What is True for the Fruit Fly is True for Arabidopsis: Quantitative Live Imaging Uncovers Principles of Transcriptional Regulation in Eukaryotes.

By

Juan Simon Alamos Urzua

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Krishna K. Niyogi, Co-chair
Assistant Professor Hernan G. Garcia, Co-chair
Professor Peter Quail
Professor David Savage

Spring 2021

What is True for the Fruit Fly is True for Arabidopsis: Quantitative Live Imaging Uncovers Principles of Transcriptional Regulation in Eukaryotes.

Abstract

What is True for the Fruit Fly is True for Arabidopsis: Quantitative Live Imaging Uncovers Principles of Transcriptional Regulation in Eukaryotes.

by

Juan Simon Alamos Urzua

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Krishna K. Niyogi, Co-chair

Assistant Professor Hernan G. Garcia, Co-chair

Due to their apparent passive behavior, plants have long been considered something in between inanimate objects and ”proper” living things, that is, animals. When an animal is incapable of moving or communicating we refer to it as being in a vegetative state. Research over the last three decades has shown that, when it comes to molecules, plants cells are bustling with activity. It was our task to just go and take a look.

These dynamic molecular events operating at the cellular level include a myriad of signaling pathways that can sense environmental inputs such as light intensity and quality, temperature, osmotic pressure, the axis of gravity, mineral nutrients and several hormones to name but a few. The output of these signaling pathways often involves changes in the mRNA abundance of responsive genes. Plant growth and development depends on quantitative aspects of this regulation -when, where and how much of each mRNA species each cell makes in response to a specific stimulus. In turn, human societies ultimately depend on plants, which makes these problems not only fascinating from a basic science point of view, they are also pressing from a very practical perspective.

In large part, we owe our current understanding of how this information transfer occurs to studies in other eukaryotic model organisms, in particular the fruit fly *Drosophila melanogaster* (Drosophila). From the first experimental description of morphogens to the recognition of transcriptional enhancers, foundational discoveries related to gene regulation in higher eukaryotes have been made using the fruit fly embryo as a model system. The fact that this inter-kingdom comparisons are possible and useful highlights one of the most important lessons of the molecular biology revolution, the notion that all life forms share fundamental principles because they are built largely of the same

type of molecules. Jacques Monod captured this view when he famously claimed that 'what is true for *E. coli* is true for the Elephant'.

While Monod's quote nicely justifies the logic behind a specific research program, it also challenges us to question whether paradigms drawn in one model organism bias and limit our understanding when applied uncritically to other biological systems. Monod certainly was not expecting us to believe that the specific details of how bacterial genes are turned on and off in response to sugar sources are conserved all the way from bacteria to animals. However, the physical principles that nature exploits to bring about a bacterium and an elephant are ultimately the same. The genomic era has nothing but confirmed the underlying unity across all life kingdoms by revealing that the way these principles are encoded in DNA derives from one common ancestor. One of the challenges we face as biologists is finding what these shared principles are. In this effort, by distinguishing what is shared from what is different, drawing from different experimental systems can be illuminating. There are more practical reasons why a trans-kingdom perspective can be beneficial. Engaging at a deep level with the technical details of an experimental method developed in one model organism allows one to more easily transfer that knowledge to other contexts. A substantial part of this dissertation builds on this premise.

It should be possible to express these biophysical principles as models that can be applied across biological contexts. Since the dawn of molecular biology, transcriptional regulation has had a central role as a test bed to combine theory and experiments to uncover generalizable biological principles. A good example of this is the Monod-Wyman-Changeux (MWC) model of allostery, which captures such seemingly different behaviors as sugar regulation of gene expression in bacteria, oxygen transport in the blood by hemoglobin and the effect of drugs on cell surface receptors. During the summer of 2016 I had the privilege of attending the Physiology Course at the Marine Biological Laboratory where, under the guidance of Professor Rob Phillips, I used statistical mechanics to derive and test the MWC model in the context of sugar sensing in *E. coli*. This type of simple and intuitive mechanistic model is of a fundamentally different nature to the statistical regression-based models that are becoming increasingly popular in biology. Because they are conceived as arbitrarily complex black box 'fits' these models can certainly 'recapitulate the data', but, for that very same reason, cannot be used to test molecular or biophysical hypotheses.

We should, however, recognize the relative failure of using theory and experiments to develop predictive models of transcription in eukaryotes. Compared to the resounding successes in phage λ and the *E. coli Lac* operon (reviewed by [103] and [206]), we are far from achieving this type of understanding in plants or in any other higher eukaryote. Recent calls to advance a unified model of transcription in eukaryotes advocate

for making an inventory of moving parts and building tools to measure their dynamics in single cells [263]. Many of the results presented here align with the goal of closing technical gaps to make it possible to measure the phenomena under study. It may well be that we have reached a technical wall beyond which the dialog between theory and experiments will come naturally. For example, there is mounting evidence that the formation of biomolecular condensates of RNA Polymerase and transcription factors is a fundamental property of transcriptional regulation [36, 121]. The theoretical framework for these processes is an active area of research and it is thus still unclear what experiments could be designed to falsify these models [27, 124, 277]. Theoretical explorations, even if frustrating, can sharpen our thinking of what experimental tools we are still missing. For example, as I argue in the following chapters, there is a still a need for live cell technologies to measure post-translational modifications of histones and other proteins at a specific locus.

Beyond this quest for building unifying models based on physical principles, we should not forget that transcriptional regulation lies at the core of a myriad of diverse biological phenomena that we have only scratched the surface of. Simply taking a quantitative, single-cell view of gene regulation can offer unprecedented insights about the way a particular biological process works and make us revisit the way we think about it.

As I have argued here, toggling between model systems offers several advantages, however, it is does not come without risk. The practice of citing faulty research in animals models to buttress claims about plants goes back to the very origins of plant biology. Theophrastus, considered the father of botany, claimed that spontaneous generation in plants should not be taken with suspicion since it had already been demonstrated in animals:

If some [plants] are generated in both ways, spontaneously as well as by seed, there is no absurdity: so some animals similarly come from two sources, both from other animals and from earth.

It would be presumptuous for me to pretend that I have avoided the same mistake, but that did not stop me from trying. Chapter 1 consists on a comparative overview of novel insights about transcriptional regulation in *Arabidopsis thaliana* (Arabidopsis) and Drosophila acquired in the course of my research. In it, I try to find unifying themes and put results in the context of the field at large. Chapter 1 is also intended to give a relatively detailed overview of this dissertation and so many of its figures are found in chapters that deal with specific topics.

The rest of the chapters are as follows. In chapter 2, I describe the motivations to establish optogenetic tools to manipulate transcription factors in Drosophila embryos and the results from our efforts. Chapter 3 contains a manuscript that is currently in prepa-

ration that describes the development of a minimal synthetic transcriptional system in Drosophila and the insights we have gained from it. Chapter 4 contains a manuscript currently in bioRxiv and under peer review that describes how transcription operates at the single cell level in plants. Chapter 5 builds upon the live imaging tools presented in Chapter 4 and describes experiments to measure the dynamics of light signaling during a dark to light transition in Arabidopsis seedlings. Finally, in Chapter 6, I outline experiments that I was unable to finish that point to potential future directions.

To the friends I left in Chile. I miss them every day.

Acknowledgments

First and foremost, I want to thank my advisors, Hernan Garcia and Kris Niyogi who gave me a great degree of intellectual freedom to do exploratory research and make my own mistakes. I owe my scientific writing skills to the relentless and critical feedback from Hernan, who essentially taught me how to write a scientific publication. Teaching introduction to biophysics with him was one of the hardest and most rewarding experiences of my Ph.D, I am hoping to carry with me his pedagogical approach to science, lecturing and writing. I am also grateful for the team of researchers that my advisors have assembled and that I had the privilege of interacting with. If I ever knew anything, I learned it during group meetings. Even the most trivial and obvious of my questions never failed to receive a good, thoughtful answer from my lab mates. My advisors also consistently encouraged and supported me financially to attend international meetings and training courses that greatly influenced the way I think and work and allowed me to meet other mentor figures. Among these, Alexander 'Xander' Jones, Dominique Bergmann and Patrick Shih stand out. Professor Rob Phillips made a lasting impact on me with his advice to "work out at least 45 minutes three times a week" which I want to think has had a positive impact on my productivity and life expectancy. My collaborators made the work presented here possible. Armando Reimer greatly improved the computational image analysis pipeline of the Garcia laboratory which enabled the live imaging experiments showed here. He is my first coauthor in the manuscript in Chapter **??**. I worked closely with Yang Joon Kim attempting to generate functional optogenetic fly transcription factors. Jake Zhao took this efforts and generated the fly lines and the exciting results related to optogenetic Knirps, it has been a lot of fun working closely with him. Armando Reimer helped me test the effect of optogenetic perturbations on the Dorsal activator. Meghan Turner provided the reagents to use the ParB-*parS* DNA labeling technology. Peter Quail and Eduardo Gonzalez-Grandio provided critical unpublished results to create live imaging reporters of Phytochrome light signaling. Gabriella Martini set up the temperature controlled microscope stage used for heat shock treatments. Two fantastic undergraduate students made significant contributions to my research, Jordan Xiao in the optogenetics project and Clayton Westrum in the Dorsal synthetic enhancers project. It was a joy mentoring both of them. Others offered their technical knowledge at critical junctions. I could not have done RT-qPCR experiments without the guidance of Setsuko Wakao. I owe my agroinfilitration prowess to Johan Andersen-Ranberg. I have created more fly lines that I would like to admit, which was only possible thanks to the training in classic fly genetics from Jacques Bothma. His big picture knowledge of the literature and shrewd instinct for scientific questions left a lasting impact on me. I also would like to thank my Niyogi lab subgroup in general for their advice and guidance, in particular Chris Gee for allowing me to pick his brain about ideas. Last but not least I want to thank Marilyn Kobayashi for her patience and for making research in the Niyogi lab possible. The research presented here was funded in part by the Howard Hughes Medical Institute through Kris Niyogi. The laboratory of Hernan Garcia was supported

by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl and Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236).

Chapter 1

Comparative lessons

1.1 Introduction

To grow, develop and respond to their environment, organisms regulate *when*, *where*, and *how much* specific genes are transcribed into mRNA, this puts transcriptional regulation at the center of multiple biological processes. When, where and how much are, in essence, *quantitative* aspects with a precise meaning in terms of time, space and production rates. Thus, to fully understand the role of transcriptional regulation in any biological process we need to answer these questions quantitatively. This quantitative data is necessary to push forward our capacity to test existing mathematical models of gene regulation. In addition, theoretical models that will enable complex engineering applications in agriculture and medicine will demand this type of quantitative understanding.

Since the beginning of the sequencing revolution a deluge of information has accumulated regarding mRNA abundance across experimental conditions, genetic backgrounds and organisms. This could easily make us believe that we are close to the goal of achieving a quantitative view of transcription. However, as we discuss in this chapter, this is far from true. Save for a spare few exceptions, our current view of transcription is a qualitative and static one. Simultaneously, technical advances have enabled studying the protein that regulate transcription with unprecedented detail. Methods like ChIP-seq [202] and other more recent genome-wide techniques [198, 11, 21], make it now possible to determine transcription factor binding sites genome-wide. Thanks to advances in structural biology, the basis of protein DNA binding is now known with atomic resolution to the extent that it can be predicted and rationally engineered [35]. Yet, again, this tsunami of data belies our ignorance about how changes in transcription factor concentration are read out by regulatory DNA to dictate quantitative changes in transcriptional activity.

This chapter is organized as a review of experiments presented in detail in the following chapters. I describe multiple efforts aimed at achieving a quantitative model of

transcription in eukaryotic systems. Even though these efforts were carried out in very different models -plants and fruit flies- here, I have attempted, to the best of my ability, to find unifying themes. These themes are put in the context of the literature at large. Outstanding questions and follow up experiments are discussed as well.

1.2 Assigning numbers to arrows

The regulation of gene expression at the level of transcription plays a critical role in adjusting plant physiology and development in response to fluctuations in their environment. On the other hand, during the early development of the fruit fly *Drosophila melanogaster* embryo, concentration gradients of maternal transcription factors dictate the activity of zygotic genes. These seemingly different contexts can be conceptualized under the umbrella of input-output functions, where a signaling stimulus or a transcription factor protein are read out by the cell to regulate the output mRNA levels. These Input-output relationships are usually depicted as logical diagrams where arrows represent negative or positive interactions. In this dissertation, I deal with three such network diagrams. In the fruit fly, I used as an experimental system the activation of target genes by the Dorsal maternal transcription factor (Fig. 1.1A). In Arabidopsis, I studied the transcriptional aspects of Phytochrome-mediated light signaling and the heat shock response (Fig. 1.1B).

To engineer complex biological functions, it will be necessary to go beyond these wiring diagrams and build predictive models of gene regulation. The first task towards this goal is conceptually straightforward, yet technically challenging. First, it entails manipulating and/or measuring the transcriptional input (Fig. 1.1D). Further, it involves measuring transcriptional outputs in terms of RNA Polymerase activity or gene products. Because gene expression is a stochastic process, these measurements should ideally be done at the single cell level (Fig. 1.1E). The manuscript in Chapter 3 describes our efforts to establish an experimental system to do these measurements in live fruit fly embryos. The manuscript in Chapter 4 details the implementation of novel methods to quantify these input-output functions in plant cells. Next, in order to determine what determines the shape of this input-output functions, it is necessary to repeat these measurements under systematic perturbations of the factors that control the readout of the input. These include the binding affinity of transcription factors, the position of binding sites, the accessibility of these sites, and the deposition of histone modifications (Fig. 1.1C). Chapter 3 describes experiments where we systematically explored the role of transcription factor binding site affinity.

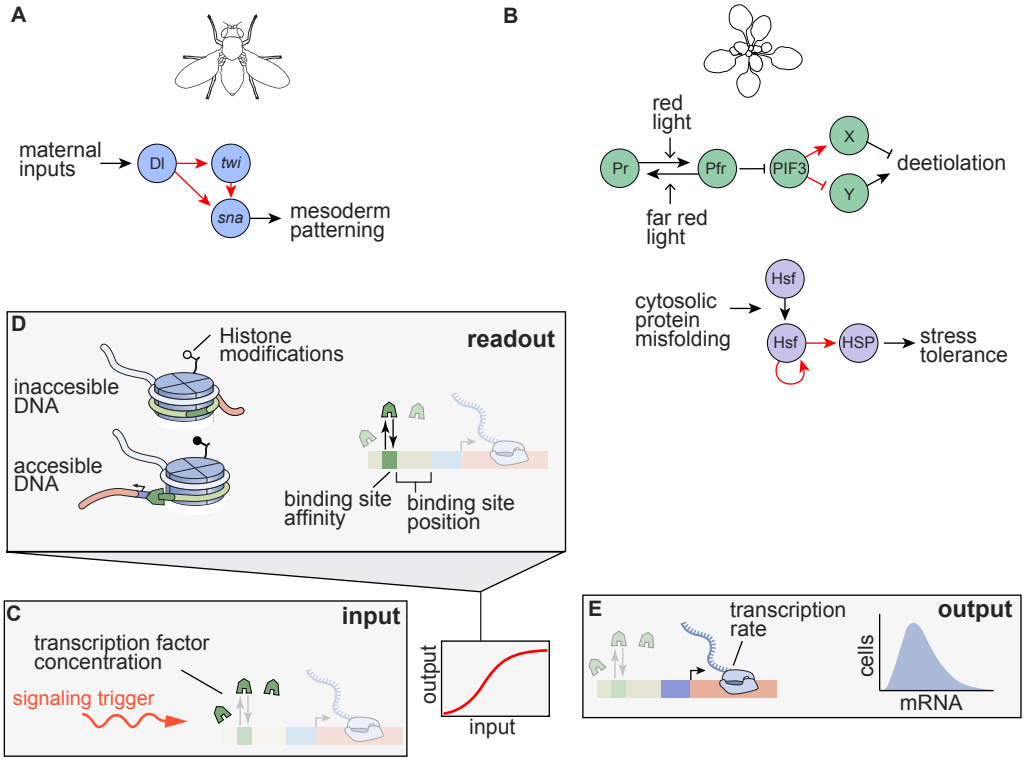

Figure 1.1: **Assigning numbers to gene regulation arrows in eukaryotes.(A)** The transcriptional network responsible for determining the mesoderm cell fate in fruit flies. The Dorsal transcription factor (Dl) activates the transcription of *twi* and *snail*, leading to the specification of the mesoderm. **(B)** Top: signaling network leading to deetiolation in response to red light in Arabidopsis. The inactive Phytochrome protein (Pr) is transformed into its active form (Pfr) by red light. Pfr represses PIF3 by targeting for degradation. PIF3 is an activator and repressor of transcription. Bottom: the cytosolic heat shock response. Protein misfolding due to heat or reactive oxygen species activates Hhat shock transcription factors (Hsf), which induce their own transcription and that of heat shock protein chaperones (HSP), which is necessary for stress tolerance. In (A) and (B), red arrows correspond to direct transcriptional regulation, black arrows correspond to positive or negative regulation that is non-transcriptional. **(C-E)** Molecular components involved in determining the quantitative aspects of transcriptional input output functions.

1.3 Making decisions in a rush: rapid transcriptional regulation in flies and plants

One of the most striking aspects of the development of the early fruit fly embryo is its speed. Nuclei in the embryo syncytium divide synchronously every $\approx$ 10 minutes. Dur-

ing an interphase of just a few minutes, nuclei regulate the transcription of patterning genes to give rise to the adult body plan in just a couple hours. The timescales relevant for transcriptional input-output functions in the fly embryo are shown in Fig. 1.2A. In plants, the speed with which transcription is regulated in response to changes in their environment is comparable (Fig. 1.2A). In response to excess light intensity, changes in mRNA abundance have been detected in a minute or less [188, 250, 58]. Similarly, transcript changes can be detected within 2 minutes of a heat shock treatment [282]. Other examples include the response to nitrogen sources [152, 6] and red light [159] which have been detected in 5 minutes. These timescales are likely to overestimate the speed of the response since our knowledge is limited by the first time point chosen in each study.

The speed of these processes requires that tools to study them are capable of reporting in similar timescales. Fluorescent proteins are typically used as reporters of gene expression dynamics, however, they mature in the order of tens of minutes and are indirect (Fig. 1.2A). In Chapters 3 and 4, I describe the implementation of the PP7 and MS2 technologies to measure gene transcription in real time in plants and fruit flies, respectively.

The speed with which plants make transcriptional decisions makes biological sense, considering how dynamic their environment is. For example, light intensity in nature can fluctuate widely in minutes or less (Fig. 1.2B). Similarly, nuclei in the fly embryo have a very narrow window during which they are exposed to gradients of morphogens such as the Dorsal transcription factor (Fig. 1.2C).

To illustrate the speed of signaling cascades impacting transcription in plants consider the response to heat shock in leaves and a dark-to-light transition in seedlings. When seedlings grown in the dark encounter light for the first time, the PIF3 transcription factor aggregates and rapidly degrade, with its concentration decreasing by a factor of 2 in just 10 minutes (Fig. 1.2D and F). When shifted from 25°C to 39°C, the HsfA1d transcription factor rushes into the nucleus and reaches half the steady state concentration in about 5 minutes (Fig. 1.2D and F). In response to these rapid changes in inputs, their target genes rapidly load RNA Polymerases, which transcribe in stochastic bursts (Fig. 1.2G).

As these examples emphasize, the regulation of transcription is an extremely dynamic process in the Drosophila embryo and in some plant signaling pathways. This imposes technical challenges to study them, which we discuss in the next sections.

1.4 Dynamic measurements for dynamic processes

It is now a well appreciated fact that bulk sampling can introduce averaging artifacts when interpreting phenomena that occur at the cellular level. Since many different frequency distributions can have the same mean but wildly different overall shapes, samples of different underlying cellular composition can superficially look similar to each other while actually being quite distinct. This notion is almost an afterthought in studies of development, perhaps because the behavior of specific cells in terms of division and

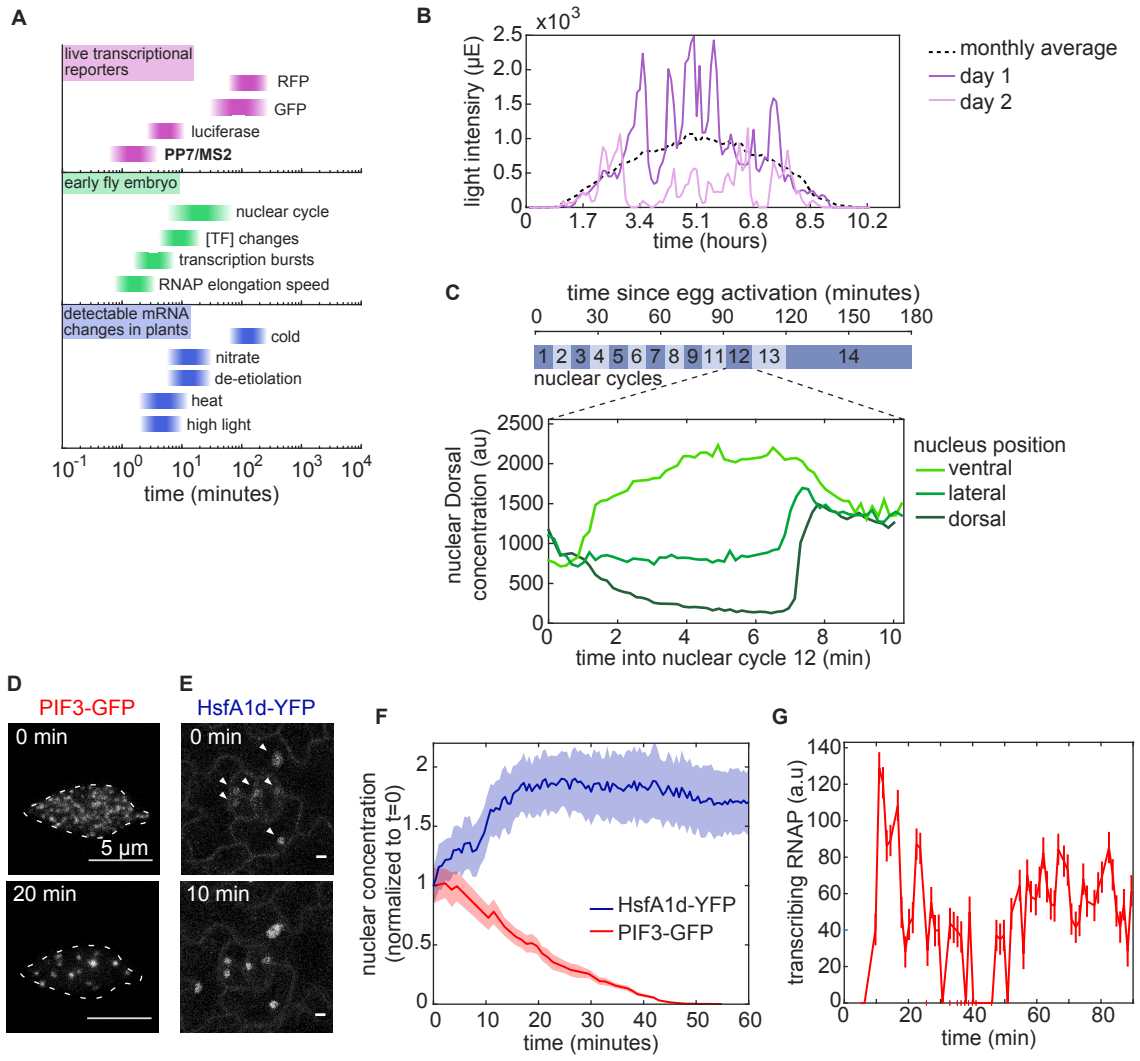

Figure 1.2: **Timescales of plant environmental responses and events in the early fruit fly embryo.(A)** Top: Detection timescales in a number of live methods. Middle: Relevant timescales during early development of the fruit fly embryo. Bottom: Timescales with which changes in mRNA abundance have been reported in Arabidopsis. **(B)** The dynamics of light intensity in nature during two individual days (colored lines). The dashed black line shows the monthly average. **(C)**. Nuclear concentration of the Dorsal transcription factor during an interphase (nuclear cycle) in Drosophila embryos. **(D)** Aggregation and degradation of the PIF transcription factor in response to a dark to light transition imaged by a GFP fusion. **(E)** Rapid nuclear translocation of a HsfA1d-YFP fusion in response to a 39°C heat shock treatment. ***Caption continues on next page***.

Figure 1.2: ***Continued from previous page*: Timescales of plant environmental responses and events in the early fruit fly embryo (F)** Quantification of the experiment in (D) and (E). The average nuclear concentration was normalized to the value prior to the treatment. **(G)** Transcription bursts in response to light exposure revealed by a PP7 reporter driven by the PIF-repressed RPT2 promoter in Arabidopsis.

gene expression lies at the core of patterning and morphogenesis. The importance of single cell resolution is, however, much less appreciated in studies of plant responses to stress and their environment. This may be due in part to the fact that, until recently, the extent of cellular heterogeneity in plants had not received much attention [222, 54].

A common experiment to determine gene expression output as a function of time involves using one sample per time point and then measuring the output, for example, as mRNA. Because each sample is measured once and destroyed in the process, these are end-point measurements (Fig. 1.3A). A series of such measurements is then used to deduce expression kinetics, even though this curve does not represent an individual sample, much less an individual cell. This type of experiment has been extremely valuable to place genes in relation to each other in pathways or networks but should be interpreted with caution, since it has little to say about the way these processes operate in a cellular context (Fig. 1.3B).

Live cell imaging and, more recently, single cell RNA sequencing have provided ample evidence for cellular individuality in plants [183, 261, 136]. It is worthwhile to recall here that experiments with single-cell resolution in which cells are destroyed are also end-point measurements. As a result, cells with very different temporal trajectories can look similar when examined only once (Fig. 1.3D). To overcome this limitation of sequencing-based experiments, a variety of approaches have recently been developed to store dynamic information in RNA that can then be retrieved in a single sequencing run. The idea here is to have a constitutive editing process going on in the cell that chemically modifies RNA at a constant rate. The extent of RNA modification then depends on the time an RNA molecule has been exposed to editing. The distribution of edits in a population of RNA molecules in a cell can then be used to reconstruct the temporal dynamics of transcription in that cell prior to the experiment [210, 220]. These high throughput approaches could replace live imaging in many applications.

To understand transcription dynamics, it is also important to consider the difference between activity and abundance. Activity of a gene refers to the production rate of its gene products -transcription or translation- which may or may not be correlated to the abundance of these products [41]. Measuring the accumulated products can introduce artifacts, particularly when they are long-lived compared to the timescale at which activity is being regulated. Conversely, it is possible to reconstruct transcriptional dynamics from protein abundance, provided that these dynamics change in timescales that are short compared to fluorescent protein maturation [31, 44].

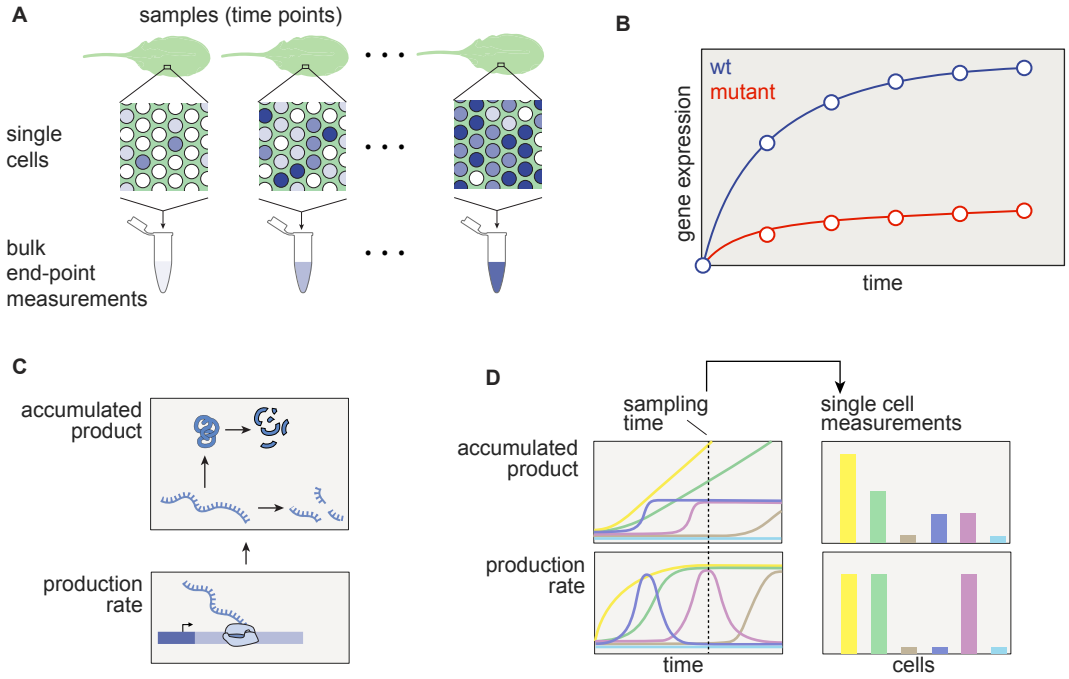

Figure 1.3: **Measurement artifacts associated with averaging in space and accumulating over time.(A)** Experiment where each time point corresponds to a different sample. Bulk sampling averages cell expression across space. **(B)** An experiment with single cell resolution where each sample corresponds to a time point. Top: At the indicated sampling time, cells with different expression trajectories look identical. Bottom: if what is measured is an accumulated product such as mRNA or protein, the history of expression has a strong influence on the measurements compared to directly measuring the production rate.

1.5 Quantifying inputs and outputs in live cells of whole organisms

We have discussed how building a predictive understanding of transcriptional regulation requires first quantifying the 'input-output' function (Fig. 1.1 between regulatory stimuli and mRNA production. We next saw how these input-output functions are embedded in time and can operate in timescales of just a few minutes in plants and fruit fly embryos. As we discussed in the previous section, this poses technical measurement challenges. In order to overcome these challenges, I established two experimental platforms -in plants and Drosophila- to study input-output functions in single living cells in their endogenous context. Chapter 4 corresponds to a manuscript where we describe the implementation of a live imaging technique to fluorescently label nascent RNAs in plants. Here, the gene of interest is tagged with 24 repeats of the PP7 sequence which,

when transcribed, forms an RNA stem loop. The loops are bound by a ubiquitously expressed PCP protein fused to GFP, resulting in the decoration of the RNAs with fluorophores (Fig. 1.4A) [155]. Multiple nascent RNAs can be resolved as diffraction-limited spots whose fluorescence is proportional to the number of RNAs. We used this system to measure the transcriptional dynamics of heat shock inducible reporters in Tobacco and Arabidopsis. As shown in Fig. 1.4B and C, induction with 39°C results in the appearance of transcription spots, but these spots are constitutively expressed when the reporter is driven by a constitutive promoter.

A similar technology was used in Drosophila embryos to label nascent RNAs where the MCP protein binds to the MS2 loop [26, 98, 174]. We used this method to visualize the expression of minimal synthetic enhancers carrying a single binding site for the Dorsal transcription factor. To simultaneously measure the concentration of the input transcription factor we generated a fly carrying a CRISPR knock-in fusion of Venus to Dorsal [107, 215] (Fig. 1.4D and E). We validated this minimal platform by demonstrating that the single Dorsal binding site is necessary and sufficient to drive transcription. Flies lacking Dorsal or carrying mutations in the reporter Dorsal binding site do not have detectable expression (Fig. 1.4F). This system allowed us to simultaneously quantify Dorsal nuclear concentration and its transcriptional interpretation in live embryos (Fig. 1.4G).

1.6 What are a cell's regulatory knobs?

Having direct access to RNAP activity allows us to answer how transcription is regulated to achieve biological functions. Using the heat shock response as a model, we asked how temporal patterns of mRNA accumulation arise from the regulation of transcription in individual cells. Bulk experiments in Arabidopsis have shown that upon exposure to $>$37°C, RNAP is rapidly recruited to the gene body of Heat Shock Proteins and heat shock transcription factors, resulting in the fast accumulation of their mRNAs [55, 170]. These dynamics represent an average that could originate from any number of possible single cell behaviors. In one extreme, the response at the single cell level could be digital, with the time under heat shock affecting the fraction of cells that switch to a high expression state. This type of regulation has long been hypothesized and has been observed in a variety of systems [148, 114, 247]. On the other hand, it is conceivable that the transcription rate operates as a continuum across all cells, increasing the rate of transcription with the time under heat stress. To distinguish between these hypotheses we decomposed the tissue-wide transcription rate into the rate of transcription across transcribing cells and the instantaneous fraction of active cells (Fig. 1.5A-C). We applied this analysis to four different Arabidopsis PP7 reporters driven by the promoters of two heat shock inducible genes (HsfA2 and HSP101), a light-inducible gene (RPT2) and a constitutive expressed gene (EF-Tu). As shown in Fig. 1.5D-F, the tissue-wide rate of transcription stems primarily from the number of cells that transition to an active state rather than from the specific activity of these cells. This behavior is reminiscent of the response to

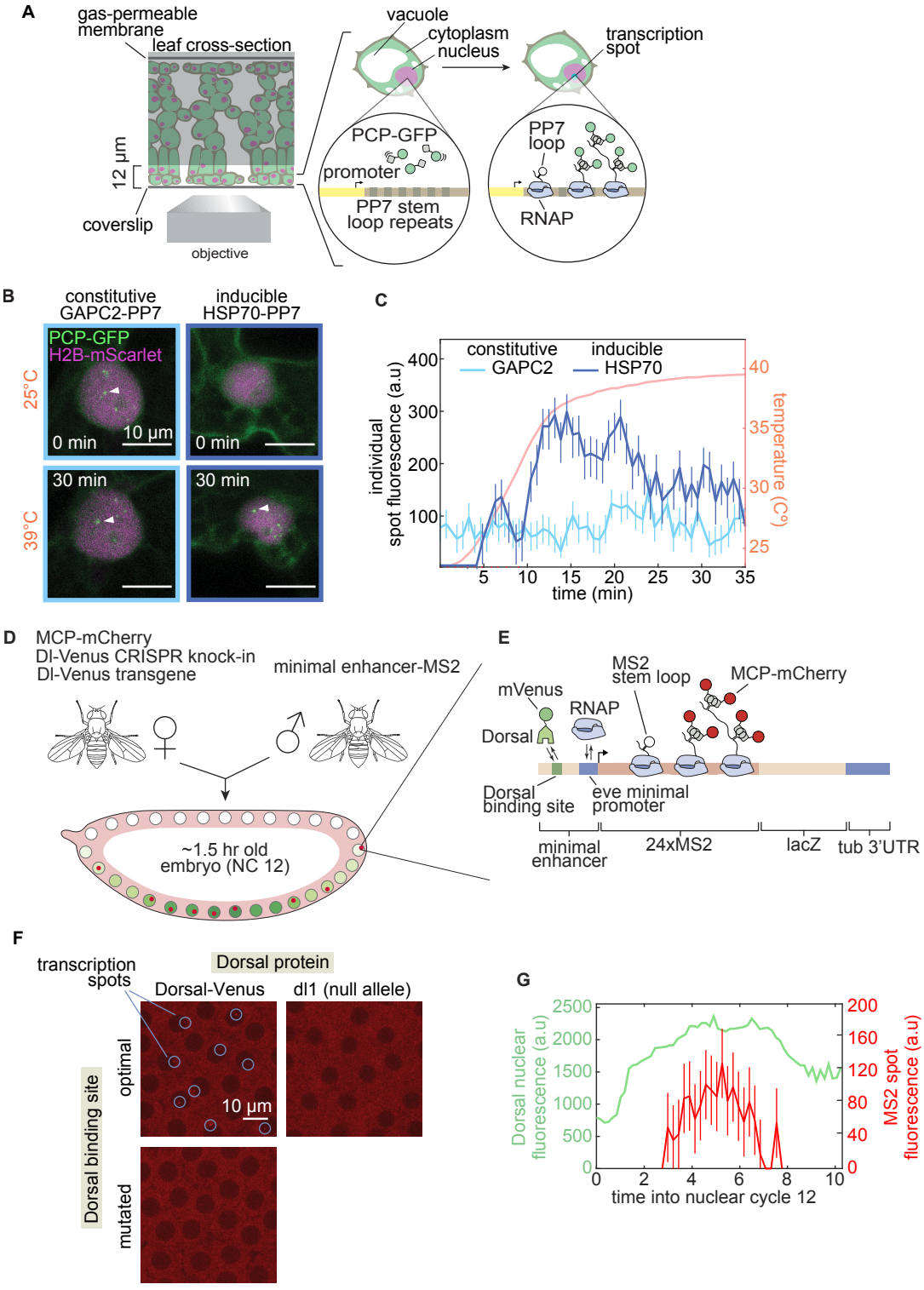

Figure 1.4: **Live imaging methods to quantify transcription inputs and outputs in plants and fruit fly embryos.** ***See caption on next page***

Figure 1.4: **From figure in previous page: Live imaging methods to quantify transcription inputs and outputs in plants and fruit fly embryos**. **(A)** Implementation of the PCP-PP7 technology in plants. Cells express a ubiquitous PCP-GFP fusion and a Histone-mScarlet nuclear marker. Transcription of PP7-tagged genes creates nascent RNAs bearing PP7 stem loops, which are bound by PCP-GFP. **(B)** Snapshots showing the system in (A) implemented in Tobacco epidermis cells. Reporters driven by a constitutive (GAPC2) and a heat-inducible (HSP70) promoters are shown. The GAPC2-PP7 reporter is already active at 25C but HSP70-PP7 is only detected at 39 °C. **(C)** Quantification of one spot per nucleus in (B). The red line shows the sample temperature plotted against the right y-axis. **(D)** Schematic of the experimental setup to simultaneously image the Dorsal transcription factor gradient with a Venus fusion and its transcriptional readout by minimal synthetic enhancers with MS2-MCP. **(E)** Venus-tagged Dorsal binds to a single site in a synthetic enhancer, driving transcription of an MS2-tagged reporter transgene. MCP-mCherry binds to the MS2 stem loops, fluorescently labeling nascent RNAs. **(F)** Validation of the setup in (D) and (E). Both the Dorsal protein and the Dorsal binding site are necessary for transcription. **(G)** Raw single-nucleus fluorescence data obtained from the setup in (D) and (E). In green, plotted against the left y-axis, is the Dorsal nuclear concentration. In red and plotted against the right y-axis is the nucleus MS2 signal. Error bars in (C) and (G) correspond to the uncertainty in spot fluorescence calculation (see Chapter 4 for details)

glucocorticoids and UV radiation in mammalian cells [148, 114].

These regulatory dynamics are dictated by the concentration of transcription factors. The transcriptional aspects of the Arabidopsis heat shock response (a special case of the cytosolic misfolded protein response) [249] as well as deetiolation, depend on changes in the nuclear abundance of Heat shock transcription factors (Hsf) and PIF transcription factors, respectively [197, 201]. The same is true for dorso-ventral patterning in Drosophila [215] and its regulation by the Dorsal protein.

To test more specific hypotheses of how the abundance of these proteins determines transcriptional outcomes, one would like to titrate their concentration at the same time that the DNA regulatory sequence is systematically perturbed. Titrating transcription factors in plants is possible using inducible transgenes [261], dCas9-mediated overeexpression and silencing [189] and promoter allelic series (reviewed by [244]). In comparison, rigorously dissecting regulatory sequences is much harder, in part, due to lack of targeted transgene insertion technologies in plants. In a promising approach, dense, targeted mutagenesis of endogenous promoter sequences was used at a modest scale to determine the quantitative contribution of regulatory sequences [221]. On the other hand, established genetic tools make the fruit fly a more suitable model system to study *cis* regulation. Moreover, transcription factor proteins, such as Dorsal, naturally occur as gradients in the Drosophila embryo, providing a natural "titration curve". This doesn't mean, however, that transcriptional regulation is anywhere near simple in this system,

which is why we developed a minimal synthetic enhancer platform (Fig. 1.4E,F, see also Chapter 3).

Fig. 1.5G shows MS2 fluorescence traces driven by a minimal Dorsal-responsive synthetic enhancer and three metrics of transcriptional activity that could potentially be regulated by Dorsal. The maximum trace fluorescence corresponds to the maximum RNA Polymerase density on the gene and is thus related to the rate at which Polymerases are loaded. The transcriptional onset time is defined as the time at which the spot was detected and the integral of the trace fluoresence over time provides a measure for how much RNA was produced by the gene [98]. A fourth metric corresponds to the fraction of responsive nuclei, defined as those nuclei in which activity can be detected at any point during the experiment. Because we used different terminology in the manuscripts in Chapters 3 and 4, in this chapter I will refer to 'responsive' and 'active' nuclei interchangeably. Similarly, 'refractory' and 'inactive' nuclei correspond to the same definition (see also Fig. 1.6A). Comparing each of these metrics across nuclei exposed to different levels of Dorsal in minimal enhancers carrying binding sites of varying affinity reveals that all metrics are relatively insensitive to Dorsal concentration with the exception of the fraction of active nuclei. (Fig. 1.5J-X).

We now turn our attention to mathematical models that can capture these behaviors.

1.7 Probing the nature of switch-like transcriptional induction

As we saw in the previous section, in a Dorsal-regulated minimal enhancer, the fraction of nuclei that ever turn on the reporter activity at any point in the experiment is regulated by Dorsal (Fig. 1.5H). We refer to these as responsive nuclei, in contrast to refractory nuclei. In Chapter 4, I show that for a given reporter gene, plant cells can too be classified into 'responsive' and 'refractory', depending on whether we detected transcription at any point during the experiment. The fraction of refractory cells is reproducible among replicates but variable across lines, indicating that it is a regulatory feature of the reporter (Fig. 1.6C). To explore how this fraction is regulated, one would need to modulate an experimental parameter, such as the intensity of the treatment, the dosage of transcriptional activators acting on the promoter, or the promoter sequence itself. In Chapter 3, we discovered that this fraction is determined by Dorsal concentration and the affinity of its binding site, demonstrating that the decision of whether to transcribe can be dictated by the interaction of a single transcription factor molecule to one DNA motif (Fig. 1.6D).

