding profiles of the posterior third segments compared to whole embryos emphasizes the need to perform genomic analysis in a spatially resolved manner across the anteroposterior axis. While we provide a proof of principle on how to perform these measurements though manual dissection, improved methods are required in order to improve the throughput of these experiments.

3.5 Formation of Bicoid hubs in the posterior embryo is dependent on Zelda

The posterior-thirds genomic data and the published evidence for Zelda's role in the regulation of chromatin accessibility (Liang et al. 2008; Harrison et al. 2011; Foo et al.

2014; Li et al. 2014; Xu et al. 2014; Schulz et al. 2015; Sun et al. 2015) and its suggested role in the modulation of TF binding at low concentrations (Xu et al. 2014; Schulz et al. 2015) naturally led us to hypothesize that the observed clustering of BCD binding events may be, in part, mediated by ZLD. We thus generated *zelda*-null embryos with BCD-eGFP and measured BCD binding at 100 msec exposure times. We found an abolishment of BCD hubs in the posterior embryo and a small decrease in RTs (Fig. 3C, Supplemental Table S1). Due to this loss of clustering the same analysis that was performed in the wild type case (Fig. 2C) could not be performed in the ZLD mutants. We thus calculated the pair-correlation function for the spatial distribution of binding events in both the wild type and ZLD mutants (Cisse et al. 2013). This analysis allows us to infer whether binding events are spatially randomly distributed or clustered (Supplemental Fig. S11). Both the magnitude and correlation length indicate a diminishment of clustering in the posterior nuclei of the ZLD mutants. We also note that due to the lower labelling density of BCD in the ZLD-null embryos, the presumably ZLD independent clustering in the anterior embryo now becomes more apparent (Supplemental Fig. S11). The loss of clustering in the ZLD mutants also confirms that the clustering we originally observed is not due to aggregation of eGFP, simply non-homogenous distribution of Bcd within nuclei or other artifactual reasons. To provide further validation that the observed clustering is not occurring by chance, we calculated the pair-correlation function for randomly distributed points in a disc approximately the diameter of the nuclei for comparison (Supplemental Fig. S11).

Although the exact mechanism by which ZLD mediates BCD hub formation and binding remains unclear, we speculate that a combination of protein-protein interactions facilitated by intrinsically disordered low-complexity domains (Hamm, Bondra, and Harrison 2015) of ZLD, and its reported role in promoting chromatin accessibility (Foo et al. 2014; Li et al. 2014; Schulz et al. 2015; Sun et al. 2015) may contribute to BCD clustering (Fig. 4).

4   Discussion

Our initial observation of the low-affinity nature of BCD binding to chromatin is partially congruent with the classical view of BCD as a concentration-dependent morphogen. High BCD concentrations in the anterior embryo lead to high on-rates along with high chromatin occupancy, with the high off-rate enabling frequent sampling of both specific and non-specific sites (Fig. 4). As the BCD concentration decreases posteriorly along the gradient, there should be progressively lower on-rates and reduced binding to specific sites, regardless of the roles of opposing repressor gradients. However, if this is all there is to BCD binding, then this would consign BCD to have no binding or function in more posterior positions, contradicting a wealth of evidence pointing to a specific role for BCD in the regulation of posterior gene expression (Rivera-Pomar et al. 1995; Small, Blair, and Levine 1996; La Rosee et al. 1997; Burz et al. 1998; Ochoa-espinosa et al. 2009;

Chen et al. 2012).

Our demonstration that BCD overcomes the combination of low concentrations and high off-rate by forming transient hubs highlights the power of dynamic, single molecule studies on living embryos. Although most vivid in posterior nuclei, this phenomenon is likely to play a significant role in regulating the large number of important known BCD targets in the middle of the embryo. Unlike BCD targets in the most anterior region, a large number of BCD targets in the middle of the embryo are dependent on ZLD, and our *in vivo* imaging data provide an elegant model for how this is accomplished, in which ZLD binds to BCD targets, triggers the clustering of BCD (likely via low-complexity region mediated interactions among ZLD molecules and between ZLD and BCD), thereby increasing the effective concentration of BCD at its targets. We speculate that this protein-protein interaction mediated clustering acts in tandem with Zelda's known role in promoting chromatin accessibility to facilitate binding to low-affinity enhancers.

The formation of such clusters or hubs has been reported for other TFs (Chen et al. 2014b; Liu et al. 2014; Crocker, Tsai, and Stern 2017) and for RNA Polymerase II (Cisse et al. 2013), indicating that such spatial organization of the nucleus may be a general mechanism to catalyze important regulatory interactions. During embryonic development, it is likely that clustering of TFs mediated by co-factors has evolved to allow exquisite spatial and segmental modulation during development through enabling interactions with low-affinity enhancers (Crocker, Ilsley, and Stern 2016).

5   Materials and Methods

5.1   Fly husbandry

All fly cages were prepared by combining males and females of the desired strains in a plastic cage left at room temperature in light at least 3 days prior to imaging. The lids on the cages were filled with agar dissolved in apple juice (2.4% g/w Bacto agar, 25% apple juice, 75% distilled water, and 0.001% of mold inhibitor from solution of 0.1 g/mL (Carolina 87-6165). A paste of dry yeast was smeared on the lids to induce egg laying. Lids were exchanged once each day.

The fly strain used for all wild type BCD imaging experiments was: *yw; his2av-mrfp1; BcdE1, egfp-bcd*. This fly line results in embryos where only the labelled BCD is expressed indicating proper functionality and expression levels (Gregor et al. 2007). For the *zld-* experiments, *bcd-egfp* heterozygous virgins with *zld-* germline cells (maternal germline clones prepared as in (Liang et al. 2008)) were crossed to *yw* males, and progeny were used for imaging 2-3 h after laying. The heterozygosity results in only half of the BCD labelled in the *zld-* embryos. For photo-bleaching controls, the line used was *yw, his2av-egfp; +/+* (Bloomington # 24163).

5.2   Live embryo collection for imaging

For embryo collection, lids on fly cages were exchanged one hour prior to imaging. After one hour, embryos were collected from the lids using an inoculation loop. A 5 mm diameter glass cover slip (#64-0700, Warner Instruments) was prepared by immersion in a small amount of glue (prepared by dissolving adhesive from about 1/5 of a roll of double-sided Scotch tape overnight in heptane) and was left to dry for 5-10 min while collecting embryos. Collection was performed on a dissection scope with trans-illumination. Embryos were bathed in Halocarbon oil 27 (Sigma) for staging and then selected between developmental stages 1 and 4. Selected embryos were placed on a small square of paper towel, then de-chorionated in 100% bleach for 1 min. Bleach was wicked off with a Kimwipe after one minute, then the square was washed with a small amount of distilled water. Excess water was wicked off the square and the square was dipped in a small water bath. Unpunctured embryos that floated to the top of the bath were selected for imaging and placed on a small paper towel square to slightly dry. To prevent excess desiccation embryos were immediately placed on the glass cover slip in rows and then immersed in a drop of phosphate-buffered saline (PBS).

5.3   Lattice Light-Sheet Microscopy

Imaging was performed using a home built lattice light-sheet microscope (Supplemental Fig. S1) following the design described by Chen et al. 2014a and detailed blueprints provided by the Betzig group at HHMI Janelia Research Campus. To perform the single molecule experiments we added a detection module containing two EMCCDs (Andor iXon Ultra) for dual color imaging. The EMCCDs provided a significant improvement in signal-to-noise over the sCMOS (Hamamatsu Orca Flash V2.0) used in the original system and made it possible to use lower excitation powers while maintaining single molecule sensitivity. In brief, the output beam from each laser is expanded and collimated independently to a size of 2.5 mm. The expanded beams for each laser are combined and input into an Acoustic Optical Tunable Filter (AOTF) to allow for rapid switching between excitation wavelengths and adjustment of power (Supplemental Fig. S1A). A pair of cylindrical lenses is then used to elongate and collimate the output Gaussian beam to illuminate a stripe on a spatial light modulator (SLM). The SLM is used to generate a coherent pattern of an array of 30 Bessel beams spaced such that they coherently interfere to create a 2D optical lattice pattern with a maximum numerical aperture (NA) of 0.6 and minimum NA of 0.505. A 500 mm lens is used to project the Fourier transform of the SLM plane onto an annular mask, conjugate to the back pupil plane (BPP) of the excitation objective, to spatially filter the pattern (Supplemental Fig. S1B). The BPP is then projected first onto a galvo scanning mirror for z-scanning and then onto a second galvo scanning mirror for x-dithering. The x-galvo scanning plane is projected onto the BPP of the excitation objective (Supplemental Fig. S1C). The excitation objective focuses the lattice pattern onto the sample, exciting any fluorophores within the

axial range ($\sim$400 nm) of the sheet. The emitted fluorescence is collected by the detection objective which is oriented orthogonally to the excitation objective and projected onto an intermediate image plane by a 500 mm tube lens (Supplemental Fig. S1D). An 80 mm and 200 mm lens pair is then used to de-magnify the image further to provide a 100 nm sampling per pixel on each of the EMCCD sensors. A dichroic mirror (Semrock FF560-FDi01) is placed between the last lens pair to allow for dual color imaging in red and green with maximal spectral separation. An emission filter is placed in the path of each camera to both reject the excitation wavelengths and also select the wavelength range of interest (Semrock FF03-525/50 for eGFP and FF01-593/46 for RFP) (Supplemental Fig. S1E). During each camera exposure the x-galvo mirror is dithered twice over a 5.1 μm range in 100 nm steps to provide uniform illumination.

The prepared coverslip, with embryos arranged in rows as described above, was then loaded into the sample holder and secured on to the positioning stages of the lattice light-sheet microscope. The sample chamber was filled with PBS for imaging and kept at room temperature. The slide was then scanned to find an embryo of suitable age (between nuclear cycles 10 and 11) and the positions of the anterior and posterior extremes of the embryo were then marked. For each data set acquired, the stage position was recorded to determine the position as a fraction of the embryonic length (EL) as the distance of the position from the anterior pole divided by the total length of the embryo.

For the residence time measurements on BCD-eGFP, a 488-nm excitation laser was used with a power of 2.9 mW measured at the back pupil plane of the excitation objective. Images were acquired with 100 msec exposure times and an EM Gain setting of 300. At each location at least 1000 frames were acquired resulting in a total time of 105 sec, with a frame rate of 105 msec. Prior to and after acquiring the BCD-eGFP data an image was taken in His2-AVmRFP using a 561 nm excitation laser at an excitation power of $\sim$0.17 mW at the back pupil plan to determine the nuclear cycle phase, data not acquired during interphase was discarded upon examination of these images. For residence time measurements at 500 msec exposure times, the 488-nm excitation laser power at the back pupil plane was reduced to 0.5 mW all other settings were the same as above. For the displacement distribution measurements the exposure time was set to 10 msec, resulting in a frame rate of 15 msec and the excitation power was increased to 8.28 mW for the 488-nm laser line. All other settings were the same as described above. Viability of the embryos was determined by allowing them to develop until gastrulation after imaging. For the Zelda- embryos, lethality was confirmed after imaging.

5.4   Curation of data for analysis

For all data sets the following procedure was followed: First, for each movie the corresponding before-and-after histone images were checked for any evidence of chromatin condensation to ensure that analysis was only performed in interphase nuclei. Data from mid-to-late nuclear cycle 14, where the nuclei exhibit an elongated shape, were

also excluded. A metadata file was then created for each movie file containing the position as a fraction of the embryonic length (0 for anterior, 1 for posterior), the nuclear cycle (determined by counting the number of mitoses before the 14th cycle). Visual examination of the data set was used to determine if there was any motion of the nuclei during the acquisition period. Movies that contained any detectable motion were discarded or cropped to only include the time interval where there was no motion. A rectangular region of interest was then drawn around each nucleus which was then used to crop areas around individual nuclei. The boundary of each nucleus was then marked using a hand-drawn polygon. A masked movie was then created for each nucleus where regions outside the nucleus were set to 0 gray scale values so that all the analyses described below were only performed on molecules within nuclear regions.

5.5 Single molecule localization and tracking using Dynamic Multiple-target tracing (MTT)

Localization and tracking of single molecules was conducting using a MATLAB implementation of the MTT algorithm (Sergé et al. 2008). In brief, the algorithm first performs a bi-dimensional Gaussian fitting to localize particles constrained by a log-likelihood ratio test subject to a localization error threshold. Deflation looping is performed to detect molecules that are partially overlapping. The parameters of the localization and tracking algorithms were empirically determined through iterative examination of the results. For all data sets the following settings were used: For localization, maximum number of deflation loops was set to 10, localization error to 10^{-6}. For tracking the maximum expected diffusion coefficient was set to 5 $\mu m^2/s$, maximum number of competitors was set to 1, and maximum off/blinking-frames was set to 1.

5.6 Residence time analysis

The residence times were estimated from the 100-msec data using the results from the single particle tracking using the MTT algorithm. The data was pooled into bins corresponding to the position along the A-P axis of the embryo in 1/3 fractions of the embryonic length (EL, 0-1 anterior to posterior), with the following number of nuclei and single molecule trajectories per position bin: Anterior: 34 nuclei, 17735, trajectories Middle: 70 nuclei, 40092 trajectories Posterior: 83 nuclei; 20823 trajectories

In the ZLD-null embryos, we measured the following number of nuclei and single molecule trajectories per position bin: Anterior: 23 nuclei, 11415, trajectories Middle: 31 nuclei, 7572 trajectories Posterior: 31 nuclei; 3606 trajectories

The survival probability distribution was then calculated as 1 - the cumulative distribution function of the trajectory lengths and was fit to both single and double exponential models (Mazza et al. 2012). The double exponential model fit the data better in all cases (Supplemental Figs. S3 and S4). The model used to fit the data and to calculate

the time constants and fraction of the population is:

$$\text{Survival Probability(t)} = A(F_a * e^{-k_s * t} + (1 - F_a) * e^{-k_{ns} * t}) \tag{1}$$

Where k_s and $k_n s$ are the low (specific) and high (non-specific) off-rates respectively. The total pooled data sets of 78650 trajectories from the MTT results from 187 nuclei were also fit in the same manner.

To correct for photo-bleaching we used a His2Av-eGFP to estimate the bleaching constant (0.00426 s^{-1}) and correct the off-rate as $k_{s,corrected} = k_s - k_{bleach}$.(Supplemental Table S1). We note that the bleaching correction has minimal effect on our estimated off-rates. The fact that we are not limited by bleaching due to the transient nature of BCD binding is further validated through even longer, 500 msec exposure time measurements on 17 nuclei which provides an estimate for the specific and non-specific off-rates that do not differ significantly from those measured at 100 msec (Supplemental Fig. S4 and Supplemental Table S1). The results of the fits to all data are shown in Supplemental Table S1.

5.7 Fluorescence Recovery After Photobleaching (FRAP)

Experiments were performed on a Zeiss LSM 800 laser scanning confocal system (coupled to a Zeiss Axio Observer Z1) using a Plan-Achromat 63x/1.4 NA Oil immersion objective and GaAsP-PMT detector and a 488-nm laser. A circular bleach region with a diameter of 1.5 μm was used and the bleach location was selected manually in each nucleus at approximately the center and a total bleach time of 78.1 msec was used. Data was acquired for at least 1 sec prior to bleaching and for at least 6 sec after bleaching, with a time interval of 0.430 msec. Experiments were performed using live embryos from the same fly-line as for lattice light-sheet imaging and which were collected and prepared in the same manner as described above with the exception of the mounting procedure. For FRAP the embryos were mounted between a semipermeable membrane (Biofolie; In Vitro Systems Services) and a coverslip and then embedded in Halocarbon 27 oil (Sigma). As in the case of the single molecule measurements, the embryos were staged using the HIS2-AV-MRFP channel and all data was acquired on embryos in nuclear cycle 13. The data was acquired on 21 nuclei all within the first 25% of the embryonic length from the anterior pole. FRAP experiments could not be performed at more anterior locations due to the low BCD concentrations and thus signal-to-noise ratio.

For analysis, the spatial and temporal location of the bleach region was retrieved from the meta-data recorded by the microscope control software and manually verified by inspecting the data. Due to the short duration of the movies and rapid recovery drift correction was not necessary. Each nucleus was manually segmented from the rest of the image by defining a polygon region of interest. A region of interest the same size as the bleach region was also marked in an area of each image outside of the nuclear region to be used to measure the background or dark intensity. The FRAP curve for each nucleus was then calculated as follows:

First, the mean intensity in the nuclear region, $I_{nuc}(t)$, bleach region, $I_{bleach}(t)$, and background region, $I_{dark}(t)$, was calculated at each frame.

The background intensity was them subtracted from both the mean nuclear and bleach region intensities. To correct for photo-bleaching from imaging the bleach region intensity was then divided by the mean nuclear intensity at each time point. The resulting photo-bleaching corrected intensity was then normalized to the mean pre-bleach intensity, calculated as the average intensity in all frames prior to bleaching ($I_{pre-bleach}$). The final FRAP curve for each nucleus was thus calculated as:

$$FRAP(t) = [(I_{bleach}(t) - I_{dark}(t)/(I_{nuc}(t) - I_{dark}(t))]/I_{pre-bleach} \qquad (2)$$

The calculated FRAP curves for each nucleus were then aligned to the bleach frame and averaged to generate the average FRAP curve shown in Fig. 1D. The averaged recovery data was then fit to both a single exponential $(1 - A * exp([-ka * t])$ and double exponential $(1 - A * exp([-ka * t] - B * exp([-kb * t])$ model (Supplemental Fig. S5 A-B), to measure the recovery time constants. There is no significant difference in the quality of fit between the two models. For comparison, the exact same experimental and analysis procedure was followed for His2AV-RFP1 (Supplemental Fig. S5C) in the same embryos with the exception of using a 561-nm bleach laser, 3 nuclei were measured with these settings. For the histone measurements, the two-exponent fit was significantly better than the one-exponent as expected.

5.8 Displacement Distribution Analysis

Single molecule trajectories were analyzed as described above. A total of 158 nuclei from 4 embryos were analyzed. The data from nuclei were binned according to their position along the A-P axis in 1/3 fractions of the embryonic length (EL, 0-1 anterior to posterior), with the following number of nuclei and single molecule displacements per bin: In the anterior most (0-0.2) positions, tracking could not be performed reliably due to the high concentrations of Bicoid at those locations at these high frame rates and presumably a large mobile population.

Anterior: 30 nuclei, 12923, trajectories Middle: 67 nuclei, 23640 trajectories Posterior: 66 nuclei; 8600 trajectories

The fraction of the population that is bound vs. mobile was estimated using two approaches. First, a cumulative distribution function of the displacements was calculated for each EL bin (Supplemental Fig. S6), displacements corresponding to distances less than a value of 225 nm were scored as part of the bound population. In the second approach the probability density functions of the displacements were fit to a two population model (Supplemental Fig. S6B).

$$P(r) = F_{Bound}\frac{r}{A}e^{\frac{r^2}{2A}} + (1 - F_{Bound})\frac{r}{B}e^{\frac{r^2}{2B}} \qquad (3)$$

Where F_{Bound} is the fraction of the population that is considered bound and r is the displacement distance. The two population model fits the displacement data with R^2 values of 0.97, 0.98, and 0.96, for the anterior, middle, and posterior distributions, respectively, and provides an estimate of the fraction present in the mobile and bound populations. The trend of an increase in the population bound estimate in the middle and posterior positions relative to the anterior is similar from both approaches. Although the bound population includes non-specific binding events as well, what we are interested in the relative change across the A-P axis.

We note that the diffusion coefficients for the bound and mobile populations can in principle be estimated from displacement distributions, however, as explained by Mazza et al. 2012, they cannot be estimated accurately from a fit to a displacement distribution at a single time step and are thus not reported. To accurately measure the diffusion coefficients more stable and photo-switchable fluorophores are necessary to be able to track single molecules for more time points and at higher temporal resolutions.

5.9 DBSCAN based analyses of clusters

The clustering of Bicoid binding events in the 100 msec wild type embryos data set is readily apparent in nuclei across the A-P axis in the projection of all localizations from the MTT algorithm (Fig. 2B). In order to automatically identify clusters and count the number of detections per cluster vs. the whole nucleus (Fig. 2C), a MATLAB implementation (Yarpiz Team 2015) of the widely used Density-based spatial clustering of applications with noise (DBSCAN) was used with a minimum points setting of 10 points and an epsilon (maximum radius of neighborhood) setting of 0.8. These settings were empirically determined by iteratively changing parameters and examining the results. A balance had to be struck in the ability to accurately identify clusters in the high density situations in anterior nuclei and also low density situations in the posterior. Only data sets with time spans of at least 105 sec were included. A total of 12, 49, and 48 nuclei and 436, 1168, and 367 clusters were analyzed in this manner for the Anterior, Middle, and Posterior positions respectively. Examples of clusters identified in nuclei at various positions are shown in Supplemental Fig. S7. For comparing distributions, outliers were removed (<5th or >95th percentile) to disregard clusters that were significantly over or under-partitioned. In the case of the ZLD-null embryos, due to the loss of apparent clustering, DBSCAN was not able to detect clusters, so instead the spatial distribution of points was compared using pair-correlation analysis as described below to provide insight on the change beyond visual examination.

5.10 Pair correlation analysis of clustering

The pair correlation function for the spatial distribution of particles computes the probability of finding a particle at the range of distances from another particles. In the case of complete spatial randomness which can be represented by a Poisson process the

pair correlation function is equal to 1. The analysis is conducted on point lists generated from the MTT algorithm using the first spatial coordinate of each trajectory. A total of 22, 48, and 42 nuclei were analyzed for the Anterior, Middle, and Posterior positions respectively for the wild type embryos and a total of 23, 31, and 31 nuclei were analyzed for the Anterior, Middle, and Posterior positions respectively for the *zld-* embryos. A MATLAB implementation was used to calculate the correlation function for the spatial distribution in each nucleus, the results were then averaged for each position (Anterior, Middle, Posterior) embryo for comparison (Supplemental Fig. S11). To generate simulations of randomly distributed points, spatial coordinates of detections were randomly generated (using a MATLAB script and random draws from a uniform distribution) to lie within a 5 μm disc, simulations were performed with 5000, 3000, and 1000 points to correspond to detections at Anterior, Middle and Posterior positions.

5.11   Chromatin Immunoprecipitation and Sequencing (ChIP-Seq)

Embryos were collected from a population cage for 90 min and then aged for 2 h in order to enrich for embryos at developmental stage five. Embryos were then fixed with formaldehyde as previously described (Li et al. 2008) and sorted by morphology for those at early stage 5. The posterior thirds of embryos were sliced off by hand with a scalpel. A pool of the embryo segments from approximately 300 embryos was combined with 20 μg of whole *Drosophila pseudoobscura* embryos at stage 5 (to serve as carrier), and homogenized in homogenization buffer containing 15 mM NaCl, 15 mM TrisHCl pH 7.5, 60 mM KCl, 1 mM EDTA, and 0.1% Triton-X100, with 1 mM DTT, 0.1 mM PMSF, and protease inhibit cocktail (Roche) added before use. After homogenization, 0.5% NP40 was added, and following a 5-min incubation, samples were spun down at a low speed. Nuclei in the pellet were then washed once with the homogenization buffer containing 0.2 M NaCl. The low-speed centrifugation was repeated, and the recovered nuclei pellet was then re-suspended in nuclear lysis buffer (10 mM TrisCl, pH 7.9, 100 mM NaCl, 1 mM EDTA, 0.5% sarkosyl) +1%SDS, +1.5% sarkosyl. The chromatin was recovered by spinning the sample at full speed in a micro-centrifuge at 4°C for 1 h and was re-suspended in a small volume of nuclear lysis buffer. Chromatin was sheared to an average size of 300 base pairs (bp) using a Covaris sonicator (peak power, 140; duty factor, 2; cycle burst, 200; time, 2:20 min). Chromatin immunoprecipitation was performed using 72 ng of chromatin and 1.5 ug of an anti-Bicoid polyclonal antibody described previously (Li et al. 2008). The Bicoid antibody was coupled to Dynabead M-280 sheep anti-rabbit magnetic beads and the immunoprecipitation was conducted with the standard protocol from the manufacturer. DNA libraries for the chromatin immunoprecitation samples were prepared using the Rubicon genomics thru-plex DNA-seq kit using 16 PCR cycles and sequenced using Illumina High-seq with 2500 rapid-run 100-bp single end reads. The sequencing reads were aligned to a combined *D. pseudoobscura* (Flybase Release 1.0) and *D. melanogaster* genome (Flybase Release 5) using Bowtie with options set as -3 70 -n 2 -m 1. The aligned reads were converted to WIG files using custom scripts available

on request. WIG files were normalized to 10 million mapped reads. The data have been deposited in the NCBI Gene Expression Omnibus (GEO) database with accession number GSE103695.

5.12   ChIP-Seq analysis

The posterior embryo ChIP-seq data was compared to published data on whole embryos from the same developmental stage (Bradley et al. 2010) downloaded from the NCBI GEO database with accession number GSM511083. To compare against Zelda binding, previously published data (Harrison et al. 2011) was downloaded from the NCBI GEO database with accession number GSM763061. Since the purpose of the analysis was to compare relative enrichment at genomic loci over each data sets respective background, the following normalization procedure was used: First, for each chromosome the average of the signal over the entire chromosome (a proxy for the background signal) was subtracted, negative values were then treated as below the background and discarded. The data was then normalized to the background subtracted average of each chromosome such that the normalized data now represents enrichment over background. For visualization data was smoothed using a Savitzky-Golay filter of order 3 over a 0.250-kbp sliding window. For analysis on CRMs the RedFly annotation database was used. For analysis centered on called peaks in either the whole embryo BCD data or the Zelda data .bed files containing peak locations were downloaded from the NCBI GEO database.

6   Acknowledgements

7   Figures

Figure 1: **Single molecule kinetics of Bicoid in living Drosophila Embryos.** (A) Raw images of BCD-eGFP molecules in a living Drosophila embryo acquired with a 100 millisecond exposure time. Scale bar is 5 μm. Positions along the A-P axis are shown as a fraction of the Embryonic Length (EL, x/L). (B) Example of a single molecule binding event. Top row shows raw images from a 1.2 x 1.2 µm area, bottom row shows corresponding surface plot representations to illustrate the signal-to-noise. (C) Uncorrected survival probability curves for Bicoid binding (markers) in the Anterior (34 nuclei), Middle (70 nuclei) and Posterior (83 nuclei) segments of the embryo and corresponding fits to a two-exponent model (solid lines) show no significant differences. (D) FRAP curve for Bicoid shows a recovery time on the order of hundreds of milliseconds, error bars show standard deviation over 21 nuclei.
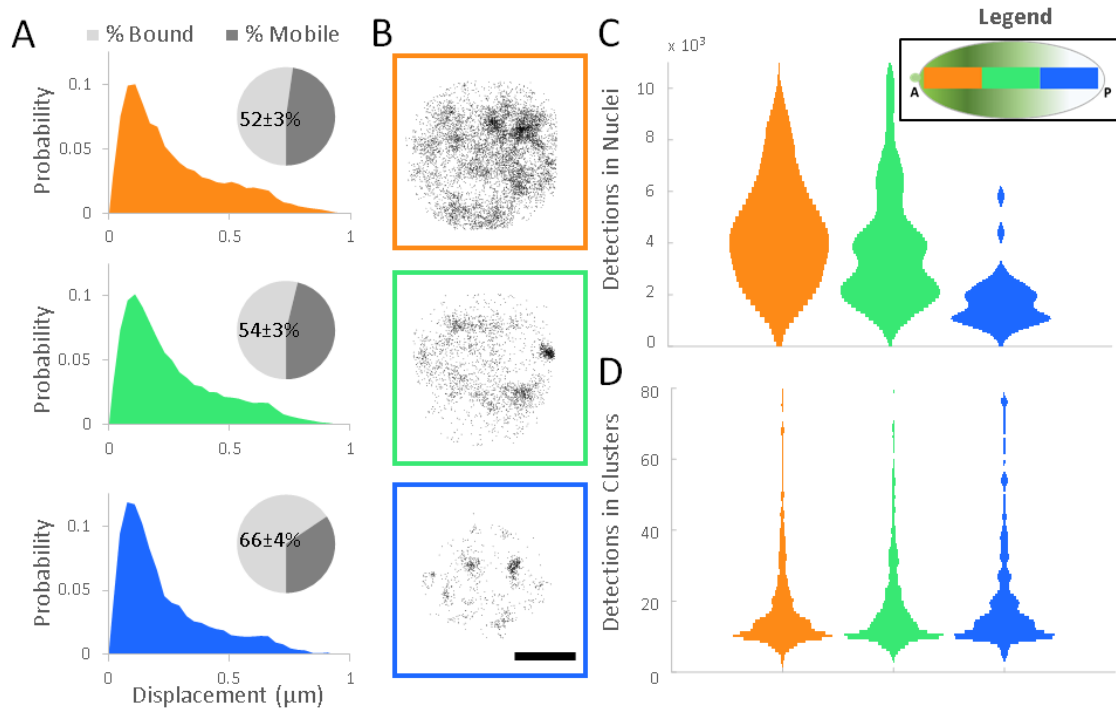
Figure 2: **Local modulation of Bicoid Concentration** (A) Normalized probability distributions of measured displacements in the Anterior (30 nuclei), Middle (67 nuclei), and Posterior (66 nuclei) positions of the embryos, pie charts show the estimated mobile and bound fractions from fits to a two population distribution with the bound population percent labelled with the standard error of the fit parameter. (B) Examples of the spatial distribution of all detections in nuclei along the A-P axis, scale bar is 2.5 μm. (C) Distribution of the number of detections in all nuclei. (D) Distributions of the number of detections within all clusters.
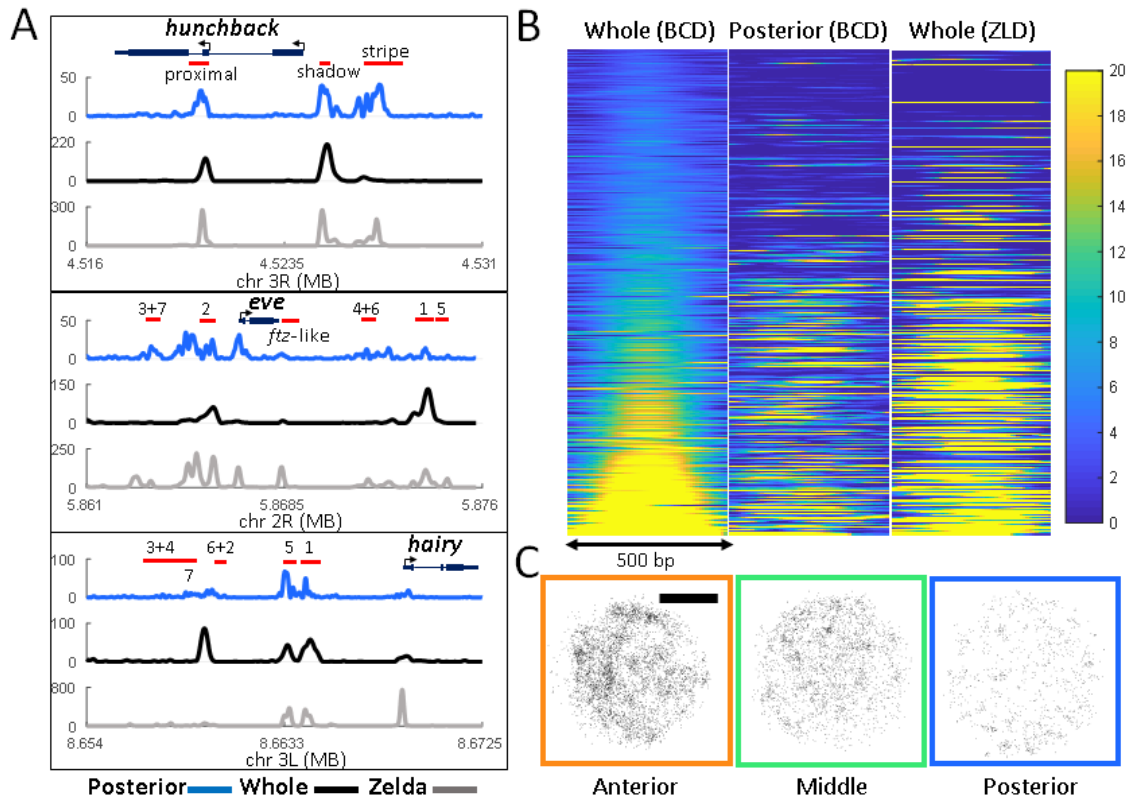
Figure 3: **Zelda mediated Bicoid binding in the posterior embryo.** (A) Posterior third (blue) and whole embryo (black) BCD, and whole embryo Zelda (grey) ChIP-seq signal normalized reads at the *hunchback*, *eve*, and *hairy* gene loci. Red bars show known enhancers as annotated in the RedFly database, for eve and hairy they are numbered according to the stripes they are thought to regulate. (B) Heat-map representation of normalized BCD ChIP-seq reads (first two panels) and ZLD-ChIP-seq reads (third panel) in a 500-bp window centered on BCD peaks called in the whole embryo data and sorted according to increasing signal of the whole embryo data, a total of 2145 peaks are shown, colors indicate enrichment over the background (blue) with all plots displayed on the same scale. (C) Examples of the spatial distribution of all detected bound molecules in nuclei along the A-P axis in ZLD- embryos, scale bar is 2.5 μm. A loss of clustering is apparent compared to the distributions shown in Fig. 2B.
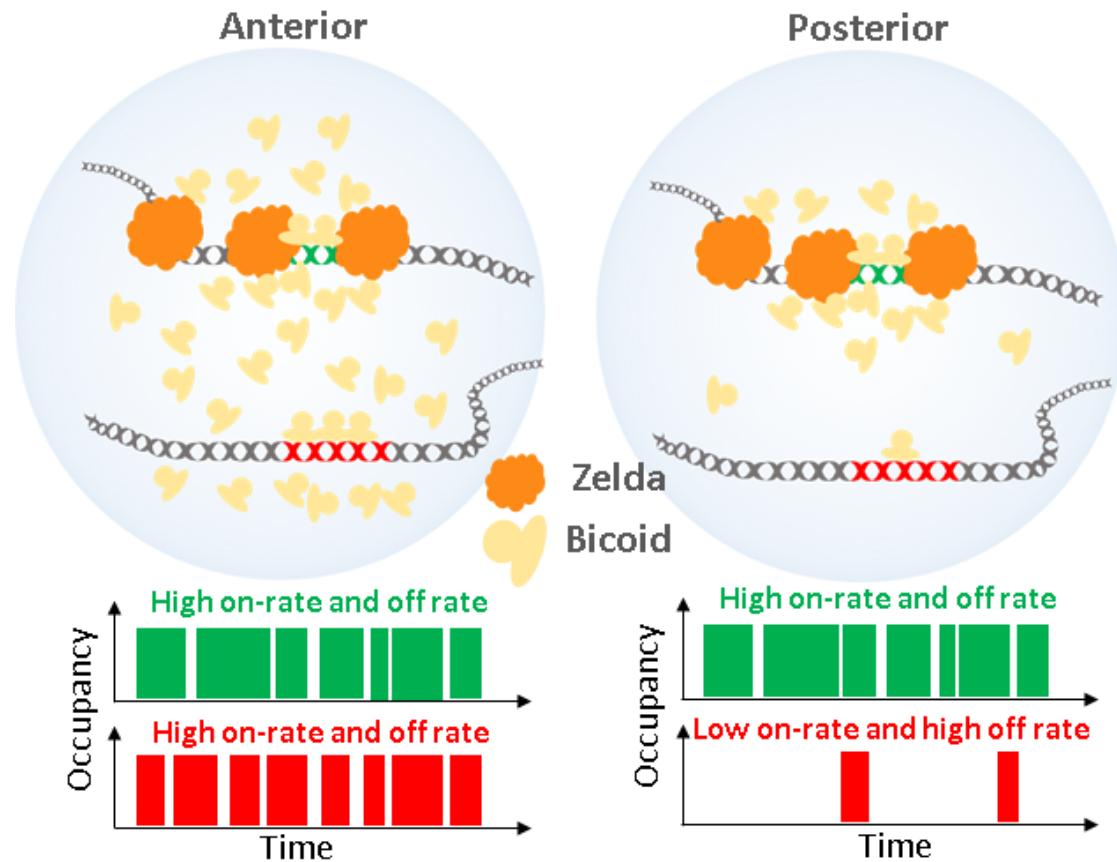
Figure 4: **Model of Zelda dependent modulation of the Bicoid on-rate at specific loci in the posterior embryo.** At high concentrations in the anterior of the embryo, all target sites are highly occupied. At low concentrations, loci with Zelda occupancy have an increased time averaged occupancy through the formation of spatiotemporal hubs that enrich local concentrations and increase the on-rate.
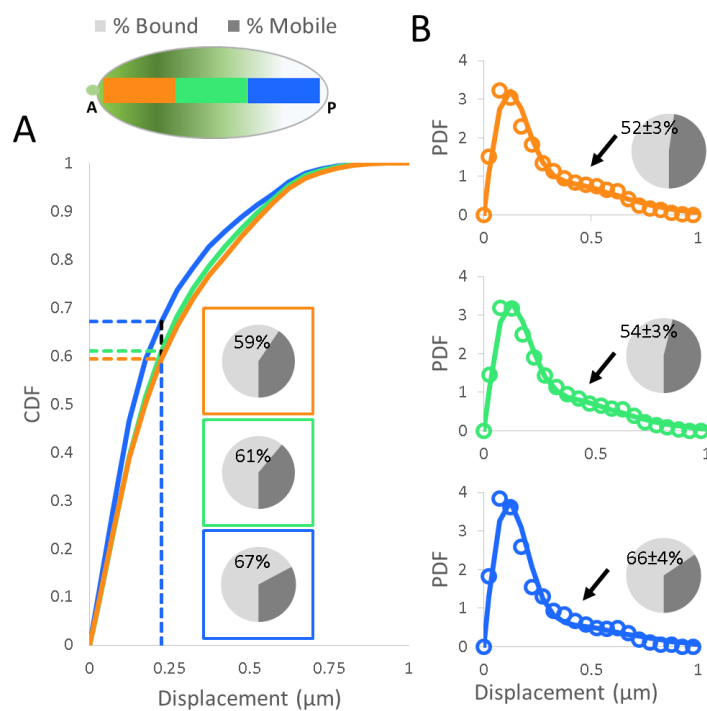
8 Supplemental Figures
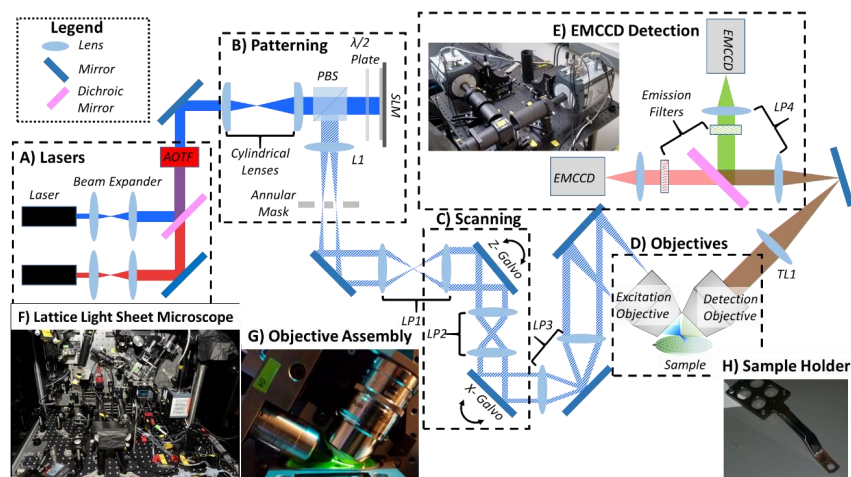
**Supplemental Figure S1. Lattice Light-Sheet Microscope Implementation**