Experiments with single molecule resolution in a variety of systems [247, 148], including plants, [71, 130] have shown that for a given gene at a given time, there is a fraction of cells in which the gene is not being transcribed. Because these experiments require cell

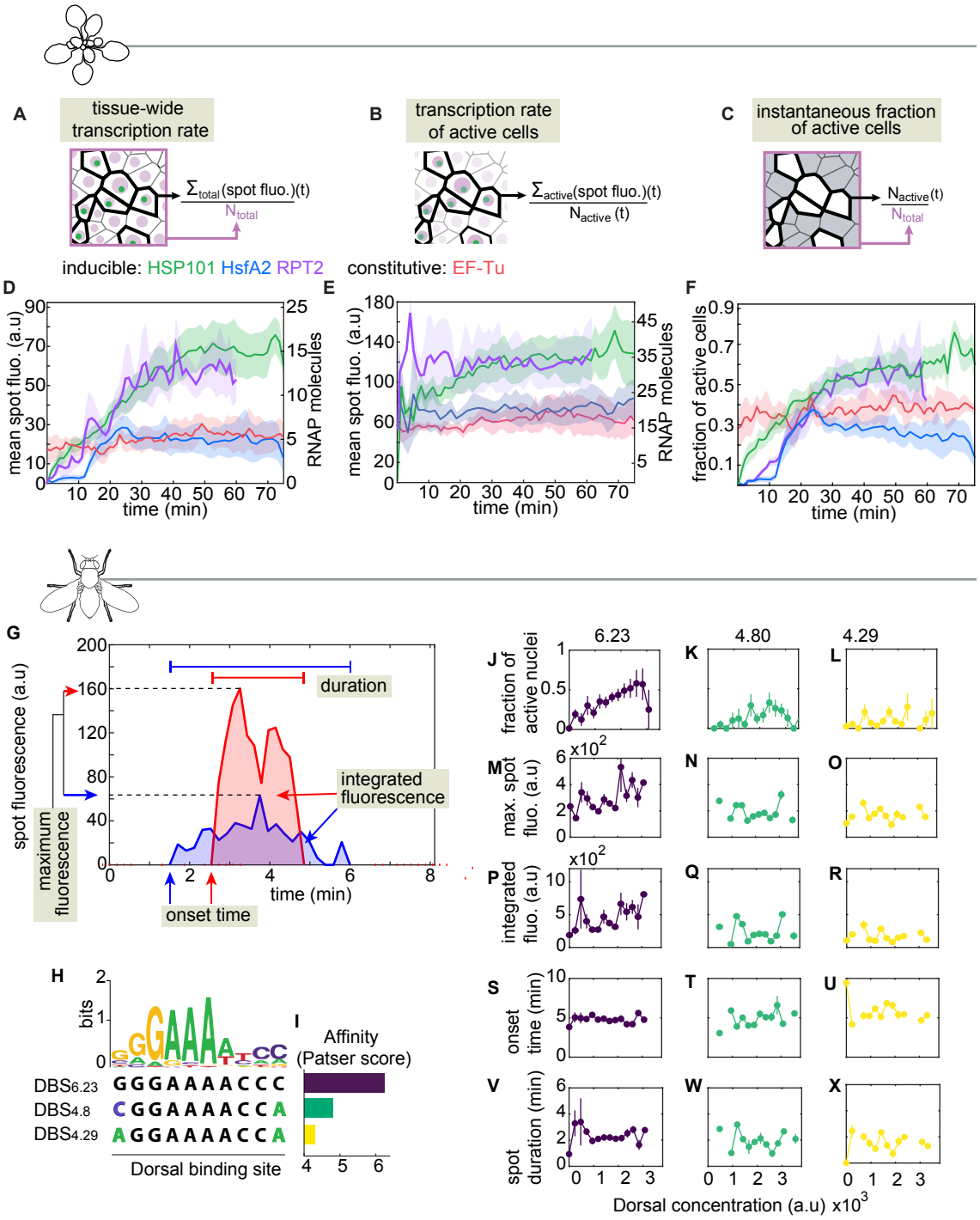

Figure 1.5: **Single cell regulatory strategies leading to patterns of mRNA accumulation in time and space**. ***See caption on next page***.

Figure 1.5: ***Continued from previous page*: Single cell regulatory strategies leading to patterns of mRNA accumulation in time and space**. **(A-C)** Schematic showing how tissue transcription metrics are calculated in a field of view of plant cells. **(D-E)** Mean of the transcription metrics shown in (A)-(C) in four different PP7 reporters in Arabidopsis. Time 0 corresponds to the time the first spot was detected. HSP101 and HsfA2 were treated with a 39°C heat shock. RPT2 was induced by transferring dark-grown seedlings to light. EF-Tu is a constitutive promoter. **(G)** Two example MS2 fluorescence traces driven by a minimal Dorsal-dependent synthetic enhancer in the Drosophila embryo. Highlighted are three different metrics used to quantify transcriptional activity: the maximum trace fluorescence, the turn on time and the integral of the fluorescence trace. **(H)** Dorsal consensus binding site ($DBS_{6.23}$) and two different variants containing point mutations ($DBS_{4.8}$ and $DBS_{4.29}$). **(I)** Bioinformatic prediction of the Dorsal DNA binding affinity to the sequences on (H) measured as Patser score. **(J-X)** Shown is the behavior of three minimal synthetic enhancers containing a single binding site for Dorsal with the sequences shown in (H). Each activity metric is shown against the Dorsal nuclear concentration. **(J,K,L)** Mean fraction of nuclei that transcribe at any point during the experiment (see Fig.1.6). **(M,N,O)** Mean in the maximum spot fluorescence over time as a function of Dorsal concentration. **(P,Q,R)** Mean in the total number of mRNA molecules produced per active nucleus measured as the integral of MS2 fluorescence traces over time. **(S,T,U)** Mean transcriptional onset time. **(V,W,X)** Average spot duration. In (D)-(F) the shaded area corresponds to the standard error of the mean across three or more plants. in (J)-(X) the error bars correspond to the standard error of the mean across three or more embryos.

fixation, they cannot distinguish between a stable and a transient transcriptionally inactive state. On the other hand, live imaging techniques using gene reporters in general do not have single molecule detection, but do provide access to single cell expression time trajectories. These experiments have consistently shown, across all kingdoms of life, that transcriptional activation operates both like a digital switch and a continuum [98, 53, 154, 115].

The notion that gene expression can operate in a binary fashion should hardly be a controversial one. Working in the 1950's, Novick and Weiner showed that, under non-saturating levels of the IPTG inducer, the level of β-galactosidase in individual *E.coli* cells exhibits an "all or nothing" behavior [195]. The expression program of phage λ is basically a switch between two stable states with no 'in between' (Reviewed in [103] and covered in Mark Ptashne's fantastic book 'A genetic Switch' [209]). This is not limited to bacterial gene expression. 30 years ago Ko and coworkers [148] found that the apparent graded response of a gene to glucocorticoid concentration stems from the fraction of alleles switching to an active state. Extensive theoretical and experimental evidence has shown that positive feedback loops, mutual repression, cooperative binding, and

allostery can be all exploited to achieve sharp on/off transcriptional responses [99, 17] (Fig. 1.6B) .

Still, these are not the only mechanisms by which transcription can operate in a digital manner at the cellular level. Over the past 15 years, evidence has mounted showing that even in cells fully committed to the "active" state of a gene or signaling pathway, the activity of RNA Polymerase is discontinuous. The simplest explanation for this punctuated transcription is that, due to the small number of molecules involved, transcription is subject to the randomness of molecular collisions and biochemical transactions in *cis* (Fig. 1.6B). In this view, a gene happens to make an RNA whenever the right molecules come together at the right place and the right time. Such a stochastic process would result in the waiting time between consecutive RNAPs being exponentially distributed, and the number of mRNAs per cell following a Poisson distribution. While there have been reports of genes falling into this category [284], the vast majority of studied genes exhibit mRNA distributions inconsistent with the so-called Poissonian promoter. This means that new RNAPs are loaded in 'bursts', separated by extended periods of inactivity. Recent, excellent reviews cover the multiple mechanisms that have been proposed to explain transcriptional bursting [260, 154, 218]. Although varying in their molecular assumptions, these models have in common that they invoke one or more rate-limiting steps that the promoter must go through before it enters a permissive state where RNA Polymerases can be loaded.

One key issue when dealing with the nature of digital transcriptional activity is that of detection artifacts. Arguably, transcriptionally inactive cells could be nothing but false negatives resulting from a technical detection threshold higher than a single RNA Polymerase. More important from a theoretical standpoint though is whether there really are distinct "states" at all instead of a continuum of activity (Fig. 1.6E). In Chapter 3, I describe an experimental method devised to tackle this problem by quantifying RNAP activity independent of reporter detection. In this scheme, the reporter locus is labeled with eGFP using the ParB-*parS* system [49] and nascent RNA is labeled with mCherry using the MCP-MS2 system [98]. Instead of relying on the detection of spots in the mCherry channel to measure transcription, the 3D localization of the locus is determined using the eGFP channel, while the mCherry fluorescence is calculated at the same position(Fig. 1.6F). If a gene could exist in two distinct states, then the distribution mCherry fluorescence would look bimodal, with the lowest mode being composed of those loci that were not detected as spots in the MS2 channel. Experiments with a reporter construct driven exclusively by the Dorsal activator in Drosophila embryos demonstrated this bimodality. Moreover, in flies lacking the Dorsal activator, the mCherry fluorescence at eGFP-labelled loci is the same of undetected spots from flies carrying wild type Dorsal. Thus, undetected MS2 spots correspond to a weaker, Dorsal-independent state (Fig. 1.6G).

1.8 Stochastic models to describe refractory cells

Having shown that the refractory fraction is not an imaging artifact, we turned our attention to models to explain the presence of two populations, i.e, nuclei that transcribe at some point and nuclei that never do. The most parsimonious explanation for this observation is that before becoming transcriptionally active, genes must stochastically transition through a number of inactive states (1.7A). Here, we propose that Dorsal acts by accelerating these transitions in a way that depends on its binding affinity and concentration (1.7 B). Since transitioning from an inactive to an active state is a random process, some alleles transition very early while others take much longer, sometimes longer than the duration of the experiment itself. If the waiting time of the gene is longer than the time limit available for it to start transcribing, then that nucleus is then classified as 'refractory' even when there was nothing essentially different about it. In plants, this limit is set by the duration of the experiment, while in flies it corresponds to a discrete permissible time window in the 12th interphase (1.7C). We termed this the "kinetic barrier model".

To determine if this model could recapitulate our observations in terms of the fraction of active nuclei and the mean transcriptional onset time, we performed a global fit to data from 7 different minimal synthetic enhancers. Here, we forced the model to use the exact same parameters to describe all enhancers, except for the Dorsal binding K_D. This exercise revealed that this simple model was sufficient to explain our observations (1.7D, E.)

1.9 Two spot experiments illuminate the causes and consequences of cellular heterogeneity in expression

The question of what dictates whether a cell is responsive or refractory points to the more general issue of what is the nature of cellular heterogeneity in transcription. Multiple variables can differ across cells that may lead to transcriptional heterogeneity, such as the stage of the cell cycle [284, 84], cell size [19, 84], the concentration of transcriptional activators [239], or the abundance of RNA Polymerase [279]. However, as mentioned before, even alleles exposed to the exact same so-called extrinsic variables can exhibit differences in transcription due to the intrinsic stochasticity associated with any process where a small number of molecules is involved. A straightforward strategy to distinguish between extrinsic and intrinsic sources of noise is measuring the activity of more than one allele per cell [76].

As shown in Fig. 1.8A, the law of total variance applied to the problem of gene expression states that the total spread in expression across a population of cells is the sum of two types of variance. The explained variance is related to cellular parameters that influence the expression of the gene of interest that differ across cells. The unexplained

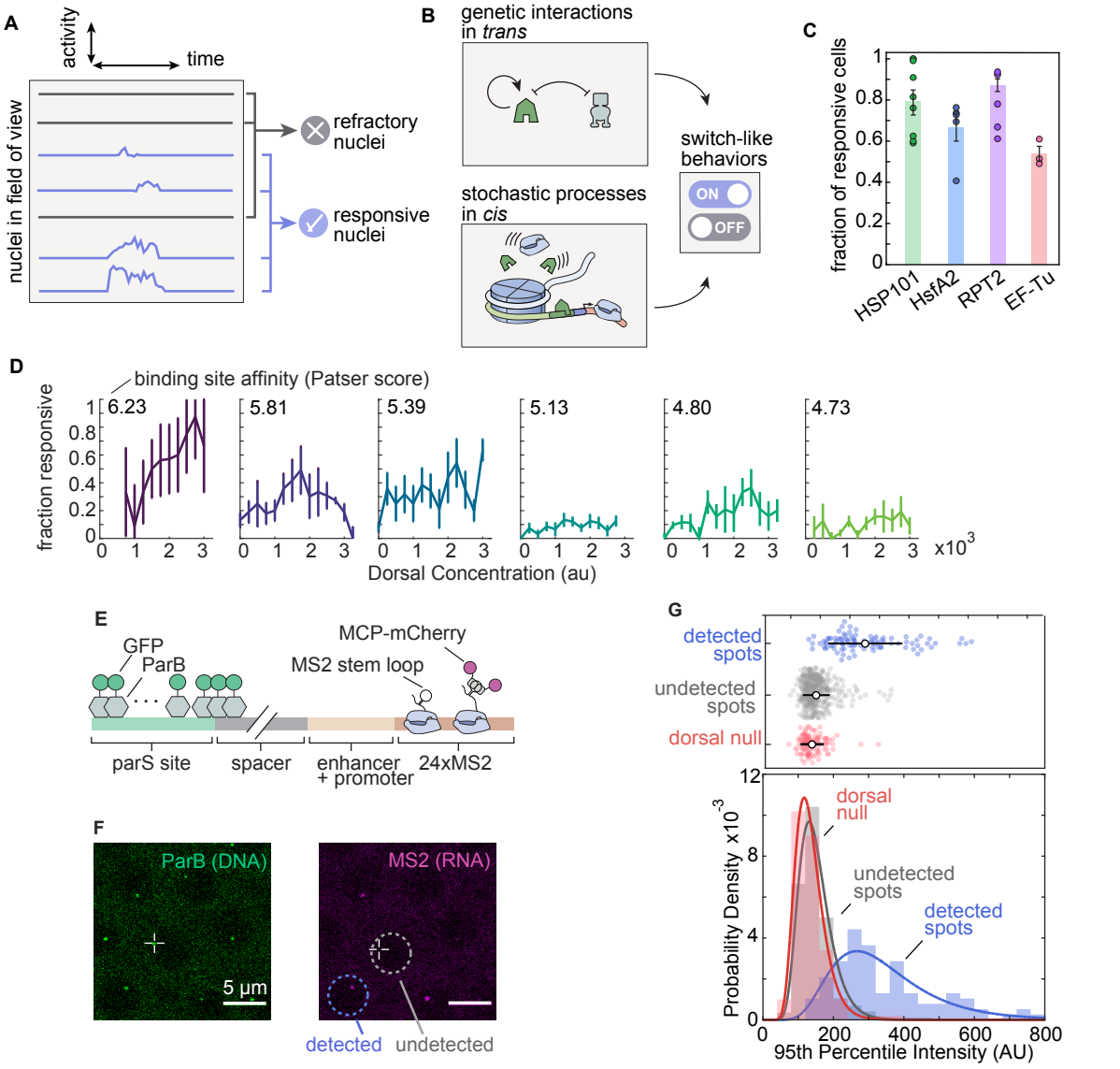

Figure 1.6: **Refractory cells in Arabidopsis and Fruit flies.(A)** Schematic showing how refractory and responsive nuclei or cells are defined. Nuclei in which a transcription spot was detected in at least one frame it is classified as responsive and classified as refractory otherwise.**(B)** Mean and standard error in the fraction of refractory nuclei in Arabidopsis based on PP7 reporters driven by the promoters of the indicated genes. EF-Tu is a constitutively active gene, HSP101 and HsfA2 were induced by heat shock (see Chapter 4) and RPT2 was induced by transferring seedlings from dark to light (see Chapter 5). **(C)** Mean and standard error in the fraction of refractory nuclei for an MS2 reporter driven by minimal enhancers activated by Dorsal. The fraction is shown for enhancers with varying affinity for Dorsal across a range of Dorsal concentrations. See Chapter 3 for more details. ***Caption continues on next page***

Figure 1.6: ***Continued from previous page*: Refractory cells in Arabidopsis and Fruit flies.** **(D)** Top: Cartoon of observed spot fluorescence distributions in refractory (grey) and inducible (blue) nuclei. Bottom: possible underlying distributions in the absence of spot detection artifacts. **(E)** Snapshot of a Drosophila embryo where the locus of a minimal Dorsal-responsive MS2 reporter was labeled with ParB-eGFP. The location of ParB-eGFP is used to guide the fluorescence calculation in the MS2 channel in nuclei without a detected MS2 spot. **(F)** Top: Fluorescence values of the the MS2 channel in loci detected (blue) and undetected (gray) by MS2. Bars correspond to the mean and standard deviation. Bottom: histogram of fluorescence values showing bimodality.

variance refers to the remaining variance that is observed even in otherwise identical cells. The challenge with this formulation is that, for a given gene, it is not possible to know *a-priori* what the set of cellular parameters affecting its expression is. Even were the parameters known, it is not guaranteed that they can all be measured with single-cell resolution. To circumvent these issues, Elowitz and coworkers [76] devised a method to calculate the explained and unexplained components of the variance -referred to in the literature as extrinsic and intrinsic noise- by measuring the expression of two alleles per cell. Differences between pairs of alleles in different cells are due to extrinsic noise while intrinsic noise captures differences between alleles in the same cell. One benefit of this approach is that it can quantify the contribution of each source of noise while being agnostic to its molecular nature. As a result, this approach can help distinguish between competing hypotheses of how cells make gene expression decisions without necessarily having to commit to a specific mechanism [176].

Using homozygous plants carrying two alleles per nucleus of heat shock inducible reporters, we measured the contribution of intrinsic and extrinsic noise to cellular heterogeneity in transcription (Fig. 1.8B). Comparing the expression of alleles belonging to the same nucleus shows that in a large proportion of nuclei only one allele transcribes. When both do, their produced mRNA tends to be loosely correlated to each other (Fig. 1.8C). This is indicative of the presence of local processes acting in *cis* to individual alleles and result in intrinsic noise being comparable or larger than extrinsic noise (Fig. 1.8D).

We have seen that gene transcription can be regulated by different strategies, such as tuning the rate of loading of new RNAPs or the probability of switching promoters to an active state. The decomposition of total noise into intrinsic and extrinsic components lumps together all of these different layers, but this need not be the case. For example, Falo-San Juan and collaborators [79] decomposed the noise in the timing of transcription in Drosophila embryos to argue that low intrinsic noise reflects the presence of a nucleus-wide 'priming' mechanism. Binary or digital transcriptional regulation can play a major role, and so it is interesting to examine the sources of cellular heterogeneity specifically at this level as well. If switching on is dictated exclusively by the presence of *trans* factors, then we should expect all alleles in the same nucleus to share their

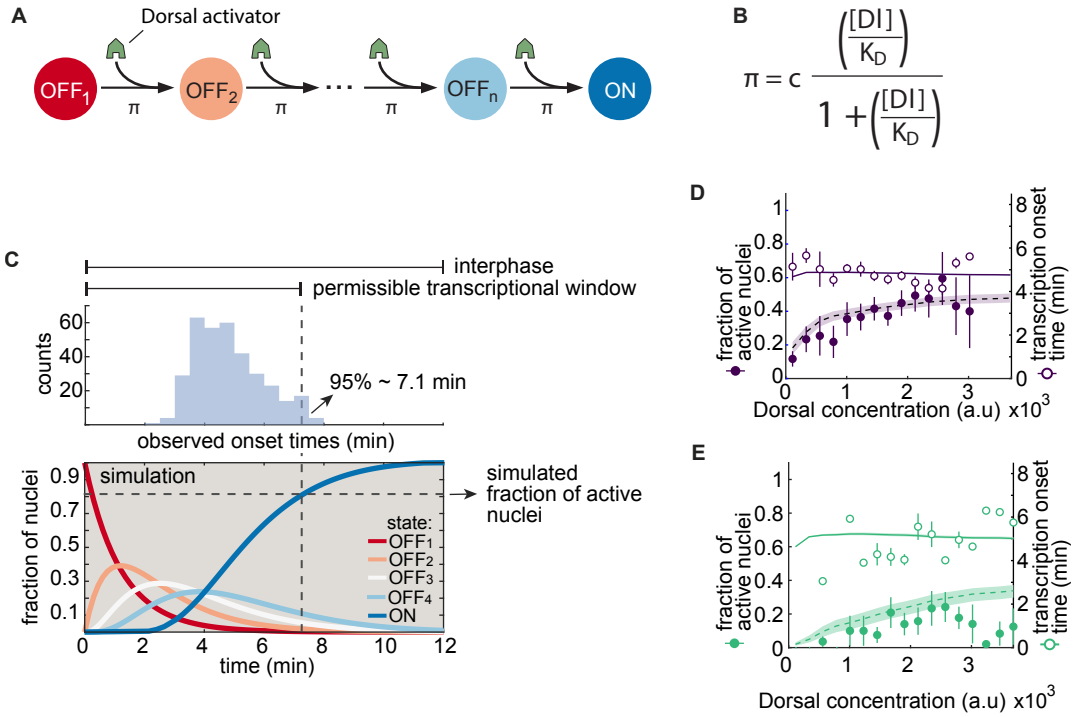

Figure 1.7: **A kinetic barrier model can recapitulate the data in terms of mean transcriptional onset times and fraction of active nuclei (A)** Schematic showing the kinetic barrier model. The promoter goes through a series of n inactive 'OFF' states before reaching an 'ON' state in which it can transcribe. Dorsal can accelerate the transition between states by a factor of π. **(B)** To make π depend on Dorsal concentration and Dorsal binding affinity (K_D) we made it proportional to a simple binding equation scaled by an arbitrary linear factor c. **(C) (C)** Top: distribution of mean spot detection times per embryo across all bins of Dorsal fluorescence and enhancers in Chapter 3. The horizontal dashed line marks the time at which 95% of spots have turned on which defines the permissible transcriptional window. Bottom: simulated fractions of nuclei in each of the states in (A) for a model with 4 OFF states. The fraction of nuclei in the ON state at the end of the permissible transcriptional window (at $\approx$ 7.1 min) corresponds to the simulated fraction of active nuclei. **(D)** Model fits for the fraction of active nuclei (plotted against the left y-axis) and mean transcriptional onset times (plotted in red against the right y-axis) for an enhnancer containing a single strong Dorsal binding site. Dashed lines correspond to the median of the posterior distribution. The shaded areas indicate the 25%-75% prediction interval. Open blue circles correspond to the experimentally observed mean onset time. Closed circles correspond to experimentally observed mean fraction of active nuclei. **(E)** Same as (D) for an enhancer containing a single weak Dorsal binding site.

state, either on or off. Conversely, if the binary promoter state depends on a local process, we should expect alleles in the same nucleus to behave independently of each other. The binomial distribution can be used to determine if two stochastic processes with a binary outcome are independent. If there are n cells with 2 alleles per cell where k total alleles are active in the population, then, assuming independence the expected probability of cells with 2 active alleles P_2 is

$$P_2 = \left(\frac{k}{2n}\right)^2. \tag{1.1}$$

Alternatively, if both alleles are completely coupled, then cells have either zero or two actively transcribing alleles,

$$P_2 = \frac{k}{2n}. \tag{1.2}$$

This analysis was used to determine if the transcriptional state of Notch-responsive genes in *C. elegans* is set by intrinsic or extrinsic factors. It was found that -for a given level of Notch- allele pairs turn on independently of each other, revealing that this is a local stochastic process [156]. Similarly, in hepatocytes, alleles turn on and off stochastically independent of each other [12]. Intriguingly, Wang and coworkers determined that in *E. coli*, the degree with which copies of the *Lac* promoter turn on independently of one another depends on the culture growth conditions [266]. In plants homozygous for heat shock-inducible PP7 reporters, there is a large fraction of nuclei in which only one allele can be detected (Fig. 1.9A), revealing that alleles are far from coupled. Comparing the fraction of cells with two alleles to the expectation based on independence in one of these lines shows that they are close to independence (Fig. 1.9B).

So far, we have dealt with cases in eukaryotic and bacterial cells with at most two alleles of each gene per cell. Through endoreduplication, plant cells can duplicate their genome without mitosis. Endoreduplicated Arabidopsis leaf cells can have up to 16 copies of each chromosome [182]. Fig. 1.9C shows a polyploid leaf cell with 5 active alleles of a PP7 reporter. Additionally, whole genome duplications are the rule in plant lineages, further increasing the number of (almost) identical alleles per cell in the form of retained paralogs.

As a result, a gene with four closely related paralogs in a cell with a genome content of 16C has an effective number of 64 alleles. This is arguably a qualitatively different scenario than the two allele problem studied in the past and leads to the intriguing possibility that carrying a large number of alleles per nucleus can lead to a dramatic increase in the reproducibility of gene expression across cells. To explore this hypothesis in more detail, I simulated alleles turning on as a binomial distribution where each allele has a probability of 0.5 of turning on in a completely independent fashion (Fig. 1.9D). There are two interesting consequences of increasing the number of alleles. First, because the mean of a binomial distribution scales linearly, while the standard deviation scales as the square root, the noise in the number of active alleles per cell decreases

with the number of alleles (Fig. 1.9A). Second, and more intuitively, increasing the number of alleles exponentially decreases the probability that a cell fails to express at least one of them (Fig. 1.9F). Thus, the presence of large number of alleles per nucleus in plants might play a role analogous to shadow enhancers in animals (Fig. 1.9G), where functional redundancy fosters reproducible cellular outcomes [205, 43, 268].

1.10 Dynamic perturbations for dynamic models

Combined with theoretical models, transcriptional dynamics can indirectly inform about the regulatory processes behind gene regulation. Testing these models is but a logical step but there is a lack of tools to precisely and rapidly perturb proteins such as transcription factors. A system where the transcription factor input can be removed while simultaneously measuring transcription of its targets could be a powerful tool to study how the timing of gene expression is regulated. This inspired the experiments shown in Chapter 5, where we used the live imaging techniques introduced in Chapter 4 to measure the dynamics of transcriptional induction in response to a dark-to-light shift in Arabidopsis seedlings. Deetiolation is the process in which plants germinated in the dark reverse the dark growth phenotype by a series of changes that lead to photoautotrophic development. Deetiolation includes changes in the mRNA abundance of $\approx$ 10% of genes in Arabidopsis [128], changes that have been detected as soon as 5 minutes after light exposure [159]. In dark-germinated seedlings, thousands of genes are bound by PIF transcription factors, which can act as repressors or activators depending on the context. Red light activates the Phytochrome protein, which then physically interacts with PIFs, leading to their degradation (Fig. 1.10A). The gradual degradation of PIF proteins in response to light provides an ideal system to study the ordering of regulatory events in a signaling pathway leading to mRNA accumulation (Fig. 1.10B).

The regulation of transcription by light in plants is not just a convenient model to study gene expression dynamics,it has also been a source of tools to control cell biology with light in heterologous systems. These tools, collectively known as optogenetics, can be used to control protein localization and function with exquisite spatiotemporal resolution. As such, optogenetics can transform the way we test models of gene regulation that include dynamics such as the ones discussed in Chapter 3. In Chapter 2, I describe the development of optogenetic tools in the early Drosophila embryo to control the nuclear concentration of the transcription factors Knirps (Fig. 1.10C,D) and Dorsal (Fig. 1.10E,F) while simultaneously imaging their downstream activity using MS2.

Interestingly, genes repressed by Knirps and PIF3 show increased H3K9ac and mRNA levels when these transcription factors are naturally or genetically removed ([248] Gonzalez-Grandio, unpublished results). On the other hand, genes activated by PIF3 lose H3K9ac and upon PIF3 light-induced degradation (Gonzalez-Grandio, unpublished results). By comparing the timescales of PIF3 removal, transcriptional initiation and loss of H3K9ac marks it should be possible to build a rank ordering model of the molecular steps lead-

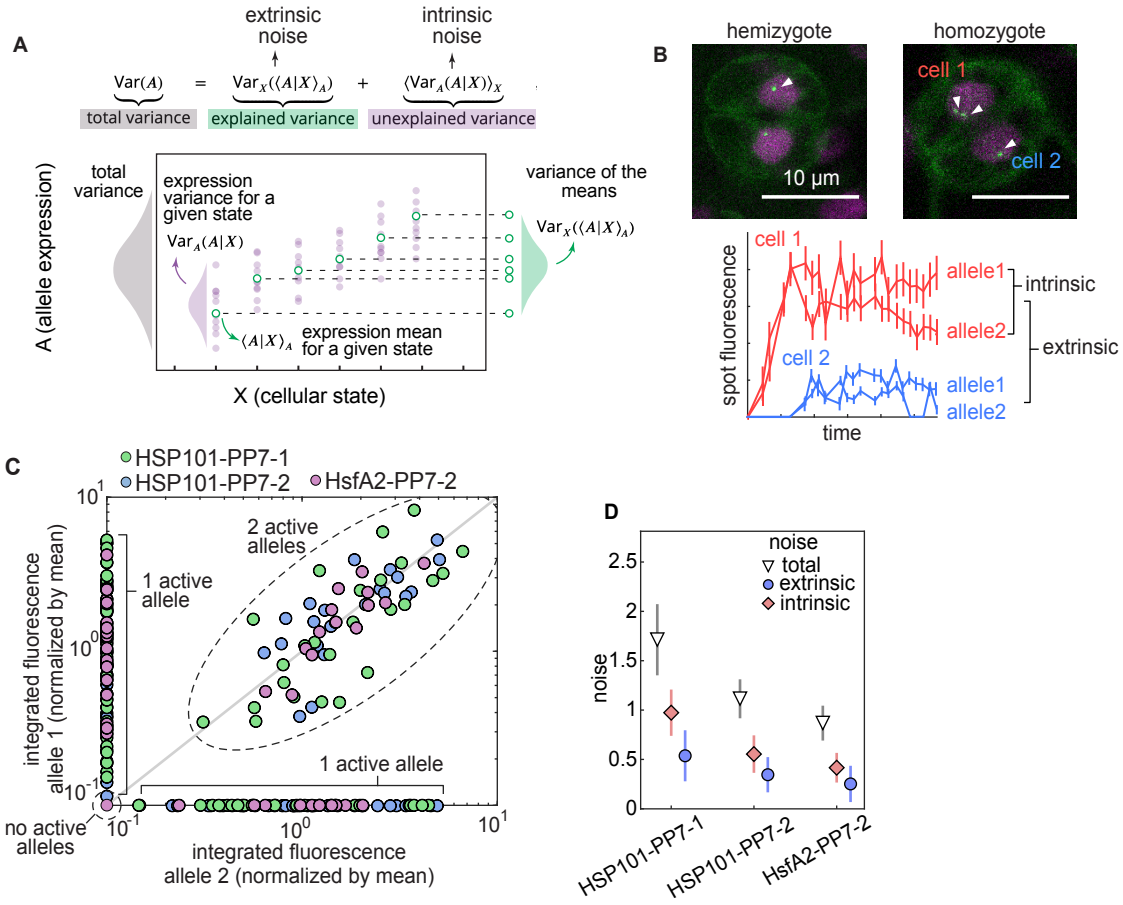

Figure 1.8: **Cellular differences in expression have an important locus-specific stochastic component.(A)** Schematic of the law of total variance applied to gene expression. The total variance in the expression of a gene (gray) is the sum of the variance across cells in an identical state (purple, intrinsic noise) and the variance across cells in different states (green, extrinsic noise).**(B)** Top: Arabidopsis guard cells carrying one or two alleles of a HSP101-PP7 reporter depending on the transgene zygocity. Bottom: spot fluorescence traces from the cells shown on top. Differences among alleles in the same cell are due to intrinsic noise. Differences between cells are captured by extrinsic noise. **(C)** Scatter plot showing the total mRNA produced by pairs of alleles belonging to the same nucleus in three different homozygous lines. The absence of intrinsic noise would result in perfectly coordinated alleles lying on top of the diagonal. **(D)** Quantification of the gene expression noise components in (C). Intrinsic noise is at least as important as extrinsic noise.

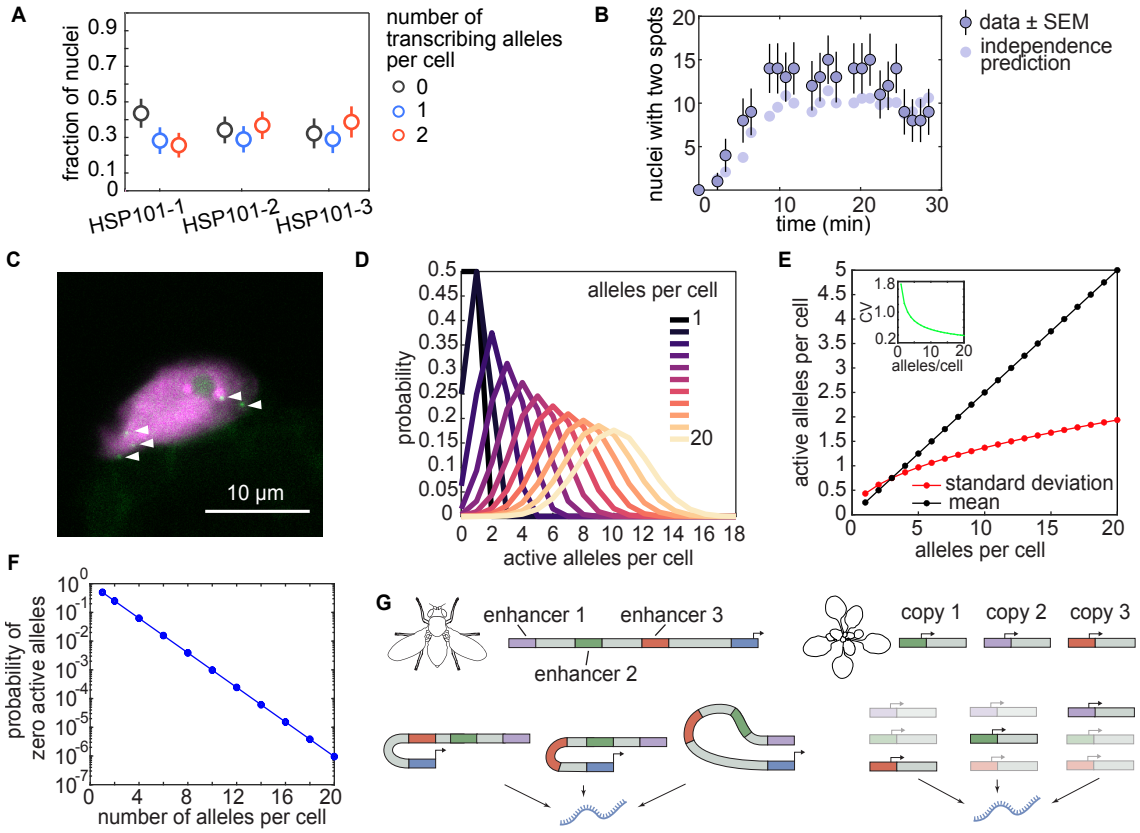

Figure 1.9: **Alleles transcribing independent of each other and potential consequences.(A)** Mean fraction of cells with zero, one or two spots after the total number of spots in the field of view had reached steady-state in heat treated HSP101-PP7 and HsfA2-PP7. **(B)** Mean number of nuclei with two spots in a heat shock treated HSP101-PP7 line compared to the expected number if alleles belonging to the same nucleus turned on independently of each other. Error bars in (A) and (B) correspond to the bootstrapped standard error of the mean. **(C)** Nucleus from a heat shock treated leaf of a single insertion homozygous HsfA2-PP7 plant. White arrowheads indicate transcription spots originating from the same endoreduplicated chromosome. **(D)** Simulated probability distributions of the number of transcriptionally active alleles per cell for cells carrying 1-20 copies of the allele. The probability of any allele turning on was set to 0.5 and alleles were assumed to behave completely independent of each other. **(E)** Mean and standard deviation of the distributions on (D). Notice that the mean scales linearly while the standard deviation scales as the square root. As a result, the coefficient of variation (CV) decreases exponentially (inset). **(F)** Fraction of nuclei in which no allele turns on in the distributions in (D). **(G)** Left: in animals genetic redundancy often occurs at the level of enhancers. Multiple enhancers with overlapping expression regulate a single gene and can compensate for the loss or fail of one another. Right: In plants, genetic redundancy is often found as multiple copies of genes with similar function which can compensate for each other.

ing to gene induction in response to transcription factor concentration. Such experiments would help addressing the long-standing 'chicken or egg' question between histone acetylation and transcription.

A major challenge in linking histone modification experiments to live imaging results is reconciling bulk data with single-cell measurements. This problem could potentially be resolved by live imaging tools with single allele resolution that report on the chromatin state in the few minutes timescale. Specific histone modifications can be targeted *in vivo* by nanobodies or "reader" proteins that recognize histone marks. Fusing GFP to these domains has been used to determine the abundance of chromatin modifications at the level of the nucleus or in large tandem DNA repeats [228, 246, 175]. Thus, measuring the dynamics of chromatin modifications in single loci remains one of the most interesting unresolved technical challenges in live cell imaging.

1.11 Conclusions and future prospects

The temporal dynamics of transcription play a functional role across organisms and biological contexts. The timescales with which nuclei in the Drosophila embryo make transcriptional decisions is comparable to the speed of gene expression changes in response to the environment plants. A dynamic and quantitative single-cell view of these events is necessary to build the type of predictive models that make engineering possible but realizing this goal requires new technologies.

On this technical front, I described efforts to develop a number of live imaging approaches to quantify activity in individual loci of live cells. The PCP/PP7 method of labeling nascent RNAs revealed an unsuspected mode of gene regulation where induction treatments such as heat shock and a dark-to-light transition increase the likelihood of alleles switching to an active and relatively stable state and the presence of transcriptionally silent refractory cells. A comparable process was shown to be at play in Drosophila embryos in response to the concentration of the Dorsal transcription factor. Changes in transcription factor concentration play a key role in the plant heat shock and deetiolation responses, prompting the question of whether a shared mechanism explains this digital behavior in both systems. Experiments in plants in which we measured two alleles per nucleus are consistent with this switch-like, digital process being a stochastic event local to the promoter.

These findings lead to the question of how changes in *trans* such as transcription factor nuclear concentration can affect stochastic events in *cis*. The combinatorial complexity of gene regulation makes this a challenging question. In order to build and test models based on physical principles that explain this connection, we engineered a minimal transcriptional regulatory system in fruit fly embryos that allows us to systematically test the role of the concentration of a single transcription factor and that of its DNA binding affinity. These experiments revealed the regulatory repertoire contained in a single base pair of a transcription factor binding site. In addition, they allowed us to

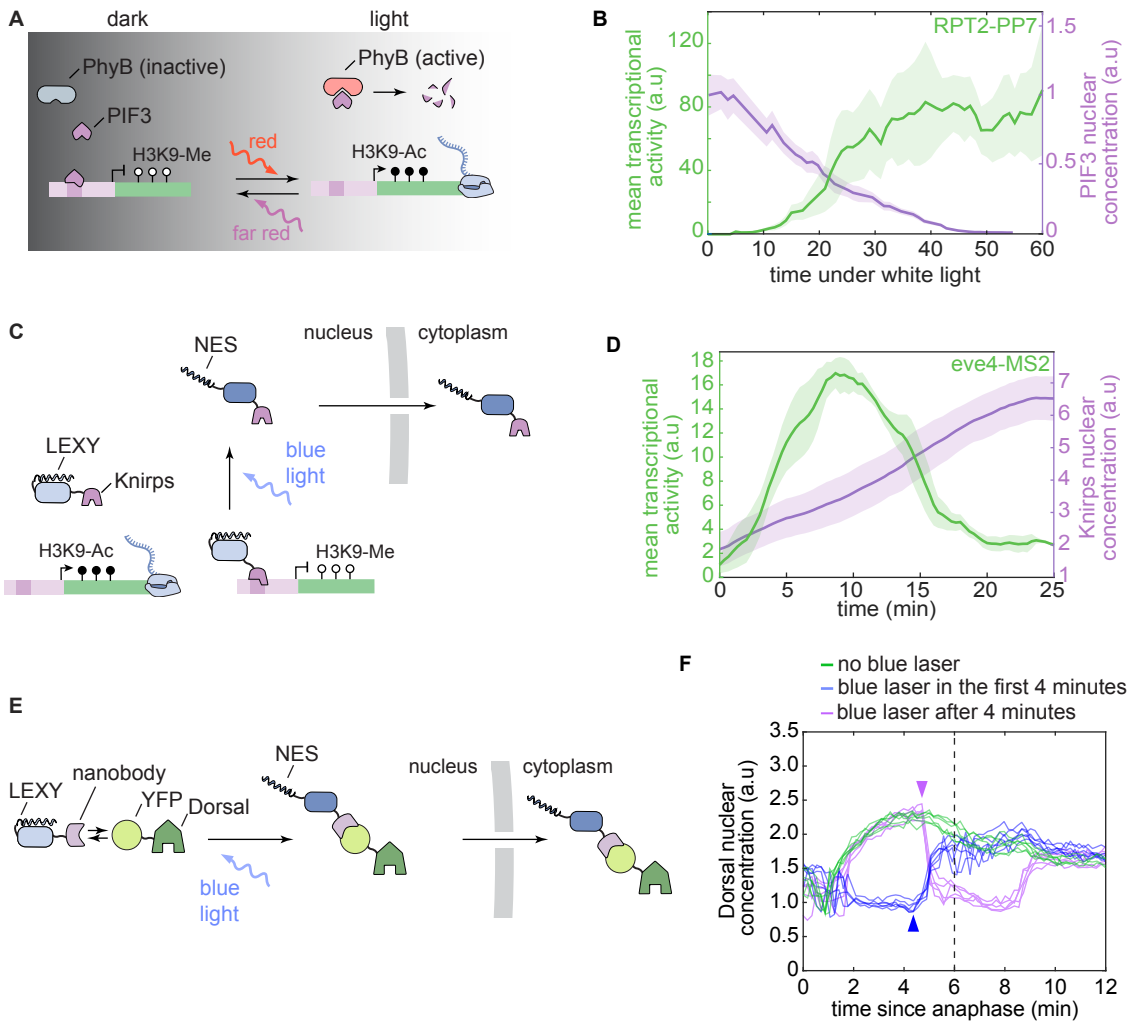

Figure 1.10: **Endogenous and engineered control of transcription factor concentration by light. (A)** Induction of transcription by red light through the PhyB-PIF3 pathway. In the dark, PIF3 is abundant in the nucleus and binds to its targets to either repress or activate transcription. Repression is associated with de-acetylation of lysine 9 of Histone 3 (H3K9-Me). Upon light exposure, red light activates Phytochrome B (PhyB) which targets PIF3 for degradation. mRNA accumulation of PIF3 repressed genes is associated with acetylation of lysine 9 of Histone 3 (H3K9-Ac). **(B)** Transcriptional dynamics of a PIF3-repressed gene measured with a PP7 reporter driven by the RPT2 promoter (green). Shown against the right y-axis is the nuclear concentration of PIF3 measured in a different PIF3-GFP plant (purple). **(C)** Scheme showing the regulation of the eve stripe 4 enhancer by Knirps in the Drosophila embryo and its control using optogenetics. Binding of Knirps leads to transcriptional repression and loss of (H3K9-Ac). Fusing Knirps to the LEXY tag allows controlling its nuclear concentration via a blue light-activated nuclear export signal (NES). ***Caption continues on next page***.