The Lattice light-sheet was built as described by Chen et al, Science, 2014. This simplified schematic shows the major modules of the microscope as follows: (**A**) Laser Module, contains 6 lasers ranging from 405 nm to 639 nm, which are independently expanded, collimated, and input into an Acousto-Optic Tunable Filter (AOTF). (**B**) The patterning module contains a pair of cylindrical lens to expand the input Gaussian beam to a stripe, and a half-wave ($\lambda/2$) plate, a polarizing beam splitter (PBS), and a Spatial Light Modulator (SLM) to perform the patterning. Lens L1 projects the Fourier Transform of the SLM pattern onto an annular mask for spatial filtering (**C**) Lens pair 1 (LP1) is then used to de-magnify the annular mask plane and project it onto the z-scan galvo and LP2 projects the z-galvo plane onto the x-scanning galvo. (**D**) LP3 magnifies the x-galvo plane and projects it onto the back pupil plane of the excitation objective, where it is focused (Fourier transform) to project the final light sheet pattern onto the sample. Emitted fluorescence is collected by the detection objective. (**E**) The tube lens (TL1) focuses an

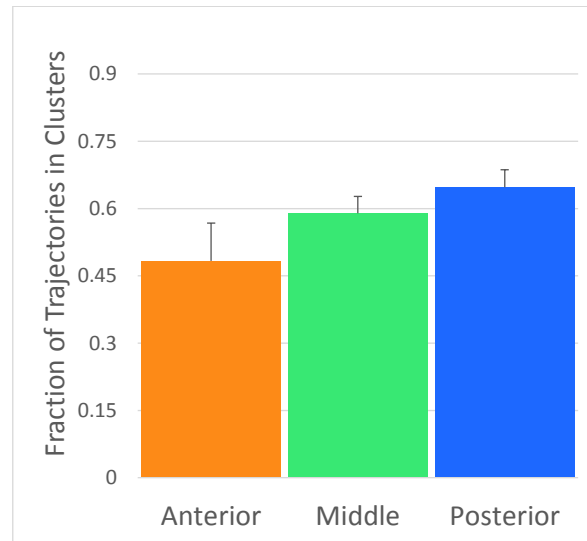Figure 5: **Supplemental Figure S1.**
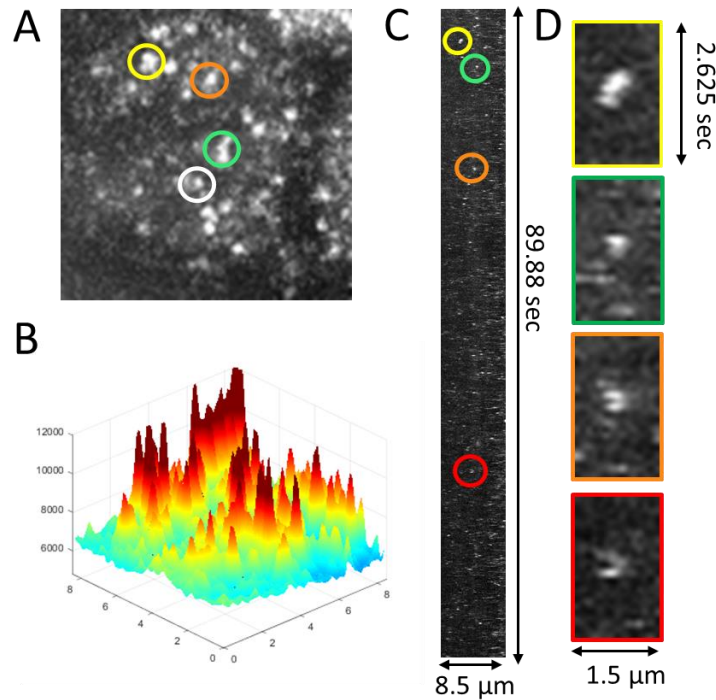
Supplemental Figure S2. **Single Molecule Imaging of BCD-eGFP at 100 milliseconds to estimate residence times**. **(A)** The max projection in time of a 90 second segment of a representative 100 millisecond dataset acquired at an anterior nucleus (EL (x/L) of 0.1), corresponding to the last frame of Video 2. **(B)** Surface plot representation of (A) to illustrate the signal-to-background ratio of single molecule binding events. The smaller peaks likely correspond to slowly diffusing molecules that are not in the imaging volume for the entire exposure time. **(C)** Maximum projection through x-t (kymograph representation) of the data shown in Video 2 from which (A-B) were calculated. Colored circles correspond to those in (A). **(D)** Zoomed in x-t view of the circled regions in (A) and (C), illustrating the transient nature of BCD binding.

Figure 6: **Supplemental Figure S3.**
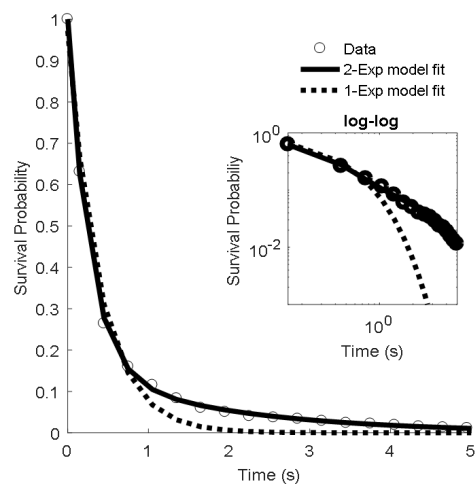
Supplemental Figure S4. Fits to the survival probability distribution of the 500 millisecond dataset. The survival probability distribution of residence times calculated from the MTT algorithm on the 500 ms data set of 17 nuclei and 1211 trajectories. The solid lines show the fit to a 2-exponent model and the dotted lines to a 1-exponent model. The insets shows the same on a log-log scale to visualize the difference between 1 and 2 exponent fits. R^2 value for the 2-exponent fit is 0.99

Figure 7: **Supplemental Figure S4.**
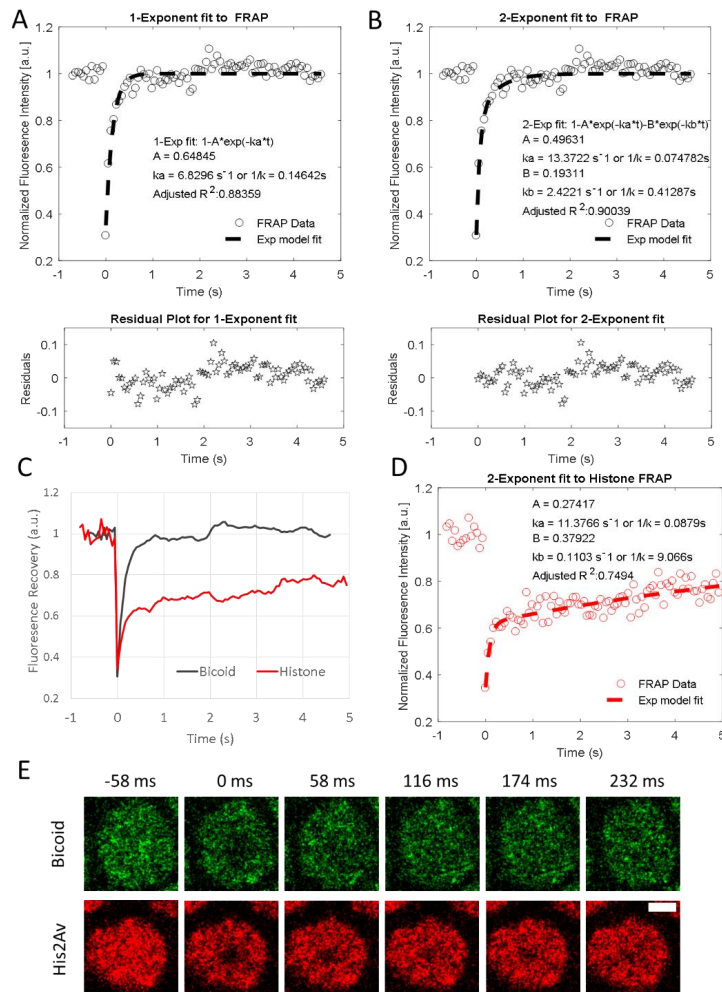
Supplemental Figure S5. Analysis of FRAP data
(**A**) Averaged Bcd-eGFP FRAP data and single exponential fit results. (**B**) Double exponential fit

results to average Bcd-eGFP FRAP. (**C**) Comparison of averaged BCD-eGFP (21 nuclei) and

Figure 8: **Supplemental Figure S5.**

Supplemental Figure S1. Lattice Light-Sheet Microscope Implementation

The Lattice light-sheet was built as described by Chen et al, Science, 2014. This simplified schematic shows the major modules of the microscope as follows: (**A**) Laser Module, contains 6 lasers ranging from 405 nm to 639 nm, which are independently expanded, collimated, and input into an Acousto-Optic Tunable Filter (AOTF). (**B**) The patterning module contains a pair of cylindrical lens to expand the input Gaussian beam to a stripe, and a half-wave (λ/2) plate, a polarizing beam splitter (PBS), and a Spatial Light Modulator (SLM) to perform the patterning. Lens L1 projects the Fourier Transform of the SLM pattern onto an annular mask for spatial filtering (**C**) Lens pair 1 (LP1) is then used to de-magnify the annular mask plane and project it onto the z-scan galvo and LP2 projects the z-galvo plane onto the x-scanning galvo. (**D**) LP3 magnifies the x-galvo plane and projects it onto the back pupil plane of the excitation objective, where it is focused (Fourier transform) to project the final light sheet pattern onto the sample. Emitted fluorescence is collected by the detection objective. (**E**) The tube lens (TL1) focuses an

Figure 9: **Supplemental Figure S1.**

Supplemental Figure S6. Analysis of Displacement Distributions (A) Cumulative distribution functions of the displacement data, pie charts indicate the Bound and Mobile fraction as measured by the CDF value at 0.225 μm (dashed lines). (B) Probability Density Function of the displacement data (markers) and fits to the two population model (solid lines), pie charts show the mobile and bound fractions from the results of the fit, labels indicate the bound fraction and the error is the standard error associated with the fit parameter. Arrows point to the mobile population in the distributions that is decreasing from anterior to posterior positions. (A-B) For each position the

Figure 10: **Supplemental Figure S6.**

EL=0.0914   EL=0.1067   EL=0.3751   EL=0.5362

EL=0.6457   EL=0.7660   EL=0.8642   EL=0.9038
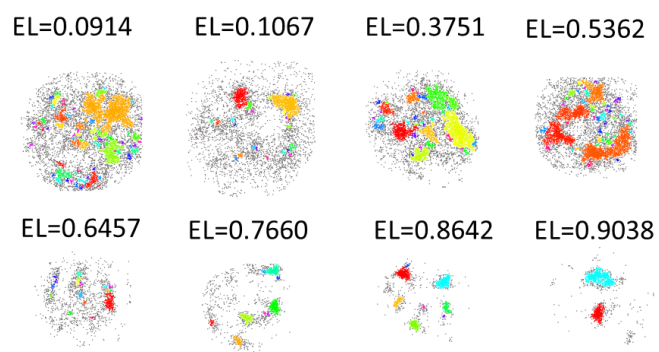
Supplemental Figure S7. Cluster identification results from DBSCAN across the A-P axis.
Examples of clusters identified along the A-P axis using DBSCAN with the position shown as fraction of embryonic length (EL). Particles included in the same clusters are represented with the same color.

Figure 11: **Supplemental Figure S7.**

Supplemental Figure S8. Fraction of Trajectories within clusters across the A-P axis.
Calculated as trajectories within clusters over total number of trajectories in each nucleus. The median for all nuclei at each spatial position is shown, error bars show standard error.

Figure 12: **Supplemental Figure S8.**
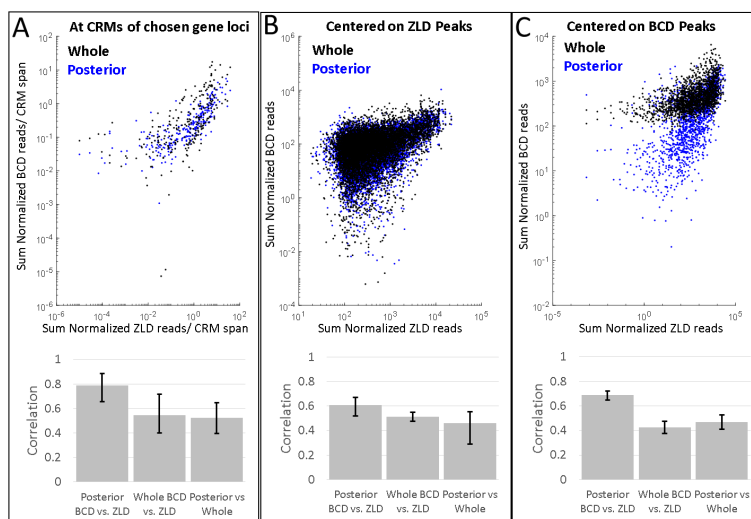
Supplemental Figure S9. BCD binding in whole and posterior thirds embryos compared to ZLD binding. (A) Sum of Normalized BCD ChIP-seq reads divided by the length of the respective CRMs for posterior thirds (blue) and whole embryo (black) vs. Sum of Normalized ZLD ChIP-seq reads divided by the length of the respective CRM at annotated cis-regulatory modules (CRMs) of *eve*, *giant*, *hunchback*, *knirps*, *hairy*, *kruppel*, *caudal*, *fushi-tarazu*, *engrailed*, *wingless*, *runt*, and *gooseberry* loci. A total of 293 CRMs from the RedFly database were analyzed. **(A-B)** Sum of Normalized BCD ChIP-seq reads for posterior thirds (blue) and whole embryo (black) vs. Sum of Normalized ZLD ChIP-seq reads over a 500 bp region centered on **(B)** 8331 ZLD peaks and **(C)** 2145 BCD peaks. **(A-C)** Corresponding bar plots show Pearson correlation coefficients and error bars show 95% confidence intervals as determined by bootstrapping.

Figure 13: **Supplemental Figure S9.**
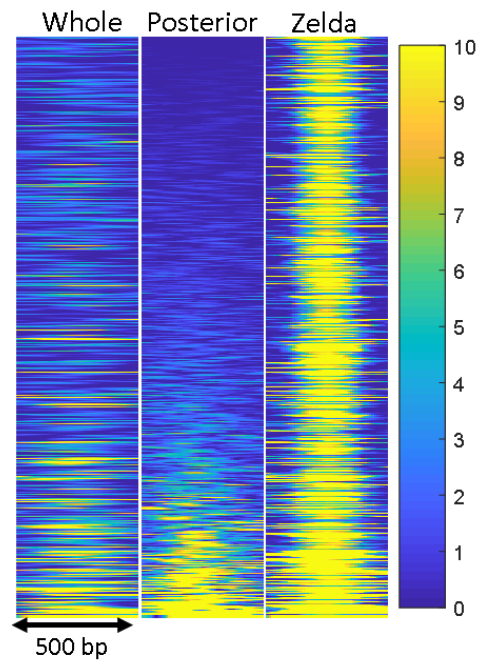
Supplemental Figure S10. BCD and ZLD binding at Zelda peaks. Heat-map representation of normalized ChIP-seq reads for BCD (1st two panels) and ZLD (3rd panel) in a 500 bp window centered on ZLD peaks sorted according to increasing BCD signal in the posterior embryo data. A total of 8331 peaks are shown.

Figure 14: **Supplemental Figure S10.**
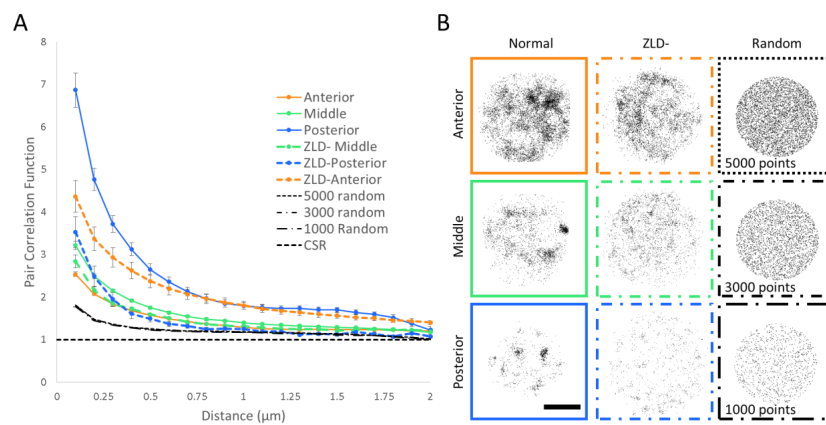
Supplemental Figure S11. Averaged pair-correlation (radial distribution) functions for nuclei in the normal, ZLD- and simulated cases and representative images. (A) Pair-correlation functions. Error bars indicate the standard error. **(B)** Representative images corresponding to pair-correlation functions for Normal, ZLD- and randomly distributed points.

Figure 15: **Supplemental Figure S11.**

9 Supplemental Tables

LIST of SUPPLEMENTARY MATERIALS

Mir et al., Dense Bicoid Hubs Accentuate Binding along the Morphogen Gradient

Supplemental Movie S1 (Related to Figure 1). Movies corresponding to the still frames shown in Figure 1A.

Supplemental Movie S2 (Related to Figure 1) Representative data from a 90 second segment of a 100 millisecond exposure time movie acquired at an anterior position (EL (x/L) of 0.1). Top left shows the raw data and top right the corresponding surface plot representation. Bottom left shows a running max projection of the data and bottom right shows a surface plot representation of the same

Supplemental Movie S3 (Related to Figure 2) Representative data acquired at 10 millisecond exposure times for 4 nuclei.

Supplemental Movie S4 (Related to Figures 2 and 3) Temporal dynamics of cluster formation for representative nuclei at Anterior, Middle, and Posterior positions.

Supplemental Table S1 (Related to Figures 1 and 3). Results from 2-exponent model fits to survival probability distributions.

Supplemental Fig. S1 (Related to Figure 1). Lattice Light-Sheet Microscope Implementation

Supplemental Fig. S2 (Related to Figure 1). Single Molecule Imaging of BCD-eGFP at 100 milliseconds to estimate residence times.

Supplemental Fig. S3 (Related to Figure 1). Fits to the survival probability distributions of the 100 millisecond datasets

Supplemental Fig. S4 (Related to Figure 1). Fits to the survival probability distribution of the 500 millisecond dataset

Supplemental Fig. S5 (Related to Figure 1). Analysis of FRAP data

Supplemental Fig. S6 (Related to Figure 2). Analysis of Displacement Distributions

Supplemental Fig. S7 (Related to Figure 2). Cluster identification results from DBSCAN across the A-P axis.

Supplemental Fig. S8 (Related to Figure 2). Fraction of trajectories within clusters across the A-P axis.

Supplemental Fig. S9 (Related to Figure 3)**.** BCD binding in whole and posterior thirds embryos compared to ZLD binding

Figure 16: **Supplemental Materials List.**

Data Set	A-P Poistion	Non-specific binding time $(1/k_{ns}$ (sec))	Specific Binding time $(1/k_s$ (sec))	Corrected Specific Binding time $1/(k_s-k_{bleach})$ (sec)	Number of trajectories
100 ms	Ant	0.175 [0.166, 0.184]	1.100 [1.013, 1.203]	1.105	17735
	Mid	0.160 [0.152,0.168]	0.940 [0.859, 1.038]	0.944	40092
	Post	0.146 [0.141, 0.151]	0.865 [0.806, 0.932]	0.868	20823
	All Together	0.159 [0.166, 0.152]	0.959 [0.882, 1.050]	0.963	78650
100 ms (Zelda-)	Ant	0.152 [0.146, 0.160]	0.777 [0.726,0.836]	0.780	11415
	Mid	0.131 [0.126, 0.137]	0.629 [0.587, 0.679]	0.631	7572
	Post	0.112 [0.108, 0.117]	0.463 [0.435, 0.495]	0.464	3606
	All Together	0.139 [0.133, 0.145]	0.694 [0.647, 0.748]	0.696	22593
500 ms	All Together	0.265 [0.255, 0.277]	1.876 [1.69, 2.107]	-	1211

Supplemental Table S1. Results from 2-exponent model fits to survival probability distributions. k_{ns} and k_s are the un-corrected off-rates for the short-lived (non-specific) and longer-lived (specific) populations respectively determined from a two-exponent fits to the survival probability distributions. The binding time (one-over the off rates) are shown with 95% confidence intervals in square brackets. The photo-bleaching corrected binding times are calculated as $1/(k_s-k_{bleach})$ where k_{bleach} is 0.0043 s^{-1}).

Figure 17: **Supplemental Table S1.**

Chapter 5

Development of modular maternal promoters as a tool for generating tunable maternally deposited protein levels in the early *Drosophila melanogaster* embryo

1 Abstract

Despite a burgeoning of live fluorescent imaging techniques in the study of the developing *Drosophila melanogaster* fruit fly embryo, there exists a paucity of methods for driving precise levels of maternally deposited proteins that are necessary for quantitative research. Here, we developed a series of modular promoters that drive levels spanning three orders of magnitude that can be useful for many studies of gene expression in the early embryo such as those requiring multiple transgenes at precise levels and stoichiometries. Additionally, we developed tools that can be used as the foundation for calibrating expression levels from these maternal promoters with desired precision into absolute numbers of molecules inspired by the standard candles ubiquitous in the field of astronomy.

2 Introduction

Maternal promoters are an essential tool for driving the gene expression—of both endogenous proteins and orthogonal labelling tools (e.g., the MS2-MCP system or fluorescent transgene fusions) in the early embryo of *Drosophila melanogaster*. The detailed studies of the spatial and temporal patterns driven by a variety of maternal promoters, such as *bicoid*, *otu*, *matrimony*, *vasa*, *nanos*, and *alpha-tubulin*, have provided us with

a wide range of possible maternal transgenic expression levels. For example, these maternal promoters have been used to test the concentration dependence of a maternally deposited transcription factor (Hannon, Blythe, and Wieschaus 2017), and to drive the expression of fluorescently tagged coat proteins for use in the MS2-MCP nascent mRNA labelling system.

However, these endogenous promoters do not always provide satisfactory expression levels. There are situations where both high and precise levels of protein expression are required. For example, in labelling nascent mRNA with the MS2-MCP system, it is important to ensure that there is enough MCP protein available so that all MS2 loops are actually bound in order to allow for quantitative measurements. For example, the commonly used *nanos* promoter coupled with the fluorophore fusion MCP-mCherry is unable saturate 24 transcribed MS2 loops with a single copy (Fig. 1 A, B, C). This necessitates the use of additional copies of the MCP-mCherry transgene, complicating fly crosses and delaying experiments. It would be worthwhile to have tools available to drive higher expression levels at precise concentrations that are known to be saturating.

Copies	Genotype
2	yw; MCP-mCherry(3) ; +
3	yw; $\dfrac{\text{MCP-mCherry(3)}}{\text{CyO}}$; MCP-mCherry(6)
4	yw; MCP-mCherry(3) ; MCP-mCherry(6)
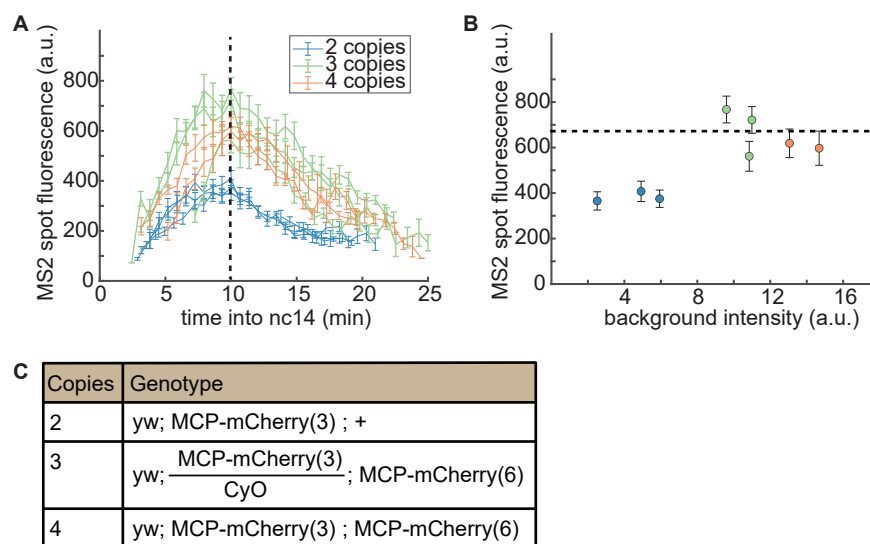
Figure 1: **Concentration of maternally deposited MCP-mCherry required for saturation of 24 MS2 loops. (A)** Plot of mRNA output from the *hunchback* P2 promoter (Garcia et al. 2013) as a function of time into nuclear cycle 14 for different maternal MCP-mCherry gene copy numbers with the vertical dashed line showing when the data points for (B) were taken in time. **(B)** Plot of mRNA output from the *hunchback* P2 promoter as a function of background intensity with horizontal dashed line showing saturation plateau; colors correspond to maternal MCP-mCherry gene copy number from (A). **(C)** Table showing genotypes of flies used in (A) and (B). Data and figure kindly provided by Yang Joon Kim, private communications.

Additionally, while these repurposed endogenous promoters often provide a fairly

broad dynamic range of expression, the large gaps between their expression levels do not allow for fine tuning of protein levels. In order to explore a wider range of expression levels, these reporters are often integrated randomly using the P element transposase method (Rubin and Spradling 1982; Spradling and Rubin 1982), which leads to expression variability. This has the significant drawback, however, that expression levels generated in one construct are not reproducible if the fly line must be made again through re-injection. If the fly line goes extinct or ages unfavorably, for instance, it is difficult to generate a new fly line with similar expression levels. Additionally, the randomness associated with P element transposition makes it more difficult to design fly crosses that rely on recombination events, the frequency of which are dependent on distance from the inserted P element transgene (unless the insertion is mapped in the genome, which is difficult and seldom done in practice). Furthermore, transgenes may be inserted in unfavorable locations in the genome, leading to unhealthier fly lines. Clearly, an alternative with non-random insertion would obviate many of these difficulties.

One alternative to P element transposition is ϕC31 integration (Groth et al. 2004) — wherein plasmids containing the *attB* sequence are injected into fly lines containing an existing *attP* sequence— at well characterized landing sites (Venken et al. 2009). However, this method requires the construction of multiple new fly lines to test each landing site for expression levels.

With no ideal integration method available, we took a closer look at existing enhancer-promoters maternal drivers that could potentially be refined intro multimerizable elements capable of predictable, precise control of gene expression in the early embryo.

Inspired by the work of (Sano, Nakamura, and Kobayashi 2002), which showed that multimerizing a minimal 40 bp region of the *vasa* enhancer drove expression in the oocyte, we developed a set of multimerized enhancers to finely tune maternally deposited protein levels that could be integrated with the ϕC31 system. We used the maternal *transposase* core promoter (abbreviated as *pTrans* in sequence names) and multimerized small (40-80 bp) regions of the enhancers from the *vasa* and *nanos* genes to drive reporter genes. We then quantified the protein expression levels with confocal microscopy. Our promoters span three orders of magnitude and are small enough to be inserted into common injection plasmids for *Drosophila*. Furthermore, we tested a subset of these promoters at ten different positions within the fly genome. Finally, we developed tools that can be used as the foundation for calibrating expression levels from these maternal promoters into absolute numbers of molecules inspired by the standard candles of the field of astronomy. Together, these tools may be very useful for future quantitative studies in early development that require variable and precise levels of transgene expression.

3 Results

3.1 Design of modular *vasa* and *nanos* enhancer elements

We began by identifying minimal enhancer elements of the *vasa* and *nanos* (*nos*) enhancers. For the *vasa* enhancer, we selected a 40 bp region (*eVasa*) identified by Sano, Nakamura, and Kobayashi 2002, which we then multimerized various times upstream of a proximal maternal *transposase* core promoter (Fig. 2 A).

For the *eNos* enhancer, we selected an 80 bp region (Fig. 2 B) of the full *nanos* enhancer-promoter in common use (e.g. used to drive MCP-eGFP expression by Garcia et al. 2013). To identify this sequence, we adopted a trial-and-error approach in order to find the smallest region capable of driving expression comparable to the full *nanos* enhancer-promoter sequence (data not shown). All truncated *nanos* promoter sequences tested can be found in Table S1. To achieve consistency with the *eVasa* enhancer lines, we also chose to place the *eNos* multimers upstream of a maternal *transposase* promoter. Although our primary goal in this study was to be able to multimerize the *eNos* enhancer to achieve higher levels of gene expression, we note that it is even useful as a monomer as a more compact alternative to the full enhancer, which could accommodate a broader range of plasmid designs.
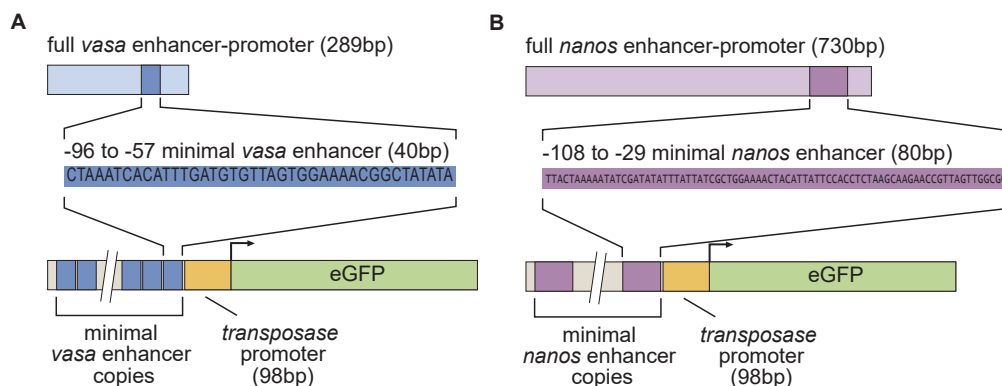
Figure 2: **Designs of the *eVasa* and *eNos* modular maternal enhancers.** **(A)** Minimal *eVasa* design. From the 289 bp *vasa* enhancer-promoter in common use (e.g. for driving Cas9 used in germline genome editing), a 39 bp portion previously identified by Sano, Nakamura, and Kobayashi 2002 was multimerized and placed upstream of the maternal *transposase* core promoter sequence to drive the expression of eGFP. **(B)** Minimal *eNos* design. From the 730 bp annotated enhancer-promoter sequence in common use (e.g. for driving MCP-eGFP in MS2 studies in the early embryo (Garcia et al. 2013), we identified a minimal 89 bp sequence sufficient to drive expression levels comparable to the full sequence. This 89 bp minimal enhancer was then multimerized and placed upstream of the maternal *transposase* promoter to drive the expression of an eGFP reporter.

3.2 Expression levels of modular *eVasa* and *eNos* driven reporters

We used a confocal laser scanning microscope to observe the expression levels of *eVasa* (Fig. 3A) and *eNos* (Fig. 3 B) driven fluorophore, tandem-MCP-eGFP (tdMCP-eGFP; Wu et al. 2012), injected at the VK33 landing site in the fly genome (Venken et al. 2009) during early embryonic development during nuclear cycles 12, 13, and 14. Tandem-MCP-eGFP was used here because one of our motivations for the earliest constructs made in this project was to drive extremely low levels of tandem-MCP-eGFP in the early embryo to be used for low background signal, high affinity MS2 loop binding. Although this application was not further pursued, we kept the tandem-MCP protein in our constructs for the sake of consistency. For each construct, we measured the mean nuclear concentrations of our reporters at approximately halfway through the nuclear cycle. From one to ten copies of multimerized *eVasa*, we observed rapidly increasing nuclear protein concentrations. Surprisingly, increasing beyond 10 copies led to dramatic reductions in expression, thus placing an upper limit on the expression levels possible from an *eVasa* multimer enhancer. This observation was our motivation for exploring the *eNos* multimers, which would allow us to have maternal promoters with levels in the range of necessary for, for example, MCP-mCherry MS2 loop saturation.

To explore higher concentration ranges, we assayed expression from the *eNos* multimer series (Fig. 3 B). While the *eNos* multimer series did not achieve as broad a range of expression levels as the *eVasa* series, we still observed a boost in expression by $\sim 50\%$ going from 1 to 3 copies, after which there was a reduction in activity, akin to the drop in activity observed after 10 copies in the *eVasa* series.
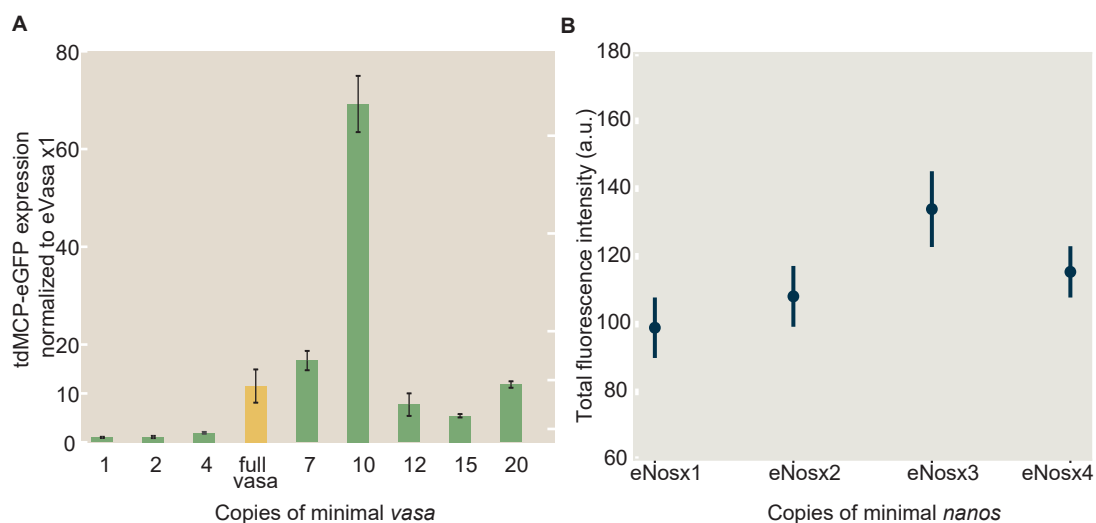
Figure 3: ***eVasa* and *eNos* multimer series gene expression levels. (A)** Mean nuclear levels of minimal *eVasa* driven tdMCP-eGFP and **(B)** minimal *eNos* driven tdMCP-eGFP during nuclear cycle 12 including all nuclei in the field of view. Error bars are SEM over ≥ 3 embryos.

3.3 Effects of genomic location on the *eNos* driven reporter

There is often a need in *Drosophila* embryo research to integrate new transgenes into specific chromosomes or chromosomal locations to accommodate fly crossing schemes. As a result, it was important to quantify the effects of genomic position on the reporter levels driven by the modular enhancer series. Thus, we placed one specific reporter, *eNosx2-pTrans-MCP-mCherry* in various locations of the genome. We chose ten VK *attP* landing sites (previously developed by Venken et al. 2009): one on the X chromosome, two on the left arm of chromosome 2, two on the right arm of chromosome 2, two on the left arm of chromosome 3, and three on the right arm of chromosome 3 (Fig. 4 A). These ten sites were selected in order to accommodate a wide variety of possible crossing and recombination schemes and to evaluate how strongly genomic position affects levels and variability of reporter expression. Subsequently, we measured the mean nuclear concentrations of reporters at these landing sites in the early embryo and plotted the results in Figure 4 B. Here, we observe that there is a ≥ 8-fold difference in mean expres-

sion between the highest and lowest expressing landing sites, with the X-chromosome landing site having the highest expression.