Figure 1.10: ***Continued from previous page*: Endogenous and engineered control of transcription factor concentration by light (D)** Mean transcriptional activity driven by the eve stripe 4 enhancer during nuclear cycle 14 in a Drosophila embryo measured with an MS2 reporter (green). Plotted against the right y-axis is the mean concentration of Knirps in the same cells without applying optogenetic perturbations. **(E)** Schematic representation of the Llama Shepherd approach. Existing YFP transcription factor fusions such as Dorsal-Venus are bound by an anti-GFP nanobody fused to LEXY. Blue light exposes the LEXY NES leading to export of the complex. **(F)** Single nucleus concentration dynamics of a Dorsal-Venus Llama Shepherd without blue light excitation (green), with blue light excitation during the first 4 minutes of interphase (blue) or after 4 minutes into interphase (purple). The arrowheads show the frame where the blue laser was turned on (purple) or off (blue). In (B) and (D), the shaded areas correspond to the standard error across cells.

show that transcriptional induction must operate as a series of irreversible stochastic steps whose rates are biased by Dorsal. Better targeted transgenesis tools in plants will make this synthetic approach possible in the future.

These findings contradict classic models of transcriptional regulation based on statistical physics but, perhaps, are not a big surprise, given that transcriptional regulation involves processes far from equilibrium, such as chromatin remodeling and covalent modification of histones. Experimental platforms to study the role of histone acetylation in transcription, such as repression by PIF3 in Arabidopsis or by Knirps in Drosophila could offer insights, since both proteins can be modulated easily with light.

Stochastic gene expression can be a nuisance for cells that must make the right decision when exposed to a stress treatment. On the other hand, examples from developing systems and microorganisms show that cells can exploit this stochasticity to drive decision-making. The phenotypic impact of cell-to-cell variation in stress response in a eukaryote has hardly been studied. The tools presented here make pursuing this question now possible in plants. Their unique cell biology in terms of genome copy number and intercellular connections make this an intriguing problem.

A future frontier in plants will be connecting transcriptional activity to downstream steps in gene expression, such as protein accumulation and, ultimately, single-cell phenotypic outcomes. This will help in our understanding of how molecular collisions at the DNA level lead to cellular decisions and, ultimately, to organism physiology.

Chapter 2

Development of optogenetic approaches to manipulate transcription factor dynamics in fruit fly embryos

2.1 Abstract

In order to test models of how transcription factor gradients are interpreted during development we sought to establish optogenetic tools compatible with live imaging in *Drosophila* embryos. These efforts led to the establishment of functional versions of two different transcription factors whose nuclear concentration can be controlled by light by a factor of ≈ 2 in a timescale of tens of seconds. Optogenetic perturbations to the Dorsal activator will be critical for testing kinetic models of Dorsal concentration readout by enhancers. An optogenetic version of the Knirps transcription factor promises to uncover the mechanisms by which repressors work.

2.2 Introduction

At an organismal level, transcriptional regulation operates in a characteristic temporal sequence where genes are connected to each other in time and space through activation or repression. This leads to regulatory cascades where a combination of transcription factors involved in a particular processes are *interpreted* by the DNA regulatory sequences, enhancers, to dictate the expression of downstream genes. How this *interpretation* operates in time still remains largely unknown. Classic models of transcriptional regulation in development implicitly assumed steady-state concentration of transcription factors [272, 68, 236], even though the spatio-temporal profile of these proteins is highly dynamic [111, 215]. The steady-state assumption is still prevalent in the field [202], although recent efforts have tried to incorporate the nuclear concentration dynamics of transcription factors [72]. Moreover, transcriptional regulation downstream of these in-

puts is itself a highly dynamic process, likely operating as a multi-step kinetic sequence involving opening of chromatin, binding of transcription factors, assembly of the general transcriptional machinery and RNAP elongation to name but a few steps [238, 160]. The fact that regulatory protein concentration varies in time while the biochemical process that responds to these inputs is also highly dynamic opens several questions. When exactly are transcription factors read out during development? Do concentration *dynamics* provide regulatory information in addition to the absolute concentration? How do distinct biochemical steps in the transcription cycle respond to transcription factor concentration?

Genetic perturbations have proven to be a powerful tool to test hypotheses and establish causality in gene regulation but traditional methods to probe gene function do not allow to precisely and rapidly manipulate transcription factor concentration (Fig. 2.1A). Optogenetics offers an ideal method to close this technical gap. Optogenetics - the manipulation of biology with light- is used in the literature in two main ways. Specifically in neurobiology, it typically refers to the use of light to manipulate the permeability of ion channels to control neuronal activity. In cell and developmental biology however, optogenetics refers to the use of light sensitive protein domains to alter the localization, dimerization, and/or structure of fused proteins. Since the Quail lab took advantage of the light-dependent interaction between Arabidopsis PhyB and PIF3 proteins to manipulate gene expression in yeast [241], a myriad other optogenetic systems have been developed and used in cell and synthetic biology.

A series of elegant optogenetics experiments performed in single cells in culture have described the regulatory role of transcription factor concentration dynamics independent of absolute concentration values: cells experiencing the same time-averaged transcription factor concentration adopt different transcriptional decisions depending on how this concentration changes over time [132, 257, 270]. Recently, using the CRY2 tag, a few groups have implemented optogenetic approaches to study the temporal aspects of transcription factor activity in development using the early *Drosophila melanogaster* (Drosophila) embryo as a model. [179] showed that Zelda activity is necessary continuously throughout development. Also using a CRY2 fusion, [129] studied the role of temporal dynamics in the interpretation of the Bicoid transcription factor concentration gradient in Drosophila. In contrast to Zelda and previous theoretical expectations [110], the duration of exposure to Bicoid that is sufficient to commit cells to Bicoid-dependent cell fates varies across different cell types.

In order to refine mechanistic models of transcriptional regulation and uncover unknown regulatory richness, we sought to develop optogenetic methods in Drosophila embryos to precisely and rapidly perturb transcription factor nuclear concentration while simultaneously measuring their activity using the MS2-MCP technique.

An ideal optogenetic method for this research program should be 1) compatible with live imaging in at least two disinct wavelengths, 2) easy to engineer, 3) fast, 4) tunable, and 5) reversible. Finally, we think it is paramount that the mechanism of this perturbation tool is known. Much in the same way that interpreting the effect of allelic mutants

depends on the nature of the allele, interpreting the effect of optogenetic perturbations depends on how they affect protein function.

2.3 Results

2.3.1 Proof of principle of LEXY tag in the fly embryo

Recently, the LEXY optogenetic tag was developed [194]. Here, amino acid substitutions were introduced in the Jα helix of the *Avena Sativa* (oat) LOV2 domain (AsLOV2) to make it resemble a nuclear export signal (NES). Excitation of the LOV domain triggers a conformational change that exposes the otherwise precluded jα helix, resulting in recognition of the NES by the nuclear export machinery (Fig. 2.1D).

The LEXY tag offers several advantages over other optogenetic tools that make it particularly attractive. First, in contrast to two-component systems like the PhyB-PIF3 module (129.4 and 59 kDa), LEXY is a single, relatively small (19 kDa) self-contained tag. In addition, the chromophore in the LOV domain is NADPH, a ubiquitous molecule in all domains if life. The absorption spectrum of AsLOV2 becomes negligible past $\approx$ 520 nm [229], potentially leaving the green and red parts of the light spectrum for simultaneous imaging in two channels without triggering undesired export (Fig. 2.1E). The half-life of dark recovery of AsLOV2 is on the order of 20 seconds [229], much faster than the PhyB and CRY2 dark relaxation half lives of $\approx$ 5 minutes in heterologous systems [73, 16, 129]. Finally, the mechanism of action of LEXY –reduction of nuclear concentration–is straightforward to measure and interpret. This is in contrast to other optogenetic methods to inactivate transcription factors such as fusions to the Arabidopsis CRY2 protein which are thought to work through the formation of aggregates.

To test LEXY in the fly embryo, we generated a line expressing a fusion of LEXY to mCherry and an NLS (Nuclear Localization Signal) to target the fusion to the nucleus in the absence of LEXY excitation (see Materials and methods). As shown in Fig. 2.1F and G this protein localizes to the nucleus and is rapidly and reversibly exported upon blue light in a timescale of 1-2 minutes, demonstrating the functionality of LEXY in the Drosophila embryo.

2.3.2 Characterization of NLS-mCherry-LEXY export dynamics

In order to determine the limits of using LEXY in fluorescence live imaging experiments, we sought to characterize the LEXY tag further than in its original publication [194]. Specifically, we aimed to measure the wavelength dependence of export and the effect of laser power in export dynamics. Determining the wavelength dependence of LEXY export is particularly important for the design of a multicolor imaging experiments where visualization must be independent of optogenetic perturbations.To this end we used the NLS-mCherry-LEXY line described above.

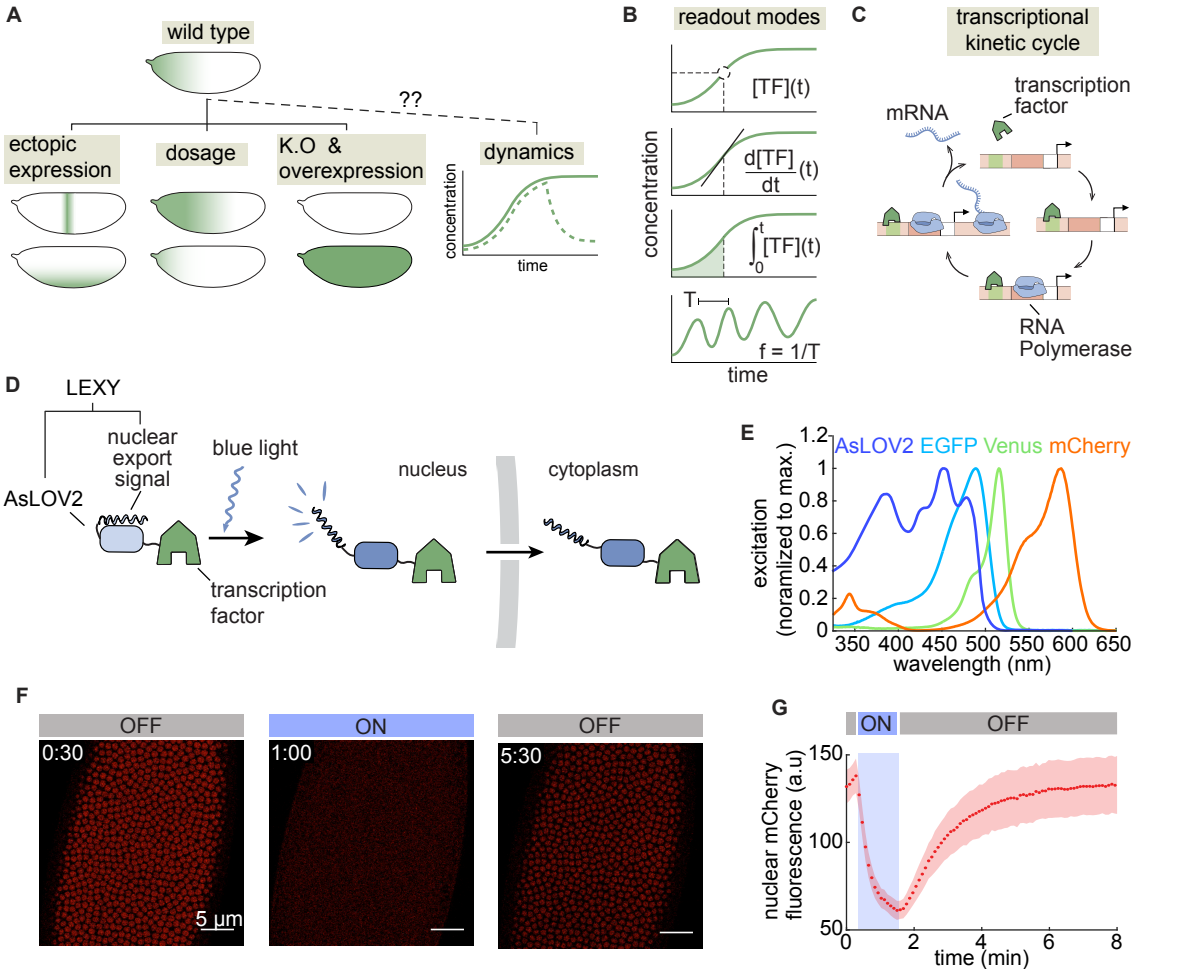

Figure 2.1: **Application of optogenetics in the fly embryo allows precise and rapid perturbation of nuclear protein concentration**. **(A)** Genetic strategies to perturb the characteristic patterns of transcription factor concentration in development. Static perturbations (ectopic expression, dosage change and overexpression) are well established but similar tools to precisely control fast temporal dynamics with the same ease and freedom are not available. **(B)** Models of how transcription factor concentration could be read out. Top: Instantaneous readout of transcription factor concentration. Middle: the rate of change of in concentration is what provides regulatory information. Bottom: The promoter integrates the concentration over a time window. **(C)**. Temporally controlling the presence of a transcription factor on the locus could help determine its role in different steps of the transcriptional cycle. **(D)** Schematic representation of the LEXY optogenetic tagging system. Upon blue light excitation, the LOV domain exposes a nuclear export signal and the fusion protein gets transported to the cytoplasm. ***Caption continues in next page***

Figure 2.1: ***Continued from previous page*: Application of optogenetics in the fly embryo allows precise and rapid perturbation of nuclear protein concentration. (E)** Excitation spectra of commonly used fluorescent proteins in live imaging and the *Avena sativa* Phototropin 1 LOV domain (AsLOV2). Note that excitation of the LOV domain becomes negligible after $\approx$ 520 nm, permitting the use of Venus and mCherry. **(F)** Snapshots of embryos expressing a ubiquitous NLS-mCherry-LEXY fusion protein. Before blue light exposure the protein is enriched in the nucleus. Turning on the 458 nm laser results in rapid nuclear export, which is reversed when the blue laser is turned off. **(G)** Quantification of the data in (F). The shaded area represents the standard deviation across nuclei.

We quantified the nuclear concentration of the NLS-mCherry-LEXY protein under different laser wavelengths from 458nm to 540nm while simultaneously imaging mCherry. To excite mCherry we used a 587nm laser, far enough from the absorption spectrum of the LOV2 domain to make export negligible. (Fig.2.1 E and G). The time trace of the nuclear concentration is shown in Fig. 2.2 A. Note that because longer export wavelengths can excite mCherry, the nuclear fluorescence was normalized to the intensity right before the export. Excitation laser with different wavelength was turned on at the end of the 5th frame, which then starts exporting of the NLS-mCherry-LEXY from the nucleus. The time trace shows steeper decrease of the fluorescence in the shorter wavelength. The export process can be modelled using a one-step kinetic model, whose dominant rate is the export rate. The equation for the decay in nuclear fluorescence intensity can be written as

$$Intensity = A * e^{-B*t} + C \tag{2.1}$$

where A is a scaling factor, B is the decay constant, and C is a basal fluorescence level.

By fitting our data to Eqn. 2.1, we estimated the export rate B as a function of wavelength (Fig. 2.2 B). As expected, the export rate resembles the absorption curve of the AsLOV2 domain, becoming effectively zero past 520nm (Fig. 2.2 B). As shown in Figure 2.2A, the fluorescence time traces for 520 nm and 540 nm export wavelengths have similar trends, with much smaller decay constants. We attribute this decrease to mCherry photobleaching. Bleaching should not significantly affect the quantification of export in shorter wavelengths as the export process is orders or magnitude faster.

Next, we reasoned that the ability to control how fast LEXY gets exported could add even finer experimental control. To test this idea, we measured the laser power dependence of the export rate. We tested 3 laser powers -2 μW, 5 μW, and 10 μW- for 3 different laser wavelengths, 458 nm, 490 nm, and 510 nm. The nuclear mCherry fluorescence traces under the 510 nm laser are shown in Fig. 2.2 C. We found that for a wavelength that is sub-optimal in terms of export like 510 nm, the stronger the laser power, the faster the export was. However, this relationship was not as clear for optimal

export wavelengths such as 458 and 490 nm, likely because they already saturate at 2 μW. By fitting these time traces to the Eqn. 2.1, we could extract the export rate as a function of the laser power for 3 different wavelengths. The result is given in Fig. 2.2 D. The export rate saturates at 2μW –the weakest laser power measurable by our equipment– when using shorter wavelengths (458nm, 490nm). There is some dependence of export on laser power for longer wavelengths such as 510nm with 5 μW exporting at a rate ≈ 2 times faster than 2 μW.

In summary, these results show that to use LEXY orthogonally with fluorescence imaging, the fluorophores must be excited by wavelengths higher than 520 nm. Thus, measuring transcription factor concentration and downstream transcriptional activity should be possible using a yellow and a red fluorescent protein such as Venus and mCherry, respectively. Additionally, this characterization demonstrated that it is possible to tune the rate of export by modulating the laser power of a laser whose wavelength is inefficient at export. In the future, the power dependence of export could be further characterized using a power meter capable of reliably measuring intensities lower than 2 μWm.

2.3.3 Application of LEXY to maternally deposited transcription factors

These results convinced us that it should be possible to design an experiment where a transcription factor input is visualized with a YFP variant while its transcriptional output is measured simultaneously with MS2 using a red fluorescent protein. This arrangement would allow controlling transcription factor nuclear dynamics independent of measuring its concentration and downstream activity (Fig. 2.3A).

In order to establish such a system, we turned our attention to three well-studied maternally-deposited transcription factors, Bicoid, Dorsal, and Zelda. Maternally deposited proteins offer the advantage that they can be precisely measured with fluorescent protein fusions, unlike zygotic transcription factors which accumulate and degrade over timescales that are short compared to fluorophore maturation [39].

There are two key considerations in this optogenetic approach. First, adding the LEXY tag in addition to a fluorescent protein means adding a total of$\approx$ 50 KDa to the transcription factor of interest, which is on the order of the molecular weight of Bicoid, Dorsal and Zelda (54.5, 75.3 and 145.4 kDa respectively). This could potentially hinder their functionality. The second consideration has to do with nuclear export itself. Since transcription factors usually carry multiple nuclear import sequences and can physically interact with chromatin and other nuclear components, it is possible that the rate and/or the dynamic range of export could be much smaller than that of NLS-mCherry-LEXY.

With these considerations in mind, we first focused on obtaining transcription factor LEXY-YFP fusions that rescue embryonic development. We generated CRISPR knock-in lines of LEXY-mCitrine-Bicoid, Dorsal-mCitrine-LEXY and Zelda-mCitrine-LEXY based on previously published, rescuing transgene fusions [110, 215, 118] (see Materials and Methods). With the exception of Zelda, homozygous females carrying these fusions failed to

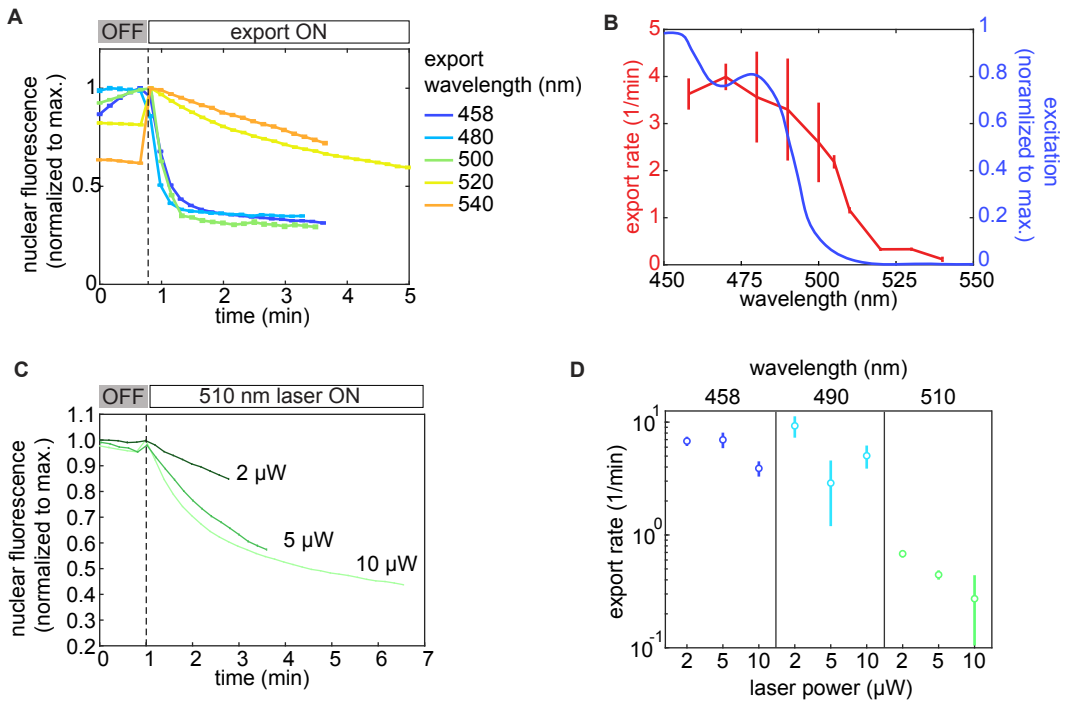

Figure 2.2: **LEXY export dynamics can be tuned by laser wavelength and power**. **(A)** Characterization of the wavelength dependence of export dynamics. NLS-mCherry-LEXY was exported using different laser wavelengths. Because longer wavelengths can result in the excitation of mCherry the data was normalized to the value right before export. The decrease in fluorescence at 520 and 540 nm is likely due to photobleaching and/or endogenous nuclear dynamics such as increasing nuclear volume. **(B)** Data from (A) was fitted to an exponential decay function. Shown in red is the fitted decay rate and the 95% confidence interval as a function of wavelength. Plotted against the right y axis is the AsLOV2 absorption spectrum. **(C)** Measurement of the laser power dependence of export dynamics. For a given wavelength three different laser power intensities were used to trigger export of NLS-mCherry-LEXY. **(D)** Quantification of the data in (C) using the same approach as in (B).

produce viable offspring in the dark, implying that the function of these proteins was compromised. In order to generate functional fusions we tried a number of different linkers and protein termini but these attempts were unsuccessful.

In terms of light-triggered export, Bicoid and Dorsal fusions worked, with an export rate on the order of minutes (Fig. 2.3B-D) although the dynamic range of Bicoid was smaller than that of NLS-mCherry-LEXY (Fig.2.3D). The Zelda-mCitrine-LEXY protein nuclear concentration did not change under blue light (data not shown), perhaps due to the known interaction of Zelda with chromatin. The results of these efforts are summarized in Table 2.1

Optogenetic Maternal Transcription Factor Fly Lines			
Gene	**Plasmid**	**Perturbation**	**Rescue**
Zelda	LEXY-mCitrine-Zelda (CRISPR)	No	Yes
Dorsal	Dorsal-mCitrine-LEXY (CRISPR)	Yes	No
Bicoid	LEXY-mCitrine-Bcd (CRISPR)	Yes	No
Bicoid	LEXY-mVenus-Bcd (CRISPR)	Yes	No
Bicoid	pBphi-mCitrine-Bcd-LEXY	Yes	No
Bicoid	pBphi-Zdk-mCitrine-Bcd	Yes	No
Bicoid	pBphi-Zdk-mVenus-Bcd	Yes	No
Bicoid	pBphi-eGFP-Bicoid-LEXY	Yes	No
Bicoid	pBphi-LEXY-Bicoid-eGFP	Yes	No
Bicoid	pBphi-LEXY-SLL-mCitrine-Bicoid	Yes	No
Bicoid	pBphi-LEXY-12xGS-mCitrine-Bicoid	Yes	No
Dorsal	pBphi-Dorsal-mCitrine-Zdk	Yes	No

Table 2.1: List of optogenetic fly lines created. The order of protein domains correspond to their order from N to C terminus. Domains were joined by a 6x Glycine linker unless stated. SLL corresponds to the 'Stadler Long Linker' (SGDSGVYKTRAQASNSAVDG TAGPGSTGSS), 12xGS corresponds to a 12 amino acid long Glycine-Serine linker.

2.3.4 Two-component optogenetic systems

Although none of our LEXY fusions of Bicoid and Dorsal were able to rescue development to a significant extent, we were encouraged by the fact that –in contrast to Zelda– it was

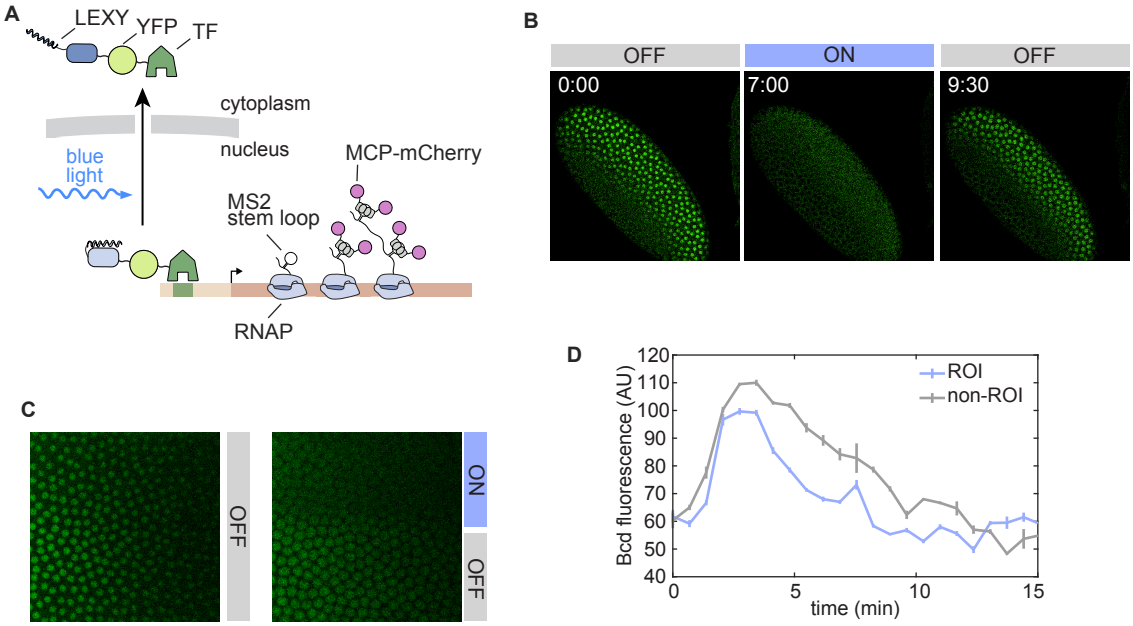

Figure 2.3: **Application of the LEXY technology to the transcription factors Bicoid, Dorsal.A** Schematic of th cartoon of LEXY fusion to the transcription factors of interest. We chose the fluorescent protein as YFP such that we can do 2-color imaging with the MCP with mCherry for the transcriptional activity of the downstream genes. **(B)** Maximum projection snapshots of embryos expressing a CRISPR knock-in fusion of the Dorsal transcription factor to the mCitrine fluorescent protein and the LEXY tag. To export Dorsal, the whole field of view was illuminated using a 458 nm laser. **(C)** Maximum projection snapshots of an embryo expressing a knock-in fusion of the Bicoid transcription factor to mCitrine and LEXY. To trigger export, only the upper half of the field of view (Region-Of-Interest, ROI) was illuminated with a 458 nm laser, and the lower half is only visualized for the Venus fluorophore (non-ROI). **(D)** Mean Bicoid nuclear concentration dynamics within the two halves of the experiment in (C) for nc 13 and nc14.

possible to export these proteins from the nucleus. We hypothesized that the reason why these proteins were rendered non-functional was due to the large size of the LEXY-fluorescent protein fusions.

Taken together, these experiments led us to believe that Venus-Bicoid and Dorsal-Venus should be minimally perturbed in the nucleus in terms of additional proteins covalently or non-covalently bound to them. To achieve this, we resorted to sequestration-based two-component optogenetic approaches. Broadly speaking, these systems rely on the light-dependent dimerization between a protein of interest (in our case, Venus-fused transcription factors) and a second binding partner that is tethered away from the site of transcription. Blue light disrupts this sequestration, releasing the transcription factor into the nucleus.

LOVTRAP is a two component system where the short Zdk peptide (6.5 kDa) binds to the AsLOV2 domain in the dark [267]. We generated CRISPR knock-in fusions of Bicoid to Zdk and Venus, and crossed this line to flies carrying an AsLOV2 protein tethered to the outer mitochondrial membrane. Embryos laid by transheterozygote females containing these two components showed normal Bicoid localization in the dark, indicating the lack of substantial AsLOV2-zdk dimerization. Notably, females homozygous for a CRISPR knock-in of zdk-Venus-Bicoid did not produce viable embryos, while Venus-Bicoid fusions are functional [112]. Thus, even the addition of small Zdk peptide is enough to compromise the transcription factor activity. We therefore concluded that we have to avoid covalently tagging the transcription factor with anything other than already validated fluorescent proteins.

With these lessons in mind, we next decided to take advantage of the strong nanobody-Venus interaction to establish an Optobody approach [86, 146]. Optobodies are split nanobodies where each half is fused to light-dependent dimerization partners. In the dimer conformation, the nanobody halves come together to form a functional nanobody capable of binding the antigen [101, 281]. By tethering anti-GFP optobody pieces to the mitochondrial outer envelope, the plasma membrane, or the nuclear lamina, it should, in principle, be possible to sequester Venus-transcription factor fusions away from enhancers. We created a single plasmid containing two transgenes, each coding for one anti-GFP nanobody half fused to either zdk or AsLOV2. Both proteins were targeted to the outer mitochondrial membrane. Transcription factors are synthesized in the cytoplasm and are exported from the nucleus during mitosis, which should enable sequestration at the mitochondrial outer envelope in the dark. We found that embryos laid by females carrying the optobody constructs and Venus-Bicoid had normal Bicoid subcellular localization and no evidence for sequestration. The lack of sequestration is likely due to the lack of interaction between the optobody and Venus, possibly because the arrangement of the optobody halves was suboptimal. This was shown to play a big role in other contexts [281].

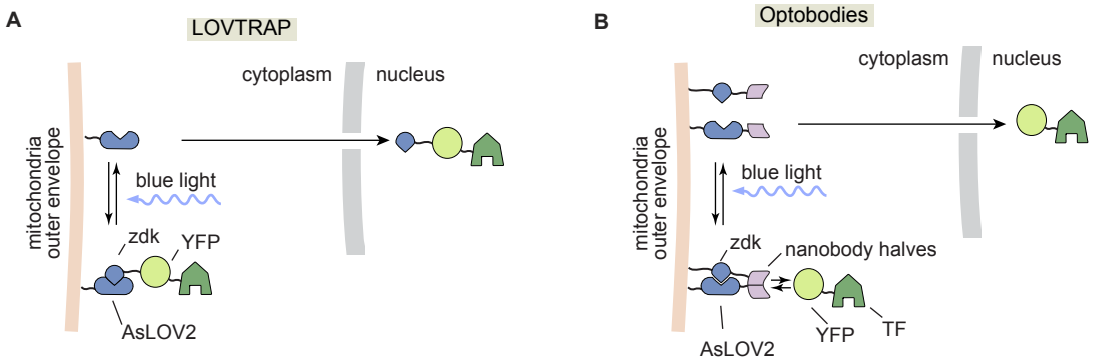

Figure 2.4: **Two component systems to sequester transcription factors away from the nucleus. (A)** In the LOVTRAP system, the zdk protein binds to the LOV domain in its dark conformation. Exposure to blue light disrupts the zdk-LOV physical interaction. Targeting the LOV domain to the mitochondrial outer membrane can be used to sequester a protein of interest fused to zdk. In this example, a transcription factor is fused to a fluorescent protein and zdk and is released into the nucleus upon blue light exposure. **(B)** Schematic of the optobody system. An anti-GFP nanobody is split into two and each half is fused to optogenetic binding partners such as zdk and LOV. In the dark, interaction between the binding partners brings the nanobody halves together, reconstituting a functional protein capable of binding YFP. Blue light exposure disrupts the zdk-LOV interaction, splitting the nanobody and releasing YFP. In (A) and (B) the optogenetic binding partners can be replaced by any other pair such as PhyB and PIF3. Alternative tethering surfaces other than the mitochondrial outer membrane are the plasma membrane and the cytoplasmic face of the ER.

2.3.5 Llama Shepherds, a novel plug-and-play approach to turn existing fluorescent protein fusions optogenetic

We and others have generated functional fusions of Dorsal, Bicoid, and other transcription factors to Venus and other GFP-derived fluorophores [112, 215, 110]. We sought to take advantage of these resources to turn them into their optogenetic versions instead of engineering them from scratch as described above. A generalizable strategy to convert any fluorescent protein fusion into a light-controllable tool would have widespread applicability.

To achieve this, we developed a novel optogenetic system inspired by the LlamaTag approach. A LlamaTag consists of an anti-GFP Llama nanobody fused to the transcription factor of interest. In the presence of GFP-derived proteins, the nanobody binds to the fluorophore, effectively labeling the transcription factor in a way that does not depend on fluorescent protein maturation [39] (Fig. 2.5A). We hypothesized that a nanobody-LEXY fusion could bind to existing Venus-TF fusions, carrying them out of the nucleus upon

blue light excitation. We termed this approach Llama Shepherds (Fig. 2.5B). We speculated that the non-covalent interaction between Venus and the nanobody-LEXY fusion might minimize perturbations to transcription factor function compared to the covalent fusions described above. We tested Llama Shepherds by generating transheterozygous females carrying the LEXY-nanobody fusion and either Venus-Bicoid or Dorsal-Venus. Illumination with a blue laser resulted in nuclear export in both cases (Fig. 2.5C,D). Next, to test whether Llama Shepherds retain protein function, we generated females homozygous for CRISPR knock in Venus-Bicoid or Dorsal-Venus fusions and heterozygous for the LEXY-nanobody transgene. The survival rate of embryos laid by these females in the dark was extremely low, on the order of 1% compared to $\approx$ 50% of females homozygous for Venus-Bicoid and Dorsal-Venus (see: Materials and Methods).

Although the rescue rate low, Llama-shepherded transcription factors do not result in complete loss of function. We reasoned that increasing the transcription factor dosage could increase the rescue rate. To test this hypothesis, we generated females homozoygous for a CRISPR knock-in Dorsal-Venus fusion and heterozygous for a Dorsal-Venus transgene whose expression is identical to that of the endogenous gene [215], thus bringing the Dorsal-Venus dosage to 1.5X that of the original Dorsal Llama Shepherd. As in the original experiment, these females were also heterozygous for a nanobody-LEXY fusion. Around 30% of the embryos laid by these females develop into larvae. Thus, increasing the dosage of the YFP-tagged transcription factor is a viable strategy to obtain rescuing Llama Shepherd flies. To more precisely define the degree of genetic complementation of the 1.5X Dorsal Llama Shepherd, we measured the activity of a minimal reporter construct driven exclusively by Dorsal (see Chapter 4). As shown in Fig. 2.5 E, comparing multiple metrics of transcriptional activity between embryos with or without the nanobody-LEXY fusion showed that the activity of this reporter is largely the same. We conclude that the 1.5X Dorsal Llama Shepherd can be used to dissect the temporal aspects of transcriptional regulation by Dorsal.

By illuminating with the 458 nm laser from the beginning of nuclear cycle 12 to 4 minutes into interphase, it was possible to reduce the Dorsal concentration by a factor of $\approx$ 2 (Fig. 2.5F). Turning off the 458 nm laser results in nearly immediate release of Dorsal into the nucleus, with concentration reaching steady-state in about 1 minute (Fig. 2.5F). Turning the 458 nm laser after 4 minutes into interphase results in the rapid decrease of Dorsal nuclear concentration (Fig. 2.5F) with timescales and dynamic range similar to sequestration and release. This demonstrates that the Llama Shepherd approach applied to Dorsal yields both reversible and fast control.

Readout of nuclear Dorsal concentration by enhancers could be instantaneous or could depend on the history of exposure to Dorsal. In Chapter 4, we explored several theoretical approaches to address this question. Tackling this question experimentally is currently not feasible because the dynamics of Dorsal nuclear concentration are highly stereotypical. The 1.5X Dorsal Llama Shepherd creates the opportunity to arbitrarily disentangle concentration from time (Fig. 2.5F).

Next, we sought to test whether the readout of the Dorsal gradient by in a minimal

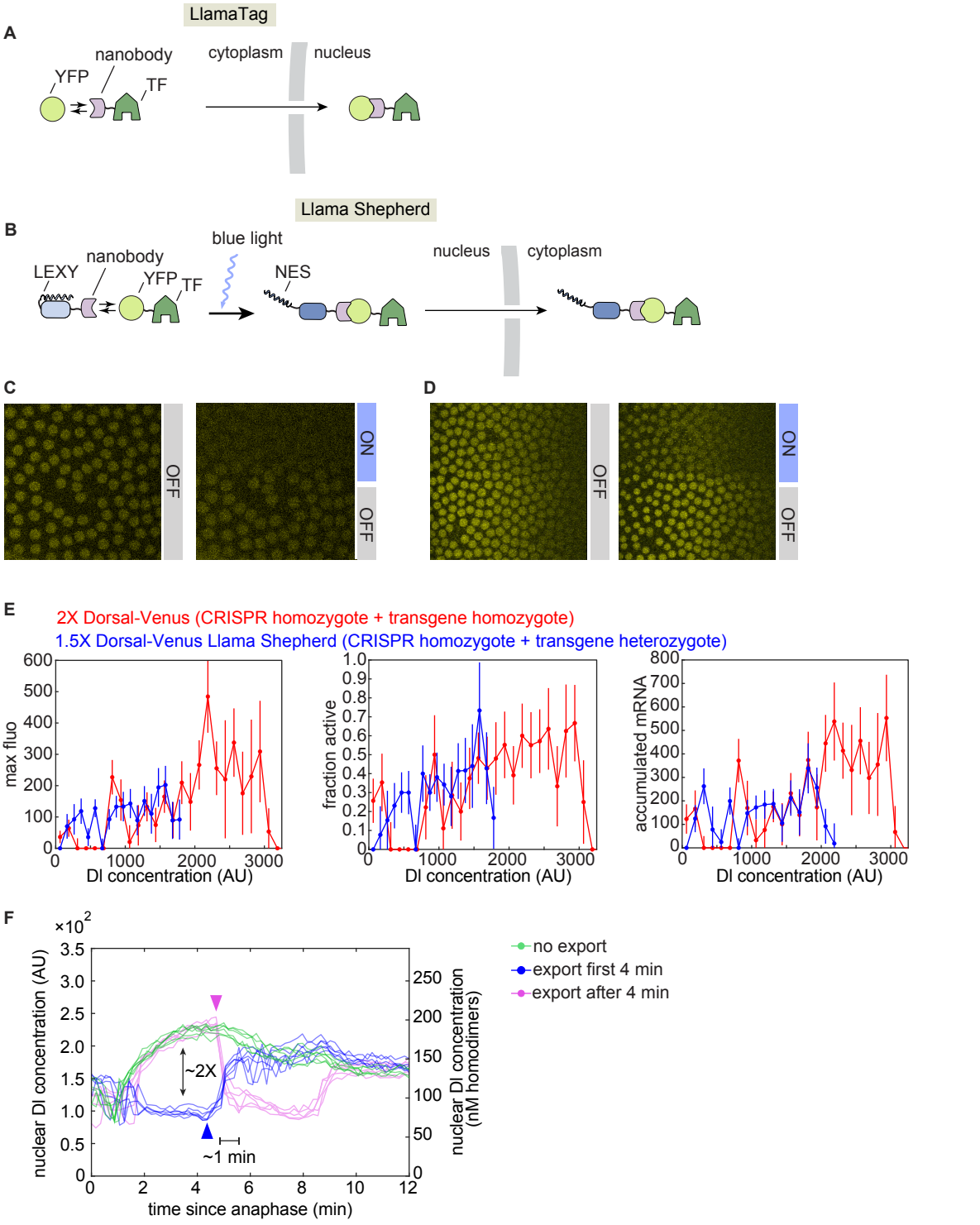

Figure 2.5: **Engineering optogenetic Dorsal using the Llama Shepherds approach.** ***See caption on next page***

Figure 2.5: **Continued from previous page: Engineering optogenetic Dorsal using the Llama Shepherds approach. (A)** Schematic representation of the LlamaTag approach to fluorescently label proteins. The transcription factor of interest (TF) is fused to a single-chain Llama nanobody, which binds to a ubiquitously expressed mature fluorescent protein such as YFP. **(B)** The Llama Shepherd approach. An existing fusion between a transcription factor (TF) and a GFP-derived fluorescent protein is bound by a nanobody fused to the LEXY optogenetic tag. Blue light exposes the nuclear export signal (NES) in LEXY, resulting in nuclear export of the whole complex. **(C)** Application of Llama Shepherds to Venus-Bicoid. **(D)** Application of Llama Shepherds to Dorsal-Venus. In (C) and (D) the blue rectangle on the right shows the region of the field of view where the 458 nm laser was on. **(E)** Comparison of the activity of a Dorsal-Venus Llama Shepherd (blue) to Dorsal-Venus (red). Three different aspects of transcriptional activity regulated by Dorsal were compared as a function of Dorsal nuclear fluorescence. Error bars correspond to the standard error across 3 or more embryos. **(F)** Nuclear fluorescence time traces of the ventral-most nuclei in a Dorsal-Venus Llama Shepherd embryo carrying 3 copies of Dorsal-Venus. The field of view was exposed to blue light for the first 4 minutes (blue), after 4 minutes (magenta) or never (green). The arrowheads indicate when the blue laser was turned on. The dynamic range of optogenetic export is around 2 and the time to maximum export is close to 1 minute.

dorsal-responsive enhancer is instantaneous or not. To this end we measured the activity driven by this enhancer using MS2 in an embryo carrying the 1.5X Dorsal-Venus Llama Shepherd (see Chapter 4 for a description of this synthetic enhancer). We used three different blue light illumination regimes. First, a control without blue light excitation, second, blue light illumination during the first 4 minutes of nuclear cycle 12 and finally, illumination only past 4 minutes. We then quantified four different regulatory metrics in the output expression: The fraction of nuclei that transcribe at any point of nuclear cycle 12, the time of first transcription spot detection, the maximum fluorescence of transcription spots over time, and their mean fluorescence over time. We chose 4 minutes because that is the time at which transcription spots are first detected on average.