Additionally, we assayed expression from the full series of *eNos* multimers (1, 2, 3, and 4 copies) at two particular landing sites from Figure 4 A, VK22 (Chr2R) and VK33 (Chr3L), and found that not only did mean levels of expression change according to landing site, but so too did embryo-to-embryo variability (Fig. 4 C, D). Indeed, within error, all *eNos* enhancers at VK22 had the same mean expression levels, and these expression levels were high variable compared to those at VK33 (Fig. 4 C, D). From these experiments, we conclude that genomic position is an important factor to account for when using the modular promoter series to drive precise levels of gene expression.

.
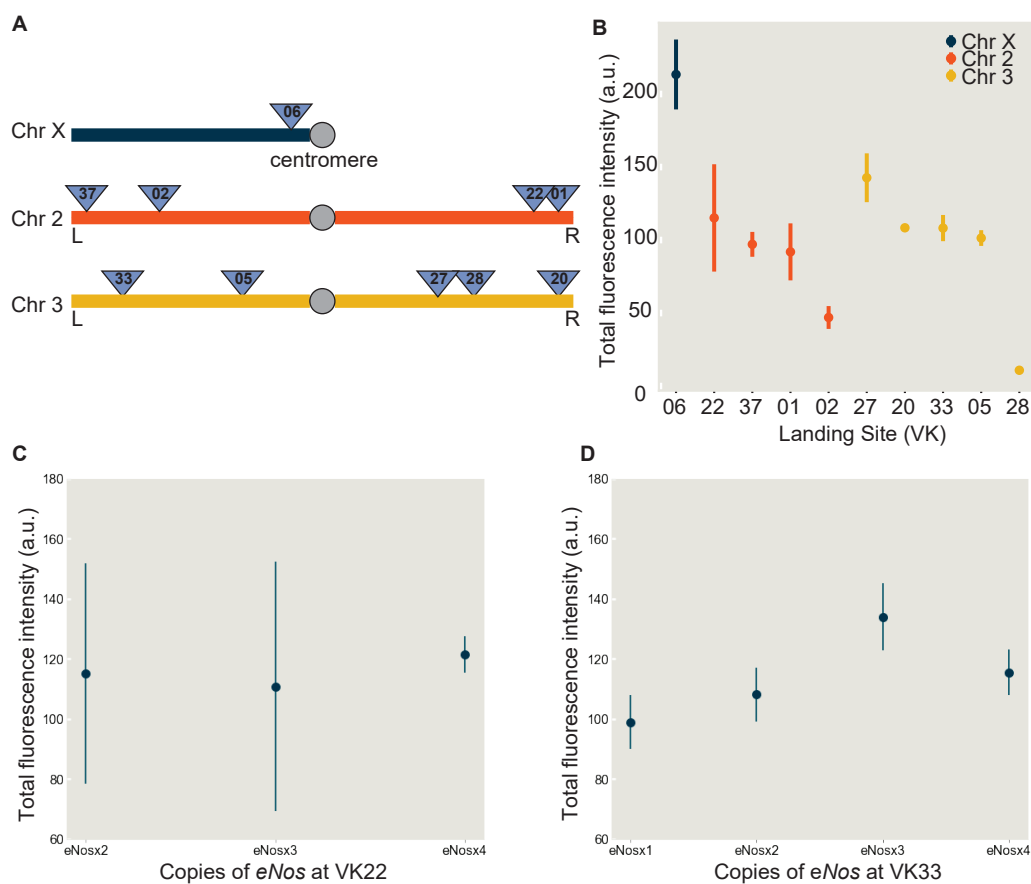
Figure 4: ***eNosx2* enhancer expression levels at different VK ϕC31 *attP* landing sites (A)** Schematic showing positions of genomic landing sites used for reporter integration. **(B)** Mean nuclear concentration of MCP-mCherry driven by *eNosx2-pTrans* at different genomic locations in (A) colored according to chromosome. **(C)** Expression of *eNosxN-pTrans* at the VK22 landing site. **(D)** Expression of *eNosxN-pTrans* at the VK22 landing site. Error bars are SEM over ≥ 3 embryos.

3.4 Developing standard candles to calibrate reporter fluorescence

In order to expand the utility of our maternal promoter series, we sought to develop a method of calibrating our fluorescence reporters into absolute numbers of molecules in order to allow for comparisons between different fluorophores. For example, knowing that a concentration of 1000 nM was desired, one could select the 1000 nM maternal promoter based on our standard curves, regardless of which fluorophore is attached to the molecule of interest. In contrast, using arbitrary units, one would have to image the old and the new fluorophore under identical conditions to convert into units comparable with the reported data. To implement this idea, we developed "standard candles" for the *Drosophila* embryo based on those developed by (Hsia et al. 2016; King et al. 2014; Akamatsu et al. 2020). Standard candles are astronomical bodies with known luminosities that can be used as a basis for comparison to determine the distances of other astronomical bodies from the observer. Likewise, the biological standard candles can be seen as objects with known fluorescence values that can be used to calibrate the molecular number of other fluorescence species. These are polypeptides that oligomerize to form regular dodecahedra with precise stoichiometries. Here, a trimer of polypeptides assembles at each of the 20 vertices of the dodecahedron (Fig. 5 A; adapted with permission from Alamos et al. 2020). By fusing each subunit to a fluorescent protein, we can determine the precise mean fluorescence value of 60 or 120 fluorophore molecules, respectively, under our experimental conditions. This allows us to straightforwardly determine a fluorescence-to-molecule number calibration factor for other molecules fused to the same fluorophore imaged under the same microscopy conditions.

In order to be able to accurately quantify the fluorescence intensity from single molecules, it is necessary to limit the diffusion of the standard candles by anchoring them to a cellular surface. However, only a small number of polyhedron subunits should be anchored to the membrane, roughly one per polyhedron, to ensure that the complete polyhedra can actually self-assemble in solution. Thus, ideally, two proteins would be expressed in cells at $\sim 1/100$ ratio: tethered and untethered subunits, both fused to fluorophores. Expressing proteins at precise ratios is not completely straightforward, but one of the simpler methods is to have both proteins be apart of a single gene and use post-translational modification to create two different final protein subunits. Following Akamatsu et al. 2020, who constructed *in vivo* standard candles for mammalian cell culture, we used the F2A system to construct our standard candle genes (Figure 5 B). F2A is a self-cleaving peptide that can be used to separate two parts of a nascent polypeptide chain if its coding sequence is inserted in between them within the gene. Crucially, the F2A peptide does not cleave with 100% efficiency. Thus, we can exploit the inefficiency to produce a subset of uncleaved polypeptides that have a tether and cleaved polypeptides without.

Since this system needed to be adapted for *Drosophila* embryos, we investigated several different possible tethers. To this end, we made multiple constructs, two of which are depicted in Figure 5D. We focused on nuclear localization to ensure that stan-

dard candles existed in the same chemical environment as the fluorophores we need to calibrate. Thus, we tried a myristoyal lipid binding tether for membranes combined with an nuclear localization signal (NLS) to tether to the interior nuclear membrane (*eVasax10-pTrans-myr-F2A-eGFP-60mer-eGFP-Tub3'UTR*), and a Lamin protein tether for localization to the nuclear lamina (*eVasax5-pTrans-60mer-mCherry-F2A-lamin-tub3'UTR*). As seen in Figure 5D, bottom, the Lamin tether was arguably better at tethering candles to the nuclear peripheries.

It was also necessary to determine the ideal expression level of standard candles for ease of image analysis and to ensure a high concentration of fully self-assembled polyhedra. Too high a concentration would make it difficult to quantify the intensity from single molecules, while too low a concentration would lead to lower formation of fully self-assembled 60mer or 120mer molecules. Thus, we made standard candle constructs driven at a range of concentrations using the *eVasa* multimer series (Figure 5 E). From this exploration, we determined that *eVasax5* struck the right balance of having enough single molecules to quantify without the field of view being inundated with fluorophores. The next milestone in the development of these standard candles will be to generate intensity histograms for both 60mer and 120mer standard candles (and perhaps others, if possible) to generate a standard curve that could actually be used for calibrating the molecular concentrations of the maternal promoters.
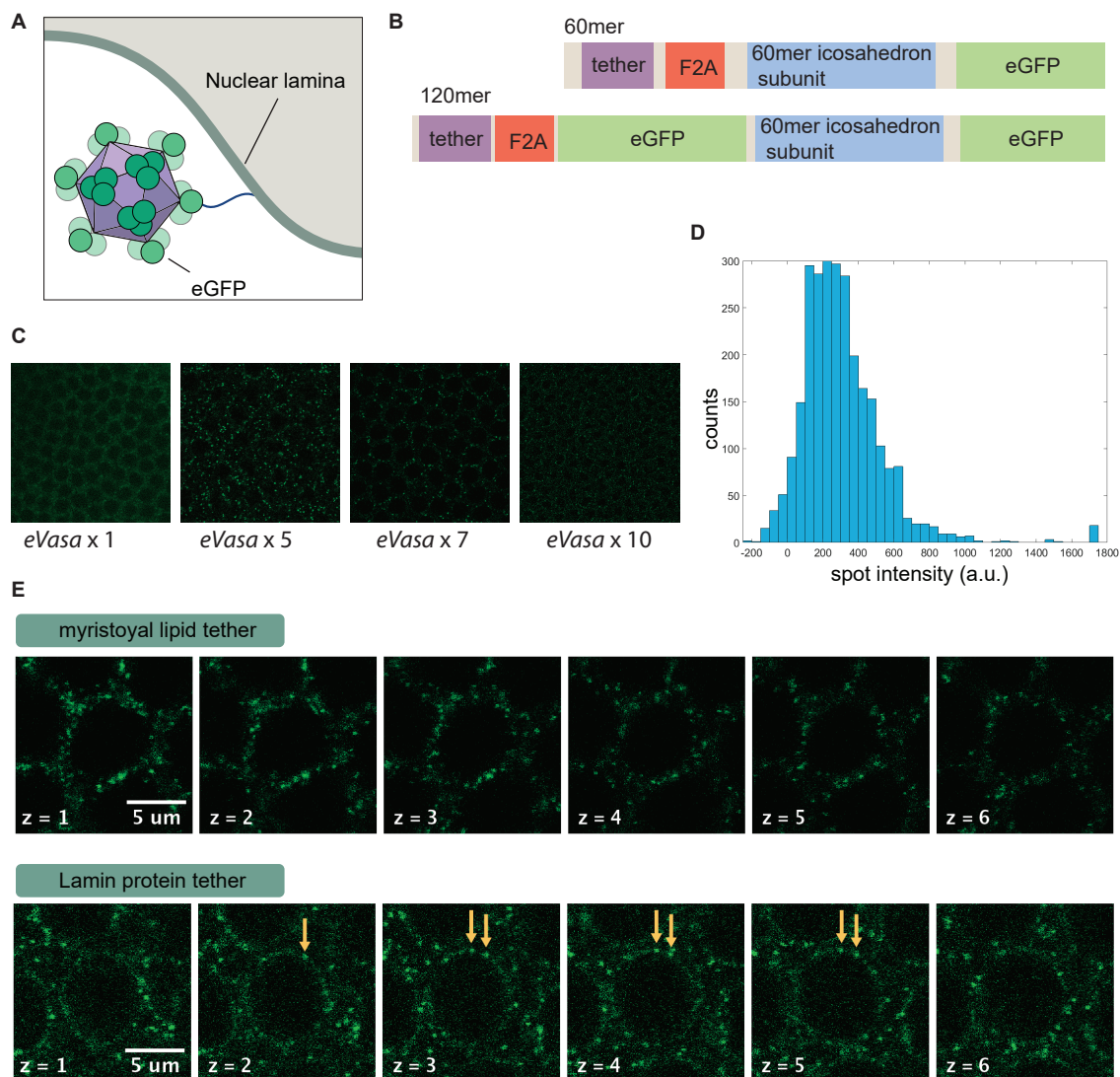
Figure 5: **Development of standard candles in *Drosophila* embryos (A)** Schematic of standard candle localization by membrane tether (adapted with permission from Alamos et al. 2020). Caption continues on next page.

Figure 5: **Continued from previous page: Development of standard candles in *Drosophila* embryos.** **(B)** Schematic of standard candle genetic construct designs (tether: either myristoyal binding protein or Lamin protein, F2A: self-cleaving peptide). **(C)** Standard candles driven at different expression levels using multimer *eVasa-pTrans-tdMCP-eGFP* reporters. **(D)** Histogram showing distribution of *eVasax5-pTrans-60mer-mCherry-F2A-lamin-tub3'UTR* standard candle fluorescence. Intensities are the sum of fluorescence from 3 z-slices centered at the brightest z-slice. Intensities only drawn from spots with 3 to 8 z-slices. **(E)** Confocal microscopy images of standard candle localization with two different types of tethers: (top) myristoyal lipid tether for membranes —*eVasax10-pTrans-myr-F2A-eGFP-60mer-eGFP-Tub3'UTR*—, and (bottom) Lamin protein tether for localization to the nuclear lamina —*eVasax5-pTrans-60mer-mCherry-F2A-lamin-tub3'UTR*.

4 Discussion

In this study, we developed a series of maternal *Drosophila melanogaster* promoters that can be used to drive genes for studies in the early embryo. This ability to quantitatively tune expression levels could prove useful for many early embryo experiments that depend on precise levels of a maternally deposited transgene. Unexpectedly, after quantifying the constructs with more than ten copies of *eVasa*, from 12 to 20, we saw a dramatic reduction in activity. While we do not actually know the reason for this sharp drop, we speculate that it could be due to a repeat-induced silencing within the genome due to the relatively high copy number of the *eVasa* element. The fact that we were unable to drive expression any higher than *eVasax10* prompted us to explore other small enhancers that could potentially be multimerized to achieve even higher levels of expression, which may be needed for some experiments (e.g., for driving MCP-mCherry, which tends to be excluded from the nucleus, at sufficient levels to fully saturate a sequence of 24 MS2 loops as in Figure 1).

Thus, we developed a series of promoters that drive at higher levels than the *eVasa* series, the *eNos* series. While we were not able to achieve as wide a range of expression levels from *eNos* multimers as we had hoped, we still believe that the *eNosx3* enhancer could be useful for experiments that require the highest levels of expression (such as the MCP-mCherry experiment mentioned above), as this is the strongest maternal promoter that we are aware of in the literature. In the future, a different enhancer altogether may be required to achieve even higher levels than *vasa* and *nanos* were capable of, as we may have achieved the limits of their expression ranges, short of incorporating multiple insertions in the fly.

In order to further our goal of precisely tuning the levels of maternally deposited protein in the early embryo, we assayed the effects of genomic location on reporter output. We found that genomic location had profound effects not just on expression levels, but

on relative expression levels of multimer promoters in the same series. Genomic location can potentially be viewed as another knob used to generate precise expression levels; however, this variability may be a nuisance for certain fly crossing schemes and must be taken into account when planning experiments. In the future, it will be useful to follow up on these observations by also varying the reporter used at different locations. It is currently unclear whether the trends observed in Figure 4B would hold similarly for different enhancers, promoters, and genes.

Finally, we pursued the development of a method to calibrate the expression levels of fluorescence protein fusions in the early embryo to absolute numbers of molecules. We tested different tethering systems, different protein cleaving systems, and different maternal drivers to engineer an optimal standard candle for the early *Drosophila melanogaster* embryo. Incidentally, this exploration also demonstrated the great utility of the modular maternal *eVasa* promoters. Our work here has provided a solid foundation and proof of principle that can be further developed in further studies. Future work on the standard candles will involve finishing the calibration experiment by imaging the standard candles at different multimer number and plotting a calibration curve. Additionally, following up by performing a similar experiment with other fluorophores, mCherry in particular, could potentially be very valuable.

5 Methods and Materials

5.1 Selection of Minimal Enhancers, Promoter

The minimal *vasa* enhancer sequence used in this chapter (referred to as *eVasa*, see Table S2) is identical to the 40 bp, -97 to -57, region of the *vasa* gene that Sano, Nakamura, and Kobayashi 2002 identified as both necessary and sufficient to drive germline-specific gene expression during oogenesis. The sequence of *eVasa* was extracted from the FlyBase Sequence Downloader for the *vasa* gene (FlyBase ID: FBgn0283442, Sequence Features: vas_minimalEnhancer). The minimal, 98 bp maternal *transposase* promoter (referred to as pTrans, see Table S2) is identical to the P element *transposase* promoter Sano, Nakamura, and Kobayashi 2002 paired with their minimal *vasa* enhancer. The minimal *nanos* enhancer sequence used in this chapter (referred to as *eNos*, see Table S2) is the -108 to -29 region of the *nanos* promoter characterized by Ali et al. 2010. This 80 bp region was selected to include all three annotated transcription factor binding sites at their endogenous location: *zen*, *DRE* (DREF binding element), and *brinker*. Other minimal versions of *nanos* were tested, but none showed protein expression in the early embryo (Table S1).

5.2 Cloning

A list of all plasmids used in this chapter is included in Table 1.

eVasa and *eNos* constructs

pBPhi-eVasax10-pTrans-tdMCP-eGFP-noUTR

The *eVasax10-pTrans* gene was synthesized by GenScript Inc. and subcloned into the NotI/NheI restriction site of *pBPhi-eVasaRx1-pTrans-tdMCP-eGFP-noUTR*.

pBPhi-eVasax10-pTrans-tdMCP-eGFP-tub3'UTR   &   pBPhi-eVasax10-pTrans-eGFP-tub3'UTR

The tub3'UTR was cloned into *pBPhi-eVasax10-pTrans-tdMCP-eGFP-noUTR* via a two-step process. First, *pBPhi-eVasax10-pTrans-tdMCP-eGFP-noUTR* was digested with the XbaI restriction endonuclease, which replaced eGFP with tub3'UTR to create the intermediate plasmid *pBPhi-eVasax10-pTrans-tdMCP-tub3'UTR*. Then, eGFP was cloned back into the AatII restriction site of the intermediate plasmid by GenScript, Inc. to generate *pBPhi-eVasax10-pTrans-tdMCP-eGFP-tub3'UTR*. Finally, the tdMCP region was removed from *pBPhi-eVasax10-pTrans-tdMCP-eGFP-tub3'UTR* via mutagenesis by GenScript, Inc. to create *pBPhi-eVasax10-pTrans-eGFP-tub3'UTR*.

pBPhi-eVasax12-pTrans-tdMCP-eGFP-noUTR was digested with NotI-HF BamHI-HF restriction enzymes to remove the *eVasax12* enhancer cassette. The 80 bp minimal *nanos* enhancer (-108 to -29, named *eNos*) was PCRed from a plasmid containing the full *nanos* enhancer-promoter (pCasper-pNos-NLS-ParB2-eGFP-tub3'UTR) using primers 31.35-synthpNos2-F-pBPhiHomo and 31.39-synthpNos4-R-pTrHomo (Table 1). *eNosx1* was incorporated into the digested backbone using Gibson assembly. The Gibson assembly products were transformed into Top10 cells, and individual colonies were screened via sequencing to select the successfully assembled plasmid,
pBPhi-eVasax12- & -eVasax15-pTrans-tdMCP-eGFP-noUTR

To generate *pBPhi-eVasax12-pTrans-tdMCP-eGFP-noUTR* and *pBPhi-eVasax15-pTrans-tdMCP-eGFP-noUTR*, *eVasax2* and *eVasax5 genes*, respectively, were synthesized by Genscript and each subcloned into the BamHI restriction site of the *pBPhi-eVasax10-pTrans-tdMCP-eGFP-noUTR plasmid*.

eNos was inserted into the *eNosx1* digested backbone using Gibson assembly to generate *pBPhi-eNosx2_notCompact-pTrans-tdMCP-eGFP-noUTR*, which then had the extra plasmid backbone overlap removed via PCR *pBPhi-eNosx2-pTrans-tdMCP-eGFP-noUTR*.

pBPhi-eNosx1/2/3/4-pTrans-NoNLS-MCP-mCherry-tub3'UTR

To make *pBPhi-eNosx1-pTrans-NoNLS-MCP-mCherry-tub3'UTR*, *pBPhi-eNosx1-*

pTrans-tdMCP-eGFP-tub3'UTR was digested with NheI and NsiI restriction endonucleases, MCP-mCherry was PCRed with primers 44.19-mCh-rv-tubUTRHomo and 44.18-MCP-fwd-pTransHomo from the genomic DNA extraction of a fly line containing an MCP-mCherry transgene, and the backbone and insert were combined via Gibson assembly. *eNosx2* (two copies of eNos separated by a BamHI restriction site), *eNosx3* (three copies of *eNos* separated by a BamHI site and an arbitrary 6 bp spacer, spacer-B), and *eNosx4* (four copies of eNos separated by a BamHI restriction site and two identical 6 bp spacer-B sequences) were synthesized and subcloned into existing vectors by GenScript Inc. The *eNosx2* fragment was subcloned into *pBPhi-eNosx1-pTrans-NoNLS-MCP-mCherry-tub3'UTR* with PmeI and BglII restriction enzymes. The *eNosx3* and *eNosx4* fragments were subcloned into *pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR* with BamHI and BglII restriction endonucleases.

Standard candles

pBPhi-eVasax5-pTrans-60mer-mCherry-F2A-lamin-tub3'UTR

The *lamin* sequence was PCR'd from expressed sequence tag (EST) cDNA obtained from the Drosophila Genomics Resource Center (DGRC) using the following primers: forward: 49.5- DmelLaminDmo F (ATGTCGAGCAAATCCCGACGTGCT), reverse: 49.6 DmelLaminDmo R (GCAGTCAAACGAGAAGTGCGCCATTATGTAA).

5.3   Generation of Fly Lines

All plasmids were sent to BestGene Inc. for injection into *D. melanogaster* embryos whose genomes contain a ϕC31 *attP* landing site. All plasmids were initially integrated at the VK00033 landing site (Chr3L, BDSC (Bloomington Drosophila Stock Center) 9750). The injected flies were crossed to *yw;+;+* and their progeny screened for successful transformants, which were balanced (*w;+;Dr/TM3*) to obtain homozygous, stable lines. The *pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR* plasmid was also integrated at the VK00037 (Chr2L, BDSC 9752), VK00002 (Chr2L, BDSC 9723), VK00022 (Chr2R, BDSC 9740), VK00001 (Chr2R, BDSC 9722), VK00005 (Chr3L, BDSC 9725), VK00027 (Chr3R, BDSC 9744), VK00028 (Chr3R, BDSC 9745), and VK00020 (Chr3R, BDSC 9738) landing sites for comparison. All *pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR* integrations on Chr2 were balanced with *yw;Gla/CyO;+*, while those with integrations on Chr3 were balanced with *w;+;sb/TM3,ser*, to obtain homozygous lines.

5.4   Microscopy

Fly cages were allowed to lay for 90 to 120 minutes prior to embryo collection. Embryos were then mounted on microscopy slides in Halocarbon 27 Oil (Sigma-Aldrich,

H8773) in between a coverslip and breathable membrane as described in (Garcia et al. 2013; Bothma et al. 2014; Garcia and Gregor 2018) . For experiments quantifying the nuclear and cytoplasmic expression levels of the respective reporter proteins, non-virgin females and males from the same maternal protein line (e.g. *yw;+;eVasax10-pTrans-tdMCP-eGFP*) were crossed together. For experiments quantifying transcription as reported by MS2-MCP signal, virgin females of the maternal protein line (e.g. *yw;+;eNosx2-pTrans-MCP-mCherry-αTub3'UTR*) were crossed to males carrying the *hbP2P-24xMS2* transcription reporter gene.

Confocal microscopy was performed on a Leica SP8 with HyD detectors and a White Light Laser. We used a 63x oil objective, and scanned bidirectionally with a scan rate of 420 Hz and a magnification of 3.4x zoom. We did not use line or frame accumulation. Time-lapse z-stacks were collected with pixel dwell time of $1.2\mu s$ and 108 nm x-y pixel dimensions and $0.5~\mu m$ separation between z-slices (12 slices). x-y resolution was 856x856 pixels. Pinhole was set to 1.0 Airy units at 600 nm. eGFP was excited by a 488 nm laser line calibrated to 5 μW using the 10x objective and detected in a 520-567 nm spectral window. mCherry was excited by a 581 nm laser line calibrated to 14 μW and detected in a 566-669 nm spectral window. In all channels, detection was performed using the counting mode of the HyD detectors.

5.5 Image processing

Image analysis of MS2 movies and movies quantifying maternal promoter levels was performed in Matlab using the custom pipeline described in Garcia et al. 2013 and Lammers et al. 2020 (this pipeline can be found in the mRNA Dynamics Github repository). Image segmentation was also aided by the Trainable Weka Segmentation plugin in FIJI (Witten et al. 2016; Arganda-Carreras et al. 2017). Further analysis of time-series and other data were likewise performed in Matlab. Movies for publication were made in FIJI (Schneider, Rasband, and Eliceiri 2012; Schindelin et al. 2012). Standard candle image processing and quantification were performed in Python using the scikit-image package (Walt et al. 2014).

6   Biological material

	Backbone	**Enhancer**	**Promoter**	**Reporter**	**3'UTR**
1	pBPhi	eVasaReversex1	pTrans	tdMCP-eGFP	*none*
2	pBPhi	eVasax2	pTrans	tdMCP-eGFP	*none*
3	pBPhi	eVasax4	pTrans	tdMCP-eGFP	*none*
4	pBPhi	eVasax7	pTrans	tdMCP-eGFP	*none*
5	pBPhi	eVasax10	pTrans	tdMCP-eGFP	*none*
6	pBPhi	eVasax10	pTrans	tdMCP-eGFP	αTubulin
7	pBPhi	eVasax10	pTrans	eGFP	*none*
8	pBPhi	eVasax10	pTrans	eGFP	αTubulin
9	pBPhi	eNosx1	pTrans	tdMCP-eGFP	αTubulin(?)
10	pBPhi	eNosx2	pTrans	tdMCP-eGFP	αTubulin(?)
11	pBPhi	eNosx1	pTrans	MCP-mCherry	αTubulin
12	pBPhi	eNosx2	pTrans	MCP-mCherry	αTubulin
13	pBPhi	eNosx3	pTrans	MCP-mCherry	αTubulin
14	pBPhi	eNosx4	pTrans	MCP-mCherry	αTubulin
15	pBPhi	endogenous vasa	pVasa	tdMCP-eGFP	αTubulin(?)
16	pBPhi	endogenous nanos	pNos	tdMCP-eGFP	αTubulin(?)
17	pBPhi	endogenous αTub67C	pαTub67C	MCP-mCherry	αTubulin(?)
18	pBPhi	eVasax10	pTrans	myr-F2A-eGFP-60mer-eGFP	αTubulin
19	pBPhi	eVasax5	pTrans	60mer-mCherry-F2A-lamin	αTubulin
?	pBPhi	endogenous otu	pOtu	tdMCP-eGFP	αTubulin(?)

Table 1: **List of all plasmids constructed for this chapter.**

Primer name	Sequence (5' to 3')
31.35-synthpNos2-F-pBPhiHomo	ccggccagatccaggtcgcagcggccgcTTAC-TAAAAATATCGATATATTTAT
31.39-synthpNos4-R-pTrHomo	tacttcggtaagcttcggctatcgacggatccCGCGC-CAACTAACGGTTC
44.18-MCP-fwd-pTransHomo	tgcggacgaattttttttttgaaaacactagtATGCTG-GCTTCTAACTTTACTCAGTTCG
44.19-mCh-rv-tubUTRHomo	gagcgttgaagtggcgcgacgcttagtgtgATGCATT-TACTTGTACAGCTCGTCCATGC
44.5-mtrmPr-fwd-pCasperHomo	caaaaaagggtttttattaacttacatacatacta-gaattcggtaccAAAGTAGTTGGCAA
44.6-mtrmPr-rv-MCPHomo	gaacgaactgagtaaagttagaagccattttaacggc-tagcattttttttAGATACTGTG
49.5- DmelLaminDmoF	ATGTCGAGCAAATCCCGACGTGCT
49.6 DmelLaminDmoR	GCAGTCAAACGAGAAGTGCGCCATTATGTAA

Table 2: **List of all primers used in this chapter.** All primers are oriented 5' to 3'. Primer regions are uppercase, while Gibson homology regions are lowercase.

S1   Supplementary Tables

Name	Description	Length	Detectable protein?	Sequence
synthNos1	2x(Zen, DRE, brinker) binding sites, condensed	56 bp	No	TAAAAATATCGATAAGTTGGCGCGT-TAAAAATATCGATAAGTTGGCGCGTG-GATCC
synthNos2	-108 to -69 of *nanos*, contains 1 Zen and 1 DRE binding site	40 bp	No	TTACTAAAAATATCGATATATTTAT-TATCGCTGGAAAACT
synthNos3	-52 to -13 of *nanos*, contains 1 brinker binding site	40 bp	No	AAGCAAGAACCGTTAGTTGGCGCG-TAGCTTTACCACAAAA
synthNos4 (*eNos*)	minimal *nanos* enhancer used in this chapter; -108 to -29 of *nanos*, contains 1 Zen, 1 DRE, and 1 brinker binding site in endogenous positions	80 bp	Yes	TTACTAAAAATATCGATATATTTAT-TATCGCTGGAAAACTACATTATTCCAC-CTCTAAGCAAGAACCGTTAGTTGGCGCG

Table S1: **All minimal *nanos* enhancers versions made and characterized.**

Name	Description	Length	Sequence
eVasa	*vasa* -97 to -57, minimal *vasa* enhancer	40 bp	CTAAATCACATTTGATGTGTTAGTG-GAAAACGGCTATATA
pTrans	basal P element *transposase* promoter	98 bp	GTCGATAGCCGAAGCTTACC-GAAGTATACACTTAAATTCAGT-GCACGTTTGCTTGTTGA-GAGGAAAGGTTGTGTGCGGACGAATTTTTTTTTGAAAAC
eNos	*nanos* -108 to -29, minimal *nanos* enhancer	80 bp	TTACTAAAAATATCGATATATTTAT-TATCGCTGGAAAACTACATTATTCCAC-CTCTAAGCAAGAACCGTTAGTTGGCGCG
spacer-A	BamHI site	6 bp	GGATCC
spacer-B	intra-eVasa & intra-eNos spacer	6 bp	CTGGAA

Table S2: **Sequences of minimal enhancer and promoter elements.** The minimal *vasa* enhancer sequence (referred to as *eVasa* throughout this article) and maternal P element *transposase* promoter.

S2 Supplementary Figures
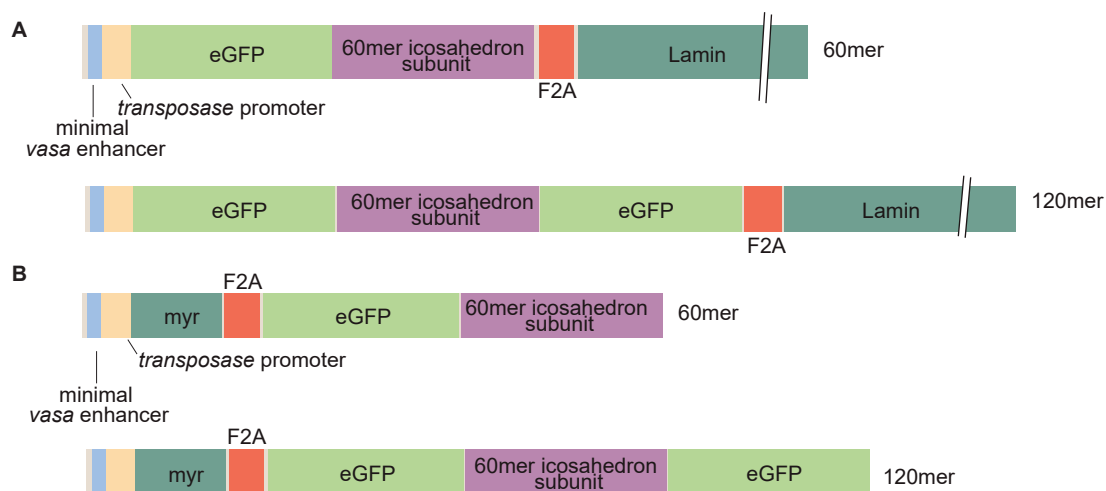
Figure S1: **Designs of the standard candle tethers. (A)** Design of the 60mer (top) and 120mer (bottom) Lamin tethered standard candle and **(B)** design of the 60mer (top) and 120mer (bottom) myristoyal tethered standard candle.

Bibliography

Akamatsu, Matthew et al. (Jan. 2020). "Principles of self-organization and load adaptation by the actin cytoskeleton during clathrin-mediated endocytosis". In: *eLife* 9. Ed. by Patricia Bassereau et al. Publisher: eLife Sciences Publications, Ltd, e49840. ISSN: 2050-084X. DOI: 10.7554/eLife.49840. URL: https://doi.org/10.7554/eLife.49840 (visited on 07/06/2021).

Alamos, Simon et al. (2020). "Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics". In: *bioRxiv*, p. 274621. DOI: 10.1101/2020.08.30.274621.

Ali, Ijaz et al. (Feb. 2010). "Cis-regulatory elements affecting the Nanos gene promoter in the germline stem cells". en. In: *Journal of Biotechnology* 145.4, pp. 323–329. ISSN: 0168-1656. DOI: 10.1016/j.jbiotec.2009.12.011. URL: https://www.sciencedirect.com/science/article/pii/S0168165609005859 (visited on 07/06/2021).

Arganda-Carreras, Ignacio et al. (Aug. 2017). "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification". en. In: *Bioinformatics* 33.15, pp. 2424–2426. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx180. URL: https://academic.oup.com/bioinformatics/article/33/15/2424/3092362 (visited on 04/09/2018).

Bai, L., A. Ondracka, and F. R. Cross (2011). "Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter". In: *Mol Cell* 42.4, pp. 465–76. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2011.03.028.

Bailey, Timothy L. et al. (July 2006). "MEME: discovering and analyzing DNA and protein sequence motifs". en. In: *Nucleic Acids Res* 34.suppl_2. Publisher: Oxford Academic, W369–W373. ISSN: 0305-1048. DOI: 10.1093/nar/gkl198. URL: https://academic.oup.com/nar/article/34/suppl_2/W369/2505578 (visited on 12/08/2020).

Bateman, Jack R., Anne M. Lee, and C.-ting Wu (June 2006). "Site-Specific Transformation of Drosophila via C31 Integrase-Mediated Cassette Exchange". en. In: *Genetics* 173.2. Publisher: Genetics Section: Investigations, pp. 769–777. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.106.056945. URL: https://www.genetics.org/content/173/2/769 (visited on 09/04/2020).

Bergman, Casey M., Joseph W. Carlson, and Susan E. Celniker (Apr. 2005). "Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster". en. In: *Bioinformatics*

21.8. Publisher: Oxford Academic, pp. 1747–1749. ISSN: 1367-4803. DOI: 10 . 1093 / bioinformatics / bti173. URL: https : / / academic . oup . com / bioinformatics / article/21/8/1747/249595 (visited on 11/21/2020).

Berrocal, A. et al. (2020). "Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene". In: *Elife* 9. ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10. 7554/eLife.61635.

Bertrand, Edouard et al. (1998). "Localization of ASH1 mRNA Particles in Living Yeast". In: *Molecular Cell* 2.4, pp. 437–445. ISSN: 10972765. DOI: 10.1016/S1097−2765(00)80143−4. URL: http://www.ncbi.nlm.nih.gov/pubmed/9809065{\%}5Cnhttp://linkinghub. elsevier.com/retrieve/pii/S1097276500801434.

Bier, Ethan et al. (Jan. 2018). "Advances in Engineering the Fly Genome with the CRISPR-Cas System". en. In: *Genetics* 208.1. Publisher: Genetics Section: FlyBook, pp. 1–18. ISSN: 0016-6731, 1943-2631. DOI: 10 . 1534 / genetics . 117 . 1113. URL: https : / / www . genetics.org/content/208/1/1 (visited on 05/06/2021).

Bintu, L. et al. (2005a). "Transcriptional regulation by the numbers: applications". In: *Curr Opin Genet Dev* 15.2, pp. 125–35.

Bintu, Lacramioara et al. (Apr. 2005b). "Transcriptional regulation by the numbers: models". In: *Curr Opin Genet Dev* 15.2, pp. 116–124. ISSN: 0959-437X. DOI: 10.1016/j.gde. 2005 . 02 . 007. URL: http : / / www . ncbi . nlm . nih . gov / pmc / articles / PMC3482385/ (visited on 02/20/2015).

Blythe, S. A. and E. F. Wieschaus (2016). "Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis". In: *Elife* 5. ISSN: 2050-084X (Electronic) 2050-084X (Linking). DOI: 10.7554/eLife.20148.

Bothma, Jacques P. et al. (2014). "Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos". In: *Proceedings of the National Academy of Sciences* 111.29, pp. 10598–10603. ISSN: 0027-8424. DOI: 10 . 1073 / pnas . 1410022111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1410022111{\% }5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/24994903.