We reasoned that, if Dorsal has a function minutes prior to active transcription, then nuclei exposed to the same Dorsal concentration at a fiducial time of 6 minutes would have different transcriptional outputs, with nuclei experiencing export of Dorsal for the first 4 minutes being weaker than the control (Fig. 2.6A). Alternatively, if exposure to Dorsal previous to transcription is irrelevant, then nuclei with the same concentration at 6 minutes would look similar in terms of their activity, regardless of whether Dorsal was exported or not prior to 4 minutes. Another way of testing the same hypothesis is to compare the no export control to nuclei where Dorsal is exported after transcription starts. Here, if nuclei exposed to the same Dorsal concentration at 6 minutes have the same transcriptional profile regardless of their past history (i.e, Dorsal exported or not)

then we conclude that Dorsal is read instantaneously (Fig. 2.6A). Alternatively, if nuclei that export Dorsal past 4 minutes are more active than their control counterparts (nuclei without export experiencing the same Dorsal concentration at 6 minutes) then Dorsal is not read instantaneously.

An important consideration for these experiments are the timescales of regulation. Dorsal has multiple targets and thus perturbing Dorsal could have indirect effects on the reporter expression. For example, Dorsal could activate the transcription of a transcriptional activator of the reporter. An argument against this idea is that the timescales of this experiment are too short for such secondary effects to take place, there is not enough time to change protein concentrations of indirect targets in a meaningful way. However, this does not completely rule out the possibility which is why using reporters that respond exclusively to Dorsal is important.

When we performed this experiment using a minimal synthetic enhancer containing a single strong binding site for Dorsal, we found that optogenetic perturbations had no discernible effect on the transcriptional output (Fig. 2.6B-E). This is likely due to the fact that in this range of Dorsal concentration, activity of the control is not modulated to an appreciable extent. For this reason, optogenetically reducing the Dorsal concentration by a factor of ≈ 2 does not have a big effect in the activity of this enhancer. In Chapter 7 I explore solutions to go around this problem.

2.3.6 Application of LlamaTags and LEXY to the Knirps repressor

After the failed attempts at creating optogenetic versions of Bicoid and Dorsal described above, we concluded that a better strategy would be to focus on the original design –a LEXY fusion– but trying multiple different transcription factors. The expectation here was that, since transcription factors can be structurally very different from each other, it is possible that fusing LEXY to some of them might not substantially alter their activity. Importantly, quantitative imaging of zygotic transcription factors is complicated by the fact that their lifetime is shorter than the maturation rate of fluorescent proteins. For this reason, in order to visualize them it is necessary to use LlamaTags [39]. In practice, this entails knocking in a LEXY-nanobody fusion into the gene of interest, whose protein product can be measured by co-expressing a free, already mature YFP (Fig. 2.7A). Candidate zigotic genes for which CRISPR reagents are available in the Garcia laboratory include the transcription factors *hunchback*, *Krüppel*, *knirps*, *giant* and *even-skipped*.

In this limited screening effort, we identified a CRISPR knock-in fusion of Knirps that can rescue development as a homozygote in the absence of free eYFP. Coexpressing eYFP makes it possible to visualize this protein, which gets rapidly exported from the nucleus under blue laser illumination (Fig. 2.7A and B). However, when eYFP is supplied, flies homozygous for this fusion die at the larval stage. Because developmental transcription factors in Drosophila tend to be reused throughout development we speculated that the Knirps-nanobody-LEXY:eYFP complex could still be functional at early embryo stages. One of the best defined roles of Knirps in the early embryo is repression of the

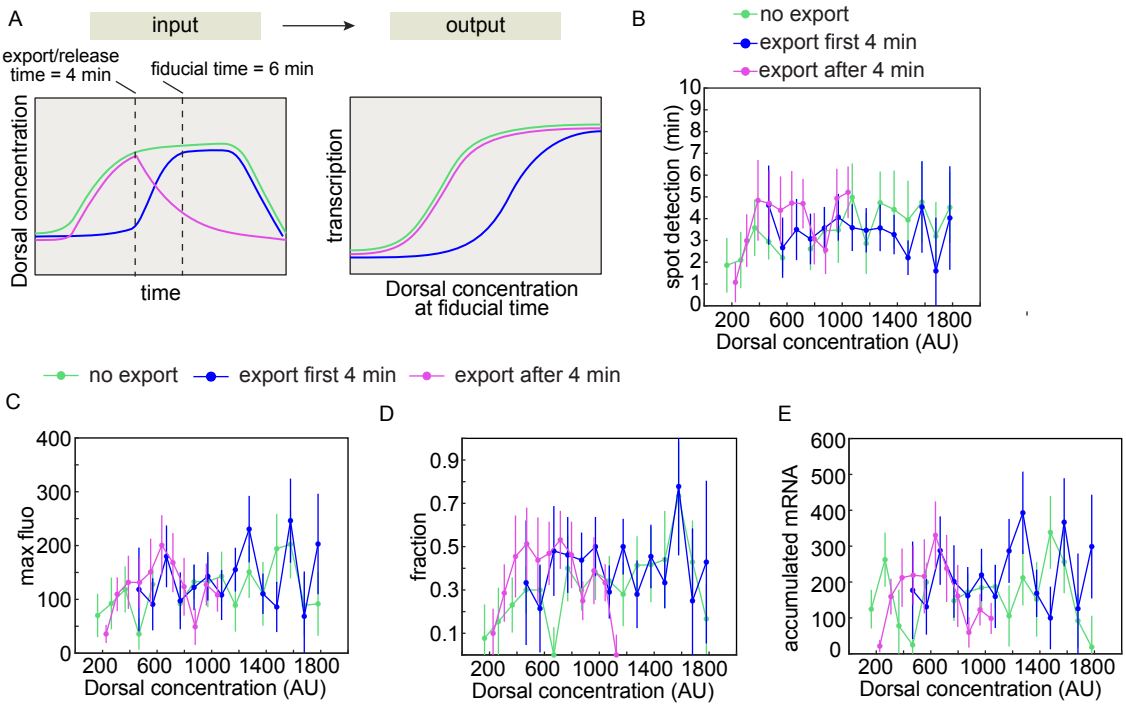

Figure 2.6: **Using optogenetics to test the time dependence of Dorsal activation (A)** Left: Cartoon showing the optogenetic perturbations used and their possible effects. Dorsal is exported from or released into the nucleus at 4 minutes (blue and magenta lines respectively). The Dorsal concentration at 6 minutes is used to group nuclei in the bins of Dorsal concentration used in the x axes in (B)-(E). Right: Schematic of the expected result if an enhancer is only responsive to Dorsal before 4 minutes. **(B)** Average time of first spot detection in nuclei sharing the same Dorsal concentration at 6 minutes (x axis) but exposed to different optogenetic perturbations. **(C)** Average maximum spot fluorescence. **(D)** Fraction of refractory nuclei. **(E)** Accumulated mRNA. In (B)-(E) the mean was taken over at least 3 embryos. The error bars correspond to the standard error of the mean.

even-skipped stripes 4 and 6. To test if optogenetic Knirps is functional at repressing *even-skipped*, we measured the expression an *eve*-MS2 reporter in embryos homozygous for Knirps-nanobody-LEXY in the presence of eYFP. As shown in Fig. 2.7E and F, nuclei experiencing high levels of optogenetic Knirps shut off expression of *eve* while nuclei with low Knirps concentration do not. Taken together, these experiments demonstrate that the quantitative aspects of the repression of *eve* by Knirps can be dissected using optogenetics.

2.4 Discussion

One of the most exciting aspects of biology is the ability to falsify hypotheses using experimental perturbations. Compared to largely observational sciences such as ecology or astronomy, in molecular biology the limits to directly testing models is set mostly by ingenuity. Many of the current models of transcriptional regulation invoke temporal dynamics [72, 234] yet the tools available to test them are static perturbations such as modifications to DNA sequence or changes in gene dosage. Here we have introduced technical developments that should close this gap between the timescales of theoretical models and timescales of perturbations.

One of the strategies that we used to create functional optogenetic proteins was increasing gene dosage. It can be argued that increasing the protein dosage is an artificial way of restoring functionality and that, as a consequence, experiments with these lines might not reflect the behavior of the endogenous system under study. Independent of the merits of this argument it should be pointed out that, in practice, this has been the established approach in the field. When protein fusions are made to complement null mutants, it is not uncommon to select lines that express the protein at higher dosage than wild type precisely because fusions tend to compromise protein activity [278].

It is useful here to distinguish between the concepts of rescue and complementation. Rescue typically refers to the ability of a line carrying a genetic perturbation to produce viable offspring. Although usually treated as a qualitative feature, rescue is, in essence, a quantitative descriptor since it can be measured as the proportion of the offspring that complete development. Because of pleiotropy, a given genetic perturbation usually has multiple phenotipic effects, not all of which are related to the specific biological phenomenon under study. When one of these discrete phenotypes is similar to wild type, it is said that the genetic perturbation complements it. This is still true even if it does not complement all of the traits impacted by the gene, meaning, if it does not rescue. So, in a way, rescue can be thought of as successful complementation of all the phenotypes necessary for finishing development.

The two optogenetic platforms presented here hold promise to refine our understanding of activation and repression and their roles in development. Optogenetic Dorsal and Optogenetic Knirps are based on tools that have applicability beyond transcription factors. Other nuclear localized proteins play important yet poorly understood roles

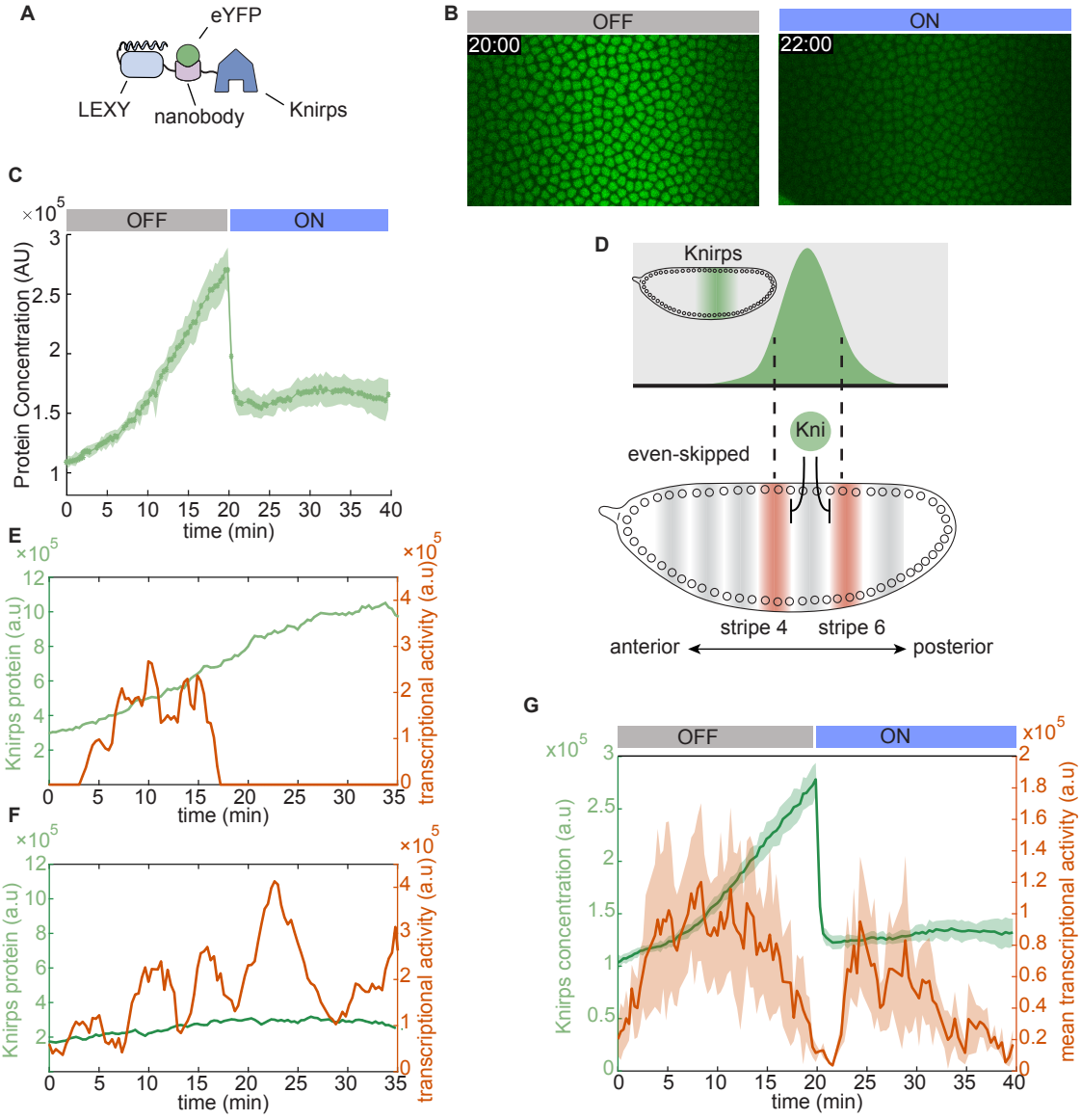

Figure 2.7: **Optogenetic Knirps regulates the activity of *eve*.** **(A)** Scheme showing the optogenetic Knirps protein. An anti GFP nanobody (LlamaTag) placed between LEXY and Knirps binds to mature eYFP, allowing the visualization of the protein in live embryos. **(B)** Snapshots of an embryo containing optogenetic Knirps and ubiquitous eYFP with and without 458 nm illumination. The time stamps correspond to the graph in (C). **(C)** Nuclear concentration dynamics of optogenetic Knirps upon illumination with blue light. **(D)** Schematic representation of the pattern of the Knirps protein in the embryo (top) and its role in repressing transcription of *eve* stripes 4 and 6 (bottom). ***Caption continues on next page.***

Figure 2.8: ***Continued from previous page*: Engineering an optogenetic Knirps repressor (E)**. Transcriptional activity of *eve* in the anterior boundary of stripe 6. Accumulation of Knirps protein shuts off transcription at around 20 minutes. **(F)** Transcriptional activity of *eve* in the middle of stripe 6. Knirps concentration is low, resulting in continued transcription. **(G)** Transcriptional effect of optogenetically removing Knirps from nuclei in the anterior end of *eve* stripe 6. Export of Knirps results in de-repression of transcription (compare with (E)).

in shaping gene expression temporal dynamics. In the future it will be interesting to apply the lessons learned here to chromatin modifiers such as Histone deacetylases and metylases.

2.5 Materials and Methods

2.5.1 CRISPR-Cas9 mediated homology directed recombination for generating knock-ins

To tag endogenous transcription factors with fluorophores and/or optogenetic tags, we used CRISPR-Cas9 mediated HDR (homology directed recombination). We designed and cloned the gRNA plasmids using the protocol described in `https://flycrispr.org` and followed the protocol by [107]. All injections were carried out by Rainbow transgenics .inc.

2.5.2 Microscopy setup

The laser lines used to excite LEXY, Venus and mCherry were 458, 520 and 587 nm respectively. The detection windows for Venus and mCherry were 530-577 and 597-650, respectively. Other imaging settings correspond to the standard in the Garcia laboratory and described in detail elsewhere [154, 72].

2.5.3 Rescue tests

To test the rescue of fly lines we crossed virgin females carrying optogenetic alleles of the gene of interest to yw males. After letting flies mate for 3 days in a cage, we flipped the agar lid of the cage and let them lay embryos in the dark for 2-3 hours. We then removed this plate from the cage and kept it in the dark for 24 hours at room temperature. This is enough time for embryos to develop to larval stage. Next we counted the number of larvae and the number of unhatched embryos on the plate. The reported rescue rate corresponds to the number of larvae divided by the number of larvae plus unhatched embryos.

2.5.4 Quantification of fluorescent protein concentration

The nuclear concentration of the fluorescent protein is quantified using an image analysis pipeline as described in [72] and Chapter 4. Briefly, the fluorescence is integrated over an area of 2 μm around the center of each segmented nucleus. The fluorescence is then averaged over nuclei in specified spatial bins for a given time point to get a reliable time series of transcription factor nuclear concentration.

Chapter 3

A minimal synthetic enhancer system to probe Transcriptional Regulation in Development

3.1 Abstract

How enhancers interpret the concentration of morphogen gradients to give rise to patterns of gene expression is a central question in animal developmental biology. Multiple models have been put forward to account for the regulation of transcription by morphogen transcription factors but we currently lack a unified model that can be experimentally tested. Here, based on previous efforts, we advance a unified model of transcriptional control by an activator transcription factor. This model posits that promoters can exist in a series of inactive states prior to RNA Polymerase loading, with transcriptional activators accelerating the transition between these states. Further, the model proposes that once the promoter reaches the active state, activators regulate transcription via a thermodynamic mechanism such that their equilibrium DNA occupancy dictates the rate of RNA Polymerase loading. In order to systematically test this model we we established a minimal synthetic enhancer system in *Drosophila melanogaster* embryos where the Dorsal transcriptional activator regulates the activity of a reporter driven by a single Dorsal binding site. Using state of the art live imaging techniques and perturbations to the Dorsal DNA binding site, we demonstrate that our unified model is consistent with the behavior of this minimal system. The minimal nature of our system sets the stage to further challenging our model and opens the door to build and understand more complex enhancers from the ground up.

3.2 Introduction

The adoption of distinct cellular identities in multicellular systems relies on the formation of spatial gene expression domains. The positional information giving rise to these mRNA patterns is provided by gradients of molecules known as morphogens whose concentration is interpreted by enhancer DNA sequences that, in turn, regulate transcription of developmental genes [272, 196]. Quantitatively predicting gene expression based on the sequence of regulatory DNA and morphogen concentration is a long-standing goal in quantitative developmental biology. This objective requires first the ability to quantitatively measure both morphogen concentration input and downstream gene transcription output [111]. In addition, these measurements need to be contrasted against predictions from theoretical models. Due to the ease of precise genetic manipulations and available imaging technology, the early embryo of the fruit fly *Drosophila melanogaster* (Drosophila) has historically served as the prime model system to study these input-output functions in development.

Previous efforts in Drosophila have been relatively successful at creating models that recapitulate native patterns of gene expression [135, 236, 231, 78]. However, these models tend to invoke a myriad of phenomenological parameters in order to fit the data, thus providing little molecular insight and not lending themselves to be rigorously falsified. This is due, in part, to the complexity of transcriptional regulation in higher eukaryotes. Developmental enhancers can contain multiple binding sites for different transcription factors, and these protein themselves are dynamic in space and time. Moreover, interactions between proteins binding to nearby sites are difficult to account for and very challenging to predict [264]. Including multiple binding sites with different affinities and their interactions leads to a combinatorial explosion of parameters [93, 95].

In bacteria and eukaryotic cells in culture, a successful approach to reduce the complexity of transcriptional regulatory systems to a level that is theoretically tractable from first principles has been engineering minimal synthetic enhancers [206, 208]. Here, a short synthetic DNA sequence containing 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 allows testing models of transcriptional regulation grounded on first molecular principles. We thus sought to establish a minimal synthetic enhancer system in Drosophila embryos.

Following in the footsteps of [78], we use the dorso-ventral patterning system as our endogenous platform for our synthetic efforts where the Dorsal transcription factor gradient specifies concentration-dependent domains of gene transcription. Using theory combined with novel live cell imaging approaches, we tested existing hypotheses of transcription factor readout. The minimal nature of our system allowed us to uncover the richness of the regulatory repertoire available to a single transcription factor molecule. We discovered that Dorsal acts pleiotropically on the promoter, affecting several aspects of the transcriptional process. Additionally, we found equilibrium statistical mechanical descriptions of the transcription cycle were lacking.

3.3 Results

3.3.1 Statistical mechanical model of a single activator binding site

Thermodynamic models [30, 29] constitute some of the simplest conceivable theoretical descriptions of transcriptional regulation by the Dorsal activator. While these models have proven successful in bacteria [206], 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 [72, 202, 77, 207, 153, 163, 13, 162, 119, 143, 164].

Following [30], [98], and [72], we assume the occupancy hypothesis that states that the loading rate of RNA polymerase onto the promoter, R, is proportional to the fraction time that RNAP molecules are bound to the promoter, p_{bound},

$$R = R_{max} \cdot p_{bound}, \tag{3.1}$$

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

In Figure 3.1A, we enumerate the states this simple system can be found in and their corresponding statistical weights [30]. Here, Dorsal binds to its target site with affinity dictated by its equilibrium dissociation constant K_D. While Dorsal is bound, it recruits RNA polymerase through a glue-like interaction described by the parameter ω. ω for an activator molecule is assumed to be > 1 so that higher Dorsal occupancy facilitates higher RNAP occupancy. Using these weights, we derive an expression for p_{bound} as a function of Dorsal concentration given by

$$p_{bound} = \frac{\frac{P}{K_P} + \frac{Dl}{K_D}\frac{P}{K_P}\omega}{1 + \frac{Dl}{K_D} + \frac{P}{K_P} + \frac{Dl}{K_D}\frac{P}{K_P}\omega}, \tag{3.2}$$

where P and Dl are the concentrations of RNAP and Dorsal in the nucleus, respectively, and K_P and K_D their corresponding equilibrium dissociation constants to their DNA binding sites. Finally, by substituting Equation 3.2 into Equation 3.1, we find

$$R = R_{max} \cdot \frac{\frac{P}{K_P} + \frac{Dl}{K_D}\frac{P}{K_P}\omega}{1 + \frac{Dl}{K_D} + \frac{P}{K_P} + \frac{Dl}{K_D}\frac{P}{K_P}\omega}. \tag{3.3}$$

Figure 3.1 illustrates how the several regulatory knobs in our model dictate the shape of R as a function of Dorsal concentration. In Figure 3.1C, we see that ω serves to control how sharply the the R curve rises from the minimum to the maximum output. Figure 3.1C shows that K_D controls the Dorsal concentration midpoint at which the output transitions from minimum output to maximum output. Finally, Figure 3.1D shows that R_{max} dictates the maximum output rate. With this model in hand, we turned to determining whether its predictions are borne out by experiment.

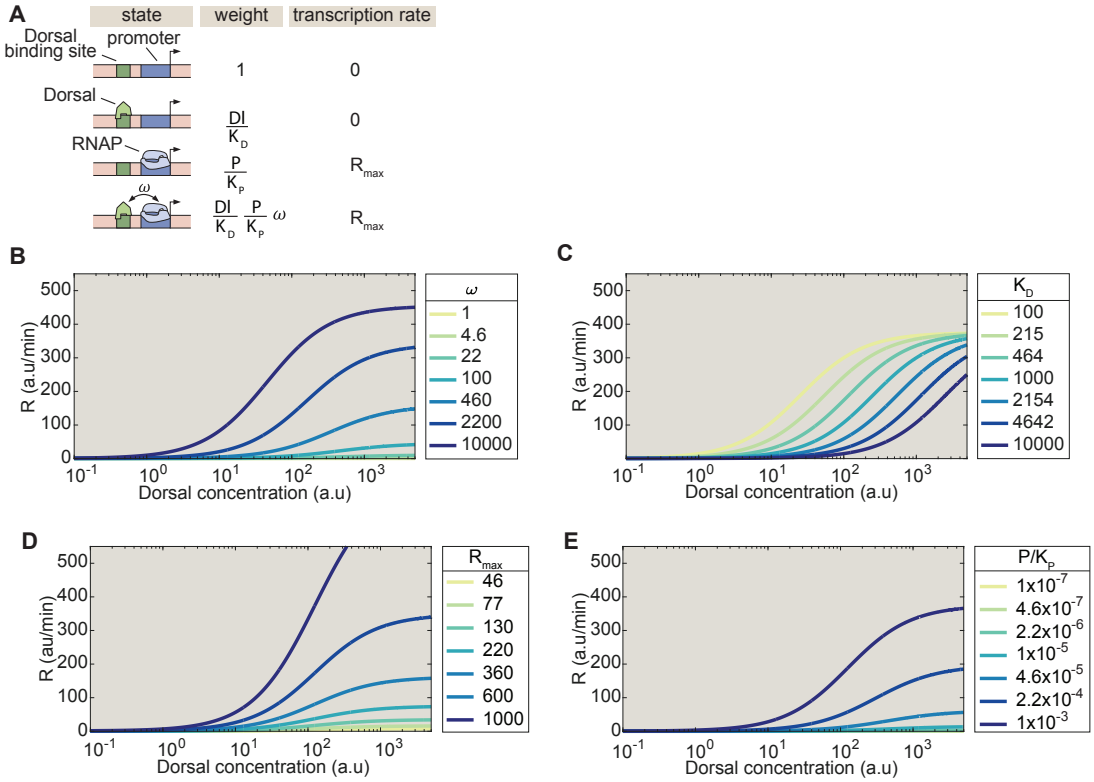

Figure 3.1: **Model of simple activation by Dorsal. (A)** Thermodynamic states and weights for the simple activator model. **(B-E)** Rate of mRNA production as a function of Dorsal concentration for different values of (B) the RNAP-Dorsal interaction term ω, (C) the Dorsal dissociation constant K_D, and (D) the maximum transcription rate R_{max} (E).

3.3.2 Single binding site measurements

To test our statistical mechanical predictions in the context of a developing embryo, we measured the transcriptional activity driven by a minimal synthetic enhancer containing a single binding site for the Dorsal transcription factor (Fig. 3.2A). In order to quantify transcriptional activity of this enhancer, we used the MS2-MCP system [98, 174]. Fusing Dorsal to mVenus [215] and MCP to mCherry allows us to simultaneously measure the Dorsal protein input and the nascent RNA output in individual nuclei (Fig. 3.2A) (see Materials and Methods for the genetic and crossing schemes used).

Dorsal is one of the best characterized transcription factors in Drosophila and a classic example of a morphogen [227, 215]. Dorsal is provided maternally and forms a dorso-ventral gradient of nuclear localization (Fig. 3.2B). Prior to the activation of the zygotic genome, the Dorsal gradient constitutes the only source of nuclear positional information along the dorso-ventral axis. During the 12th mitotic cycle zygotic proteins with dorso-ventral gradients are not present at appreciable levels [70, 230] making Dorsal

an ideal input for a minimal synthetic reporter system. Dorsal has been shown to act as an activator when bound to enhancers [256, 137], and as a repressor in the presence of nearby corepressors [147, 200]. Based on previous Dorsal fusions [215], we created a CRISPR knock-in Dorsal-mVenus fusion allele that rescues embryo development (see Materials and Methods). Dorsal-mVenus nuclear fluorescence time traces quantified over nuclear cycle 12 confirm the dynamic nature of Dorsal concentration and are quantiatively similar to previous measurments (Fig. 3.2C); [215]).

To validate this minimal transcriptional regulatory system, we first set out to show that our minimal synthetic enhancers could drive detectable levels of transcription, and that their activity is mainly determined by Dorsal and not other transcription factors. To this end, we compared transcription of our minimal construct in the presence of Dorsal with its expression in a dorsal null background lacking Dorsal protein. As shown in in Figure 3.2D, transcription spots are not present in dorsal null embryos, showing that Dorsal is necessary for transcriptional activity in our reporter constructs. We note that we did not detect a single transcription spot in the field of view during nuclear cycle 12 in any of 4 replicates, containing more than ≈ 60 nuclei in total. Further, we sought to demonstrate that activation is only due to Dorsal interacting with the binding site explicitly engineered into the construct and not due to cryptic Dorsal binding sites. To this end we measured transcriptional activity of an enhancer where the Dorsal binding site was strongly mutated. Transcription is hardly detectable in this construct, with the average transcriptional activity per nucleus being less than 10% that of the optimal enhancer across Dorsal concentration (Fig. 3.2D and Fig. 3.10). This demonstrates that the Dorsal site placed in the synthetic enhancer is necessary for activation.

To determine whether the expression levels from our synthetic construct were quantifiable and could be used to test predictions stemming from our theoretical model, we performed image analysis of the MS2 movies in Matlab using a custom data analysis pipeline ([154]; see Materials and Methods for additional details). In Figure 3.2E, we show example time traces of MCP-mCherry fluorescence at transcriptional loci during nuclear cycle 12 along with four different metrics we employ to quantify transcriptional activity: First, the maximum spot fluorescence corresponds to the 95th percentile of intensity over time and is assumed to be proportional to the transcription rate (see Appendix section 3.7.3 for explicit calculations). Second, the transcriptional onset time corresponds to the time since mitosis at which a transcription spot is first detected (see also). Third, the spot duration is calculated as the time since mitosis at which a spot was last detected minus its corresponding transcription onset time. Finally, the integrated spot fluorescence corresponds to the time integral of the fluorescence time trace and is directly proportional to the amount of mRNA produced by the locus [98]. (see Materials and Methods for a deeper discussion of how these magnitudes are calculated).

Note that, as was observed in other genes [98, 154, 24], not all nuclei exposed to the same Dorsal concentration exhibit detectable transcription. As a result, we defined the fraction of active nuclei as an additional metric. This magnitude is defined as the ration between the number of nuclei in which reporter transcription was detected in at least

one frame throughout nuclear cycle 12 and the total number of nuclei in the field of view (Fig. 3.2F).

3.3.3 Perturbing the Dorsal binding site affinity

Having established an experimental setup to test the predictions from the model in Figure 3.1, we set out to compare its predictions against our experimental data. To make this possible, we sought to systematically tune the physical parameters that dictate the transcription rate according to Equation 3.18 taking advantage of the endogenous Dorsal protein gradient to provide a natural titration curve.

One of the major predictions of the simple activation model in Section 3.3.1 is the dependence of the transcriptional rate on Dorsal on binding affinity. Specifically, our model (Fig. 3.1B-E) predicts that, if the affinity of the Dorsal binding site is experimentally perturbed, the only feature of the transcriptional output curve as a function of Dorsal concentration that will change is the concentration at which the promoter reaches half maximal transcription rate; the maximum transcription rate and slope of the transition are predicted to remain unaltered. Thus, tuning binding affinity constitutes a powerful means to test our model.

In order to test the hypotheses derived from the simple activator model, we constructed a series of enhancers containing a single binding site. Building on the optimal and mutated sites shown in Figure 3.2D, we created 5 more enhancers of varying strength by introducing point mutations to the consensus Dorsal binding motif. To guide the design of these binding sites, we used an already existing position weight matrix computed with the MEME algorithm [14] using motifs generated by DNAse I footprinting assays [23] and quantified the information content of each base pair using Patser [122], resulting in the scores and rank-ordering of binding sites shown in Figures 3.3A and B. Hereafter we refer to these enhancers as DBS for Dorsal Binding Site, followed by their corresponding Patser score. In this scheme, the optimal and mutated binding sites from Figure 3.2D are DBS_6.23 and DBS_4.29.

In Figure 3.3C, we show the integrated mRNA output over time of each enhancer as a function of Dorsal concentration across all nuclei exposed to a given Dorsal concentration. An appreciable trend across concentration is only observed for the three strongest affinities, while the integrated mRNA output driven by the weaker enhancers barely changes across the dorso-ventral axis. An alternative way to summarize our data is shown in the inset to Figure 3.3C. Here, we plot the total mRNA integrated across the entire embryo as a function of Patser score. This plot clearly demonstrates that affinity (as measured by Patser score) is strongly correlated with transcriptional output in single binding site enhancers, but that four of our binding sites are too weak to present any significant differences in the levels of transcriptional activity that they drive.

Next, we asked how the metrics of transcriptional activity described in Figure 3.2F and G are regulated by Dorsal concentration and binding affinity. In Figure 3.3D-H, we plot individual metrics of transcriptional activity for each enhancer. The fraction of active nu-

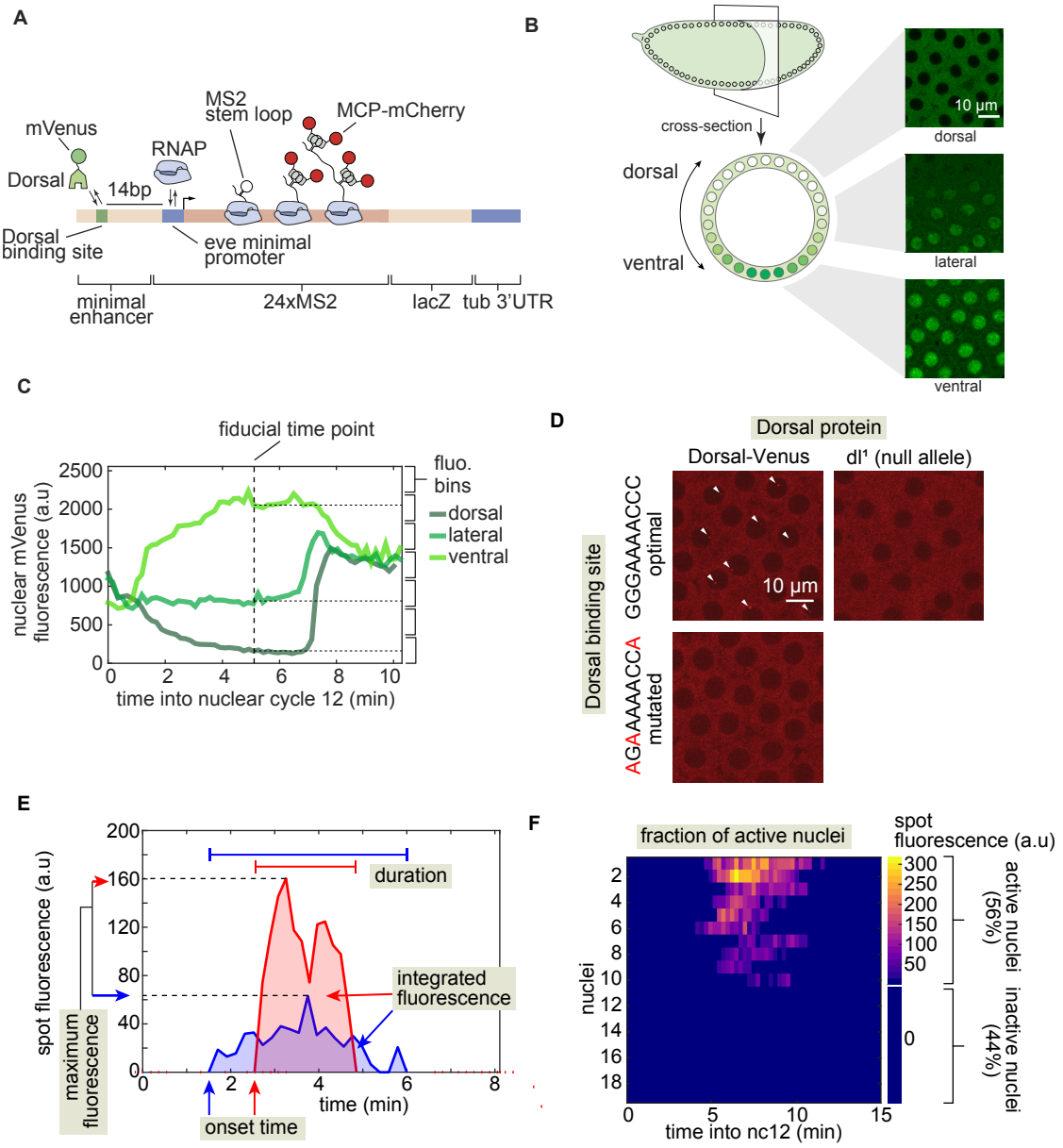

Figure 3.2: **Experimental setup to measure transcription factor protein input and transcriptional outputs. (A)** Schematic of the experimental system used in this study. Minimal synthetic enhancers containing a single binding site for Dorsal drive transcription of a reporter tagged with MS2. mVenus and mCherry are used to visualize the protein input and the transcriptional output, respectively. The Dorsal binding site is placed 14 bp upstream of the even-skipped minimal promoter. **(B)** Schematic of the Dorsal protein gradient in early *Drosophila* embryos. The Dorsal protein accumulates in ventral nuclei and is progressively excluded from more dorsal nuclei. and example snapshots showing Dorsal-Venus in different positions along the dorso-ventral axis. ***Caption continues on next page***.

Figure 3.2: ***Continued from previous page*: Experimental setup to measure transcription factor protein input and transcriptional outputs. (C)** Representative time traces of nuclear Dorsal-Venus fluorescence. To calculate transcriptional activity as a function of Dorsal protein, we sort nuclei into Dorsal concentration bins based on the the Dorsal-Venus fluorescence at a single fiducial time point. **(D)** Snapshots from embryos containing an optimal (DBS_6.23) binding site reporter in the presence or absence of the Dorsal activator, and with a mutated Dorsal binding site. Mutated bases are highlighted in red. White arrows point to sites of active transcription. **(E)** Example traces and quantitative metrics of transcriptional activity used throughout this work. **(F)** Diagram illustrating how the fraction of active nuclei is obtained. If we detect a spot at any point during the interphase of nuclear cycle 12, that nucleus is considered active, otherwise, we classify it as inactive. Representative images of active and inactive nuclei on the right. Scale bars in (F) are 5 μm.

clei (Fig. 3.3D) shows a clear strong correlation with Dorsal concentration, particularly for the stronger binding sites. The maximum spot fluorescence (Fig. 3.3E) and the integrated fluorescence (Fig. 3.3F) are only weakly regulated by Dorsal concentration in the highest affinities and insensitive to the Dorsal gradient in low affinity minimal enhancers (Fig. 3.3E and F). On the other hand, we identified metrics of transcriptional activity that were insensitive to Dorsal concentration in all enhancers. The mean transcriptional onset times (Fig. 3.3G) are essentially constant across Dorsal concentration while spot durations (Fig. 3.3H) are also relatively insensitive to Dorsal concentration in all enhancers.

Thus, we find that, in contrast to strong binding sites, changes in Dorsal concentration spanning more than two orders of magnitude, from $\approx$ 2 nM to more than 250 nM according to our calibration of Dorsal-Venus (Fig. 3.8), have little effect on activity in enhancers carrying weak sites (Fig. 3.3D-F). Importantly, this does not mean that these enhancers are insensitive to Dorsal. Indeed, they drive detectable transcription while the optimal DBS_6.23 enhancer in an embryo lacking Dorsal protein does not. Even when they appear to be insensitive to changes in Dorsal concentration, the affinity of these weaker enhancers is positively correlated with activity. This is made clear by averaging their activity over Dorsal concentration bins (Fig. 3.3J-N). Here we note that, fraction active (Fig. 3.3J), maximum spot fluorescence (Fig. 3.3K), and accumulated fluorescence (Fig. 3.3L) are clearly positively correlated with affinity. In contrast, mean transcriptional onset times (Fig. 3.3M) and spot durations (Fig. 3.3N) are relatively independent of affinity, consistent with the fact that they are not affected by Dorsal concentration even in strong enhancers. Hence, while weak sites are able to detect the presence of Dorsal in an affinity-dependent manner, they are not capable of interpreting the positional information encoded in the Dorsal gradient, even when the concentration range of this gradient is increased by doubling the genetic dosage (Figure 3.8).

These results reveal significant qualitative disagreements between theory and ex-

periment. First, note that our thermodynamic model cannot account for the fact that some nuclei will not transcribe throughout the nuclear cycle, and that the average rate of transcription of nuclei will be given by Equation 3.18. However, as shown in Figure 3.2E, only a fraction of nuclei become transcriptionally active during the nuclear cycle, and this fraction depends on the nuclear Dorsal concentration. Second, the simple activator model predicts that mean transcriptional onset times decrease monotonically with Dorsal concentration since these should depend on the first occurrence of Dorsal binding to its enhancer. In contrast, we found the onset times to be effectively constant.

While these disagreements cast doubt on the set of assumptions underlying our theoretical model, it was also important to demonstrate that the inconsistency between theory and experiment did not stem from experimental limitations such as weakly transcribing nuclei being classified as inactive due to detection artifacts.

3.3.4 Simple activator model could explain the maximum fluorescence observations.

Having established a set of minimal synthetic enhancers with varying affinity for Dorsal, we next sought to test the predictions of the simple activator model with regards to activator binding affinity. As noted above, one of the most straightforward expectations from this model is that the transcription rate increases with the activator binding K_D. More specifically, while the maximum possible transcription rate does not change with K_D at saturating levels of Dorsal concentration, the concentration at which the rate of transcription is half maximum decreases with decreasing K_D (Fig. 3.1C). The maximum spot fluorescence is directly related to the rate of transcription (Fig. 3.13).