Bradley, Robert K. et al. (Mar. 2010). "Binding Site Turnover Produces Pervasive Quantitative Changes in Transcription Factor Binding between Closely Related Drosophila Species". en. In: *PLOS Biology* 8.3. Publisher: Public Library of Science, e1000343. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1000343. URL: https://journals.plos.org/ plosbiology/article?id=10.1371/journal.pbio.1000343 (visited on 08/03/2021).

Brewster, Robert C. et al. (2014). "The transcription factor titration effect dictates level of gene expression". In: *Cell* 156.6, pp. 1312–1323. ISSN: 10974172. DOI: 10.1016/j.cell. 2014.02.022. URL: http://dx.doi.org/10.1016/j.cell.2014.02.022.

Briscoe, James and Stephen Small (Dec. 2015). "Morphogen rules: design principles of gradient-mediated embryo patterning". In: *Development* 142.23, pp. 3996–4009. ISSN: 0950-1991. DOI: 10.1242/dev.129452. URL: https://doi.org/10.1242/dev.129452 (visited on 05/31/2021).

Burz, B.S. et al. (1998). "Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo". In: *EMBO J* 17, pp. 5998–6009.

Chen, B. C. et al. (2014a). "Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution". In: *Science* 346.6208, p. 1257998. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1257998.

Chen, Hongtao et al. (2012). "A System of Repressor Gradients Spatially Organizes the Boundaries of Bicoid-Dependent Target Genes". In: *Cell* 149.3, pp. 618–629. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.03.018. URL: http://dx.doi.org/10.1016/j.cell.2012.03.018.

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

Chen, Jiji et al. (Mar. 2014b). "Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells". English. In: *Cell* 156.6. Publisher: Elsevier, pp. 1274–1285. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2014.01.062. URL: https://www.cell.com/cell/abstract/S0092-8674(14)00197-4 (visited on 08/03/2021).

Cheng, Quen J. et al. (June 2021). "NF-B dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages". en. In: *Science* 372.6548. Publisher: American Association for the Advancement of Science Section: Report, pp. 1349–1353. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abc0269. URL: https://science.sciencemag.org/content/372/6548/1349 (visited on 06/23/2021).

Cisse, II et al. (2013). "Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells". In: *Science*. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1239053.

Combs, P. A. and M. B. Eisen (2017). "Genome-wide measurement of spatial expression in patterning mutants of Drosophila melanogaster". In: *F1000Res* 6, p. 41. ISSN: 2046-1402 (Print) 2046-1402 (Linking). DOI: 10.12688/f1000research.9720.1.

Combs, Peter A. and Michael B. Eisen (Aug. 2013). "Sequencing mRNA from Cryo-Sliced Drosophila Embryos to Determine Genome-Wide Spatial Patterns of Gene Expression". en. In: *PLOS ONE* 8.8. Publisher: Public Library of Science, e71820. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0071820. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0071820 (visited on 08/03/2021).

Crocker, Justin, Garth R. Ilsley, and David L. Stern (Mar. 2016). "Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors". en. In: *Nat Genet* 48.3, pp. 292–298. ISSN: 1061-4036. DOI: 10.1038/ng.3509. URL: http://www.nature.com/ng/journal/v48/n3/abs/ng.3509.html (visited on 04/11/2016).

Crocker, Justin, Albert Tsai, and David L. Stern (2017). "A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote". In: *Cell Reports* 18.1, pp. 287–296. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.12.025.

Desponds, Jonathan, Massimo Vergassola, and Aleksandra M Walczak (July 2020). "A mechanism for hunchback promoters to readout morphogenetic positional information in less than a minute". In: *eLife* 9. Ed. by Michael B Eisen. Publisher: eLife Sciences Publications, Ltd, e49758. ISSN: 2050-084X. DOI: 10.7554/eLife.49758. URL: https://doi.org/10.7554/eLife.49758 (visited on 05/05/2021).

Desponds, Jonathan et al. (2016). "Precision of Readout at the hunchback Gene : Analyzing Short Transcription Time Traces in Living Fly Embryos". In: pp. 1–31. DOI: 10.1371/journal.pcbi.1005256.

Driever, W. and C. Nusslein-Volhard (1988a). "A gradient of bicoid protein in Drosophila embryos". In: *Cell* 54.1, pp. 83–93. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

— (1988b). "The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner". In: *Cell* 54.1, pp. 95–104. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

Driever, W., V. Siegel, and C. Nusslein-Volhard (1990). "Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen". In: *Development* 109.4, pp. 811–20. ISSN: 0950-1991 (Print) 0950-1991 (Linking).

Dufourt, Jeremy et al. (Dec. 2018). "Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs". en. In: *Nature Communications* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 5194. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07613-z. URL: https://www.nature.com/articles/s41467-018-07613-z (visited on 11/16/2020).

Dufourt, Jeremy et al. (2020). "Imaging translation dynamics in live embryos reveals spatial heterogeneities". In: *bioRxiv*, p. 2020.04.29.058974. DOI: 10.1101/2020.04.29.058974.

Eck, Elizabeth et al. (Oct. 2020). "Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility". In: *eLife* 9. Ed. by Pierre Sens. Publisher: eLife Sciences Publications, Ltd, e56429. ISSN: 2050-084X. DOI: 10.7554/eLife.56429. URL: https://doi.org/10.7554/eLife.56429 (visited on 11/13/2020).

Ephrussi, Anne and Daniel St Johnston (Jan. 2004). "Seeing is believing: the bicoid morphogen gradient matures". eng. In: *Cell* 116.2, pp. 143–152. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(04)00037-6.

Estrada, Javier et al. (2016). "Information Integration and Energy Expenditure in Gene Regulation". In: *Cell* 166.1, pp. 234–244. ISSN: 10974172. DOI: 10.1016/j.cell.2016.06.012.

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

Foo, S. M. et al. (2014). "Zelda potentiates morphogen activity by increasing chromatin accessibility". In: *Curr Biol* 24.12, pp. 1341–6. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: 10.1016/j.cub.2014.04.032.

Fritzsch, Christoph et al. (2018). "Estrogen-dependent control and cell-to-cell variability of transcriptional bursting". In: *Molecular Systems Biology* 14.2, pp. 1–17. ISSN: 1744-4292. DOI: 10.15252/msb.20177678.

Fukaya, T., B. Lim, and M. Levine (2016a). "Enhancer Control of Transcriptional Bursting". In: *Cell* 166.2, pp. 358–368. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2016.05.025.

Fukaya, Takashi (May 2021). "Dynamic regulation of anterior-posterior patterning genes in living Drosophila embryos". en. In: *Current Biology* 31.10, 2227–2236.e6. ISSN: 0960-9822. DOI: 10.1016/j.cub.2021.02.050. URL: https://www.sciencedirect.com/science/article/pii/S0960982221002943 (visited on 05/31/2021).

Fukaya, Takashi, Bomyi Lim, and Michael Levine (July 2016b). "Enhancer Control of Transcriptional Bursting". English. In: *Cell* 166.2. Publisher: Elsevier, pp. 358–368. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.05.025. URL: https://www.cell.com/cell/abstract/S0092-8674(16)30573-6 (visited on 06/17/2021).

Fuqua, Timothy et al. (Apr. 2020). "Dense encoding of developmental regulatory information may constrain evolvability". en. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 2020.04.17.046052. DOI: 10.1101/2020.04.17.046052. URL: https://www.biorxiv.org/content/10.1101/2020.04.17.046052v1 (visited on 09/04/2020).

Fussner, Eden, Reagan W. Ching, and David P. Bazett-Jones (Jan. 2011). "Living without 30 nm chromatin fibers". English. In: *Trends in Biochemical Sciences* 36.1. Publisher: Elsevier, pp. 1–6. ISSN: 0968-0004. DOI: 10.1016/j.tibs.2010.09.002. URL: https://www.cell.com/trends/biochemical-sciences/abstract/S0968-0004(10)00169-6 (visited on 01/19/2021).

Garcia, H. G., R. C. Brewster, and R. Phillips (2016). "Using synthetic biology to make cells tomorrow's test tubes". In: *Integr Biol (Camb)* 8.4, pp. 431–50. ISSN: 1757-9708 (Electronic) 1757-9694 (Linking). DOI: 10.1039/c6ib00006a.

Garcia, H. G. et al. (2020). "Lighting up the central dogma for predictive developmental biology". In: *Curr Top Dev Biol* 137, pp. 1–35. ISSN: 1557-8933 (Electronic) 0070-2153 (Linking). DOI: 10.1016/bs.ctdb.2019.10.010.

Garcia, Hernan G. and Thomas Gregor (2018). "Live Imaging of mRNA Synthesis in Drosophila". In: *RNA Detection: Methods and Protocols*. Ed. by Imre Gaspar. New York, NY: Springer New York, pp. 349–357.

Garcia, Hernan G. and Rob Phillips (July 2011). "Quantitative dissection of the simple repression input–output function". en. In: *PNAS* 108.29, pp. 12173–12178. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1015616108. URL: http://www.pnas.org/content/108/29/12173 (visited on 12/11/2017).

Garcia, Hernan G. et al. (Nov. 2013). "Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning". English. In: *Current Biology* 23.21. Publisher: Elsevier, pp. 2140–2145. ISSN: 0960-9822. DOI: 10.1016/j.cub.2013.08.054. URL: https://www.cell.com/current-biology/abstract/S0960-9822(13)01113-5 (visited on 09/04/2020).

Germier, Thomas et al. (Oct. 2017). "Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement". eng. In: *Biophys J* 113.7, pp. 1383–1394. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2017.08.014.

Gilbert, Scott F. (2010). *Developmental biology*. 9th. Sunderland, Mass.: Sinauer Associates, xxi, 711, 80 p. ISBN: 9780878933846 (casebound) 0878933840 (casebound) 9780878935642 (hbk.) 0878935649 (hbk.)

Gratz, Scott J. et al. (2015). "CRISPR-Cas9 genome editing in Drosophila". In: *Current Protocols in Molecular Biology* 2015.July, pp. 31.2.1–31.2.20. ISSN: 19343647. DOI: 10.1002/0471142727.mb3102s111.

Gregor, T. et al. (2005). "Diffusion and scaling during early embryonic pattern formation". In: *Proc Natl Acad Sci U S A* 102.51, pp. 18403–7.

Gregor, Thomas et al. (July 2007). "Probing the Limits to Positional Information". English. In: *Cell* 130.1. Publisher: Elsevier, pp. 153–164. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2007.05.025. URL: https://www.cell.com/cell/abstract/S0092-8674(07)00662-9 (visited on 07/09/2021).

Groth, Amy C. et al. (Apr. 2004). "Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage C31". en. In: *Genetics* 166.4. Publisher: Genetics Section: Investigations, pp. 1775–1782. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.166.4.1775. URL: https://www.genetics.org/content/166/4/1775 (visited on 07/14/2021).

Haario, Heikki et al. (Dec. 2006). "DRAM: Efficient adaptive MCMC". en. In: *Stat Comput* 16.4, pp. 339–354. ISSN: 1573-1375. DOI: 10.1007/s11222-006-9438-0. URL: https://doi.org/10.1007/s11222-006-9438-0 (visited on 06/07/2021).

Hafner, Antonina et al. (2020). "Quantifying the Central Dogma in the p53 Pathway in Live Single Cells". In: *Cell Systems*, pp. 1–11. ISSN: 24054712. DOI: 10.1016/j.cels.2020.05.001.

Hamm, Danielle C., Eliana R. Bondra, and Melissa M. Harrison (Feb. 2015). "Transcriptional Activation Is a Conserved Feature of the Early Embryonic Factor Zelda That Requires a Cluster of Four Zinc Fingers for DNA Binding and a Low-complexity Activation Domain *". English. In: *Journal of Biological Chemistry* 290.6. Publisher: Elsevier, pp. 3508–3518. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M114.602292. URL: https://www.jbc.org/article/S0021-9258(20)49197-5/abstract (visited on 08/03/2021).

Hannon, Colleen E, Shelby A Blythe, and Eric F Wieschaus (Sept. 2017). "Concentration dependent chromatin states induced by the bicoid morphogen gradient". In: *eLife* 6. Ed. by Joaquín M Espinosa. Publisher: eLife Sciences Publications, Ltd, e28275. ISSN: 2050-084X. DOI: 10.7554/eLife.28275. URL: https://doi.org/10.7554/eLife.28275 (visited on 07/06/2021).

Hansen, Anders S. and Erin K. O???Shea (2015). "Limits on information transduction through amplitude and frequency regulation of transcription factor activity". In: *eLife* 4.MAY, pp. 1–19. ISSN: 2050084X. DOI: 10.7554/eLife.06559.

Hansen, Eric R et al. (2017). "The 5 End of the Pea Ferredoxin-1 mRNA Mediates Rapid and Reversible Light-Directed Changes in Translation in Tobacco 1". In:

Harden, Timothy T., Ben J. Vincent, and Angela H. DePace (2021). "Defining kinetic roles of transcriptional activators in the early Drosophila embryo". In: *bioRxiv*, p. 2021.02.25.432925. DOI: 10.1101/2021.02.25.432925.

Harrison, Melissa M. et al. (2011). "Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition". In: *PLoS Genetics* 7.10. ISSN: 15537390. DOI: 10.1371/journal.pgen.1002266.

Hertz, G. Z. and G. D. Stormo (July 1999). "Identifying DNA and protein patterns with statistically significant alignments of multiple sequences." en. In: *Bioinformatics* 15.7. Publisher: Oxford Academic, pp. 563–577. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/15.7.563. URL: https://academic.oup.com/bioinformatics/article/15/7/563/278226 (visited on 09/03/2020).

Hong, Joung Woo, David A. Hendrix, and Michael S. Levine (2008). "Shadow enhancers as a source of evolutionary novelty". In: *Science* 321.5894, p. 1314. ISSN: 00368075. DOI: 10.1126/science.1160631.

Hsia, Yang et al. (July 2016). "Design of a hyperstable 60-subunit protein icosahedron". en. In: *Nature* 535.7610, pp. 136–139. ISSN: 1476-4687. DOI: 10.1038/nature18010. URL: https://www.nature.com/articles/nature18010 (visited on 07/06/2021).

Huang, Anqi et al. (2017). "Decoding temporal interpretation of the morphogen Bicoid in the early Drosophila embryo". In: pp. 1–21. DOI: 10.7554/eLife.26258.

Ip, Y. Tony et al. (1992). "dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo". In: *Genes and Development* 6.8, pp. 1518–1530. ISSN: 08909369. DOI: 10.1101/gad.6.8.1518.

Irizarry, Jihyun et al. (2020). "Twist-dependent ratchet functioning downstream from Dorsal revealed using a light-inducible degron". In: *Genes and Development* 34.13-14, pp. 965–972. ISSN: 15495477. DOI: 10.1101/GAD.338194.120.

Ivan, Andra, Marc S. Halfon, and Saurabh Sinha (Jan. 2008). "Computational discovery of cis-regulatory modules in Drosophila without prior knowledge of motifs". In: *Genome Biology* 9.1, R22. ISSN: 1474-760X. DOI: 10.1186/gb-2008-9-1-r22. URL: https://doi.org/10.1186/gb-2008-9-1-r22 (visited on 05/03/2021).

Jiang, J. et al. (1991). "The dorsal morphogen gradient regulates the mesoderm determinant twist in early Drosophila embryos". In: *Genes and Development* 5.10, pp. 1881–1891. ISSN: 08909369. DOI: 10.1101/gad.5.10.1881.

Jiang, Jin et al. (1993). "Conversion of a dorsal-dependent silencer into enhancer : evidence for dorsal corepressors". In: 12.8, pp. 3201–3209.

Kim, Harold D. and Erin K. O'Shea (2008). "A quantitative model of transcription factor-activated gene expression". In: *Nature Structural and Molecular Biology* 15.11, pp. 1192–1198. ISSN: 15459993. DOI: 10.1038/nsmb.1500.

King, Neil P. et al. (June 2014). "Accurate design of co-assembling multi-component protein nanomaterials". en. In: *Nature* 510.7503. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7503 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Molecular self-assembly;Protein design Subject_term_id: molecular-self-assembly;protein-design, pp. 103–108. ISSN: 1476-4687.

DOI: 10.1038/nature13404. URL: https://www.nature.com/articles/nature13404 (visited on 07/06/2021).

Kirov, Nikolai et al. (1993). "Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in Drosophila". In: 12.8, pp. 3193–3199.

Kremers, G. J. et al. (2006). "Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Forster radius". In: *Biochemistry* 45.21, pp. 6570–80. ISSN: 0006-2960 (Print) 0006-2960 (Linking). DOI: 10.1021/bi0516273.

La Rosee, A. et al. (1997). "Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the Drosophila embryo". In: *EMBO J* 16.14, pp. 4403–11. ISSN: 0261-4189 (Print) 0261-4189 (Linking). DOI: 10.1093/emboj/16.14.4403.

Lam, F. H., D. J. Steger, and E. K. O'Shea (2008). "Chromatin decouples promoter threshold from dynamic range". In: *Nature* 453.7192, pp. 246–50. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature06867.

Lambert, Talley J. (Apr. 2019). "FPbase: a community-editable fluorescent protein database". en. In: *Nature Methods* 16.4. Number: 4 Publisher: Nature Publishing Group, pp. 277–278. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0352-8. URL: https://www.nature.com/articles/s41592-019-0352-8 (visited on 01/22/2021).

Lammers, Nicholas C. et al. (Jan. 2020). "Multimodal transcriptional control of pattern formation in embryonic development". en. In: *PNAS* 117.2. Publisher: National Academy of Sciences Section: Physical Sciences, pp. 836–847. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1912500117. URL: https://www.pnas.org/content/117/2/836 (visited on 09/04/2020).

Lebrecht, Danielle et al. (2005). "Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila." In: *Proceedings of the National Academy of Sciences of the United States of America* 102.37, pp. 13176–81. ISSN: 0027-8424. DOI: 10.1073/pnas.0506462102. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0506462102{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/16150708{\%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1201621.

Levine, Mike (Sept. 2010). "Transcriptional Enhancers in Animal Development and Evolution". English. In: *Current Biology* 20.17. Publisher: Elsevier, R754–R763. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.06.070. URL: https://www.cell.com/current-biology/abstract/S0960-9822(10)00856-0 (visited on 01/19/2021).

Li, G. W. et al. (2014). "Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources". In: *Cell* 157.3, pp. 624–35. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2014.02.033.

Li, X. Y. et al. (2008). "Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm". In: *PLoS Biol* 6.2, e27. ISSN: 1545-7885 (Electronic) 1544-9173 (Linking). DOI: 10.1371/journal.pbio.0060027.

Li, Xiao-Yong and Michael B. Eisen (2018). "Zelda potentiates transcription factor binding to zygotic enhancers by increasing local chromatin accessibility during early ¡em¿Drosophila melanogaster¡/em¿ embryogenesis". In: *bioRxiv*, p. 380857. DOI: 10.1101/380857.