As discussed in Appendix 3.7.3, we can approximately relate the initiation rate predicted by the simple activator model (Equation 3.2) to the maximum fluorescence by a constant factor (Equation 3.15). This allows us to directly compare theoretical predictions with experimental data. We performed a global fit of the maximum spot fluorescence data across all enhancers shown in Figure 3.3E to the simple activator model such that all enhancers share the exact same parameters except for K_D, which is left as a free parameter specific to each enhancer (Fig. 3.4A).

We found that the data agrees well with the model requiring all enhancers to share the same maximum transcription rate at saturating Dorsal, R. As shown in Figure 3.4A and B, this parameter is predicted with high confidence to be $\approx$ 250 a.u. On the other hand, we find that this model cannot account for the relatively constant rate of transcription across Dorsal concentration and enhancer affinities. The model predicts a hyperbolic relationship between Dorsal concentration and transcription rate, while our data suggests that these two values are largely independent (Figure 3.3E). As a result, the fitted K_Ds have an extremely low confidence, with the standard errors in the same order of the mean (Fig. 3.4B).

Thus, we conclude that the simple activator model cannot capture the behavior of

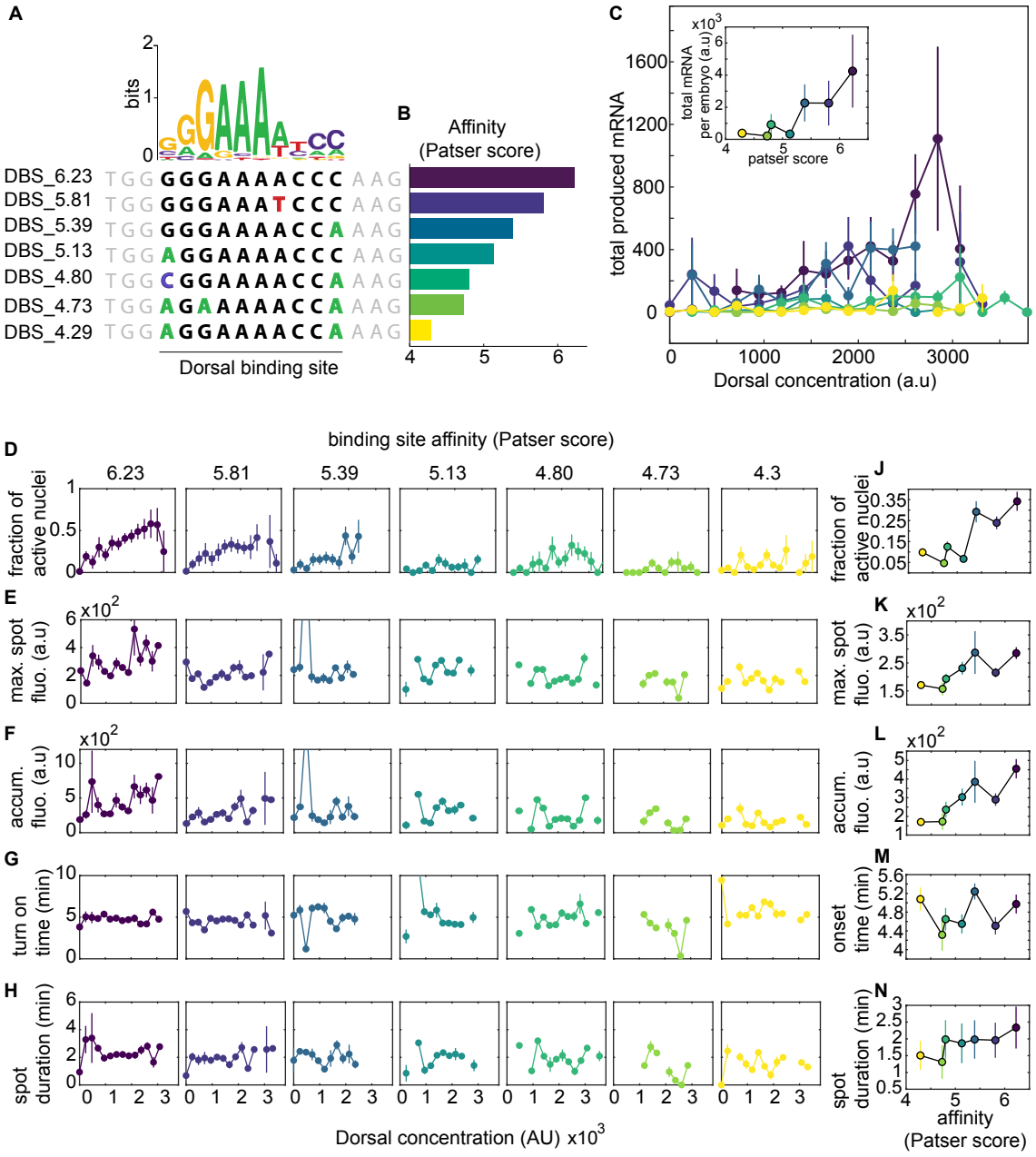

Figure 3.3: **Dorsal binding site affinity dictates several transcriptional dynamics features. (A)** Sequence of the Dorsal binding site in different minimal synthetic enhancers. Bold letters represent the 10 bp Dorsal motif. Black letters correspond to consensus bases, which are colored when mutated. Gray letters show the sequence context. **(B)** Affinities of different Dorsal binding sites estimated from the Patser algorithm using the Dorsal position weight matrix. ***Caption continues on next page***.

Figure 3.3: ***Continued from previous page*: Dorsal binding site affinity dictates several transcriptional dynamics features. (C)** Overall transcriptional activity driven by the enhancers in (A) measured as the total produced mRNA as a function of Dorsal concentration. The inset shows the mean total mRNA produced per embryo integrated across all Dorsal concentrations. **(D)** Mean fraction of active nuclei, **(E)** mean maximum spot fluorescence, **(F)** mean accumulated fluorescence across active nuclei, **(G)** mean time of first spot detection, and **(H)** fluorescence spot lifetime as a function of Dorsal concentration and binding site affinity. **(J)-(N)** For each of the metrics shown in (D)-(H), the mean across Dorsal fluorescence was calculated for each enhancer based on a moving average. (C)-(H), error bars correspond to the standard error of the mean calculated over at least 3 embryos; (J)-(N), error bars correspond to the standard deviation).

our regulatory system in terms of the transcription rate across active nuclei. This is in contrast to E.coli [97]. Because the regulation of the fraction of active nuclei is the main regulatory layer in response to Dorsal concentration and affinity, we turned our attention to theoretical models to capture this behavior.

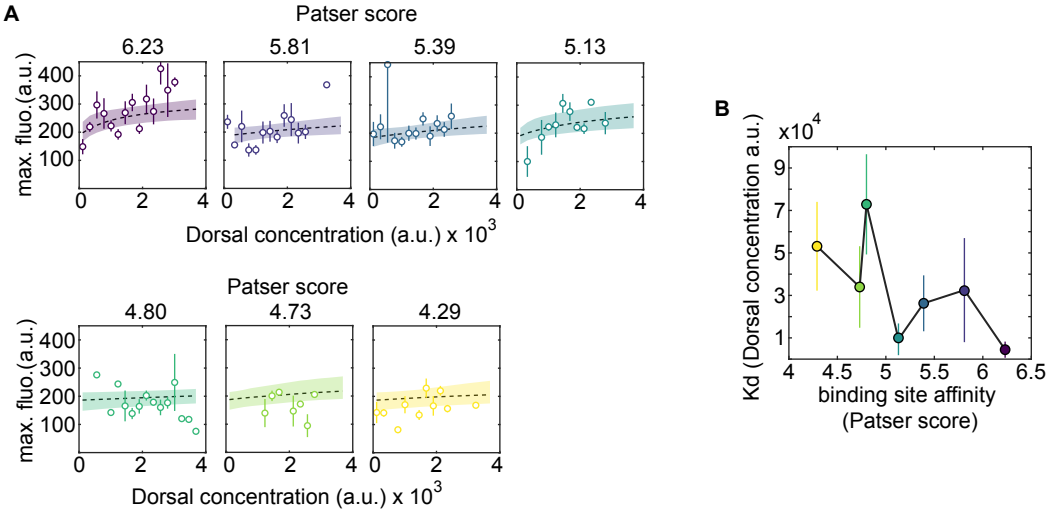

Figure 3.4: **Comparison of equilibrium model predictions with observed maximum fluorescence values. (A)** Prediction for maximum fluorescence as a function of Dorsal concentration for each affinity. Red circles are experimentally observed data with standard error of the mean. Orange curves are solutions obtained from a single simultaneous Markov Chain Monte Carlo curve fit. All solutions share all parameters except for K_D. Red dashed curves are 95% prediction intervals. **(B)** Inferred parameter values.

3.3.5 Dorsal regulates the partition of loci into two distinct states of activity giving rise to a fraction of inactive nuclei

The presence of transcriptionally silent nuclei that remain inactive throughout interphase has been observed using MS2 (and its sister PP7) in live imaging experiments in flies [98, 154, 24], plants [4], and mammalian cells [114]. As in our setup, the nascent RNA labeling system in these experiments did not have single-molecule resolution, it is thus conceivable that these nuclei correspond to false negatives transcribing below the fluorescence detection threshold.

We sought to determine whether nuclei that never transcribe (i.e. 'inactive nuclei'; Fig. 3.2F, Fig. 3.3D,J) are just below our detection limit or constitute a separate population transcribing at a Dorsal-independent, weaker level. In order to do this, it is necessary to compare the activity in the absence of Dorsal protein to that of inactive nuclei in the presence of Dorsal. Clearly, this experiment is not possible using only MS2 since the quantification of spots necessitates a detectable mCherry signal. Hence, we developed a method to quantify the MCP-mCherry fluorescence in all loci, independent of whether an MS2 spot was detected or not.

To realize this experiment, we turned to the ParB-ParS DNA labelling system [100, 49]. In this system, fluorescently labelled ParB proteins bind to the ParS DNA sequence, resulting in a fluorescence spot at the site of the locus (Fig. 3.5A). We created flies with and without functional Dorsal expressing ParB2-eGFP and MCP-mCherry to label our locus DNA and nascent RNA, respectively. In addition, we added a ParS sequence followed by a 400 bp spacer to our DBS_6.23 enhancer. We then crossed flies containing ParS-DBS_6.23-MS2 to flies carrying ParB2-eGFP and MCP-mCherry to create embryos that have our locus of interest labelled with eGFP colocalized with transcriptional loci in the MCP-mCherry channel (Fig. 3.5A and B).

Guided by the positions reported by ParB-eGFP, we measured the mCherry signal at all DBS_6.23 reporter loci in embryos carrying wild type Dorsal and the dl^1 null allele (Fig. 3.5C). As shown in Figure 3.5B, we classified loci into two categories, detected and undetected, depending on whether they were identified as spots in the MCP-mCherry channel.

In Figure 3.5D we show example fluorescence traces of mCherry corresponding to a transcriptional event detectable as a spot in both the ParB-GFP and MCP-mCherry channel and an event only visible as a spot in the ParB-eGFP channel. We assigned a fluorescence value to each locus corresponding to the 95th percentile of intensity over its mCherry fluorescence dynamics. As a result from this analysis, we obtain three populations: all loci in Dorsal null embryos, undetected loci in wild-type Dorsal embryos and detected loci in wild-type Dorsal embryos. The values of mCherry fluorescence at the reporter locus of these three populations are shown in Figure 3.5E.

We find that the mCherry fluorescence of undetected spots in wild-type Dorsal embryos is identical to that of all spots in Dorsal null embryos (Fig. 3.5E). Moreover, this distribution is distinct from that of detected spots in wild-type Dorsal embryos. This is

strong evidence that nuclei exposed to a similar Dorsal concentration can belong to two distinct populations: those that transcribe at a high, Dorsal-dependent level and those that are transcriptionally inactive or active at a low, Dorsal-independent level.

Thus, we conclude that the presence of a fraction of active nuclei is not simply a result of our spot detection limit. This shows that the decision made at each locus to transcribe is an additional regulatory feature dictated by the Dorsal activator, separate from its role in modulating transcription rate once loci become transcriptionally active.

As noted in Section 3.3.3, equilibrium-based models of transcriptional activation such as the one captured by Equation 3.18 cannot explain the presence of these two populations, nor how they are regulated by Dorsal concentration. Thus, we view the results presented in this section as a clear demonstration of the inability of equilibrium models to recapitulate the full transcriptional dynamics of this minimal synthetic system. As a result, in the next section, we turned to kinetic models of transcription to explore alternative mechanisms that could explain how inactive populations might arise.

3.3.6 Dorsal-dependent kinetic barriers can explain the transcription onset dynamics and the fraction of active nuclei

As shown in the previous section, most inactive nuclei are not an artifact of the detection limit of our experimental setup. However, our equilibrium model from Section 3.3.1 cannot account for the fact that only some nuclei become transcriptionally active throughout the nuclear cycle. In addition, we also note that the this model cannot recapitulate the constant transcriptional onset time across Dorsal concentration and affinity (Fig. 3.3G). Hence, we sought to revise our theoretical model in order to account for the presence of active and inactive nuclei and for the uniform transcriptional onset time across Dorsal concentrations.

To posit a feasible mechanism behind the presence of a Dorsal-dependent fraction of inactive nuclei and a constant transcriptional onset time, we turned our attention to kinetic models that invoke multi-step transitions for the initiation of transcription. Specifically, inspired by recent modeling efforts in *Drosophila* by [69] and [72], we proposed a 'kinetic barrier' model (Fig. 3.6A). Here, after exiting mitosis, all promoters are in an inactive state where transcription is not possible (labeled 'OFF$_1$' in Fig. 3.6A). Promoters must then traverse a series of n distinct inactive states (labeled 'OFF$_2$' to 'OFF$_n$' in Fig. 3.6A) before reaching an active state in which transcription proceeds (labeled 'ON' in Fig. 3.6A). The transition between states is modelled as a stochastic irreversible one-step reaction characterized by a rate π. Although we make no assumptions about the molecular identity of these kinetic barriers, they can be identified with, for example, the step-wise opening of previously closed, inaccessible chromatin [72, 77, 143, 161, 202].

The necessity of individual nuclei to traverse a set of intermediate states before transcription can ensue results in a delay in activation in the nucleus; if given enough time,

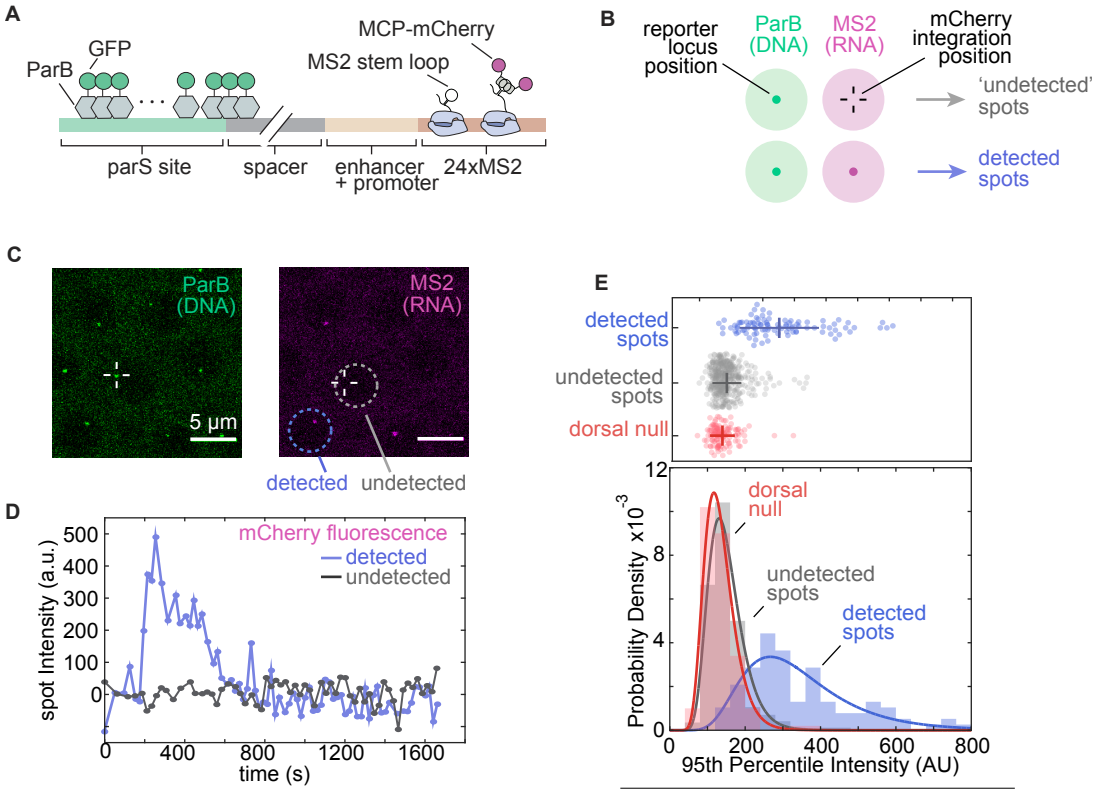

Figure 3.5: **Transcriptionally-independent ParB labeling confirms the presence of active and inactive nuclei.** **(A)** Cartoon schematic of our ParB-eGFP construct. **(B)** Schematic illustration of data analysis in this experiment. The gene locus was located by detecting a spot in the ParB-eGFP channel. The positional information for the ParB-eGFP spot was used to fit a 2D Gaussian to the same x-y region in the MS2-mCherry channel to estimate background signal, regardless of whether an MS2-mCherry signal was detected. **(C)** Example images from the experiment showing the ParB-eGFP and MCP-mCherry channels. Detected and undetected transcriptionally active nuclei solely, based on the MCP-mCherry signal alone, are shown. **(D)** Example time traces showing the MS2-mCherry fluorescence over time at the ParB-GFP loci in nuclei with (blue) and without (grey) detected MS2-mCherry spots of the DBS_6.23 enhancer. ***Caption continues on next page***.

Figure 3.5: ***Continued from previous page*: Transcriptionally-independent ParB labeling confirms the presence of active and inactive nuclei. (E)** Top: swarm plots showing the maximum MS2-mCherry fluorescence over time at loci with detected (blue) and undetected MS2-mCherry transcription (grey) in the DBS_6.23 enhancer. The maximum fluorescence is defined as the 95^{th} percentile of intensity over time such that the top 5% brightest spot intensities were included. Bars correspond to the mean $\pm$ standard deviation. Bottom: Histograms of the data on top. Solid lines correspond to log-normal fits. Image acquisition was done sequentially for GFP and mCherry to avoid fluorescence bleed-through.

all promoters would eventually reach the ON state. Yet, if the time it takes a promoter to reach the ON state is longer than the time during which initiation is permissible (during interphase), this promoter will not have enough time to turn on at all. A simulation of the transcriptional dynamics predicted by this model for individual nuclei is shown in Figure 3.6B.

Since we observe a very strong dependence of the fraction of active nuclei on binding site affinity and Dorsal concentration (Fig. 3.3D and J), we assume that π—the probability per unit time that the promoter will transition from one state to the next—is determined by Dorsal concentration and binding affinity K_D. One of the simplest functional forms for such dependence can be given by the equilibrium occupancy of our activator on the promoter, such that

$$\pi = c \cdot \frac{\frac{[Dl](t)}{K_D}}{1 + \frac{[Dl](t)}{K_D}}, \tag{3.4}$$

where c is a rate constant and $[Dl](t)$ is the Dorsal concentration at time t since the previous anaphase, obtained from the Dorsal-Venus time traces.

Next, we sought to explore the fraction of active nuclei and mean transcription onset times predicted by this model. To this end, we performed numerical simulations of the probability of nuclei being in each state as a function of time (details of this simulation can be found in Appendix 3.7.2). One free parameter in our model is the total duration of time when initiation is permissible. We decided to fix this value by looking at when we last see spots turning on within our data. To determine the end of the window during which initiation occurs, we calculated the 95th percentile of the observed spot onset times across all data sets from all affinities, resulting in ≈7.1 minutes after the previous anaphase (Fig. 3.6C, top). An example result of this simulation can be seen in Figure 3.6C, bottom. Here, the probability of a nucleus being in the final 'ON' state increases over time to eventually reach 1. Based on our experimentally observed onset times, we posit that there is a discrete window of time during which nuclei can become transcriptionally active (Figure 3.6C, top), since we do not see nuclei turning on past ≈ 7.5 minutes. As a result, nuclei that would transition to the 'ON' state past this permissible window never get the chance to do so and are counted as inactive. Thus, the combination of delayed

turn on times and a discrete permissible window to turn on result in only a fraction of nuclei ever transcribing.

To simultaneously explore the whole parameter space of this model and determine if it can recapitulate the data, we performed a simultaneous fit to the fraction of active nuclei and mean transcription onset times from Figure 3.3 across all enhancers. Here, we forced all enhancers to share the exact same value for all parameters except for the Dorsal binding K_D. We found that a model with 4 'OFF' states is capable of capturing the qualitative behavior of our observations: a Dorsal and affinity dependent fraction of active nuclei and a mean turn on time that is mostly constant across Dorsal concentrations and affinities (dashed lines in Figure 3.6D).

Thus, we conclude that a model where Dorsal accelerates the transition through a number of kinetic barriers through affinity-dependent binding can explain our observations.

3.4 Discussion

In this paper, we have demonstrated that there is enough information in a single transcription factor binding site to specify a developmental boundary. This result should not be taken lightly since previous theoretical results have pointed out that the opposite should be the case [276]. In addition, a single transcription factor molecule binding near the promoter is able to provide positional information that determines several transcriptional features such as the switch to a transcriptionally active state and the transcription rate once in that state. Single base pair mutations to the Dorsal binding site are sufficient to simultaneously alter these and other parameters. This single base pair pleiotropy has been reported in recent quantitative studies of developmental enhancers [90].

While equilibrium thermodynamic models have been successful in modelling gene regulation in prokaryotes, they have been found lacking in eukaryotes. Nonetheless, these are a natural starting point for our attempts to model the behavior of our reporter. In essence, the equilibrium model is our null hypothesis against which any more complicated models must be contrasted. It is intuitively clear that an equilibrium thermodynamic model can not explain the fraction of active nuclei and constant onset times. The equilibrium model cannot explain onset times because, in this model, all nuclei start transcribing nearly immediately after the cycle begins once a Dorsal molecule finds its binding sequence. In other words, in order for only a fraction of nuclei to become active in a purely equilibrium model, Dorsal would need to have a binding rate, $k_o n$ on the order of the duration of the nuclear cycle. This, then, would not even be an equilibrium model since the equilibration period would be as long as our timescale of observation. The fraction of active nuclei suffers a similar problem- the equilibrium model predicts that *all* nuclei will transcribe at the same *average* rate, not that nuclei would be divided into two distinct populations: active and inactive nuclei.

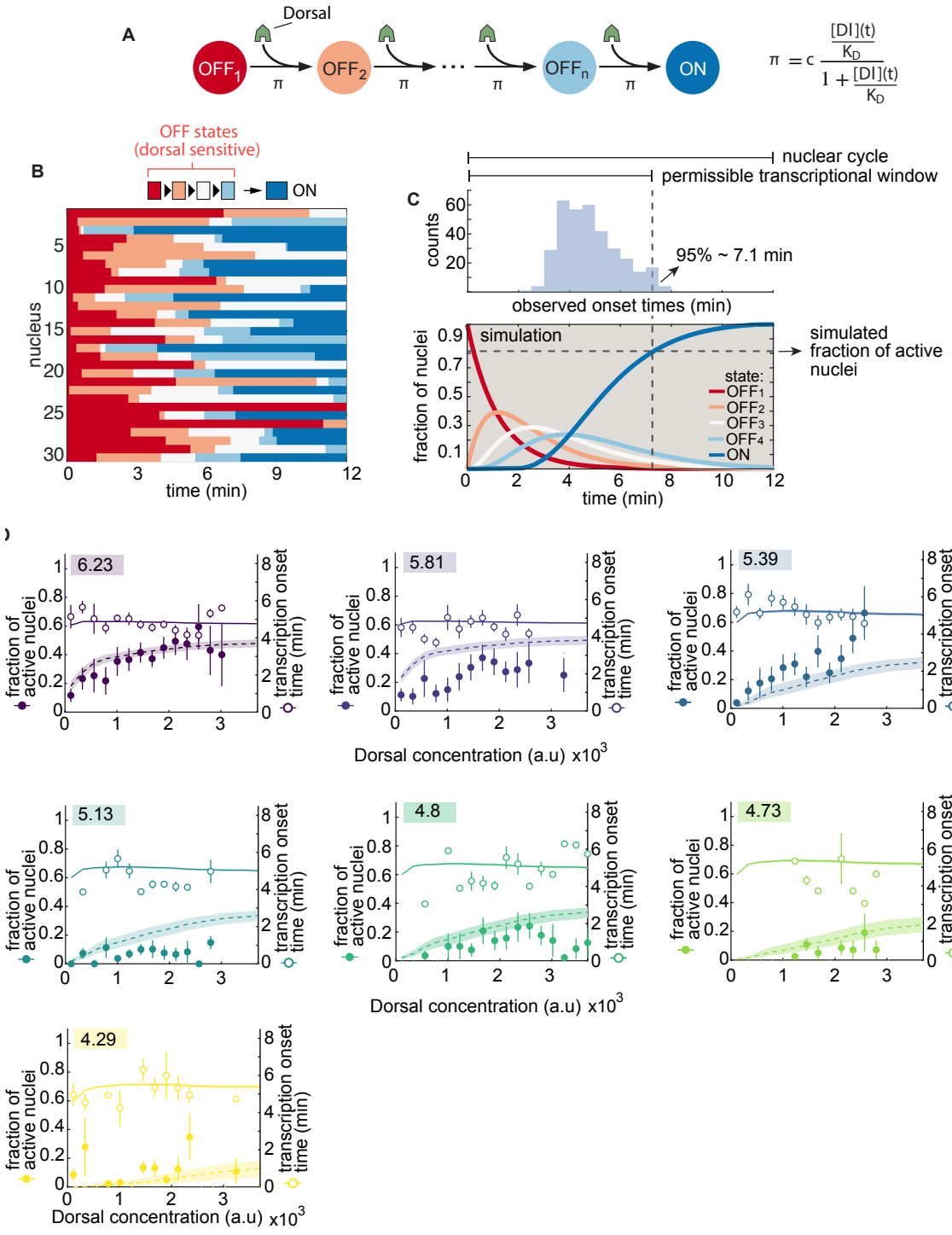

Figure 3.6: **A multi-step kinetic barrier model can explain a Dorsal-dependent fraction of active nuclei with constant mean transcriptional onset times.** ***See caption on next page.***

Figure 3.6: ***Continued from previous page:*** **A multi-step kinetic barrier model can explain a Dorsal-dependent fraction of active nuclei with constant mean transcriptional onset times. (A)** Model where the promoter undergoes kinetic transitions from transcriptionally inactive states (OFF_1 to OFF_n) to an active state (ON). π is the transition rate between states and is proportional to Dorsal occupancy at the promoter. **(B)** Visualization of a simulation of the model in (A) showing trajectories through states for one set of parameters (Dl = 1000 a.u., K_D = 1000 a.u., c = 10/min). Each row corresponds to an individual nucleus, columns correspond to time steps. The color indicates the state each nucleus is at at a given time according to the legend on top. **(C)** Top: distribution of mean spot detection times per embryo across all bins of Dorsal fluorescence and enhancers in Figure 3.3. The horizontal dashed line marks the time at which 95% of spots have turned on. Bottom: simulated fractions of nuclei in each of the states in (A) for a model with 4 OFF states. The fraction of nuclei in the ON state at $\approx$ 7.1 min corresponds to the simulated fraction of active nuclei. **(D)** Model fits for the fraction of active nuclei (plotted against the right y-axis) and mean transcriptional onset times (plotted in red against the left y-axis) for all affinities. Dashed lines correspond to the median of the posterior distribution. The shaded areas indicate the 25%-75% prediction interval. Open blue circles correspond to the experimentally observed mean onset time as in Figure 3.3G. Closed circles correspond to experimentally observed mean fraction of active nuclei as in Figure 3.3D.

However, because live cell imaging techniques typically lack single molecule resolution, it was unclear whether these undetected nuclei correspond to a distinct population or are simply a detection artifact. Here, by simultaneously labeling DNA and nascent RNA, we were able to demonstrate that the fraction of active nuclei is a biological phenomenon and not a shortcoming of our experimental techniques.

For all enhancers tested, we found that the fraction of active nuclei and mRNA production rates were highly modulated by Dorsal concentration, while transcriptional onset times and durations were not. The insensitivity of transcriptional onset times to activator concentration in early embryonic enhancers has been reported before in the P2P promoter [72], so this is not a surprise.

Intriguingly, weak binding sites were enough to drive Dorsal-dependent transcription, albeit with little to no concentration dependence across range of Dorsal concentration spanning three orders of magnitude. This points to a role for binding affinity in determining how much positional information a morphogen gradient can provide. It would be interesting to know if these weak enhancers with "flat" expression can become modulated in the presence of higher Dorsal concentrations that those explored in this study. In this manuscript, we used a reporter with two times the Dorsal dosage of wild type, but going as high as 3-4 times would help resolve some lingering questions about the behavior of low affinity enhancers. For example, by imaging the low affinity constructs at very high Dorsal concentrations, we may see them become more active as

they pass their K_D values.

Our initial attempt at using an existing model, the kinetic barrier model of [72] didn't work. Firstly, transition rates had to be modified to be functions of K_D, instead of just functions of $[activator]$ as in their model. Even with this modification, it was found that this model could not simultaneously recapitulate our observations of fraction of active nuclei and transcriptional onset times. This is due to the fact that in this model, higher onset times necessarily lead to lower fraction active. This tight coupling made it impossible for the model to explain a constant onset time with a fraction active that spanned the full range between 0 and 1.

To successfully explain the dynamics in our model, we advanced a theoretical model, the window model, which constrains the readout of Dorsal concentration to a relatively short widow of time during interphase. To achieve this we invoked repression from the previous mitosis and the existence of an irreversible and constitutive inactivating process that acts in opposition to Dorsal activation.

To falsify a dynamic model like this, it will be necessary to employ dynamic perturbation tools such as optogenetic control of Dorsal concentration. According to this model, if Dorsal were absent or inactive during the period when the enhancer is not sensitive to it (i.e for the first few minutes of the nuclear cycle) the fraction of active nuclei would be relatively unchanged. In contrast, there should be a discrete time window during which perturbing the Dorsal concentration has a drastic effect on the probability that nuclei turn on the reporter. Several optogenetics systems have been successfully deployed in the early fly embryo to inactivate transcription factors during discrete time widows [179, 129, 133]. In the future, a version of one of these systems may be capable of dissecting how the temporal dynamics of Dorsal affect transcriptional activation.

A better understanding of how the chromatin environment affects transcription levels, timing, and fraction of active nuclei, and how to make precise perturbations of local chromatin would greatly help in designing synthetic enhancers and with interpreting the results of the experiments. Measuring the chromatin state and state of histone modifications in individual loci remains an unmet goal in live cell imaging.

Going forward, minimal enhancers could be the foundation for exploring the behavior of more complex regulatory regions. Concretely, parameters such as binding constants can help constrain models of Dorsal interacting with other transcription factors, such as those studied by [78, 231]. Dorsal is the sole maternal input specifying dorsoventral position in *Drosophila*, but Dorsal rarely acts alone in endogenous enhancers [126]. The interaction of Dorsal with Twist is a classic example of positive cooperativity in development [251]. Dorsal can also act as a repressor depending on the presence of nearby Capicua binding sites [242]. 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.

In this study, we pushed the technical limits of what experiments are possible in a developing system. These efforts highlight the main technical limitations that still remain. Perhaps the most glaring limitation is the low throughput nature of the measurements.

This limitation stems from two sources: generation of transgenic embryos and the need to continuously image the sample over $\approx$ 20 minutes of development. The first of these issues could perhaps be remedied by using a higher throughput approach to generating a library of reporter embryos with varied enhancer regions such as those used in e.coli and yeast [145]. The second issue could be partially improved by switching to fixed tissue measurements and using an automated imaging setup [90, 102]. However, these snapshot measurements would be devoid of the dynamical information presented here that was fundamental in constraining theoretical models of Dorsal action. For example, when using fixed samples it would be impossible to determine if transcriptionally inactive nuclei represent a transient or a stable state [98]. An alternative higher throughput, dynamic method would perhaps need to be sequencing based. A variety of approaches have recently been developed to store dynamic information in RNA that can then be retrieved in a single sequencing run. The idea here is to have a constitutive editing process in the cell that chemically modifies RNA at a constant rate. The extent of RNA modification then depends on the time an RNA molecule has been exposed to editing. The distribution of edits in a population of RNA molecules in a cell can then be used to reconstruct the temporal dynamics of transcription in that cell prior to the experiment [210, 220]. These high throughput approaches could replace live imaging in many applications.

Better microscopy techniques and fluorescent imaging tools would improve temporal and molecular resolution of the experiment, increasing our ability to measure transcriptional activity. For instance, newer generation microscopes like the lattice light-sheet microscope [48, 185] could improve signal to noise and enlarge the field of view of the measurement. Better fluorescent fusions that take advantage of, perhaps, organic dyes injected into the embryo could potentially lead to brighter, non-bleaching reporter [59]. On the MS2 front, there are many potential improvements. A greater number of stem loops in the reporter like the 128 used by [253] would in principle make each transcription spot brighter and allow us to see it for a longer duration. Additionally, MCP proteins that bind to MS2 with higher affinity could help reduce the signal from background MCP fusions in the nucleoplasm, such as the tandem-MCP system [274]. Techniques to improve the fluorescence of individual MCP molecules such as SunTag [252] could improve resolution of single polymerases. In addition, conducting all transcription measurements with DNA loci labelled by ParB would improve the ability to follow transcription spots that have low activity. Further, speculatively, ParB may allow us to correlate the activity of enhancers with how "open" the local chromatin environment is, as measured by the intensity of the ParB signal. Moreover, because the nucleus is not a homogeneous environment, measuring the Dorsal concentration specifically around the locus could collapse some of the variability we observe between loci exposed to the same bulk, average nuclear concentration.

3.5 Materials and Methods

3.5.1 Measuring Dorsal-Venus concentration

Dorsal-Venus concentration was calculated as in (Figure 3.7). We measured the average Venus 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 measured the concentration at the plane corresponding to the fitted parabola's vertex. (Figure 3.7B). We then plotted this value over time and selected a single time point for each trace corresponding to the middle of each nucleus's lifetime (Figure 3.7C).

3.5.2 Measuring transcriptional onset times

We measured time to turn on as the first time point where a spot was detected. Thus, we needed a reliable way to estimate the beginning of the nuclear cycle. Typically, fluorescently labeled histones are used to determine the timing of anaphase, however, only a small fraction of our embryos had measurable levels of visible Histone-iRFP. When the Histone-iRFP signal was not good enough to determine anaphase, we relied on the Dorsal-Venus channel. This is possible because, like Histone-RFP, the nuclear Dorsal fluorescence also shows a characteristic pattern during mitosis. To precisely determine what features of the Dorsal-Venus channel to use for mitosis timing, we imaged Dl-Venus and Histone-RFP simultaneously (Fig.3.11). This exercise showed that the edges of nuclei become fuzzier as they enter mitosis and then elongate during anaphase (Fig.3.11). In this way, we could find anaphase frames in movies where no visible Histone-iRFP was present. Despite using this method, we still estimate that there may be a 2-3 frame error (i.e 20-30 s) in our anaphase estimate. Given the short length of nuclear cycle 12, half a minute error is non-negligible. Fortunately, the lack of a systematic trend in turn on times across K_D or Dorsal concentration (Figure 3.6F and H)renders this point irrelevant. Moreover, because this method depends on the maternal input and not on the transcriptional output, this error does not have any systematic correlation with enhancer identity.

3.5.3 Reporter design

To design our minimal construct (Figure 3.2, we placed the the 10 bp consensus Dorsal binding site upstream of the Eve core promoter.This enhancer-promoter construct drives the expression of MS2V5x24-lacZ-tub3'UTR [98, 262].

3.5.4 Plasmids

Minimal synthetic enhancers were synthesized from complementary oligos obtained from IDT. Reporter plasmids were then constructed by either by Gibson assembly or by Genscript inc. The sequences of all plasmids generated in this study were deposited in a Benchling folder.

3.5.5 Fly lines

Enhancer constructs were injected into the 38F1 site on chromosome 2 of *yw* flies by Bestgene Inc. or Rainbow Transgenic Flies, Inc. Since recombination mediated cassette exchange (RMCE) [18] permits integration into one of two orientations in the genome, we selected a single orientation for each line to stay consistent. To prepare cages, *yw;Dl-Venus (CRISPR), MCP-mCherry; Dorsal-Venus, MCP-mCherry, His2Av-iRFP* (unless otherwise stated) virgins were crossed to *yw; enhancer-MS2;+* males. Flies were allowed to lay for 90-120 minutes, then embryos were mounted for confocal imaging.

We used *yw; Dl-Venus,MCP-mCherry; Dl-Venus, MCP-mCherry, His-2AV moms*. The majority of our data was taken from embryos with moms that had 2 copies of the Dorsal gene. The reason for this is that we needed the extra dynamic range of Dorsal concentrations to better see the titration curve of the different enhancers we used in the study. In addition, we have data demonstrating that the version with 2 copies actually more closely matches the activity of a fly with wild type Dorsal than the mom with 1 copy (Figure **??**). We explain this by inferring that the Dorsal-Venus transgene is a weaker activator than wild type Dorsal. The copy on 2 was a Dorsal-Venus CRISPR insertion we made and the copy on 3 was a transgene from [215]. There was a small difference in the sequence of each transgene. The CRISPR contains the full length Dorsal gene, while the transgene is missing a few amino acids from the N-terminus. Despite this difference, there does not seem to be a difference in activity between the CRISPR knock-in and the transgene Dorsal-Venus alleles (Fig. 3.9).

3.5.6 Microscopy

Embryos were mounted on slides as described in [98, 37, 96]. Confocal microscopy was performed on the Leica SP8 with HyD detectors. We used a 63x oil objective and scanned bidirectionally. Time-lapse z-stacks were collected with $\sim$10 s frame rate and 106 nm xy pixel dimensions and .5 μm separation between z-slices (7 μM range, 16 slices). XY resolution 512x512 pixels. Pinhole was set to 1.0 Airy units at 600 nm. The 510 nm laser was calibrated to 5 μW and detected in a 520-567 nm spectral window. The 585 nm laser was calibrated to 25 μW and detected in a 597-660 nm spectral window. The 700 nm laser was set to 10% and detected in a 700-799 nm spectral window. All in counting mode. Frame rate was 10.2 s per stack. We used 1x line and frame accumulation with a scan speed of 420 Hz and a magnification of 3.4x zoom. All movies were taken at $\sim$50%

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

3.5.7 Image and time-series analysis

Image analysis was performed in Matlab using the custom pipeline described in [98, 154]. Image segmentation was also aided by the Trainable Weka Segmentation plugin in FIJI [10]. Further analysis of time-series and other data were likewise performed in Matlab.

Movies for publication were made in FIJI [232, 233].

3.6 Acknowledgments

We are grateful to Jacques Bothma, Rob Phillips, Michael Eisen, Nipam Patel, Xavier Darzacq,for their guidance and comments on our manuscript. We thank Greg Reeves for providing the Dorsal-Venus and dl^1 fly line. This work was supported by the Burroughs Wellcome Fund Career Award at 595 the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl and Kay Curci Foundation, the Hellman Foundation, the NIH Directors New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236) (HGG), an NSF GRFP (DGE 1752814) (AR).

3.7 Supplementary Material

3.7.1 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 ParB2 binding sites
Dl-Venus-dsRed	Donor plasmid for Dorsal-Venus CRISPR knock-in fusion
pU6-DlgRNA1	synthetic guide RNA for Dorsal-Venus CRISPR knock-in fusion
pBPhi-eNosx2-pTrans-NoNLS-MCP-mCherry-tub3'UTR	maternally deposited MCP-mCherry
pCasper4-His2Av-iRFP	Histone fusion to infrared RFP
pCasper4-Pnos-NoNLS-MCP-mCherry-TUB3'UTR	maternally deposited MCP-mCherry
pCasper-pNos-NoNLS-ParB2-GFP-TUB3'UTR	ParB-eGFP

Table 3.1: List of plasmids used in this study.

Fly lines	
Genotype	**Usage**
yw; parB-GFP; eNosx2-MCP-mCherry; +	Label reporter DNA and nascent RNA
yw; Dorsal-Venus, pNos-MCP-mCherry; pNos-MCP-mCherry, His2Av-iRFP	Females to visualize Dorsal protein, label nascent RNA, label nuclei
yw; Dorsal-Venus, pNos-MCP-mCherry; Dorsal-Venus, 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 visualize label nascent RNA 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 2 ParB2 binding sites

Table 3.2: Fly lines used in this study

3.7.2 Calculations

Kinetic model simulations

The problem presented in Figure 3.6A, namely that the time evolution of the probability of nuclei occupying a discrete number of consecutive steps, can be described by the following system of linear differential equations (also known as the 'master equation')

$$\frac{d\overrightarrow{P}}{dt} = \mathbf{\Pi}(t)\overrightarrow{P}, \tag{3.5}$$

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. $\boldsymbol{Pi}$ 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 $\pi(t)$, Equation 3.5 can be written as

$$\begin{bmatrix} \frac{dP(OFF_1,t)}{dt} \\ \frac{dP(OFF_2,t)}{dt} \\ \dots \\ \frac{dP(OFF_n,t))}{dt} \\ \frac{dP(ON,t)}{dt} \end{bmatrix} = \begin{bmatrix} -\pi(t) & 0 & \dots & 0 & 0 \\ \pi(t) & -\pi(t) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\pi(t) & 0 \\ 0 & 0 & \dots & \pi(t) & 0 \end{bmatrix} \times \begin{bmatrix} P(OFF_1,t) \\ P(OFF_2,t) \\ \dots \\ P(OFF_n,t) \\ P(ON,t) \end{bmatrix}, \tag{3.6}$$

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

Note that, as described for Figure 3.6, Π is itself a function of time since the transition rate π depends on the Dorsal concentration, which changes with time and is given by

$$\pi(t) = c \cdot \frac{\frac{[Dl](t)}{K_D}}{1 + \frac{[Dl](t)}{K_D}}, \tag{3.7}$$

where K_D is the Dorsal binding equilibrium constant and c is a rate constant.

Because $\pi(t)$ depends on the empirical Dorsal-Venus fluorescence, it does not have a concrete functional form. This makes solving the system in Equation 3.6 analytically intractable. Thus, in order to obtain $\overrightarrow{P}$ and calculate the fraction of active nuclei and the mean transcription onset times, we solve the system in Equation 3.6 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.

As before, we express the probability of a nucleus being in state s at time t as $P(t,s)$. To calculate $P(t,s)$ we need to consider the previous time step $t-1$ and take into account three possible processes:

1. Nuclei that were already in state s at time $t-1$ and stay in this state at time t.
2. Nuclei that were in state $s-1$ at $t-1$ that transition into state s at time t.
3. Nuclei 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 nucleus jumping from one state to the next at time t during an arbitrarily small time window of dt is given by the transition rate $\pi(t) \times dt$.

The probability of nuclei being in state s at time t can then be calculated as

$$\underbrace{P(t,s)}_{\text{Probability of state } s \text{ at time } t} = \underbrace{P(t-1,s)}_{\text{stay in state } s} + \underbrace{\pi(t)dtP(t-1,s-1)}_{\text{enter from state } s-1} - \underbrace{\pi(t)dtP(t-1,s)}_{\text{leave for state } s+1}. \tag{3.8}$$

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

As described above, $\pi(t)$ is a function of time because Dorsal concentration changes with time. In contrast to analytical solutions, this numerical approach makes it possible to use time-variant Dorsal concentration (measured from Dorsal-Venus fluorescence) to obtain the transition rate $\pi(t)$ at each time point.

To obtain the fraction of active nuclei we initialize the system to $P(1,1) = 1$ and calculate $P(T/dt, m+1)$ where T is the duration of the transcriptional window.

$$\text{Fraction of active nuclei} = P(T/dt, m+1). \tag{3.9}$$

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

$$\text{Mean transcriptional onset time} = \mathrm{E}[onset] = \frac{\sum_{t=0}^{t=T} \left[(P(t, n+1) - P(t-1, n+1)) \times t\right]}{\sum_{t=0}^{t=T} \left[P(t, n+1) - P(t-1, n+1)\right]}. \tag{3.10}$$

Exploring limits of simple activation model

Here we explore the weak promoter limit and the strong activator limit and show that they don't reduce complexity in the model very much.

We assume the weak promoter limit, where $\frac{P}{K_P} \ll 1$. This limit is justified by the fact that, in the absence of Dorsal, we see no transcriptional activity (Figure 3.2A and B). In other words, P is far smaller than K_P and, consequently, the state in which only RNAP is bound has a much lower occupancy than any other state. Furthermore, we assume the weak RNAP limit, where RNAP's loading rate in the absence of Dorsal is negligible. These assumptions allow us to remove one parameter from our simple activator model such that Equation 3.2 simplifies to

$$p_{bound} \approx \frac{\frac{Dl}{K_D}\omega'}{1 + \frac{Dl}{K_D} + \frac{Dl}{K_D}\omega'}, \tag{3.11}$$

where we have defined $\omega' = \frac{P}{K_P}\omega$ which describes the interaction between Dorsal and RNAP.

Finally, by substituting Equation 3.11 into Equation 3.1, we find

$$R = R_{max} \cdot \frac{\frac{Dl}{K_D}\omega'}{1 + \frac{Dl}{K_D} + \frac{Dl}{K_D}\omega'}. \tag{3.12}$$

We plot predictions for R in Figure 3.1.

3.7.3 Relating MS2 signal to the statistical mechanical model

In this section, we will address how to obtain the RNAP loading rate R by measuring experimentally accessible quantities. To make this possible, we begin by defining a calibration factor $alpha$ that relates the time-integrated fluorescence, $F(t)$, and accumulated mRNA, $mRNA(t)$ such that

$$F(t) = \alpha \cdot mRNA(t). \tag{3.13}$$

In Figure 3.12B, we estimate the calibration factor, α, between fluorescence units and completed MS2 transcripts.

By differentiating Equation 3.13, we find that this same calibration factor relates the instantaneous fluorescence and mRNA production rate,

$$f(t) = \alpha \cdot \frac{dmRNA}{dt}. \tag{3.14}$$

When we observe the MS2 signal in a nucleus, the instantaneous fluorescence is not a simple readout of $\frac{dmRNA}{dt}$. In order to understand how the maximum fluorescence of a trace relates to the average RNAP loading rate, we need a model of the fluorescence trajectory during a nuclear cycle. To model the MS2 signal trajectory of a single locus, we start by assuming that polymerase molecules begin loading at time t_0 into the nuclear cycle and continue to load at a constant rate proportional to R as shown in Equation 3.1 and Figure 3.13. The signal will increase linearly until the first polymerase terminates transcription. At this point, the signal plateaus at the value f_{max} because polymerase molecules continue to load at a constant rate while simultaneously terminating at the same rate.

Then, the maximum fluorescence we see in a trace is proportional to the loading rate from Equation 3.1 and is estimated as

$$f_{max} \approx \alpha \cdot R \cdot \Delta t_{elongation}. \tag{3.15}$$

Here, we assume that $\Delta t_{elongation}$ is a constant in our system and is given by

$$\Delta t_{elongation} = \text{elongation rate} \cdot \text{gene length}. \tag{3.16}$$

For our measurements, we have

$$\Delta t_{elongation} \approx \frac{5\ kbp}{2\ kbp/min} = 2.5\ min, \tag{3.17}$$

where the elongation rate comes from previous measurements [98]. Thus, we now have an expression for f_{max} that allows us to relate our data to models.

By comparing Figure 3.1C and Figure 3.1D, we note that there may be some redundancy between the effects of varying ω and R_{max}' for finite Dorsal concentrations. That is, a

lower ω can always compensate for a lower R_{max} value to achieve approximately the same curve. One way to see this redundancy is to look at Equation 3.3 when ω is small. Then,

$$R \approx \frac{R_{max}\omega \cdot \frac{Dl}{K_D}}{1 + \frac{Dl}{K_D}}. \tag{3.18}$$

In this regime, it becomes clear that inferring R_{max} and ω from any data separately may be difficult as their product is the only observable quantity.

3.7.4 Supplementary Figures

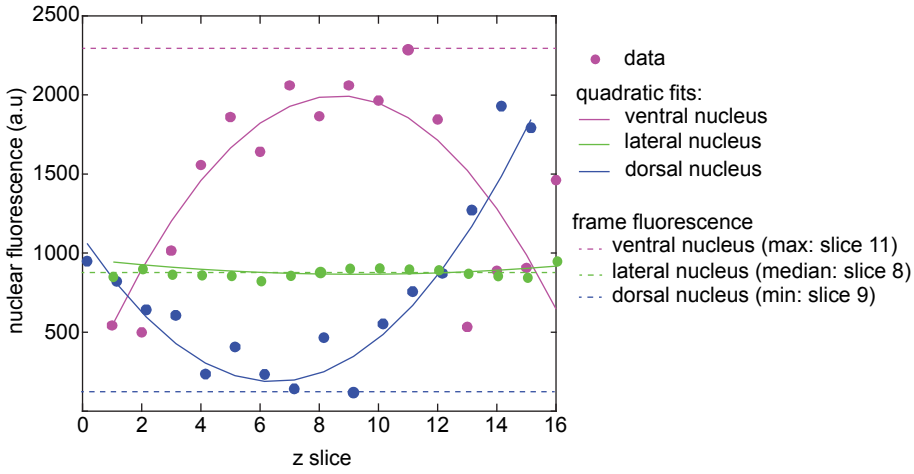

Figure 3.7: **Measuring Dorsal-Venus nuclear fluorescence across the dorso-ventral axis** In each frame, the Dorsal-Venus fluorescence is measured in each z-slice of each nucleus. This creates a series of fluorescence values as a function of z-slice (filled circles). Slices at the top and the bottom correspond to cytoplasmic fluorescence. Thus, in ventral nuclei, the brightest slice correspond to the true nuclear fluorescence (magenta circles). On the other hand, dorsal nuclei have a lower Dorsal concentration than the cytoplasm and so the darkest slice correspond to the true dorsal concentration (blue circles). In lateral nuclei, the nuclear fluorescence is similar to that of the cytoplasm (green circles). To decide which z-slice to use for nuclear fluorescence calculation, we fit the fluorescence f over z-slices z to a quadratic equation $f = az^2 + bz$ where a and b are the quadratic coefficients. 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-Venus 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 frame Dorsal-Venus fluorescence (dashed horizontal blue line).

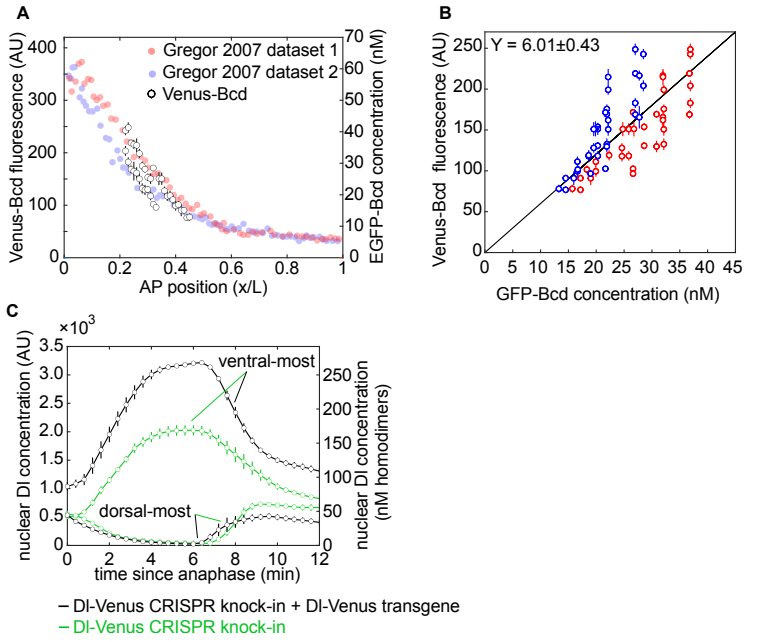

Figure 3.8: **Absolute calibration of Dorsal-Venus fluorescence using Venus-Bcd and previously measured eGFP-Bcd concentration.** Three embryos derived from *yw;Venus-Bcd;BcdE1* homozygous mothers were imaged in nuclear cycle 14 using the imaging conditions of MS2 experiments. The nuclear fluorescence was calculated 15 minutes into nuclear cycle 14 for cross comparison with data from [108] Figure 2B. The mapping between Venus-Bcd fluorescence and eGFP-Bcd absolute concentration was based on position along the AP axis of the embryo. **(A)** Linear fit forced through the origin. The slope error corresponds to the 95% confidence interval. **(B)** Mean and SEM across four embryos of the Dorsal nuclear concentration in the ventral-most and dorsal-most nuclei. 1X and 2X correspond to embryos from homozygous females containing one or two Dl-Venus alleles respectively. The right y-axis shows the concentration of Dorsal homodimers assuming 6 fluorescence a.u. per Venus molecule based on (A) and (B). In (A), each data point corresponds to the mean $\pm$ standard deviation of the fluorescence of all nuclei belonging to the same 1% AP bin in the same embryo. These data were compared to two different absolute measurements of eGFP-Bcd, shown in red and blue. Because Venus-Bcd is less abundant than eGFP-Bcd by a factor of $\approx 2/3$ (based on cephalic furrow position), Venus fluorescence values were multiplied by 1.5.

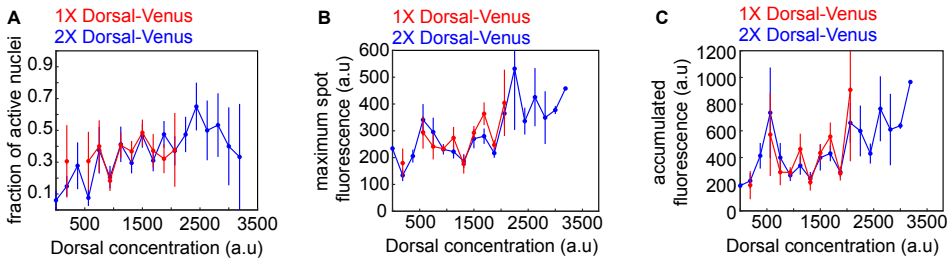

Figure 3.9: **Comparing the activity of Dorsal-Venus transgene to that of two copies of Dorsal Venus provided by a transgene plus a CRISPR knock-in**. For the DBS_6.23, we imaged embryos laid by two different mothers. 1X mothers (labeled in red) carry dl1 (a null Dorsal allele) and a Dl-Venus transgene created by [215]. 2X mothers (blue) carry a Dorsal Venus CRISPR knock-in and the aforementioned Dl-Venus transgene. We wondered whether it is fair to combine embryos laid by these different mothers. We reasoned that, if for a given Dorsal concentration -regardless of its allele of origin- the activity is the same within error, then we can pool data from 1X and 2X mothers. To test this, nuclei from these different mothers were binned according to their Venus fluorescence and different activity metrics were measured for each bin. Error bars correspond to the standard error across at least three embryos per bin. We conclude from this analysis that, taken together, these results show that these two Dorsal-Venus populations are not different within error. Thus, it is valid to treat embryos laid by these different mothers as equivalent.

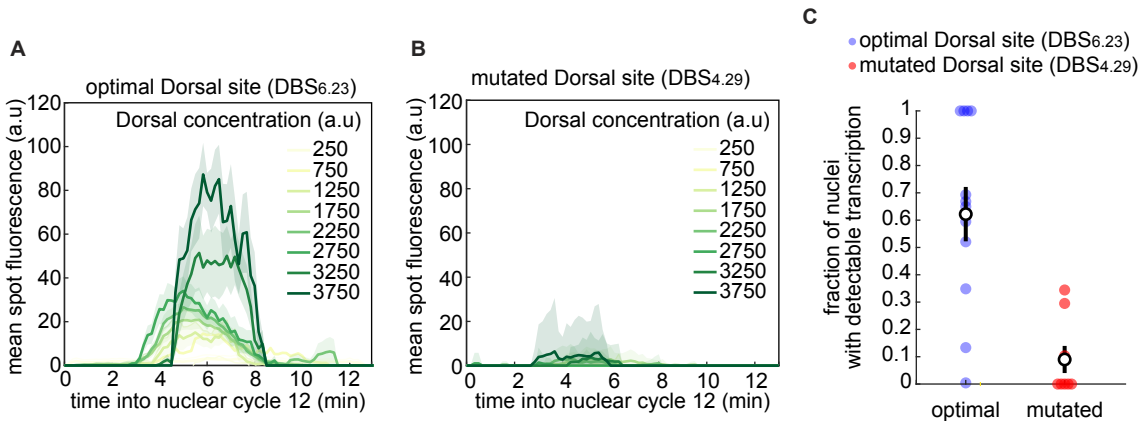

Figure 3.10: **bf Related to Figure 3.2D: Transcription of a minimal Dorsal synthetic enhancer with a mutated Dorsal binding site.** The activity of minimal synthetic enhancers carrying a single optimal (DBS_6.29, blue) or mutated (DBS_4.29, red) Dorsal binding site was measured in nuclear cycle 12 embryos across the dorso-ventral axis. **(A)** Mean fluorescence over time across all spots in the field of view of an embryo carrying a minimal synthetic enhancer with a single optimal Dorsal binding site. **(B)** Same as (A) for a mutated Dorsal binding site. **(C)** Fraction of nuclei in the field of view 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). 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.

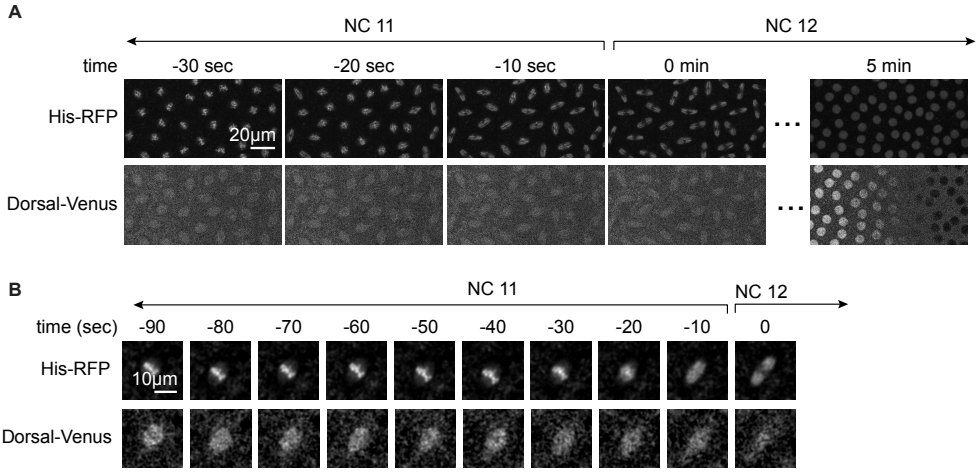

Figure 3.11: **Using the Dorsal-Venus channel to determine the timing of mitosis**. **(A)** Maximum projection snapshots of Histone-RFP (top) and Dorsal-Venus (bottom) during the mitosis leading from NC 11 to NC 12. **(B)** Same as (A) for an individual nucleus. We define as the beginning of the cycle the frame in which anaphase occurs. In the Histone-RFP channel this is characterized by the separation of sister chromatids. In the Dorsal-Venus channel nuclei elongate and their boundary becomes fuzzier.

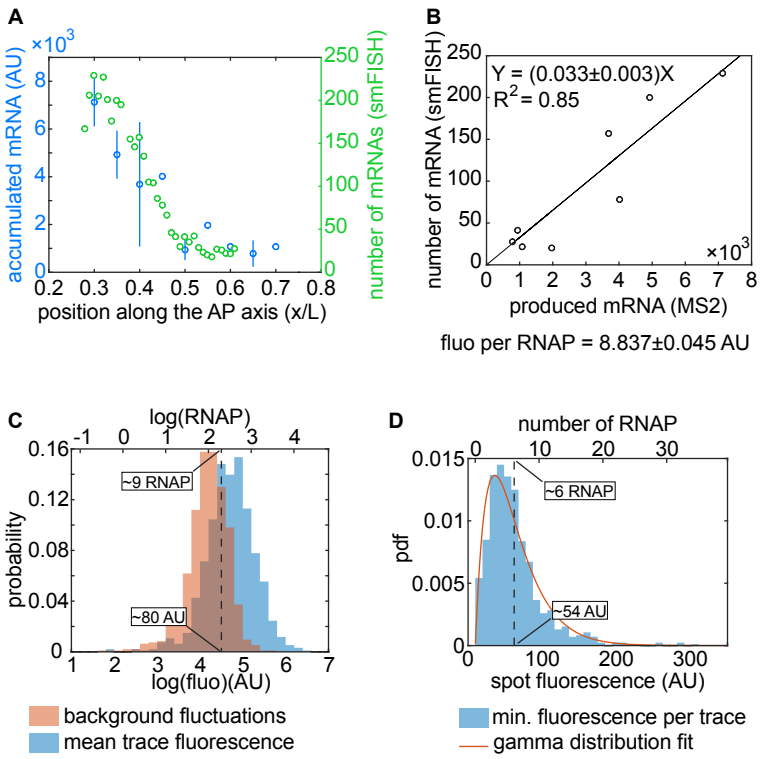

Figure 3.12: **Absolute calibration of MS2 using single molecule FISH**. **(A)** Left: mean accumulated mRNA per nuclear cycle 13 nucleus based on integrated fluorescence of P2P-MS2 using the imaging conditions of Dorsal synthetics (n = 6 embryos). Right: number of mRNAs per nucleus in nuclear cycle 13 from [98]. **(B)** Scatter plot showing data from (A) corresponding to the same anterior-posterior (AP) bin. The solid line shows the best linear fit to all of the 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. **(C)** Histograms of mean trace fluorescence in all particles across all experiments and the error in the fluorescence of these particles. Because the spot fluorescence was obtained integrating over three slices, the corresponding error was propagated by multiplying by $\sqrt{3}$. The dashed line indicates the center of where the two distributions overlap. **(C)** Histogram of the minimum spot fluorescence per trace across all experiments. The dashed line indicates the mean of the distribution. Note that in (C) and (D) the top x-axis is expressed in terms of absolute number of RNAP using the calibration from (B).

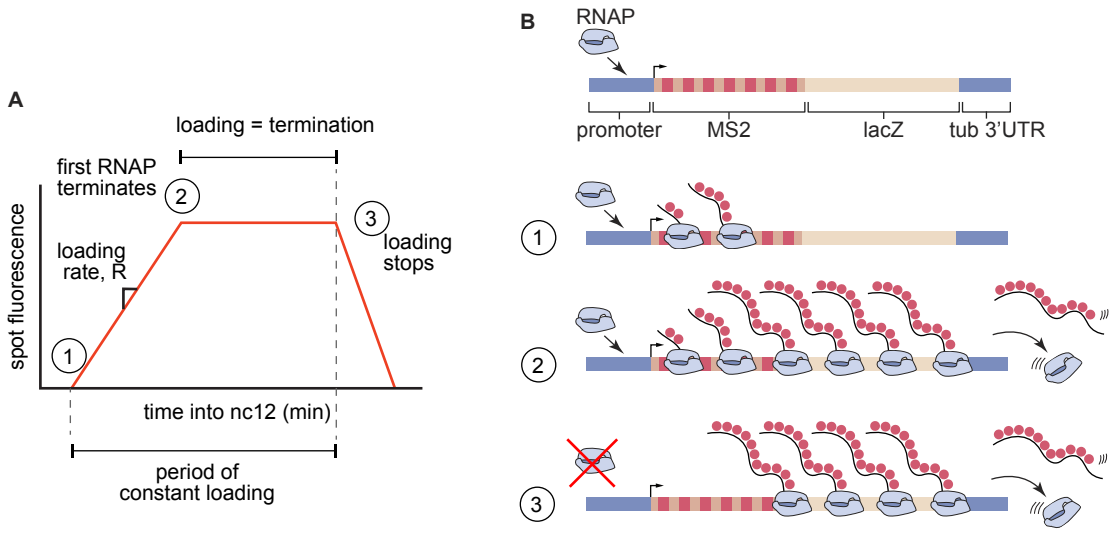

Figure 3.13: **Trapezoid model of transcription dynamics. (A)** Schematic showing an idealized spot fluorescence time trace originating from an MS2 reporter in the 12th nuclear cycle in the early fruit fly embryo. Between 1 and 2, fluorescence increases linearly as RNAP molecules are loaded onto the gene with a rate R. At 2, the first RNAP terminates. Because the rate of loading and termination are equal the gene is at steady state and the fluorescence is constant. At 3, the promoter stops loading new RNAP molecules and thus fluorescence decreases with a rate equal to the termination rate. **(A)** Cartoons showing the molecular picture corresponding to each of the steps in (A).

Chapter 4

Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics

4.1 Abstract

The responses of plants to their environment often hinge on the spatiotemporal dynamics of transcriptional regulation. While live-imaging tools have been used extensively to quantitatively capture rapid transcriptional dynamics in living animal cells, lack of implementation of these technologies in plants has limited concomitant quantitative studies in this kingdom. Here, we applied the PP7 and MS2 RNA-labeling technologies for the quantitative imaging of RNA polymerase II activity dynamics in single cells of living plants as they respond to experimental treatments. Using this technology, we count nascent RNA transcripts in real-time in *Nicotiana benthamiana* (tobacco) and *Arabidopsis thaliana* (Arabidopsis). Examination of heat shock reporters revealed that plant tissues respond to external signals by modulating the proportion of cells that switch from an undetectable basal state to a high transcription state, instead of modulating the rate of transcription across all cells in a graded fashion. This switch-like behavior, combined with cell-to-cell variability in transcription rate, results in mRNA production variability spanning three orders of magnitude. We determined that cellular heterogeneity stems mainly from stochasticity intrinsic to individual alleles instead of variability in cellular composition. Taken together, our results demonstrate that it is now possible to quantitatively study the dynamics of transcriptional programs in single cells of living plants.

4.2 Introduction

Plant growth and development depends on rapid and sensitive signaling networks that monitor environmental fluctuations and transduce this information into transcriptional changes that lead to physiological adaptation. Gene regulation in plants can be extremely fast, with changes in mRNA abundance detectable in minutes or less, for example in response to modulations in light intensity [250, 58], light quality [159], the axis of gravity [144], nutrient concentration [152], temperature [283], or presence of pathogens [33]. A first step toward understanding how plant transcriptional programs unfold in time and space is to quantify gene activity in individual living cells as they respond to external stimuli. Protein reporters have been used in plants to measure the dynamics of single-cell gene activity in live tissues over hours to days [106]. However, fluorescent proteins mature at timescales that are long ($>$30 min) compared to the rates that characterize stress-responsive transcription ($\sim$1 min) [151], particularly in organisms grown at moderate temperatures such as plants [15]. In addition, protein reporter signals convolve processes such as transcription, RNA processing, RNA transport, translation, and protein degradation, often making it challenging to precisely identify where and how regulatory control is being applied along the central dogma.

Over the last few years, our understanding of transcriptional regulation in animals has been transformed by techniques that have made it possible to quantify transcriptional activity in single cells of living fruit fly embryos [82, 174, 109, 95], in the nematode *Caenorhabditis elegans* [156], and in adult mouse tissue [63]. Here, nascent RNA is fluorescently labeled by tagging genes of interest with RNA aptamers such as MS2 or PP7 that recruit fluorescent proteins to transcriptional loci, revealing real-time transcriptional activity at the single-cell level. However, research into the equally diverse and important gene regulatory aspects of plant development and physiology has remained relatively isolated from these technological breakthroughs. Indeed, previously, MS2 and other similar approaches based on RNA-binding proteins were used in plants to visualize the movement and localization of cytoplasmic RNAs [117, 285, 235], but not their nuclear transcriptional dynamics.

Here we bridged this technological gap by developing and implementing the PP7 and MS2 technologies for labeling nascent RNA in *Arabidopsis thaliana* (Arabidopsis) and *Nicotiana benthamiana* (tobacco). Through state-of-the-art quantitative imaging, we counted the absolute number of elongating RNA polymerase II (RNAP) molecules at individual genes and measured how this number is regulated dynamically in response to heat stress. We used this stress response in leaves as a model to determine how tissue-level patterns of mRNA accumulation arise from the dynamical transcriptional behavior of individual cells. Using this technology, we also uncovered previously unmeasurable modes of gene regulation in plants by which tissues respond to external signals by modulating the fraction of cells engaged in transcription, but leave the single-cell transcription rate unchanged. Further, we determined how these regulatory layers give rise to high cell-to-cell variability—spanning three orders of magnitude—in mRNA production.

The single-locus resolution afforded by PP7 and MS2 made it possible to characterize the sources of this cell-to-cell variability, revealing that stochastic processes intrinsic to individual alleles are the main contributors to this variability independent of differences in cellular composition. Together, these results highlight the potential of live-imaging techniques for uncovering and quantitatively describing regulatory processes with spatiotemporal resolutions that cannot be achieved with methods such as traditional protein reporters or single-cell RNA sequencing. We envision that this approach will open new avenues of inquiry in plant physiology, and cell and developmental biology.

4.3 Results

4.3.1 Establishment of the PP7 and MS2 systems for single-cell live imaging of transcription in plants

To quantitatively measure transcriptional dynamics in tobacco and Arabidopsis, we implemented an mRNA fluorescent-tagging approach previously used in cells in culture [104, 52, 62, 155], *D. melanogaster* embryos [98, 174], the mouse brain [203], and *C. elegans* [156] in which the gene of interest is tagged with tandem repeats of the PP7 DNA sequence that, when transcribed, form RNA stem-loops (Fig. 4.1A) [46, 155]. The PP7 loop RNA is bound by the PP7 bacteriophage coat protein (PCP) [46] expressed under a ubiquitous promoter. Fusing PCP to a fluorescent protein results in the fluorescent labeling of nascent RNA molecules. By virtue of the relatively slow movement of genomic loci in the nucleus and the accumulation of fluorophores in the diffraction-limited volume of the gene, sites of active transcription appear as bright fluorescent puncta over the background of nuclear PCP fluorescence in a laser-scanning confocal microscope (Fig. 4.1A). The fluorescence intensity of these spots reports on the number of RNAP molecules actively transcribing the gene at any given time [98] and is proportional to the instantaneous rate of transcription [154, 37].

To optimize this imaging strategy for plants, we generated two classes of constructs (Fig. 4.1B): First, coat protein constructs that fuse PCP to a fluorescent protein such as GFP under a constitutive and ubiquitously expressed Arabidopsis promoter, and, second, reporter constructs that contain a neutral DNA sequence consisting of a firefly luciferase-β-glucoronidase fusion with 24 PP7 stem loop repeats [88] inserted in the 5' end of this gene, under the control of the promoter of interest. To aid in the automated segmentation of nuclei, reporter constructs also contain a nuclear label consisting of the mScarlet red fluorescent protein [28] fused to the Arabidopsis histone 2B coding region driven by a ubiquitous promoter [81]. These two constructs confer resistance to different antibiotics, allowing sequential and combinatorial transformation into plants.

We tested this system in tobacco by simultaneously infiltrating leaves with two Agrobacterium strains, one strain carrying a PCP-GFP plasmid and a second strain carrying a reporter plasmid lacking a functional promoter, yielding homogeneous GFP nuclear and

cytoplasmic fluorescence (Fig. 4.1C, top left). When the strong and constitutive 35S promoter was used to drive the reporter construct, nuclear GFP puncta became visible (Fig. 4.1C, top right). These results suggest that spots correspond to sites of active transcription and are not an artifact of PCP-GFP aggregation in the nucleus. Analogous results were obtained in stably transformed transgenic Arabidopsis plants (Fig. 4.1C, bottom).

We next sought to confirm that spot fluorescence constitutes a dynamical readout of transcriptional activity. To this end, we asked whether spot fluorescence dynamics in tobacco qualitatively recapitulate previous observations performed on the same promoters in Arabidopsis with orthogonal techniques. This comparison is made possible by the strong conservation of transcriptional regulation in plants [269], in particular the heat shock response [186]. We measured the transcriptional activity of two well-known constitutive and heat shock-inducible Arabidopsis genes (*GAPC2* and *HSP70*, respectively [60, 66]) before and during a heat shock treatment. GAPC2-PP7 expression was detectable at 25°C (Fig. 4.1D, top left, Movie 1). The presence of multiple spots per nucleus is likely due to multiple transgene transfer events; the number of spots did not change with treatment (Fig. 4.1D, bottom left and Fig. 4.6). Further, the fluorescence over time of these spots did not change upon heat shock (Fig. 4.1E and Fig. 4.6), in accordance with the constitutive expression of *GAPC2* in Arabidopsis [60]. Consistent with the heat shock inducibility of the *HSP70* gene in Arabidopsis [66], HSP70-PP7 transcription was hardly detectable at 25°C in tobacco (Fig. 4.1D, top right and Fig. 4.6). However, upon increasing the temperature to 39°C, multiple fluorescent puncta rapidly appeared (Fig. 4.1D, bottom right and Fig. 4.6, Movie 2), and their fluorescence increased with time (Fig. 4.1E, Fig. 4.6). A reporter construct where the PP7 cassette is inserted in an intron of Arabidopsis *HSP70* fused in its C-terminus to mCherry, confirmed that appearance of transcriptional spots is associated with the accumulation of the gene products (Fig. 4.7). Thus, we conclude that the PP7 system reliably recapitulates previous qualitative knowledge of transcriptional dynamics in plants.

Simultaneously tagging multiple mRNA species or multiple locations of the same mRNA species with different fluorescent proteins has revealed regulatory and physical interactions between loci and uncovered the regulation of distinct steps of the transcription cycle in cells in culture and animals [125, 56, 88, 89, 167, 166]. To enable such multiplexing in plants, we also implemented the MS2 system, which is analogous and orthogonal to the PP7 system. Here, MS2 loops are specifically recognized by an MS2 coat protein (MCP) [26]. We tested the MS2 system in tobacco and obtained results comparable to those obtained for PP7 (Fig. 4.8), allowing us to track the expression dynamics of two transgenes in a single cell (Fig. 4.1F).

4.3.2 Quantitative characterization of the PP7 system in Arabidopsis

To study transcriptional regulation at the single-cell level in populations of genetically identical leaf cells, we next generated stably transformed lines of Arabidopsis carrying

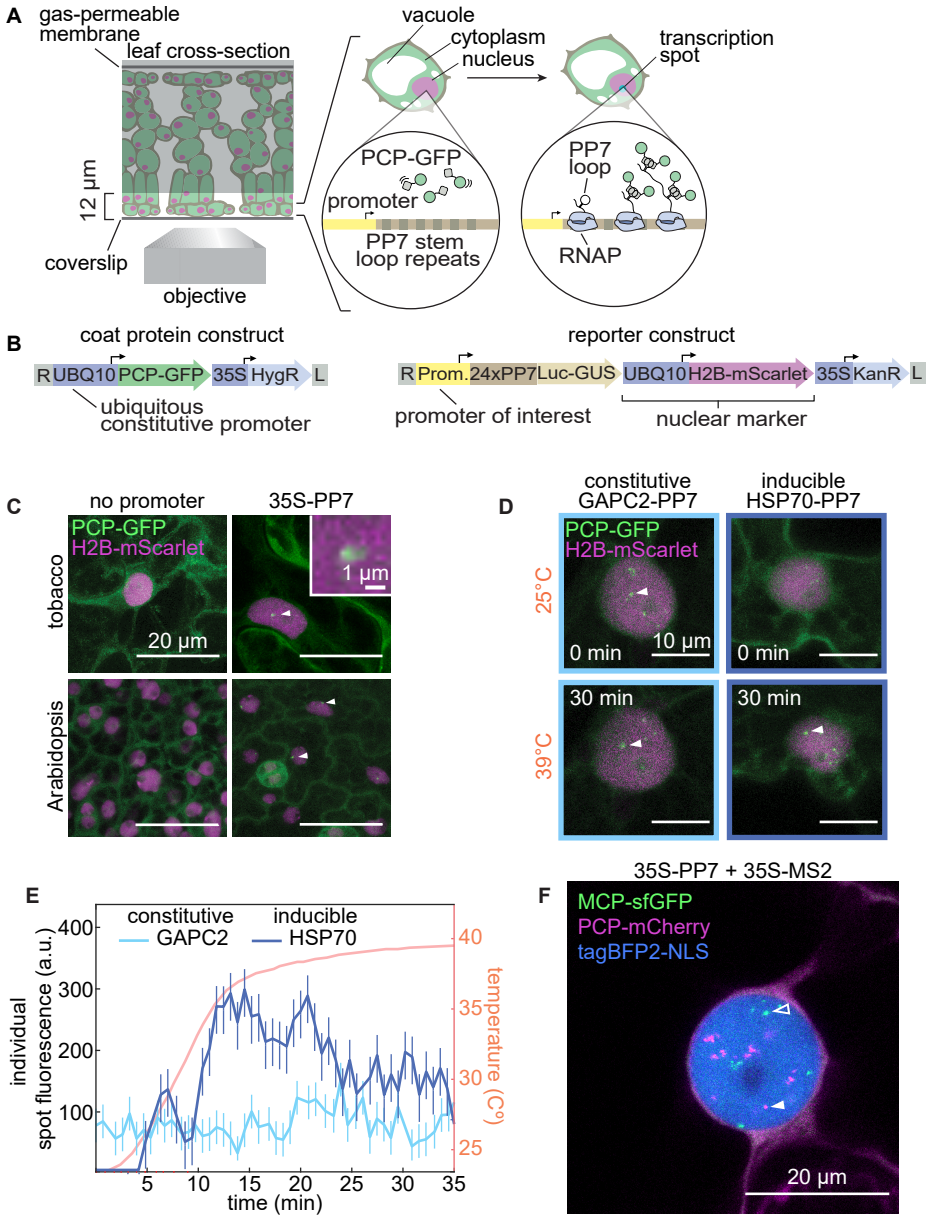

Figure 4.1: **Fluorescence labeling of nascent RNA in tobacco and Arabidopsis reveals single-cell transcriptional dynamics in real time. (A)** Schematic of the live-imaging experimental setup in leaves and diagram of the PP7 RNA labeling system. **(B)** Schematic of the constructs used in this study. (UBQ10, Arabidopsis ubiquitin 10 promoter; 35S, CaMV 35S promoter; HygR, hygromycin resistance; Luc-GUS, firefly luciferase-β-glucoronidase fusion; H2B, Arabidopsis histone 2B coding sequence; KanR, kanamycin resistance; L, T-DNA left border; R, T-DNA right border). **(C)** Maximum projection of snapshots of cells expressing PCP-GFP and the reporter construct with or without the constitutive 35S promoter driving expression of the PP7-tagged Luc-GUS gene. White arrows indicate nuclear fluorescent puncta corresponding to transcription spots. Inset: magnification of PP7 fluorescence. ***Caption continues on next page***.

Figure 4.1: ***Continued from previous page*: Fluorescence labeling of nascent RNA in tobacco and Arabidopsis reveals single-cell transcriptional dynamics in real time. (D)** Maximum projection snapshots of tobacco cells expressing PCP-GFP and reporter constructs driven by the promoters of the Arabidopsis *GAPC2* and *HSP70* genes. Time under heat shock is indicated. White arrowheads indicate the fluorescent spots quantified in (E). **(E)** Fluorescence time traces of single nuclear GFP puncta in tobacco leaf epidermis cells expressing PCP-GFP and reporter constructs driven by the promoters of the GAPC2 and HSP70 Arabidopsis genes. Each blue line corresponds to a single spot tracked over time. The orange line corresponds to the temperature experienced by the sample and is plotted on the right y-axis. Prior to spot detection, spots are assigned a fluorescence value of zero. Error bars represent the uncertainty in the spot fluorescence extraction (see Materials and Methods: Image analysis: spot fluorescence and tracking) **(F)** Maximum projection snapshot of tobacco leaf epidermal cell expressing PCP-mCherry, MCP-GFP, H2B-tagBFP2, and two reporter constructs driven by the 35S promoter and tagged with PP7 (magenta) or MS2 (green). Open and closed arrowheads indicate MCP-tagged and PCP-tagged nascent RNAs, respectively (see also Fig. 4.8).

PCP-GFP and a PP7 reporter construct driven by the promoter of the stress-inducible HSP101 gene. A line carrying a single reporter locus (hereafter referred to as HSP101-PP7-1) was used for the following experiments unless stated otherwise; for details, see Materials and Methods: Generation of transgenic Arabidopsis lines.

A key step toward establishing PP7 as a reporter of single-cell transcriptional activity in Arabidopsis is to demonstrate that the observed spot fluorescence dynamics quantitatively recapitulate this activity. We therefore sought to cross-validate PP7 measurements with RT-qPCR quantifications of reporter transgene mRNA abundance in our stably transformed Arabidopsis plants. The *HSP101* mRNA is hardly detectable across vegetative tissues under standard growth conditions [211, 170] and accumulates to high levels as quickly as 2 minutes following treatments inducing cytosolic protein misfolding such as heat shock [47, 283, 170]. As previous experiments have shown that, upon induction, *HSP101* is expressed uniformly throughout plant tissues [271, 136], we compared the average transcriptional activity of a few hundred leaf cells obtained by microscopy with that of the whole plant in bulk reported by RT-qPCR.

As expected, we did not detect actively transcribing cells in HSP101-PP7-1 plants imaged for 1 h at room temperature (Fig. 4.9), but shifting the microscope stage from 22°C to 39°C resulted in the rapid appearance of transcription spots (Fig. 4.2A, Movie 3). To compare the instantaneous metric of transcriptional activity reported by spot fluorescence with the number of accumulated reporter mRNA molecules captured by RT-qPCR, we converted spot fluorescence to number of produced mRNA molecules by integrating the fluorescence of all spots in the field of view over time [98](Fig. 4.10 and associated calculations in Section 4.7.1).

Controls for GFP photobleaching ruled out the possibility that we underestimated

the produced mRNA calculated by microscopy (Fig. 4.11). Finally, we measured HSP101 reporter mRNA abundance by RT-qPCR using whole plants treated with heat shock (see Materials and Methods: Heat shock treatments). These measurements were strongly correlated with each other ($R^2 = 0.98$ Fig. 4.2B), confirming that spot fluorescence directly reports on the rate of mRNA production. This conclusion held regardless of the magnitude of the mRNA degradation rate (Fig. 4.7 and associated calculations in Section 4.7.1).

While our measurements so far have shown that PP7 fluorescence is proportional to the number of actively transcribing RNAP molecules, this fluorescence does not, by itself, report on the absolute number of RNAP molecules. Expressing measurements in terms of absolute number of active RNAP molecules instead of arbitrary fluorescence units is necessary for directly comparing data across microscopy setups and laboratories, and for integration with other quantitative measurements and theoretical models [224, 42, 94, 98, 278]. In order to turn the PP7 system into such a precision tool, we calibrated its arbitrary fluorescence units to report on the number of RNAP molecules transcribing the reporter. We followed a recently established approach to measure the fluorescence of individual GFP molecules arranged in 60-meric nanocages in vitro [127] and in vivo [2]. We fused GFP to a monomer that forms these 60-meric nanocages and expressed it in tobacco leaves (Fig. 4.2C) to obtain a distribution of fluorescence intensity values for the resulting GFP punctae (Fig. 4.2D, left, see also Figure 4.8). Fusing two GFP molecules to each nanocage monomer yielded the fluorescence distribution of nanocages containing 120 GFP (Fig. 4.2D, left). To further validate this approach we imaged a genetically encoded multimeric nanoparticle (GEM) containing 60 GFP-tagged monomers [64]. A linear fit of the means of these distributions passing through the origin shows that the mean fluorescence of the 120 GFP nanocage is almost exactly twice that of 60 GFP nanocage and the 60 GFP GEM (Fig. 4.2D, right), confirming the validity of this approach. The slope of this fit is an estimate of the average number of arbitrary units of fluorescence corresponding to a single GFP molecule in our microscopy setup, making it possible to report PP7 measurements in absolute units.