Liang, Hsiao-Lan et al. (2008). "The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila." In: *Nature* 456.7220, pp. 400–403. ISSN: 0028-0836. DOI: 10.1038/nature07388. arXiv: 1111.6600. URL: http://www.nature.com/nature/journal/v456/n7220/pdf/nature07388.pdf.

Liu, Jonathan et al. (May 2021). "Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage". en. In: *PLOS Computational Biology* 17.5. Publisher: Public Library of Science, e1008999. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008999. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008999 (visited on 06/17/2021).

Liu, Yin et al. (2014). "Nucleolus-tethering system (NoTS) reveals that assembly of photobodies follows a self-organization model." In: *Molecular biology of the cell* 25.8, pp. 1366–73. ISSN: 1939-4586. DOI: 10.1091/mbc.E13-09-0527. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3983000{\&}tool=pmcentrez{\&}rendertype=abstract.

Liu, Zhe, Luke D. Lavis, and Eric Betzig (May 2015). "Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level". en. In: *Molecular Cell* 58.4, pp. 644–659. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.02.033. URL: https://www.sciencedirect.com/science/article/pii/S1097276515001653 (visited on 08/03/2021).

Lucas, Tanguy et al. (2013). "Live imaging of bicoid-dependent transcription in Drosophila embryos." In: *Current Biology* 23.21, pp. 2135–9. ISSN: 1879-0445. DOI: 10.1016/j.cub.2013.08.053. URL: http://dx.doi.org/10.1016/j.cub.2013.08.053http://www.ncbi.nlm.nih.gov/pubmed/24139736.

Lucchetta, E. M. et al. (2005). "Dynamics of Drosophila embryonic patterning network perturbed in space and time using microfluidics". In: *Nature* 434.7037, pp. 1134–8. ISSN: 1476-4687 (Electronic). DOI: nature03509[pii]10.1038/nature03509.

Markstein, Michele et al. (2002). "Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo". In: 99.2, pp. 763–768.

Mazza, Davide et al. (Aug. 2012). "A benchmark for chromatin binding measurements in live cells". In: *Nucleic Acids Research* 40.15, e119–e119. ISSN: 0305-1048. DOI: 10.1093/nar/gks701. URL: https://doi.org/10.1093/nar/gks701 (visited on 08/03/2021).

McDaniel, Stephen L. et al. (2019). "Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation". In: *Molecular Cell* 74.1, 185–195.e4. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.01.014. URL: https://doi.org/10.1016/j.molcel.2019.01.014.

Meijsing, S. H. et al. (2009). "DNA binding site sequence directs glucocorticoid receptor structure and activity". In: *Science* 324.5925, pp. 407–10. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1164265.

Mir, M. et al. (2017). "Dense Bicoid hubs accentuate binding along the morphogen gradient". In: *Genes Dev* 31.17, pp. 1784–1794. ISSN: 1549-5477 (Electronic) 0890-9369 (Linking). DOI: 10.1101/gad.305078.117.

Morrison, Alexander H. et al. (Apr. 2012). "Quantifying the Bicoid Morphogen Gradient in Living Fly Embryos". In: *Cold Spring Harbor protocols* 2012.4, pp. 398–406. ISSN: 1940-3402. DOI: 10.1101/pdb.top068536. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001741/ (visited on 08/03/2021).

Nien, C. Y. et al. (2011). "Temporal coordination of gene networks by Zelda in the early Drosophila embryo". In: *PLoS Genet* 7.10, e1002339. ISSN: 1553-7404 (Electronic) 1553-7390 (Linking). DOI: 10.1371/journal.pgen.1002339.

Normanno, Davide et al. (2015). "Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher." In: *Nature communications* 6, p. 7357. ISSN: 2041-1723. DOI: 10.1038/ncomms8357. URL: http://www.ncbi.nlm.nih.gov/pubmed/26151127.

Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila". In: *Nature* 287.5785, pp. 795–801. ISSN: 0028-0836 (Print) 0028-0836 (Linking).

Ochoa-Espinosa, A. et al. (2005). "The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila". In: *Proc Natl Acad Sci U S A* 102.14, pp. 4960–5. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.0500373102.

Ochoa-espinosa, Amanda et al. (2009). "Anterior-posterior positional information in the absence of a strong Bicoid gradient". In: 106.10, pp. 3823–3828.

Papagianni, Aikaterini et al. (2018). "Capicua controls Toll/IL-1 signaling targets independently of RTK regulation". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.8, pp. 1807–1812. ISSN: 10916490. DOI: 10.1073/pnas.1713930115.

Park, Jeehae et al. (2019). "Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity". In: pp. 1–25.

Perry, M. W. et al. (2012). "Precision of hunchback expression in the Drosophila embryo". In: *Curr Biol* 22.23, pp. 2247–52. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: 10.1016/j.cub.2012.09.051.

Phillips, Rob et al. (2019). "Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression". In:

Planchon, Thomas A. et al. (May 2011). "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination". en. In: *Nature Methods* 8.5. Number: 5 Publisher: Nature Publishing Group, pp. 417–423. ISSN: 1548-7105. DOI: 10.1038/nmeth.1586. URL: https://www.nature.com/articles/nmeth.1586 (visited on 08/03/2021).

Polach, K. J. and J. Widom (1995). "Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation". In: *J Mol Biol* 254.2, pp. 130–49. ISSN: 0022-2836 (Print).

Popp, Achim P et al. (2020). "Transcription factor residence time dominates over concentration in transcription activation". In: *bioRxiv*, p. 2020.11.26.400069. URL: https://doi.org/10.1101/2020.11.26.400069.

Porcher, Aude and Nathalie Dostatni (2010). "The Bicoid Morphogen System". In: *Current Biology* 20.5, pp. 249–254. ISSN: 09609822. DOI: 10.1016/j.cub.2010.01.026.

Razo-Mejia, M. et al. (2018). "Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction". In: *Cell Syst* 6.4, 456–469 e10. ISSN: 2405-4712 (Print) 2405-4712 (Linking). DOI: 10.1016/j.cels.2018.02.004.

Reeves, Gregory T. et al. (Mar. 2012a). "Dorsal-Ventral Gene Expression in the Drosophila Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient". English. In: *Developmental Cell* 22.3. Publisher: Elsevier, pp. 544–557. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2011.12.007. URL: https://www.cell.com/developmental-cell/abstract/S1534-5807(11)00573-9 (visited on 09/03/2020).

— (2012b). "Dorsal-Ventral Gene Expression in the Drosophila Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient". In: *Developmental Cell* 22.3, pp. 544–557. ISSN: 15345807. DOI: 10.1016/j.devcel.2011.12.007. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.devcel.2011.12.007.

Reimer, Armando et al. (July 2021). "Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient". en. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 2021.07.10.451524. DOI: 10.1101/2021.07.10.451524. URL: https://www.biorxiv.org/content/10.1101/2021.07.10.451524v1 (visited on 07/12/2021).

Rivera-Pomar, R. et al. (1995). "Activation of posterior gap gene expression in the Drosophila blastoderm". In: *Nature* 376.6537, pp. 253–6. ISSN: 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/376253a0.

Roth, Siegfried, David Stein, and Christiane Nüsslein-Volhard (1989). "A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo". In: *Cell* 59.6, pp. 1189–1202. ISSN: 00928674. DOI: 10.1016/0092-8674(89)90774-5.

Rubin, G. M. and A. C. Spradling (Oct. 1982). "Genetic transformation of Drosophila with transposable element vectors". en. In: *Science* 218.4570. Publisher: American Association for the Advancement of Science Section: Articles, pp. 348–353. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.6289436. URL: https://science.sciencemag.org/content/218/4570/348 (visited on 07/13/2021).

Rushlow, Christine A. and Stanislav Y. Shvartsman (2012). "Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient". In: *Current Opinion in Genetics and Development* 22.6, pp. 542–546. ISSN: 0959437X. DOI: 10.1016/j.gde.2012.08.005. URL: http://dx.doi.org/10.1016/j.gde.2012.08.005.

Sandler, Jeremy E and Angelike Stathopoulos (2016). "Stepwise Progression of Embryonic Patterning". In: 32.7, pp. 432–443.

Sano, Hiroko, Akira Nakamura, and Satoru Kobayashi (Mar. 2002). "Identification of a transcriptional regulatory region for germline-specific expression of vasa gene in Drosophila melanogaster". en. In: *Mechanisms of Development* 112.1, pp. 129–139. ISSN: 0925-4773. DOI: 10.1016/S0925-4773(01)00654-2. URL: https://www.sciencedirect.com/science/article/pii/S0925477301006542 (visited on 07/06/2021).

Sayal, Rupinder et al. (May 2016). "Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early Drosophila embryo". In: *eLife* 5. Ed. by Naama Barkai. Publisher: eLife Sciences Publications, Ltd, e08445. ISSN: 2050-084X. DOI: 10.7554/eLife.08445. URL: https://doi.org/10.7554/eLife.08445 (visited on 01/25/2021).

Schindelin, Johannes et al. (July 2012). "Fiji: an open-source platform for biological-image analysis". en. In: *Nature Methods* 9.7, pp. 676–682. ISSN: 1548-7105. DOI: 10.1038/nmeth.2019. URL: https://www.nature.com/articles/nmeth.2019 (visited on 04/09/2018).

Schneider, Caroline A, Wayne S Rasband, and Kevin W Eliceiri (June 2012). "NIH Image to ImageJ: 25 years of image analysis". In: *Nature Methods* 9, p. 671. URL: http://dx.doi.org/10.1038/nmeth.2089.

Scholes, Clarissa and Angela H Depace (2017). "Math j Bio Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle Math j Bio Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle". In: pp. 97–108. DOI: 10.1016/j.cels.2016.11.012.

Schulz, K. N. et al. (2015). "Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo". In: *Genome Res* 25.11, pp. 1715–26. ISSN: 1549-5469 (Electronic) 1088-9051 (Linking). DOI: 10.1101/gr.192682.115.

Schulze, Sandra R. and Lori L. Wallrath (Dec. 2006). "Gene Regulation by Chromatin Structure: Paradigms Established in Drosophila melanogaster". In: *Annu. Rev. Entomol.* 52.1. Publisher: Annual Reviews, pp. 171–192. ISSN: 0066-4170. DOI: 10.1146/annurev.ento.51.110104.151007. URL: https://www.annualreviews.org/doi/10.1146/annurev.ento.51.110104.151007 (visited on 01/19/2021).

Selby, Christopher P. et al. (Feb. 1997). "RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair". In: *Nucleic Acids Research* 25.4, pp. 787–793. ISSN: 0305-1048. DOI: 10.1093/nar/25.4.787. URL: https://doi.org/10.1093/nar/25.4.787 (visited on 06/09/2021).

Sergé, Arnauld et al. (Aug. 2008). "Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes". en. In: *Nature Methods* 5.8. Number: 8 Publisher: Nature Publishing Group, pp. 687–694. ISSN: 1548-7105. DOI: 10.1038/nmeth.1233. URL: https://www.nature.com/articles/nmeth.1233 (visited on 08/03/2021).

Shermoen, Antony W. and Patrick H. O'Farrell (Oct. 1991). "Progression of the cell cycle through mitosis leads to abortion of nascent transcripts". English. In: *Cell* 67.2. Publisher: Elsevier, pp. 303–310. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/0092-8674(91)90182-X. URL: https://www.cell.com/cell/abstract/0092-8674(91)90182-X (visited on 01/18/2021).

Shin, Dong Hyeon and Joung Woo Hong (2014). "Capicua is involved in Dorsal-mediated repression of Zerknüllt expression in Drosophila embryo". In: *BMB Reports* 47.9, pp. 518–523. ISSN: 1976670X. DOI: 10.5483/BMBRep.2014.47.9.122.

Simpson-Brose, M., J. Treisman, and C. Desplan (1994). "Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila". In: *Cell* 78.5, pp. 855–65. ISSN: 0092-8674 (Print) 0092-8674 (Linking).

Small, S., A. Blair, and M. Levine (1996). "Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo". In: *Dev Biol* 175.2, pp. 314–24.

Spradling, A. C. and G. M. Rubin (Oct. 1982). "Transposition of cloned P elements into Drosophila germ line chromosomes". en. In: *Science* 218.4570. Publisher: American Association for the Advancement of Science Section: Articles, pp. 341–347. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.6289435. URL: https://science.sciencemag.org/content/218/4570/341 (visited on 07/13/2021).

Stormo, G. D. and 3rd Hartzell G. W. (1989). "Identifying protein-binding sites from un-aligned DNA fragments". In: *Proc Natl Acad Sci U S A* 86.4, pp. 1183–7. ISSN: 0027-8424 (Print) 0027-8424 (Linking).

Stormo, G. D., T. D. Schneider, and L. Gold (1986). "Quantitative analysis of the relationship between nucleotide sequence and functional activity". In: *Nucleic Acids Res* 14.16, pp. 6661–79. ISSN: 0305-1048 (Print) 0305-1048 (Linking).

Struhl, Gary, Kevin Struhl, and Paul M. Macdonald (1989). "The gradient morphogen bicoid is a concentration-dependent transcriptional activator". In: *Cell* 57.7, pp. 1259–1273. ISSN: 00928674. DOI: 10.1016/0092-8674(89)90062-7.

Sun, Yujia et al. (2015). "Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation". In: *Genome Research* 25.11, pp. 1703–1714. ISSN: 15495469. DOI: 10.1101/gr.192542.115.

Szymanski, P and M Levine (1995). "Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo." In: *The EMBO Journal* 14.10, pp. 2229–2238. ISSN: 0261-4189. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC398329/{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC398329/pdf/emboj00034-0107.pdf.

Thisse, C et al. (1991). "Sequence-specific transactivation of the Drosophila twist gene by the dorsal gene product." In: *Cell* 65, pp. 1191–1201. ISSN: 00928674. DOI: 10.1016/0092-8674(91)90014-P.

Tokunaga, Makio, Naoko Imamoto, and Kumiko Sakata-Sogawa (Feb. 2008). "Highly inclined thin illumination enables clear single-molecule imaging in cells". en. In: *Nature Methods* 5.2. Number: 2 Publisher: Nature Publishing Group, pp. 159–161. ISSN: 1548-7105. DOI: 10.1038/nmeth1171. URL: https://www.nature.com/articles/nmeth1171 (visited on 08/03/2021).

Tran, H. et al. (2018). "Precision in a rush: Trade-offs between reproducibility and steepness of the hunchback expression pattern". In: *PLoS Comput Biol* 14.10, e1006513. ISSN: 1553-7358 (Electronic) 1553-734X (Linking). DOI: 10.1371/journal.pcbi.1006513.

Turing, Alan Mathison (Aug. 1952). "The chemical basis of morphogenesis". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237.641. Publisher: Royal Society, pp. 37–72. DOI: 10.1098/rstb.1952.0012. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.1952.0012 (visited on 08/03/2021).

Tutucci, Evelina et al. (Jan. 2018). "An improved MS2 system for accurate reporting of the mRNA life cycle". In: *Nat Methods* 15.1, pp. 81–89. ISSN: 1548-7091. DOI: 10.1038/nmeth.4502. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5843578/ (visited on 09/03/2020).

Venken, Koen J. T. and Hugo J. Bellen (Mar. 2005). "Emerging technologies for gene manipulation in Drosophila melanogaster". en. In: *Nature Reviews Genetics* 6.3. Number: 3 Publisher: Nature Publishing Group, pp. 167–178. ISSN: 1471-0064. DOI: 10.1038/nrg1553. URL: https://www.nature.com/articles/nrg1553 (visited on 05/06/2021).

Venken, Koen J. T. et al. (June 2009). "Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster". en. In: *Nature Methods* 6.6. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6 Primary_atype: Research Publisher: Nature Publishing Group, pp. 431–434. ISSN: 1548-7105. DOI: 10.1038/nmeth.1331. URL: https://www.nature.com/articles/nmeth.1331 (visited on 07/14/2021).

Vincent, Ben J., Javier Estrada, and Angela H. DePace (2016). "The appeasement of Doug: a synthetic approach to enhancer biology". In: *Integr. Biol.* ISSN: 1757-9694. DOI: 10.1039/C5IB00321K. URL: http://xlink.rsc.org/?DOI=C5IB00321K.

Walt, Stéfan van der et al. (June 2014). "scikit-image: image processing in Python". en. In: *PeerJ* 2. Publisher: PeerJ Inc., e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453. URL: https://peerj.com/articles/453 (visited on 07/09/2021).

Witten, Ian H. et al. (Oct. 2016). *Data Mining: Practical Machine Learning Tools and Techniques*. en. Google-Books-ID: 1SylCgAAQBAJ. Morgan Kaufmann. ISBN: 978-0-12-804357-8.

Wolpert, L. (1969). "Positional information and the spatial pattern of cellular differentiation". In: *Journal of Theoretical Biology* 25.1, pp. 1–47. ISSN: 10958541. DOI: 10.1016/S0022-5193(69)80016-0.

Wu, Di et al. (2012). "Structural basis of ultraviolet-B perception by UVR8." In: *Nature* 484.7393, pp. 214–219. ISSN: 1476-4687. DOI: 10.1038/nature10931. URL: http://www.ncbi.nlm.nih.gov/pubmed/22388820.

Wunderlich, Zeba and Leonid A. Mirny (Oct. 2009). "Different gene regulation strategies revealed by analysis of binding motifs". English. In: *Trends in Genetics* 25.10. Publisher: Elsevier, pp. 434–440. ISSN: 0168-9525. DOI: 10.1016/j.tig.2009.08.003. URL: https://www.cell.com/trends/genetics/abstract/S0168-9525(09)00165-6 (visited on 02/04/2021).

Xu, Xiaosa et al. (2015). "Illuminating Progress in Phytochrome-Mediated Light Signaling Pathways". In: *Trends in Plant Science* 20.10, pp. 641–650. ISSN: 13601385. DOI: 10.1016/j.tplants.2015.06.010. URL: http://dx.doi.org/10.1016/j.tplants.2015.06.010.

Xu, Zhe et al. (2014). "Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila". In: *Genes and Development* 28.6, pp. 608–621. ISSN: 15495477. DOI: 10.1101/gad.234534.113. arXiv: arXiv:1011.1669v3.

Yeung, Chuck, Matthew Shtrahman, and Xiao-lun Wu (Apr. 2007). "Stick-and-Diffuse and Caged Diffusion: A Comparison of Two Models of Synaptic Vesicle Dynamics". In: *Bio-*

physical Journal 92.7, pp. 2271–2280. ISSN: 0006-3495. DOI: 10.1529/biophysj.106. 081794. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1864819/ (visited on 08/03/2021).

Zhou, Jumin et al. (Nov. 1998). "TAFII mutations disrupt Dorsal activation in the Drosophila embryo". en. In: *PNAS* 95.23. Publisher: National Academy of Sciences Section: Biological Sciences, pp. 13483–13488. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.95.23.13483. URL: https://www.pnas.org/content/95/23/13483 (visited on 05/24/2021).