Our absolute calibration also provided the opportunity to determine the limits of applicability of the PP7 technology. Specifically, there is a minimum number of actively transcribing RNAP molecules below which no reliable detection is possible. Figure 4.2E compares histograms of the calibrated number of RNAP molecules in the weakest detectable spots across all spots from all replicates from Figure 4.1F and their corresponding fluctuations in background fluorescence. This calibration is based on the assumption that each PP7 loop is bound by two PCP-GFP molecules and each fully loaded RNAP carries 24 PP7 loops. Consistent with previous measurements [98, 154], these background and signal histograms overlap at approximately 3 RNAP molecules, marking the level at which PP7 fluorescent spots become undetectable (see also Fig. 4.9). An alternative way to view this detection limit is to consider the minimum detectable rate of transcription initiation. Given an elongation rate of 1.5 kbp/min [9] and the average unspliced transcript length in Arabidopsis of about 2.2 kbp [255], a RNAP molecule takes 3 min to

transcribe an average Arabidopsis gene. Thus, to ensure at least 3 RNAP molecules on the gene and signal detectability at any time point, transcription needs to initiate at a minimum rate of 1 RNAP/min.

It is also informative to determine the dynamic range of our measurements in terms of the number of actively transcribing RNAP molecules. Given a footprint of an elongating RNAP molecule of ≈40 bp [216, 258], an average Arabidopsis gene can accommodate a maximum of 2.2 kbp/40 bp≈55 RNAP molecules (or a maximum density of 25 RNAP molecules/kbp), well above the minimum 3 RNAP molecules that constitute our detection limit. The strongest transcribing loci in our HSP101-PP7 experiment have a fluorescence of ≈1000 a.u., corresponding to ≈250 RNAP molecules (Fig. 4.2E). According to our qPCR analysis (Figure 4.10 and Section 4.7.1), the insertion locus of line HSP101-PP7-1 contains two copies of the reporter construct. Since our reporter has a length of approximately 4.8 kbp, the strongest loci have an RNAP density of about 25 RNAP molecules/kbp, showing that they are likely transcribing at the maximum possible rate.

4.3.3 Uncovering single-cell transcriptional responses to heat shock

While static snapshots of tissues have provided profound lessons about the spatial control of transcription in animals and plants alike [32, 254], these approaches have not revealed how single-cell transcriptional dynamics dictate the temporal modulation of gene expression patterns. We sought to bridge this gap between single-cell and tissue-wide transcriptional dynamics by tracking individual nuclei and measuring the fluorescence of their corresponding transcription spot over time. To expand our range of inquiry, we generated two additional reporter lines under the control of a second heat shock-inducible promoter (HsfA2-PP7, Movie 4) or of a constitutive promoter (EF-Tu-PP7, Movie 5). In order to simplify our experiments, we focused on cells containing at most one spot per nucleus. We achieved this by imaging cells close to the base of the leaf which, according to their nuclear volume (Fig. 4.11) and developmental stage, should be predominantly diploid [182, 217, 91]. Consequently, young epidermis cells in hemizygous Arabidopsis derived from the first generation of single-insertion transgenic plants (i.e., T2 individuals) contained at most one spot per nucleus (Fig. 4.12, see also Materials and Methods: Microscopy setup and image acquisition).

A striking feature of the single-cell response is the existence of a reproducible fraction of nuclei that does not show detectable expression of the reporter transgene throughout the experiment in all three assayed promoters, which we define as transcriptionally refractory cells (Fig. 4.3A, Fig. 4.13). The presence of these transcriptionally refractory cells was surprising given that endogenous *HSP101* and *HsfA2* are strongly induced and are necessary to survive heat stress in a dose-dependent manner [211, 47]. Similarly, as a highly expressed constitutive gene, the *EF-Tu* promoter would also be expected to drive transcription in every cell. Yet, this constitutive transgene also presents a substantial fraction of refractory cells (Fig. 4.3A, right). Such refractory cells have also been identi-

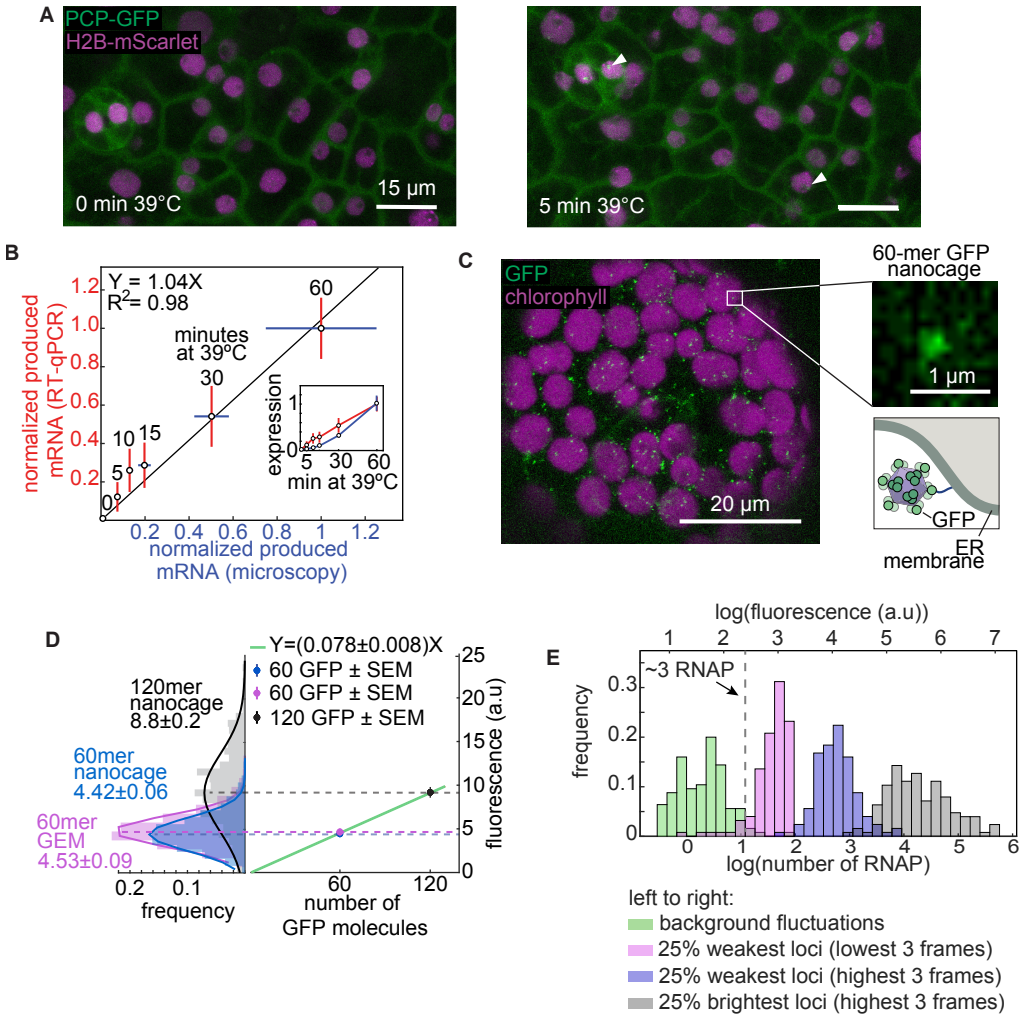

Figure 4.2: **Cross validation, absolute calibration, and sensitivity of the PP7 reporter system. (A)** Maximum fluorescence projections of leaf epidermal tissue of an Arabidopsis line stably transformed with PCP-GFP and a reporter construct driven by the HSP101 promoter under heat shock. Time stamps indicate time under heat shock. Arrowheads point to transcription spots. **(B)** Comparison between total mRNA produced as reported by RT-qPCR and PCP-GFP. PCP-GFP error corresponds to the standard error of the mean over 8 biological replicates; RT-qPCR error corresponds to the standard error of the mean (SEM) across three biological replicates. Data are normalized to each corresponding signal at 60 min. The solid black line shows a linear fit to the data going through the origin. The inset shows the normalized mean and SEM of expression level as a function of time for RT-qPCR and microscopy. **(C)** Maximum fluorescence projection of a tobacco mesophyll cell expressing a construct encoding a 60 GFP nanocage tethered to the outer membrane of the endoplasmic reticulum (ER). ***Caption continues on next page***.

Figure 4.2: ***Continued from previous page*: Cross validation, absolute calibration, and sensitivity of the PP7 reporter system. (D, left)** Absolute calibration of GFP fluorescence. Histograms and Gaussian fit of fluorescence values of individual spots for the 60-GFP nanocage (blue), 60-GFP GEM (magenta) and 120-GFP nanocage (black) transiently expressed in tobacco leaves. The mean of each distribution is shown next to each histogram. As expected, the means are related by a factor of two. **(D, right)** Mean and standard error of the mean (SEM) of the nanocages and GEM fluorescence as a function of number of GFP molecules per structure. The green line is a linear fit passing through the origin, revealing a calibration factor of 0.078 ± 0.008 a.u./GFP molecule (error reporting on the 95% confidence interval of the fit). See also Materials and Methods: Absolute calibration using nanocages. **(E)** Histograms of the calibrated number of transcribing RNAP molecules in the weakest three frames of the weakest 25% of HSP101-PP7 fluorescence time traces (magenta) and their associated fluorescence background fluctuations (green) from all spot fluorescence time traces across all 8 replicates from (B). The point where the distributions overlap, at 3 RNAP molecules (vertical dashed line), can be considered the detection threshold. Also shown are the brightest 3 frames of the weakest 25% of all time traces (blue) and the brightest 3 frames of the strongest 25% of spot fluorescence time traces (grey).

fied in live-imaging studies of the early development of the fruit fly [98, 154, 24] and in in vitro cultures of animal cells [114].

To confirm that the presence of refractory cells was not an artifact of our construct or of the PP7 technology, we examined a transgenic plant containing a HSP101-GFP fusion driven by the *HSP101* promoter that fully complements the heat-susceptibility phenotype of a hsp101 knockout [181]. Treatment of HSP101-GFP plants with the conditions used in our PP7 experiments revealed the presence of two types of cells: cells whose fluorescence was close to that of untreated cells and highly induced cells (Figure 4.14). These low-fluorescence cells, which can be located right next to highly expressing ones, support the existence of transcriptionally refractory cells and the ability of the PP7 technology to detect them.

This cellular heterogeneity in the response could arise from uneven heating across the field of view, however, a gradient of temperature with biologically relevant scales is unlikely to arise at a microscopic level (see Materials and Methods: Heat shock treatments). Consistent with this, we found that the spatial distribution of actively transcribing cells can be well described by a random distribution (Fig. 4.15).

Within responsive nuclei, we also found substantial heterogeneity in the instantaneous number of actively transcribing RNAP molecules. For example, at any given time, not all responsive nuclei harbored fluorescent spots; the fraction of active nuclei is modulated in response to heat shock, but remains constant for the constitutive promoter (Fig. 4.3B). In addition, individual spots do not turn on synchronously and present periods of high transcriptional activity interspersed by periods of low to no detectable

activity (Fig. 4.3C, see also Fig. 4.16, Fig. 4.17 and Fig. 4.18). This single-cell behavior is consistent with the presence of transcriptional bursts, which have been identified across organisms and are believed to emerge from the intrinsically stochastic nature of the biochemical process of transcription [193]. The only plant gene (to our knowledge) probed before in such detail lacked such bursts [130].

Finally, in order to demonstrate the applicability of this technique to other plant tissues, we imaged EF-Tu-PP7 and HsfA2-PP7 in Arabidopsis roots. The rapid rate of cell division in roots allowed us to capture the halting of transcription during mitosis (Figure 4.19 A-C) [240]. In addition, consistent with its behavior in leaves, HsfA2 was expressed in only a fraction of nuclei at any given time (Fig. 4.19 D and E).

4.3.4 Tissue-wide transcriptional dynamics arise from the switch-like regulation of the instantaneous fraction of transcribing cells

How do tissue-level patterns of mRNA arise from the transcriptional activities of individual cells? Such tissue-level control could be implemented in two possible ways [149, 265, 34, 83]. One strategy consists of modulating the single-cell rate of transcription across all cells in a graded, analogue fashion (Fig. 4.4A, top). Alternatively, transcriptional control could work like a switch, where the fraction of cells transcribing above basal uninduced levels is modulated across the tissue (Fig. 4.4A, bottom). Several *Drosophila* enhancers invoke both strategies simultaneously [98, 37, 154, 24]. Single time-point measurements in plants [261, 7] and live-imaging studies in cell culture [114] have also provided evidence for switch-like control.

We found that, as transcriptional induction ensues, the instantaneous fraction of cells actively transcribing increases (Fig. 4.3B). In addition, the level of transcription in active cells can also fluctuate (Fig. 4.3C). We therefore sought to determine the extent to which each regulatory strategy gives rise to tissue-wide control of the mean mRNA production rate. To this end, we expressed the total bulk transcriptional activity in terms of the quantitative contribution of each regulatory strategy as

$$\underbrace{\frac{\sum_i fluo_i(t)}{N_{total}}}_{\substack{\text{mean tissue}\\ \text{transcription rate}}} = \underbrace{\frac{\sum_i fluo_i(t)}{N_{active}(t)}}_{\substack{\text{mean transcription rate}\\ \text{of active cells}}} \times \underbrace{\frac{N_{active}(t)}{N_{total}}}_{\substack{\text{instantaneous fraction}\\ \text{of active cells}}} . \quad (4.1)$$

Here, $fluo_i(t)$ is the fluorescence of the i-th cell at time point t, $N_{active}(t)$ is the instantaneous number of active cells, and N_{total} is the total number of cells.

In order to determine how the resulting tissue-level transcriptional dynamics arises from the two contributions on the right-hand side of Equation 4.1, we first determined the tissue-wide transcription rate at each time point corresponding to the left-hand side of the equation, by adding the fluorescence of all spots in each frame and then dividing by the total number of nuclei in the field of view (Fig. 4.4B). This calculated

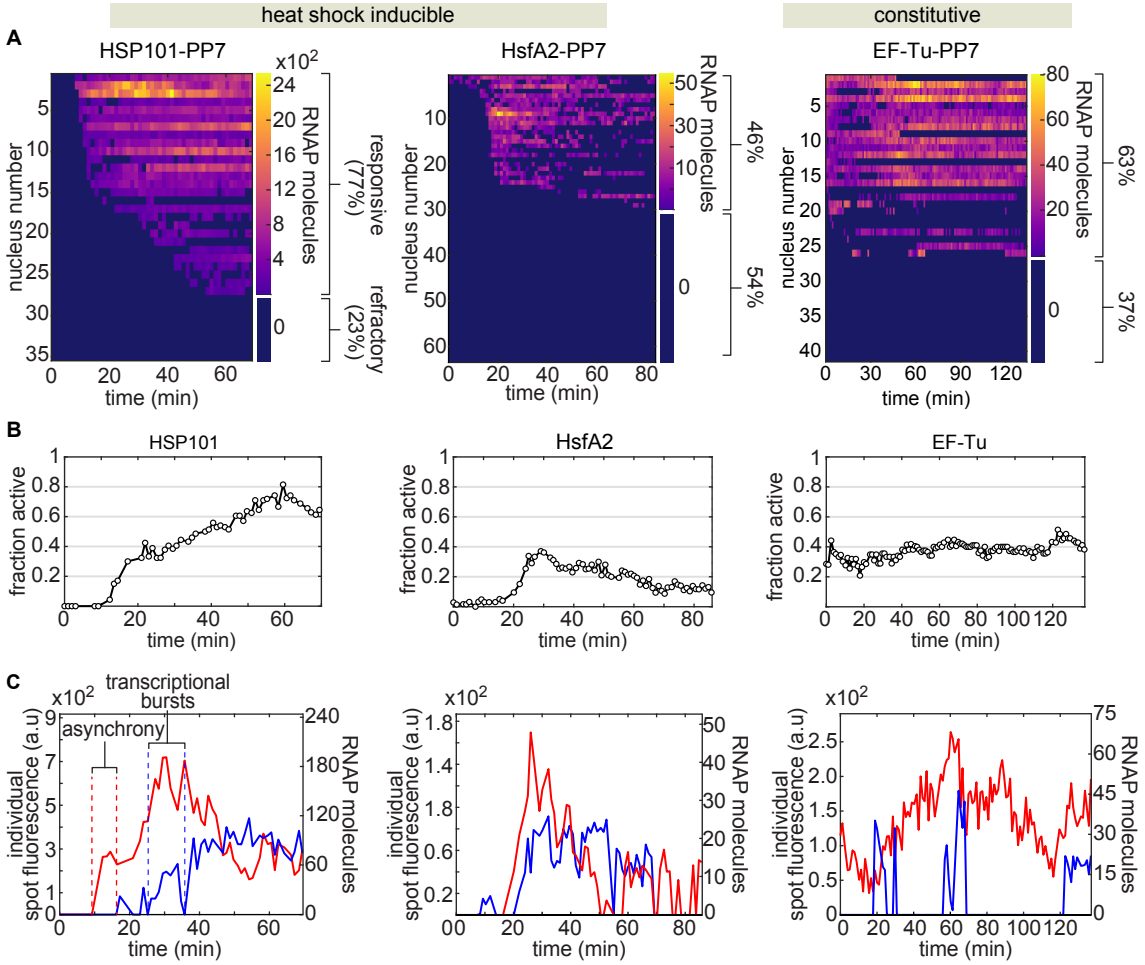

Figure 4.3: **Single-cell control of transcriptional activity in response to heat shock in Arabidopsis. (A)** Heat maps of spot fluorescence in all nuclei (rows) over time (columns) across the the field of view in HSP101-PP7-1, HsfA2-PP7-1, and EF-Tu-PP7-1 plants. Dark blue represents the absence of detectable signal. The size of the colorbar on the right of each heatmap shows the proportion of nuclei that exhibited activity in at least one frame during the experiment (¿68 min) to refractory cells that presented no spots. **(B)** Instantaneous fraction of actively transcribing nuclei measured as the number of nuclei with spots divided by the total number of nuclei in the field of view. **(C)** Fluorescence time traces of two representative transcription spots in the same field of view shown in red and blue. Upon induction, transcriptional onset can occur asynchronously and transcriptional activity occurs in bursts, modulating the instantaneous fraction of transcriptionally active nuclei in (B).

tissue-wide transcription rate is akin to the data typically obtained using a time series of bulk sampling experiments. The tissue-wide transcription rate of HSP101-PP7-1 and HsfA2-PP7-1 rose upon induction, while that of the constitutive EF-Tu-PP7-1 reporter line remained constant throughout the experiment (Fig. 4.4E).

To determine whether the graded modulation of the transcription rate among active cells contributes to the mean tissue transcription rate, we calculated the mean spot fluorescence across actively transcribing cells only (first term on the right-hand side of Eq. 4.1, Fig. 4.4C). Further, to determine the contribution of the switch-like type of regulation, we computed the instantaneous fraction of cells in which we detect reporter activity (second term on the right-hand side of Eq. 4.1, Fig. 4.4D). Our calculations revealed that the temporal modulation of the transcription rate among active cells remained relatively constant throughout induction (Fig. 4.4F). In contrast, the fraction of active nuclei was strongly modulated as a result of induction (Fig. 4.4G). The dynamics of the fraction of active cells were qualitatively comparable to the mean tissue transcription rate (compare Fig. 4.4E and G).

To quantify the relative contribution of each of these regulatory strategies to the overall transcriptional dynamics, we measured the fold-change of each term in Equation 4.1. We defined this fold-change as the ratio between the value of each magnitude at peak induction (blue and green arrowheads in Fig. 4.4E-G) and at 10 min, shortly after the beginning of the response (grey arrowhead in Fig. 4.4E-G). For both heat-inducible promoters, the fold-change in the mean transcription rate across active cells was close to one, indicating no significant change over time (Fig. 4.4H). In contrast, the fold-change in the instantaneous fraction of active cells was almost identical to that of the total activity (Fig. 4.4H).

To determine the generality of our results, we performed these experiments and analysis on a second set of independent transgenic lines of all three promoters. Our analyses yielded similar results (Fig. 4.20). In addition, we asked if these findings also apply to other tissues. Measurements of HsfA2-PP7 expression in root tips showed that, indeed, the rate of transcription of responsive cells is stable while the number of active nuclei is modulated over time (Fig. 4.4E and Fig. 4.20).

Thus, the duration of the treatment does not impact the rate of transcription of individual actively transcribing cells: when an individual cell transcribes, it tends to do so, on average, at a characteristic, relatively stable level regardless of induction time (Fig. 4.21). Instead, the time under stress modulates the tissue-wide transcription rate by increasing the probability that each individual cell switches from basal undetectable transcription to a high activity state.

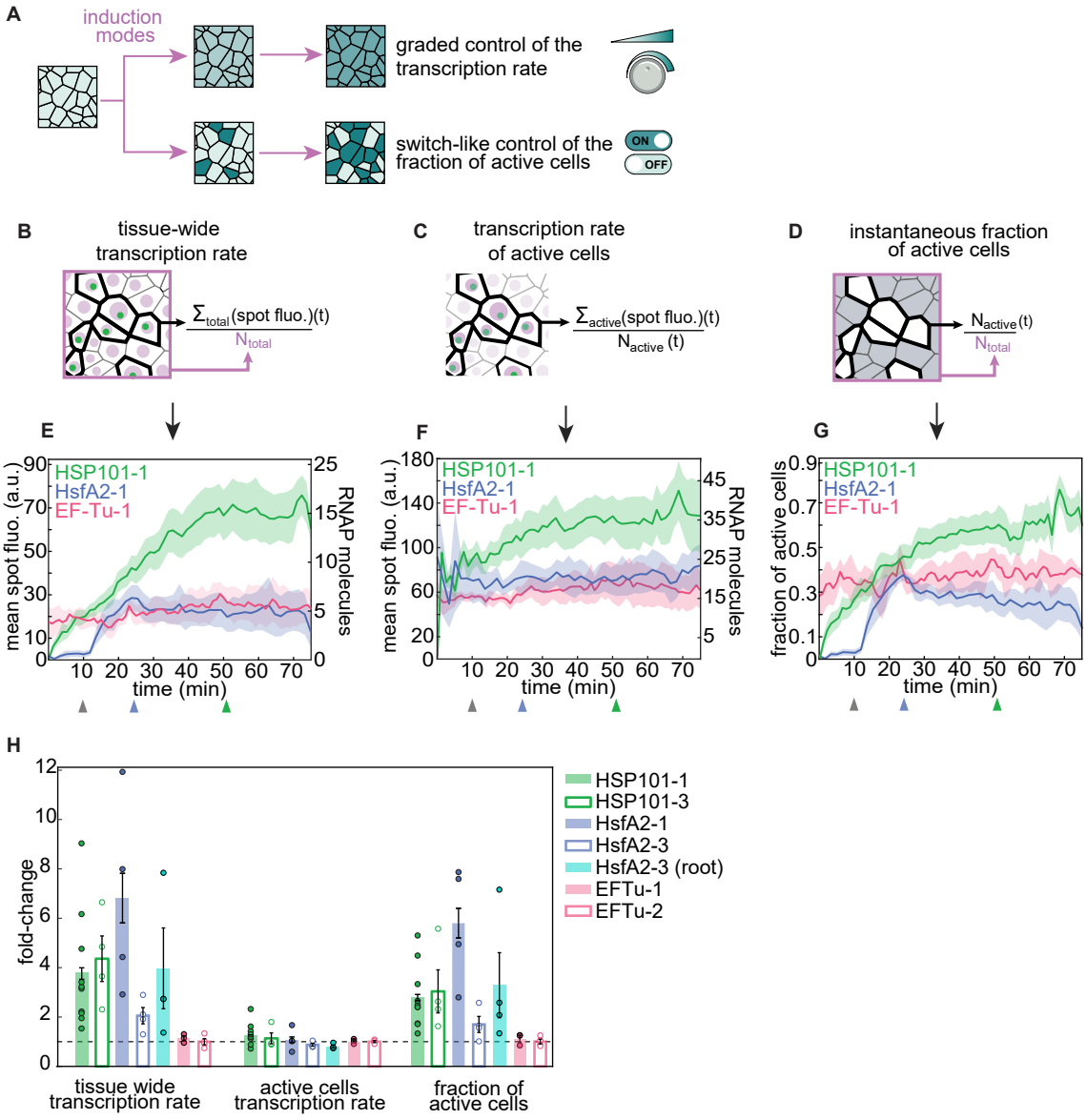

Figure 4.4: **Single-cell regulatory strategies determining tissue-wide transcriptional dynamics.** **(A)** Tissue-wide transcriptional control can be achieved through two non-exclusive regulatory modes: the graded modulation of the rate of transcription across cells, or the switch-like regulation of the fraction of actively transcribing cells. **(B)** The tissue-wide transcription rate is obtained by—at each time point—adding up the fluorescence of all spots and dividing by the total number of nuclei N_{total}, regardless of their transcriptional state. **(C)** The transcription rate of active cells is calculated by, in each frame, adding the fluorescence of all spots and dividing by the number of nuclei with spots in that frame $N_{active}(t)$. **(D)** The fraction of active cells corresponds to the number of nuclei that have detectable reporter transcription at a given time $N_{active}(t)$ divided by the total number of nuclei N_{total}. ***Caption continues on next page***.

Figure 4.4: ***Continued from previous page*: Single-cell regulatory strategies determining tissue-wide transcriptional dynamics. (E-G)** Arabidopsis lines carrying inducible promoters HSP101-PP7-1 (green) and HsfA2-PP7-1 (blue), and a line with the constitutive reporter EF-Tu-PP7-1 (red). Time $t = 0$ corresponds to the frame at which spots were first detected. **(E)** Mean tissue transcription rate. **(F)** Mean transcription rate across active cells. **(G)** Mean instantaneous fraction of actively transcribing cells. **(H)** Fold-change in the mean tissue-wide transcription rate compared to the fold-change in the mean transcription rate of active cells and in the fraction of active cells, defined as the ratio between the value at its peak and at $t = 10min$ for HSP101-PP7-1 (gray vs. green arrowheads in B) and HsfA2-PP7 (gray vs. blue arrowheads in B). For EF-Tu-PP7 10 and 30 minutes were used to calculate the fold-change. The empty and light blue bars correspond to data obtained from independent transgenic lines shown in Figure 4.20. The horizontal dashed line indicates a fold-change of 1. (E-H, shaded regions and error bars are SEM calculated across 8, 5, and 3 experimental replicates for HSP101-PP7-1, HsfA2-PP7-1, and EF-Tu-PP7-1, respectively.)

4.3.5 Allele-specific regulation underlies most tissue-wide heterogeneity in mRNA production in living plants

Although physiological responses occur at the tissue level, each cell must bear the phenotypic consequences of its individual gene regulatory behavior in response to stress. Studies of microorganisms and mammalian cells in culture have revealed that single-cell transcriptional responses to outside stimuli are often highly variable, leading researchers to posit that organisms possess mechanisms to buffer this "noise" or to leverage variability to drive the adoption of cellular fates that, for example, provide resistance against environmental insults such as antibiotics [212, 177, 75]. However, remarkably little is known about the level, functional roles, and underlying molecular mechanisms of transcriptional noise in shaping stress responses in multicellular systems like plants [54, 183].

Although, on average, the rate of transcription of our heat-responsive reporters in active cells did not substantially change with the duration of the heat treatment (Fig. 4.4C), at any given time point, the levels of activity across cells spanned more than two orders of magnitude (Fig. 4.5A). This behavior of actively transcribing cells, combined with asynchronous activation (Fig. 4.3A) and the presence of cells that are transiently or permanently transcriptionally inactive (Fig. 4.4G, Fig. 4.13), gives rise to a wide distribution in the inferred mRNA produced per cell (Fig. 4.5B). This distribution spans more than three orders of magnitude, with a coefficient of variation (CV, standard deviation divided by the mean) of approximately 1.6. While this variability might seem exceedingly high, it is on the same order that other eukaryotic systems [245, 213, 19]. Simulating a constant, homogeneous mRNA degradation rate does not considerably alter the spread of these distributions (Fig. 4.22).

What are the molecular sources of this cell-to-cell variability in the amount of mRNA produced (Figure 4.5C)? A traditionally held view invokes differences in composition across cells [40, 259]. For example, differences in cell cycle stage [224, 214], concentration of general transcriptional machinery [279], or concentration of specific transcription factors [183] can generate cellular heterogeneity (Fig. 4.5C, left). Alternatively, because at the local gene level transcription depends on a relatively small number of molecules, it is subjected to the stochasticity inherent to biochemical reactions. This can lead to variability even among otherwise identical cells (Fig. 4.5C, right).

To distinguish between these two types of sources of noise, it is necessary to compare the expression of alleles belonging to the same cell with that of alleles in nearby cells [76, 214]. Intuitively, factors extrinsic to the gene that operate at the cellular level will lead to alleles in a cell behaving similarly to each other but differently to those in other cells. In contrast, processes intrinsic to the gene operating at the local level will lead to alleles in the same cell behaving different to each other even if they are exposed to the same extrinsic factors. By decomposing the total variability into variability across allele pairs within each cell and variability across cells, extrinsic and intrinsic sources can be quantified without *a priori* knowledge of their molecular identity [76] (see Section 4.7.1).

A previous measurement of gene expression noise in Arabidopsis using constitutively expressed fluorescent proteins found that extrinsic factors explain most of their cellular heterogeneity [8]. However, it is unclear how noise in accumulated protein relates to transcriptional variability that we can now measure using PP7, and whether there are differences between constitutive and regulated promoters.

To determine the contribution of each type of transcriptional noise, we imaged T2 Arabidopsis individuals homozygous for the reporter, which display up to two fluorescent spots per nucleus in diploid cells (Fig. 4.5D, top; Movie 6). Four traces originating from two nuclei indicate that the transcriptional activity of alleles in the same nucleus can be more similar to each other than the activity of alleles in different nuclei (Fig. 4.5D, bottom), suggesting a significant role of extrinsic noise in transcriptional variability. However, our measurements also revealed that allele pairs in the same nucleus are not necessarily in the same transcriptional state: nuclei are approximately equally divided between populations where two, only one, or no alleles exhibit a transcription spot (Fig. 4.5E). This suggests that the decision of alleles to become active is intrinsic to each allele. Thus, qualitatively, we have identified that both intrinsic and extrinsic contributions can potentially underlie the total transcriptional noise.

In order to determine the quantitative contribution of each source of variability to the single-cell distribution of mRNA produced, we followed [76] (see details about this calculation in Section 4.7.1 and Figure 4.23). Transgenes in Arabidopsis are often inserted as tandem repeats [140, 141], which cannot be optically resolved from each other. We used qPCR to determine the number of tandem insertions in HSP101-PP7-1 and HsfA2-PP7-1 and found that these lines are likely to contain 2 and 3 transgenes per locus, respectively. To show that the results from this noise analysis do not qualitatively depend on the number of transgene copies per insertion, we identified additional single

insertion Arabidopsis lines (HSP101-PP7-2 and HsfA2-PP7-2) for which we confirmed the presence of a single transgene copy per insertion locus using qPCR (see Figure 4.10 and associated calculations in Section 4.7.1).

Figure 4.5F presents the integrated spot fluorescence of alleles pairs belonging to the same nucleus in homozygous plants of HSP101-PP7-1 and two additional lines with a single transgene copy per insertion. Our calculation of the noise components revealed that intrinsic sources explain most ($\sim 2/3$) of the variability in all of the lines tested (Fig. 4.5G).

We next sought to further investigate possible sources contributing to the extrinsic noise. Studies in plants [130, 183], mammalian cell culture [199], and yeast [178] have shown that cell size is positively correlated with gene expression, making it a potential source of extrinsic noise. We found that nucleus volume (a good proxy for cell size [217]) explains only 10-30% of the cell-to-cell variability in expression (Fig. 4.24). The lack of a strong correlation between transcription and nucleus size might be due to all nuclei being relatively similar in size (Fig. 4.11). An additional source of extrinsic noise could be cell type identity. For example, the expression dynamics of guard cells and non-guard cells, both present in our field of view, could contribute to this noise. As shown in Figure 4.26, we did not find a consistent, statistically significantly difference between guard cells and the rest of the cells. Thus, the molecular identity of the sources of extrinsic noise remain to be identified.

In sum, despite the presence of extrinsic noise, our results demonstrate that most of the cellular heterogeneity in the transcriptional response to heat shock is not due to cells having a different chemical composition. Instead, stochastic processes at the level of each individual allele explain most of the cell-to-cell differences in the amount of mRNA produced per cell. Importantly, while here we have focused on the noise in the amount of produced mRNA, further insights can be drawn from examining the sources of molecular variability in, for example, instantaneous transcriptional activity (Fig. 4.27).

4.4 Discussion

Over the last few decades, it has become clear that the averaging resulting from bulk tissue sampling obscures important details about the spatial control of cellular processes in plants and animals alike. In plants, this limitation has motivated recent advances in single-cell RNA sequencing [180]. However, these measurements depend on the previous history of RNA transcription and degradation and thus obscure information about regulatory dynamics. Further, single-cell sequencing technologies tend to sacrifice spatial information. While enabling technologies to light up the process of transcription and its control in real time, in single cells or whole animals, have been developed [190, 262], plants have remained surprisingly sidelined.

Here, by implementing the PP7 and MS2 systems to fluorescently label nascent RNA molecules in plants, we have shown, to our knowledge for the first time, that it is pos-

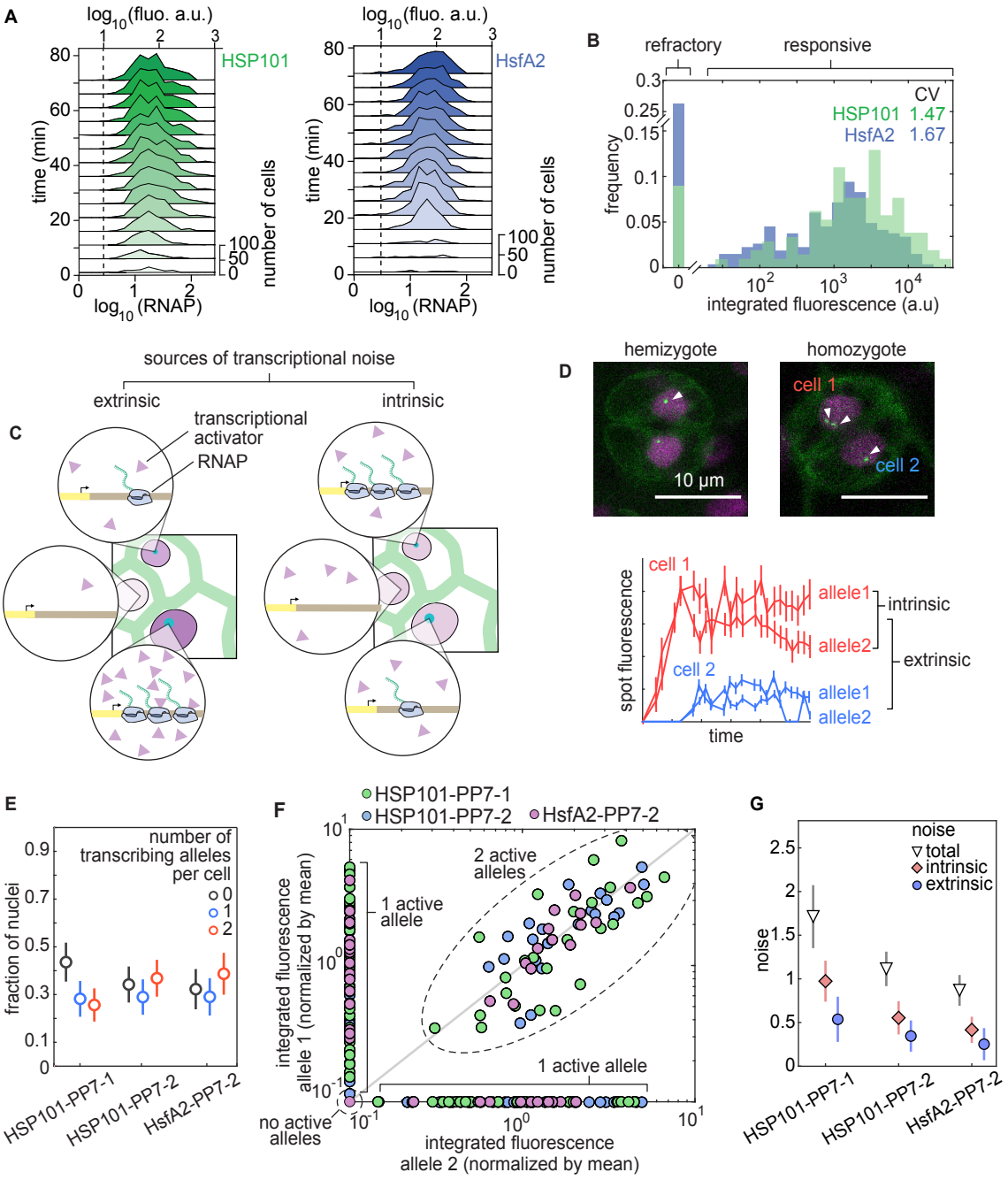

Figure 4.5: **Allele-specific processes explain most of the cellular heterogeneity in produced mRNA in Arabidopsis.** ***See caption on next page.***

Figure 4.5: **Continued from previous page: Allele-specific processes explain most of the cellular heterogeneity in produced mRNA in Arabidopsis. (A)** Histograms of spot fluorescence over time for the combined replicates of Figure 4.4. The dashed line indicates the detection threshold determined in Figure 4.2D. **(B)** Histograms of predicted total produced mRNA per cell across all replicates from Figure 4.4. **(C)** Schematic of extrinsic (left) and intrinsic (right) sources of transcriptional noise. Extrinsic noise arises from cellular differences in the abundance of regulatory molecules (purple triangles), such as transcription factors, whose abundance is common to all alleles, while intrinsic noise captures differences among cells with identical composition due to local processes at each allele such as the inherent stochasticity of biochemical reactions. **(D)** Two-allele experiment to decompose the total transcriptional variability into intrinsic and extrinsic noise. Top: guard cells—which are obligate diploids [182]—expressing HSP101-PP7. White arrowheads indicate transcription spots corresponding to one or two alleles of the reporter transgene in homologous chromosomes. In the homozygote, it is possible for only one allele to be active in different cells. Bottom: spot fluorescence traces from homozygous cells shown on top, the error bars correspond to the uncertainty in fluorescence quantification as described in the Materials and Methods. **(E)** Fraction of nuclei with zero, one, or two spots in heat shock-treated homozygous plants at the frame with the maximum number of visible spots. **(F)** Scatter plot of the integrated spot fluorescence normalized by the mean for pairs of alleles belonging to the same nucleus. Undetected spots were assigned a value of zero and plotted on the x- and y-axes. **(G)** Decomposition of the total variability in (F) into extrinsic and intrinsic components shows comparable contributions of both components to the total noise, with the intrinsic component explaining most of the variability. Error bars in (E) and (G) are bootstrapped errors.

sible to count the number of RNAP molecules actively transcribing individual alleles in single living cells of tobacco and Arabidopsis as they respond to their environment. This technical advance yielded unprecedented access to the temporal history of activity of individual alleles, making it possible to uncover distinct modes by which single-cell transcriptional activity in plants leads to tissue-wide gene expression dynamics.

Using this technique, and consistent with similar observations in other systems [98, 154, 114], we discovered a fraction of transcriptionally refractory cells that do not transcribe above our detection limit of approximately three active RNAP molecules per gene, regardless of induction conditions (Fig. 4.4D). Single-molecule RNA FISH experiments in Arabidopsis roots found that, at any given time, $\approx 20\%$ of cells are transcriptionally inactive for the constitutively expressed *PP2C* gene [71]. However, unlike the live-imaging approach developed here, single-molecule RNA FISH relies on fixed samples; it cannot determine whether this inactive state was transient or stable. Arguably, what we refer to as inactive nuclei might be transcribing at a low, basal rate and not be completely transcriptionally silent. However, in cells that divide slowly such as plant cells, extremely

infrequent transcription is sufficient to sustain low mRNA levels, particularly if these mRNA molecules have long half lives. Thus, it is not rare for lowly expressed genes to be free of polymerases for tens of hours in any given cell, even though their mRNA is detected at the population level [225, 219, 243].

We also found that tissue-wide transcriptional induction dynamics are the result of the temporal modulation in the fraction of cells that switch to a transcriptionally active state, and not of the graded control of the transcription rate of active cells (Fig. 4.4C). This form of regulation has been hypothesized to be at play in the regulation of the FLC gene in response to temperature [7] and in the commitment to xylem cell fate in response to the VND7 transcription factor [261]. Using our technologies, it should now be possible to directly test these models.

These single-cell behaviors may seem hard to reconcile with previous bulk time course experiments showing that the mRNA molecules of inducible genes are present under control conditions and accumulate gradually in response to stress treatments [170, 55]. Yet, ample evidence from single-cell studies has shown that single cell observations rarely match the average cell behavior captured by bulk experiments [22, 195, 212, 57].

Gene expression can vary significantly from cell to cell in microbial and animal species [212]. By making it possible to measure cell-to-cell transcriptional variability in real time in living plant cells, we confirmed that plants are no exception to this widespread presence of transcriptional variability. The single-locus resolution of our method allowed us to determine that cell-to-cell variability in mRNA production arises mainly from stochastic processes instrinsic to each allele (Fig. 4.5G). Studies in *in-vitro* cell cultures have found that gene-expression noise can have profound consequences for cellular survival [74, 237]; however, the role of transcriptional noise in plant stress responses remains an open question [54, 222, 1]. We envision that the strategy applied here to systematically dissect transcriptional heterogeneity in Arabidopsis and tobacco will shed light on this interplay between transcriptional variability and stress response. Further, it will be interesting to examine how some unusual aspects of plant cell biology and genetics can buffer transcriptional noise. For example, cytoplasmic connections could play a role in short-range sharing of gene products [80], averaging out extrinsic noise as observed in syncytial systems [168]; multiple genome copies per nucleus in mature plant cells may provide further opportunities to average out intrinsic noise across alleles [156]. Similarly, we speculate that the conspicuous retention of large numbers of seemingly redundant gene paralogs in plants may also help buffer intrinsic fluctuations in individual genes [165].

Our approach requires access to a confocal microscope and to transgenesis tools, and should therefore be relatively easy to apply to many biological problems in plant development and physiology. However, imaging deep into live tissues with the resolution necessary to resolve diffraction-limited spots remains a challenge, particularly in plants [113]. Advances such as multiphoton imaging, lattice light-sheet microscopy, and adaptive optics will overcome this limitation [171].

Lacking single-polymerase resolution currently limits the applicability of MS2 and PP7 to genes transcribed at relatively high rates. A transcription initiation rate of 1 RNAP/min, corresponding to our detection limit of 3 elongating RNAP molecules on an average Arabidopsis gene, could be sufficient to sustain slow transcriptional processes operating at long developmental timescales. For example, the FLC gene, a key seasonal developmental regulator in Arabidopsis is rarely occupied by more than one elongating RNAP at a time [130] which may explain why previous attempts at visualizing nascent FLC mRNAs in live Arabidopsis plants have failed [275]. Increasing the number of stem loops repeats could be a viable strategy to enable the measurement of weakly expressed genes [253]. A growing interest in live imaging of transcription combined with advances in fluorophore chemistry [134] as well as in the PP7 and MS2 technologies themselves [274] offer hope for breaking this detection threshold.

It will undoubtedly be of interest to correlate the activities of genes by visualizing their transcription simultaneously. This multiplexing is already possible for two genes using MS2 and PP7. A third color could be added by implementing interlaced MS2 and PP7 loops [125]. To further extend the palette, it should be possible to engineer other orthogonal RNA-binding proteins-RNA aptamer pairs [61, 142].

Finally, and more generally, the random integration of transgenes in plants and their associated genomic rearrangements [141] makes it challenging to dissect the role(s) of regulatory sequences at their endogenous genomic locations. In addition, if the goal is to study the behavior of endogenous genes, reporter constructs might not be sufficient since they may not faithfully recapitulate all aspects of endogenous regulation. Delivery of DNA to specific genomic locations using CRISPR/Cas9 [45, 184] or sequence-specific recombinases [5] promise to address these problems and unleash the potential of quantitative reporters of gene expression.

In this study, we focused on a simple step in the plant's use of temperature as a signaling input. More complex treatments have been previously used to show that plants can mount specific responses to inputs, such as memory in response to pulses of heat shock [47] and nonlinear integration of combinations of high light and temperature stress [283]. By administering experimental treatments while simultaneously measuring their effects on gene regulation, it will be possible to determine how these operations are performed at the cellular level. In addition, the sub-nuclear resolution of nascent RNA tagging could make it possible to resolve long-standing issues in plant signaling, such as the role of protein aggregates or "nuclear speckles" that are pervasive in light-responsive signaling pathways in plants [223].

In conclusion, by enabling the measurement of transcription at high spatiotemporal resolution, the PP7 and MS2 methods introduced here close a critical technological gap in plant biology. These new technologies open new avenues of inquiry and will make it possible to quantitatively interrogate transcriptional control in living plants and to engage in the discourse between theory and experiment that has characterized the study of gene regulation in single cells and animal tissues over the last two decades[206].

4.5 Materials and Methods

4.5.1 Plasmids and Agrobacterium strains

All plasmid sequences used in this study can be accessed from a public Benchling folder. All plasmids used in this study are available from Addgene. All vectors were based on pCambia derivatives [116] and transformed into the GV3101::pMP90 Agrobacterium strain by electroporation. Plasmids confering Kanamycin resistance in plants (i.e., reporter constructs) were based on pCambia2300. Plasmids confering Hygromycin resistance in plants (i.e., PCP, MCP and nanocages constructs) were based on pCambia1300. A list of all the plasmids used in this study along with their link to Benchling and Addgene can be found in Table 4.1. The Arabidopsis gene identifiers associated with genomic sequences used in these plasmids are listed in Table 4.2.

4.5.2 Plant growth conditions

Nicotiana benthamiana (tobacco) plants were grown in a greenhouse under natural light conditions prior to agroinfiltration. Following infiltration, tobacco plants were kept under 30 μE of constant light. Arabidopsis plants used for experiments were grown in 1/2 strength MS agar containing 50 μg/μl of Kanamycin under short day conditions (8 hours of 30 μE light per day) for four to six weeks prior to imaging.

4.5.3 Agroinfiltration

Agrobacterium glycerol stocks were streaked on LB plates containing 50 μg/μl Kanamycin and 50 μg/μl Gentamycin. Fresh colonies were grown overnight in liquid LB containing the same antibiotic concentrations, spun down and resuspended in an equal volume of infiltration buffer (10 mM MES pH5.6, 10 mM $MgCl_2$, 150 μM Acetosyringone). Cells were incubated for 2-4 hours in infiltration buffer shaking at room temperature after which the cultures were diluted 1:3 to an OD_{600} of approximately 0.3. In experiments that required combining strains, coat protein and reporter strains were mixed in a 3:1 ratio (the exact ratio does not qualitatively affect the results). In PP7 and MS2 experiments, infiltrated leaves were imaged approximately 2 days after infiltration. For absolute calibration experiments, plants were imaged 12-18 hours after infiltration.

4.5.4 Generation of transgenic Arabidopsis lines

To generate lines carrying both PCP-GFP and PP7 reporters we followed a sequential transformation approach. We first selected PCP-GFP lines in 35 μg/ml of Hygromycin and kept lines exhibiting moderate levels of fluorescence and no obvious growth phenotype. Next, we transformed T1 or T2 PCP-GFP individuals with PP7 reporter Agrobacterium strains and selected transformants in 50 μg/ml Kanamycin and 35 μg/ml Hygromycin.

Individuals T1 for the PP7 construct were screened for nuclear mScarlet fluorescence and presence of transcription spots matching previous knowledge about the activity of the corresponding endogenous gene. In all cases, to select for antibiotic resistance we followed the protocol by [120]. A list of the lines used in this study can be found in Table 4.3.

4.5.5 Determining the number of unlinked reporter transgene insertions

To select lines carrying a single insertion reporter locus we plated approximately 60 T2 seeds in MS plates containing Kanamycin and counted the ratio of survivors. This ratio was divided by the survival ratio in plates containing no antibiotics. A χ^2 test was used to determine whether the product of these two ratios was statistically different from the expected ratio of 3/4. To confirm the absence of two or more unlinked reporter loci we examined transcription spots in guard cells. Unlike other leaf cell types, these cells are exclusively diploid [182] and therefore the presence of a single spot per guard cell nucleus in a T1 individual confirms the absence of unlinked insertions.

4.5.6 Heat shock treatments

To control the sample temperature in the microscope stage we used an OkoLabs H101-LG temperature chamber calibrated to achieve a maximum of ≈ 39°C. The temperature experienced by the sample was calibrated using an electronic probe. The walls of the chamber were kept at 54°C for the sample to reach a steady state temperature of 39°C. To estimate the difference in temperature between the center of the field of view and its edges, we simplify the problem by approximating it to a radial temperature gradient going outwards from the center of the sample, with the center being at the sample temperature (39°C) and the edge at the temperature of the walls of the chamber (54 °C), located 5cm away from the center. We can then use a linear approximation for the temperature gradient, which results in a gradient of 0.0003°C/um. This means that the difference of temperature from the center of the field of view to its edge is 0.0003°C/um x 45 um $\approx$ 0.015°C. The heat shock treatment used for the RT-qPCR experiment in Figure 4.2A was performed as follows: whole 4-6 week-old plants were placed in 1.7 ml plastic tubes containing 200 μl of water. The sample corresponding to time = 0 minutes was immediately taken out of the tube, quickly tapped dry, transferred to a new tube containing silica beads and frozen in liquid nitrogen. The rest of the samples were transferred to a 39°C heat block and removed at set times. Plants were then quickly tapped dry and frozen in liquid nitrogen.

4.5.7 Microscopy setup and image acquisition

In tobacco experiments, a piece of infiltrated leaf spot was mounted in water between a glass slide and a glass coverslip with the abaxial (bottom) side facing the objective. In Arabidopsis experiments, full 2-4 day old leaves from 4-6 week old plants were detached and mounted in tap water between a gas permeable cellophane membrane (Lumox film; Starstedt) and a glass coverslip with the adaxial (top) side facing the objective. All samples were imaged close to the base of the leaf blade immediately after mounting. All data was taken in a Leica SP8 confocal microscope with a white light laser using a 63X oil objective. The dimensions of the field of view were 92.26 x 46.09 μm using 1052 x 512 pixels, resulting in a pixel size of 90 nm. Z stacks consisting of 25 slices of 0.5 μm each were taken every 60 seconds accumulating fluorescence 3 times over lines. The beginning of each stack was set to the upper-most nucleus in the leaf epidermis. For GFP, excitation 488 nm and emission 498-559 nm. For mScarlet, excitation 569 nm, emission 579-630 nm. For Chlorophyll, excitation 488 nm, emission 665-675 nm. To ensure quantitative consistency across experiments, the 488 nm laser power was calibrated to 10.5 μW ($\approx$ 5% laser power) at the beginning of each imaging session using a power meter. The percentage intensity of the 569 nm laser line was kept consistent across experiments at 5%. To minimize the background signal from endogenous plant fluorophores we used the gating function of the HyD detectors to limit detection to a time window between 0.3–6 ns after excitation.

4.5.8 RT-qPCR

Total RNA was extracted using the Quiagen RNeasy kit following the manufacturer instructions. Reverse transcription was performed using the Qiagen Omniscript kit with a primer mix of random 10mers (10 μM final concentration) and 15mer oligo dT primers (1 μM final concentration). A negative control was performed adding water instead of reverse transcriptase. mRNA abundance was calculated by the $\triangle$CT method [172]. Primers for endogenous HSP101 were 5'GGTCGATGGATGCAGCTAAT and 5'CTTCAAGCGTTGTAGCACCA from [280]. Primers for the Actin2 standard were 5'CGCTCTTTCTTTCCAAGCTCAT and 5'GCAAATCCAGCCTTCACCAT from [169]. Primers for the reporter mRNA were 5'GGGTTCATCAGAGTGCCAGAG and 5'AGGCAGAGCGACACCTTTAG. A negative control was performed under identical conditions replacing the RT enzyme with water.

4.5.9 Image analysis: Spot fluorescence and tracking

Raw image stacks of the coat protein channel were used to identify fluorescent punctae corresponding to transcription spots using the ImageJ implementation of the 3D Trainable Weka Segmentation toolbox [10]. Following [154], after segmentation, spots in each z-slice were fitted to a 2D Gaussian. The z-slice with the largest Gaussian amplitude was selected for the spot fluorescence calculation. Spot fluorescence corresponds to the

sum of pixel intensity values in a circle with a radius of 1.08 μm centered around the center of the fitted Gaussian minus the background fluorescence offset.

The fluorescence error per spot shown in Figure 4.1E and Figure 4.5D was obtained based on the approach from [98]. First, in each frame we calculate the fluorescence offset from the fitted baseline obtained from the Gaussian fitting procedure described above. This results in a time trace of offset values for each spot time trace (see Fig. 4.11G for an example). Next, we fit a spline to this time trace and calculate the root-mean-square deviation of offset values with respect to the spline. This value represents the fluctuations of the background intensity *per pixel*. Finally, we multiply this deviation by the same integration area used for transcription spots to obtain an error in the same magnitude. False negative and false positive spots were corrected manually.

4.5.10 Image analysis: Nuclear segmentation and spot tracking

Maximum intensity projections of the nuclear marker channel were used for nuclear segmentation using the ImageJ implementation of the 2D Trainable Weka Segmentation toolbox [**arganda-carreras2017**] or a custom-written Matlab pipeline. False negative and false positive nuclei were then manually corrected. Spots were assigned to nuclei based on physical overlap. Tracking of spots over time was based on nuclear tracking and manually corrected whenever errors were found.

4.5.11 Image analysis: Nucleus fluorescence

A binary mask of segmented nuclei was applied to the PCP-GFP or Histone 2B- mScarlet channel. For each z-slice in each frame, the mean fluorescence across pixels within each nucleus area was calculated. As a result, in each frame, the fluorescence intensity of a given nucleus has the form of a ”column” of intensities over z. Next, in each frame we take the brightest z-slice in this column as the fluorescence value corresponding to the concentration of bright fluorescent protein in a given nucleus at a given time point.

4.5.12 Determining transgene copy number by qPCR

Genomic DNA was extracted from leaf tissue using CTAB and phenol:chlorophorm precipitation. Primers used to amplify the reporter transgene were 5’gacgcaagaaaaatcagagagatcc and 5’ggtttctacaggacggaccatacac. Primers used to amplify a region near the *Lhcb3* gene used as an internal genomic control were 5’acaggtttggtcaagtcaattacga and 5’atggtttccatgaatactgaacacg. The final concentration of genomic DNA per reaction was 0.75 ng. For a more detailed explanation of the calculations and controls related to this experiment see Section 4.7.1.

4.5.13 Absolute calibration using nanocages

Tobacco leaves were infiltrated with agrobacterium strains containing plasmids where the promoter of the Arabidopsis UBC1 gene (1138bp upstream of the AT1G14400 start codon) was used to drive the 60mer monomer fused to either one or two mGFP5 coding sequences. The same scheme was used to express the monomers of the 60meric GEM. The N-terminus of the rabbit Cytochrome P450 CII1 was added as an N-terminal tag to target the protein fusions to the cytosolic side of the ER in order to slow down their diffusion. Samples were imaged no later than 16 hours after infiltration since long incubation periods resulted in the appearance of large GFP aggregates. To image the GFP nanocages in mesophyll cells, the abaxial epidermis was first removed. This is necessary to obtain a large number of structures in the field of view. The fluorescence of nanocages was calculated with the same analysis pipeline used for transcription spots. The imaging conditions were identical to the ones used in transcription experiments except that a 5 times stronger laser power was used for the 488 line in order to improve the signal. After obtaining the fluorescence of individual nanocages their fluorescence was divided by 5 prior to calculating their mean fluorescence. The validity of this operation is due to the linearity of fluorescence intensity and laser power under our conditions for both nanocages and PP7 spots (Figure 4.8).

4.6 Acknowledgments

We would like to thank Rob Phillips, Adrienne Roeder, Peter Quail, Setsuko Wakao, Christopher Gee, Avi Flamholz and Jake Brunkard for comments on the manuscript. We are very thankful of members of the Garcia lab, Meghan Turner and Gabriella Martini for sharing their knowledge related to the nanocages experiment and Yang Joon Kim for discussing results. We are particularly thankful to Gabriella Martini for setting up the microscope temperature chamber. We also acknowledge Allison Schwartz, Jose O'brien and Fernan Federici for sharing plasmids, and Albert Lin and Jonathan Liu for their feedback regarding calculations. HGG was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface, the Sloan Research Foundation, the Human Frontiers Science Program, the Searle Scholars Program, the Shurl and Kay Curci Foundation, the Hellman Foundation, the NIH Director's New Innovator Award (DP2 OD024541-01), and an NSF CAREER Award (1652236). KKN is an investigator of the Howard Hughes Medical Institute.

4.7 Supplementary Information

4.7.1 Calculations

Incorporating a constant degradation rate into the calculation of total produced mRNA from microscopy

As noted by Garcia2013 and explained in Figure 4.10, the total number of transcripts produced by a locus can be obtained by integrating the area under the curve of a time trace of spot fluorescence. Here, we show how we incorporate mRNA degradation to estimate the mRNA abundance at a given time point.

The rate of change in mRNA, dM/dt, can be described by the sum of a production rate r and a degradation rate γ

$$\underbrace{\frac{dM}{dt}}_{\text{change in mRNA}} = \underbrace{r(t)}_{\text{production}} - \underbrace{\gamma(t)M(t)}_{\text{degradation}} . \tag{4.2}$$

As demonstrated by Bothma et al. [38] and Lammers et al. [154], the rate of mRNA production is proportional to the spot fluorescence. In addition, for the sake of simplicity we will assume that the degradation rate is constant. Hence, Equation 4.2 becomes

$$\frac{dM}{dt} = kFluo(t) - \gamma M(t), \tag{4.3}$$

where k is the proportionality constant between spot fluorescence and transcription rate. Equation 4.3 indicates that, to calculate the change in the number of mRNAs between two time points t and $t + \Delta t$, we need to know the number of mRNAs produced between these time points and subtract the number of mRNAs degraded. The mRNAs added between t and $t + \Delta t$, for time steps shorter than the transcriptional dynamics of the system are

$$\text{mRNA added} = \int_0^{t+\Delta t} Fluo(t) - \int_0^{t} Fluo(t) = \int_t^{t+\Delta t} Fluo(t), \tag{4.4}$$

which is equivalent to the sum of spot fluorescence values per frame up to time $t + \Delta t$ minus the sum up to time t. On the other hand, the number of mRNAs degraded between t and $t + \Delta t$ corresponds to the number of mRNAs at time t that decay with a rate γ (with units of $time^{-1}$)

$$\text{mRNA degraded} = \gamma \times mRNA(t). \tag{4.5}$$

The change in mRNA from time t to $t + \Delta t$ is therefore

$$\text{mRNA change} = \text{mRNA added} - \text{mRNA degraded} \tag{4.6}$$

$$\text{mRNA change} = \int_t^{t+\Delta t} Fluo(t) - \gamma mRNA(t). \tag{4.7}$$

This formula was applied to spot fluorescence data to infer the total mRNA produced in Figure 4.2B and Figure 4.5B, F and G. Note that, to calculate averages across spots, it is necessary for their sampling times to be identical. This might not the case when averaging across data sets due to sample adjustments during imaging, in which case the spot fluorescence traces were linearly interpolated to a rate of $\approx 7\ s$ per observation.

Determining transgene copy number by qPCR

In this section, we present our calculation for determining the number of transgene insertions from the ΔCT values resulting from qPCR taking the amplification efficiency into account. Given a starting number of DNA molecules N_0, the total number of molecules after C amplification cycles is given by

$$N(C) = N_0(2\epsilon)^C, \tag{4.8}$$

where ϵ corresponds to the amplification efficiency, or the fraction of molecules that are duplicated in each cycle. The number of amplification cycles CT necessary to amplify the number of DNA molecules from N_0 to N_{ct} can be described by

$$CT = \log_{2\epsilon}\left(\frac{N_{ct}}{N_0}\right). \tag{4.9}$$

Changing the logarithm base and rearranging leads to

$$CT = \frac{\log_2\left(\frac{N_{ct}}{N_0}\right)}{1+\log_2(E)}. \tag{4.10}$$

We now define an amplification efficiency constant K as

$$K = \frac{1}{1+\log_2(E)}. \tag{4.11}$$

Equation 4.10 then becomes

$$CT = K\log_2\left(\frac{N_{ct}}{N_0}\right). \tag{4.12}$$

To experimentally obtain K (and therefore ϵ), we perform qPCR on serial dilutions of template DNA, thus varying N_0. We then plot CT as a function of the $\log_2$ of the template concentration in order to obtain K from the slope (Fig. 4.10A,B). We used genomic DNA from a transgenic Arabidopsis plant to perform this amplification on the PP7 transgene as well as on an internal control genomic sequence. We measured both PCR reactions to have an efficiency of $K = 1$ within experimental error. As a result, we can determine the ratio between the initial number of transgene molecules N_0^t and the initial number of internal control molecules N_0^c by calculating the ΔCT

$$\Delta CT = CT^t - CT^c = K\log_2\left(\frac{N_{ct}}{N_0^t}\right) - K\log_2\left(\frac{N_{ct}}{N_0^c}\right) = \frac{N_0^c}{N_0^t} \tag{4.13}$$

If the transgene occurs in a single insertion locus containing a single transgene copy per insertion, then in a T1 individual

$$\frac{N_0^c}{N_0^t} = 0.5, \tag{4.14}$$

which corresponds to a ΔCT value of -1. Using this approach we were able to identify transgenic Arabidopsis individuals with a single insertion locus containing a single transgene insertion (Fig. 4.10C).

Decomposition of total variability into extrinsic and intrinsic noise

In this section we derive the formulas for the total, intrinsic and extrinsic noise (η^2_{tot}, η^2_{int}, and η^2_{ext}, respectively) based on the two-reporter approach developed by Elowitz et al. [76]. As noted by Hilfinger et al.[123] and explained at length by Fu et al. [87], these expressions stem from the law of total variance, which states that, for a random output variable A and a random input variable X, the total variance of A can be decomposed as the sum

$$\underbrace{\mathrm{Var}(A)}_{\text{total variance}} = \underbrace{\mathrm{Var}_X(\langle A|X\rangle_A)}_{\text{explained variance}} + \underbrace{\langle \mathrm{Var}_A(A|X)\rangle_X}_{\text{unexplained variance}}, \tag{4.15}$$

where the subscripts X or A indicate that the average or the variance is taken over different values of X or A, respectively.

Applied to the problem of gene expression variability, A represents the expression level of the gene of interest and X corresponds to the cellular state indicating, for example, the concentration in each given cell of all molecules that affect the expression of that gene such as RNAP. The first term on the right-hand side of Equation 4.15 is referred to as the explained variance and captures how much the average value of A varies across different values of X. The second term is referred to as the unexplained variance and captures how much the expression of A varies in cells that share the same value of X. See Figure 4.23 for a visual explanation of the law of total variance and Equation 4.15.

Because the identity and values of X are typically not known and/or not experimentally accessible, Elowitz et al. [76] devised a two-reporter system to determine the explained and unexplained components of the total normalized variance, which they termed extrinsic (η^2_{ext}) and intrinsic (η^2_{int}) noise, respectively. In this approach, each cell has two identical but distinguishable alleles of the gene of interest. In their statistical model, these two alleles are identical in all respects meaning that their distribution over cells and over time are the same. For the purpose of this derivation, let us call A_i and B_i the expression level of each allele in the i-th cell and normalize A and B to their means such that

$$\frac{A_i}{\langle A\rangle} = 1 + \delta A_i, \tag{4.16}$$

where δA_i is the fractional deviation of the expression level A_i from the mean $\langle A \rangle$. Similarly, for B we normalize to

$$\frac{B_i}{\langle B \rangle} = 1 + \delta B_i. \tag{4.17}$$

In the following calculations we will make use of the measurable quantities δA_i and δB_i to eliminate the unknown quantity X from Equation 4.15. We start by deriving an expression for η^2_{ext} defined here as the explained component of the total variance of the normalized δA distribution

$$\eta^2_{ext} = \mathrm{Var}_X(\langle \delta A_i | X \rangle_A). \tag{4.18}$$

Note that, since X is a random variable, so is $\langle \delta A_i | X \rangle_A$, and we can write its variance as

$$\eta^2_{ext} = \langle \langle \delta A_i | X \rangle^2_A \rangle_X - \langle \langle \delta A_i | X \rangle_A \rangle^2_X. \tag{4.19}$$

Because both alleles are identical, $\langle \delta A_i | X \rangle_A$ is equal to $\langle \delta B_i | X \rangle_B$, which allows us to write Equation 4.19 as

$$\eta^2_{ext} = \langle \langle \delta A_i | X \rangle_A \langle \delta B_i | X \rangle_B \rangle_X - \langle \langle \delta A_i | X \rangle_A \rangle_X \langle \langle \delta B_i | X \rangle_B \rangle_X. \tag{4.20}$$

Note that, in this model, the variability in the values of A_i and B_i for cells with the same X are independent of each other since we assume that they are not explained by X. Because of this independence, $\langle A_i \rangle \langle B_i \rangle = \langle A_i B_i \rangle$ for a given X. Applied to the first term in Equation 4.20, the extrinsic noise can be written as

$$\eta^2_{ext} = \langle \langle \delta A_i \delta B_i | X \rangle_{A,B} \rangle_X - \langle \langle \delta A_i | X \rangle_A \rangle_X \langle \langle \delta B_i | X \rangle_B \rangle_X. \tag{4.21}$$

We now note that the double angle brackets in the first term in the right-hand side of Equation 4.21 call for averaging the value of $\delta A_i \delta B_i$ in cells with the same X and then averaging again over all possible values of X. Similarly, the second term in the equation calls for averaging over A_i or B_i for a given X, and then averaging over X. This allows us to eliminate X in the equation and simplify our expression to

$$\eta^2_{ext} = \langle \delta A \delta B \rangle - \langle \delta A \rangle \langle \delta B \rangle, \tag{4.22}$$

which is the definition of covariance. Thus,

$$\eta^2_{ext} = \mathrm{Cov}(\delta A, \delta B). \tag{4.23}$$

This makes intuitive sense, as the model assumes that, since A and B are identical genes that respond to X in the exact same way, the variance in the expression of A that is explained by X is identical to the variance in the expression of B that is explained by X. As a result, the extrinsic noise measures how A and B coordinately vary across cells.

We now turn our attention to the derivation of the intrinsic noise, which we define as the unexplained component of the variance in the normalized A distribution, namely

$$\eta^2_{int} = \langle \mathrm{Var}_A(\delta A_i | X) \rangle_X. \tag{4.24}$$

Replacing the unexplained variance in Equation 4.15 with η^2_{int}, the explained variance by its formulation as extrinsic noise from Equation 4.23, and rearranging leads to

$$\eta^2_{int} = \mathrm{Var}(\delta A_i) - \mathrm{Cov}(\delta A_i, \delta B_i). \tag{4.25}$$

Because this equation does not involve X we don't need the subscripts anymore: all variances are calculated across values of δA and δB. We now note that the total variance of δA and δB must be the same since they have the same distribution over cells and over time. Therefore we are allowed to express the first term in the right-hand side of Equation 4.25 as the average variance of the δA_i and δB_i distributions

$$\eta^2_{int} = \frac{1}{2}\left[\mathrm{Var}(\delta A_i) + \mathrm{Var}(\delta B_i)\right] - \mathrm{Cov}(\delta A_i, \delta B_i). \tag{4.26}$$

Rearranging Equation 4.26 leads to

$$\eta^2_{int} = \frac{1}{2}\left[\mathrm{Var}(\delta A) + \mathrm{Var}(\delta B) - 2\mathrm{Cov}(\delta A, \delta B)\right]. \tag{4.27}$$

Now, using the identity stating that the variance of a sum is the sum of the variances minus twice their covariance, Equation 4.27 becomes

$$\eta^2_{int} = \frac{1}{2}\mathrm{Var}(\delta A_i - \delta B_i). \tag{4.28}$$

Finally, we define the total noise η^2_{tot} as the total variance of the normalized δA_i distribution. As noted before, because the distributions of δA_i and δB_i are identical, so are their variances. Therefore, the total noise can be calculated from the average

$$\eta^2_{tot} = \frac{1}{2}\left[\mathrm{Var}(\delta A_i) + \mathrm{Var}(\delta B_i)\right], \tag{4.29}$$

which satisfies

$$\eta^2_{tot} = \eta^2_{ext} + \eta^2_{int}. \tag{4.30}$$

Note that, here, we considered δA loosely as the "expression level" of gene A. This analysis can be applied to any metric of gene expression such as the instantaneous transcription rate, or the total amount of produced mRNA.

4.7.2 Biological material

Plasmids			
Plasmid Name	**Codes for**	**Function**	**Addgene**
UPG	AtUBQ10p::PCP-mGFP5 (hyg resistance in plants)	Ubiquitous expression of PCP-GFP fusion	161003
UPmCh	AtUBQ10p::PCP-mCherry (hyg resistance in plants)	Ubiquitous expression of PCP-mCherry fusion	161004
UMsfG	AtUBQ10p::MCP-sfGFP (hyg resistance in plants)	Ubiquitous expression of MCP-sfGFP fusion	161005
AL13Rb	PP7-Gus-Luc + AtUBQ10p::H2B-mScarlet (kan resistance in plants)	Promoterless PP7 reporter and red nuclear marker	161006
AL12R	AtUBQ10p::H2B-mScarlet + PP7-Gus-Luc (kan resistance in plants)	Promoterless PP7 reporter and Histone-mScarlet RFP nuclear marker	161007
AL13Rb-35S	35S-PP7 reporter in AL13Rb	Reports on 35S promoter activity and labels nuclei red	161008
AL13Rb-GAPC2	GAPC2-PP7 reporter in AL13Rb	Reports on Arabidopsis GAPC2 promoter activity and labels nuclei red	161009
AL12R-HSP70	HSP70-PP7 reporter in AL12R	Reports on Arabidopsis HSP70 promoter activity and labels nuclei red	161010
HSP70-pp7i-mCh-UPG	Arabiopsis HSP70 C-terminal mCherry fusion, intronic PP7	Reports on Arabidopsis HSP70 transcription activity and protein abundance	161011
AL13Rb-HsfA2	HsfA2-PP7 reporter in AL13Rb	Reports on Arabdiopsis HsfA2 promoter activity and labels nuclei red	161012

Table 4.1: List of Agrobacterium plasmids for expression in plants used in this stud. *Table continues on next page.*

Plasmids			
Plasmid Name	**Codes for**	**Function**	**Addgene**
AL12R-EF-Tu	EF-Tu-PP7 reporter in AL12R	Reports on Arabidopsis EF-Tu promoter activity and labels nuclei red	161013
AL12R-HSP101	HSP101-PP7 reporter in AL12R	Reports on Arabdiopsis HSP101 promoter activity and labels nuclei red	161014
UtB2N7	AtUBQ10p::tagBFP2-NLS	Nuclear localized blue fluorescent protein marker	161015
UBC1cer60G	AtUBC1::60mer-mGFP5	Weak ubiquitous expression of an ER-targeted 60mer monomer fused to mGFP5	161016
UBC1cer120G	AtUBC1::mGFP5-60mer-mGFP5	Weak ubiquitous expression of an ER-targeted 60mer monomer fused to two mGFP5	161017
UBC1cer40GEM	AtUBC1::40nmGEM-mGFP5	Weak ubiquitous expression of an ER-targeted monomer of a 40nm GEM fused to mGFP5	

Table 4.1: *Continued from previous page*: List of Agrobacterium plasmids for expression in plants used in this study.

Arabidopsis Gene Identifiers		
Gene abbreviation	**Gene name**	**AGI**
UBQ10	Polyubiquitin 10	AT4G05320.2
H2B	Histone 2B	AT5G22880.1
GAPC2	Glyceraldehyde-3-phosphate dehydrogenase C2	AT1G13440.1
HSP70	Heat shock protein 70	AT3G12580.1
UBC1	Ubiquitin carrier protein 1	AT1G14400.1
HSP101	Heat shock protein 101	AT1G74310.1
HsfA2	Heat shock transcription factor A2	AT2G26150.1
EF-Tu	GTP binding Elongation factor Tu family protein	AT1G07920.1

Table 4.2: Arabidopsis gene identifiers of the genes used for the constructs in this study.

Arabidopsis lines generated in this study		
Name	**Transgenes (refer to the 'Plasmids' table)**	**Usage**
UPG-6	UPG	For transformation with reporter constructs
UPG-9	UPG	For transformation with reporter constructs
AL13Rb-35S	UPG and AL13Rb-35S	Image 35S promoter activity in Figure 4.1
AL12R-HSP101-1	UPG and AL12R-HSP101	Image AtHSP101 promoter activity in Figures 4.2 to 4.5
AL13Rb-HSP101-2	UPG and AL13Rb-HSP101	Image AtHSP101 promoter activity in Figure 4.5
AL13Rb-HSP101-3	UPG and AL13Rb-HSP101	Image AtHSP101 promoter activity in Figure 4.20
AL13Rb-HsfA2-1	UPG and AL13Rb-HsfA2	Image AtHsfA2 promoter activity in Figures 4.3 and 4.4
AL13Rb-HsfA2-2	UPG and AL13Rb-HsfA2	Image AtHsfA2 promoter activity in Figure 4.5
AL13Rb-HsfA2-3	UPG and AL13Rb-HsfA2	Image AtHsfA2 promoter activity in Figure 4.20
AL12R-EF-Tu-1	UPG and AL12R-EF-Tu	Image AtEF-Tu promoter activity in Figures 4.3, 4.4 and Fig. 4.11
AL12R-EF-Tu-2	UPG and AL12R-EF-Tu	Image AtEF-Tu promoter activity in Figures 4.20

Table 4.3: List of transgenic Arabidopsis lines used for experiments.

4.7.3 Supplementary Figures

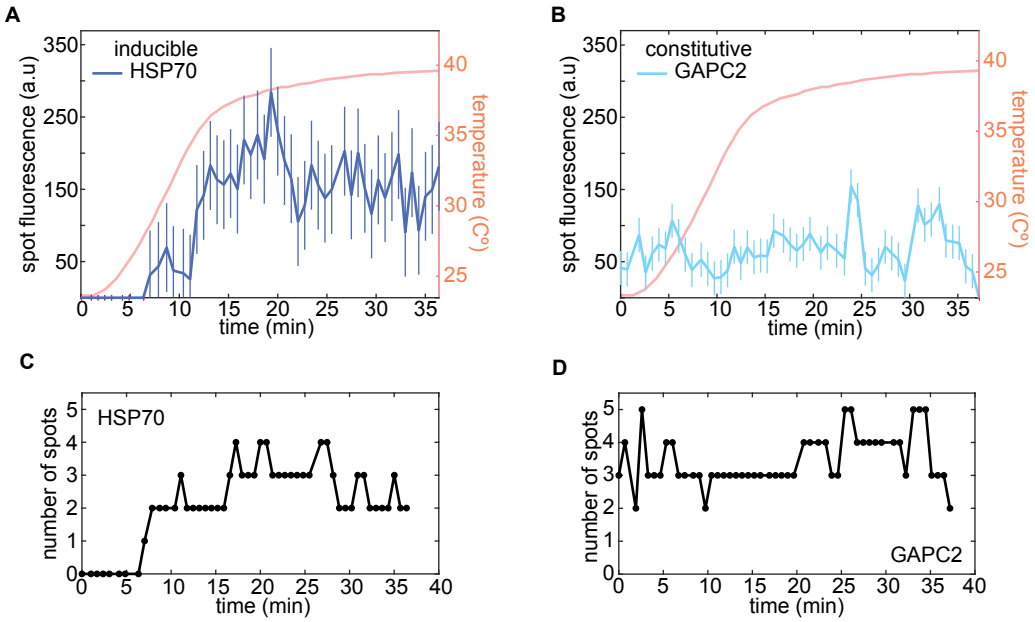

Figure 4.6: **Related to Figure 4.1D and E. Additional transcription spots in tobacco show the same qualitative transcritpional dynamics. (A)** HSP70-PP7 fluorescence time trace of a second transcription spot in the same nucleus as in Figure 4.1E. **(B)** GAPC2-PP7 fluorescence time trace of a second spot in the same nucleus as in Figure 4.1E. **(C)** Number of spots as a function of time in the nucleus shown in Figure 4.1D, left. **(D)** Number of spots as a function of time in the nucleus shown in Figure 4.1D, right.

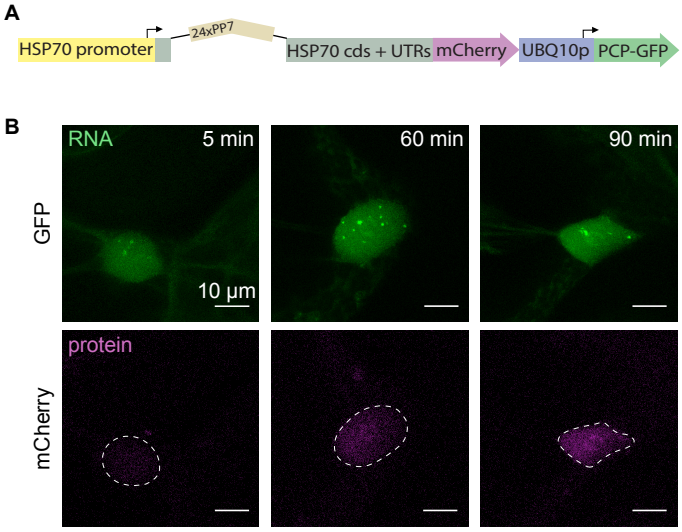

Figure 4.7: **Related to Figure 4.1. Simultaneous imaging of transcriptional activity and protein product in tobacco. (A)** Schematic of the construct used where the PP7 cassette is inserted into an intron in the Arabidopsis *HSP70* gene, which is fused in its C-terminus to mCherry. The same plasmid encodes a ubiquitously expressed PCP-GFP fusion. **(B)** Maximum fluorescence projection snapshots of a tobacco cell expressing the construct in (A) under heat shock. Nuclear mCherry fluorescence increases over time, consistently with the reported nuclear localization of HSP70 family proteins in plants [51, 150].

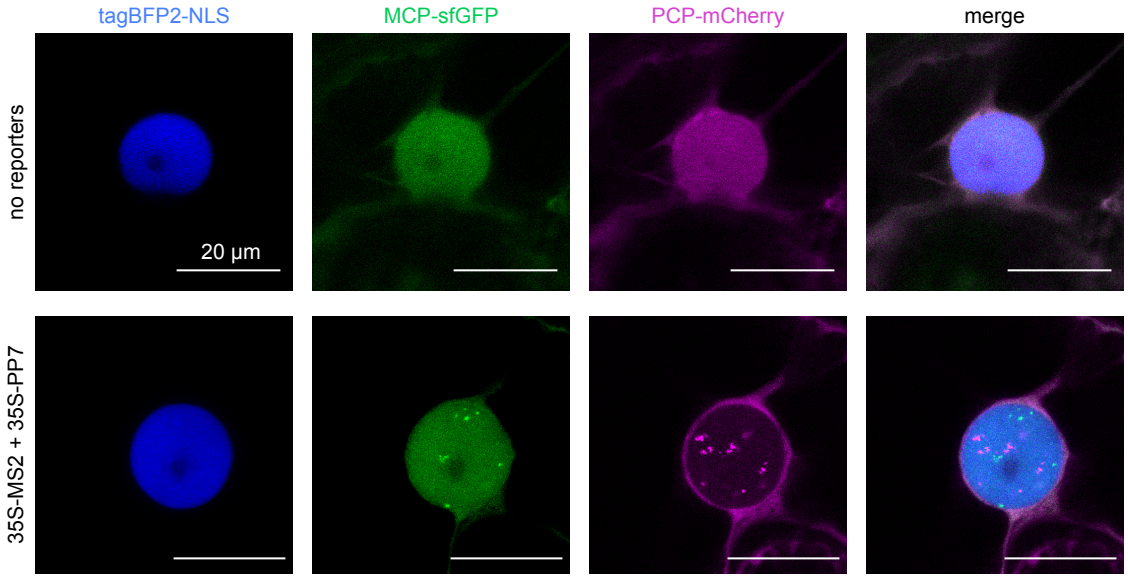

Figure 4.8: **Related to Figure 4.1F. MCP-sfGFP and PCP-mCherry are homogeneously distributed in the nucleus in the absence of transcription.** Maximum fluorescence projection snapshot of the nucleus of a tobacco cell expressing MCP-sfGFP, PCP-mCherry and nuclear localized tagBFP2. No nuclear puncta appear in the absence of PP7 and MS2 reporters.

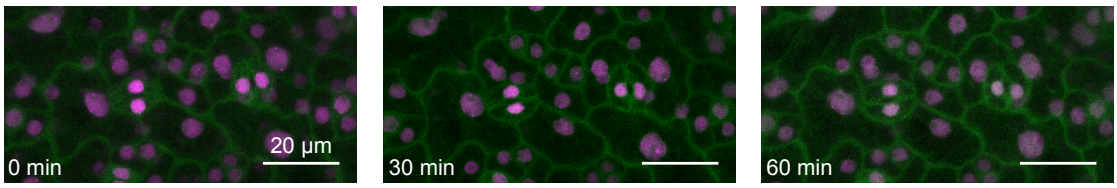

Figure 4.9: **Related to Figure 4.2A. Lack of HSP101 induction at room temperature.** Maximum z-projected image snapshots of the PCP-GFP/HSP101-PP7 Arabidopsis line imaged at room temperature. No spots were detected after continuous imaging for 60 minutes. Scale bar = 20 μm.

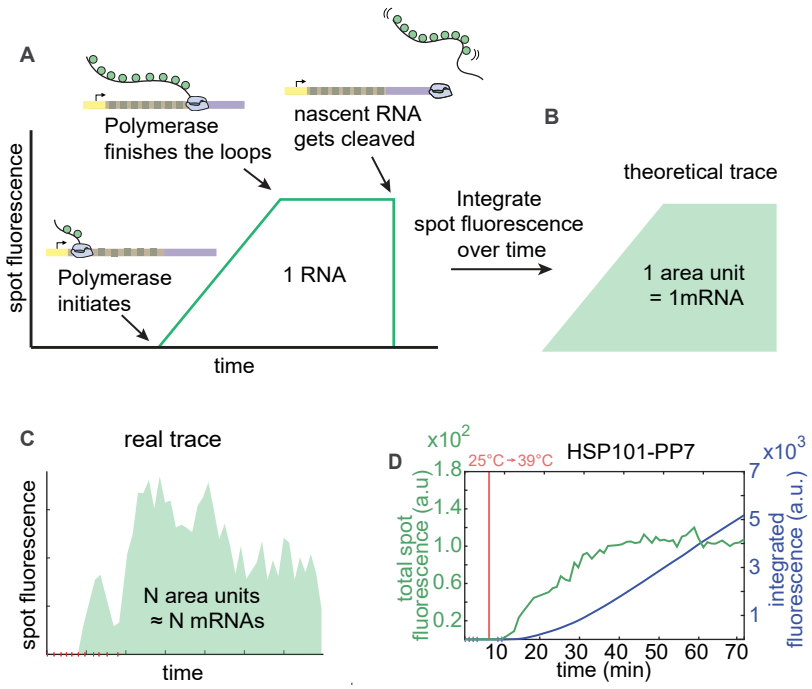

Figure 4.10: **Related to Figure 4.2B. Integrated fluorescence as a metric for total mRNA produced.** (A) Fluorescence profile of a single RNAP molecule as it traverses the gene. (B) Integrating this curve over time yields a unit of area associated with the production of a single mRNA molecule. (C) In the case of an actual transcription spot—resulting from the activity of multiple polymerase molecules—the integrated fluorescence over time will correspond to a number of area units equal to the number of produced mRNA molecules. **(D)** Data from a HSP101-PP7 replicate from Figure 4.2. Total spot fluorescence normalized by the number of cells in the field of view (green) and time integral of this signal (blue). The red horizontal line indicates when the stage temperature was shifted from room temperature to 39°C.

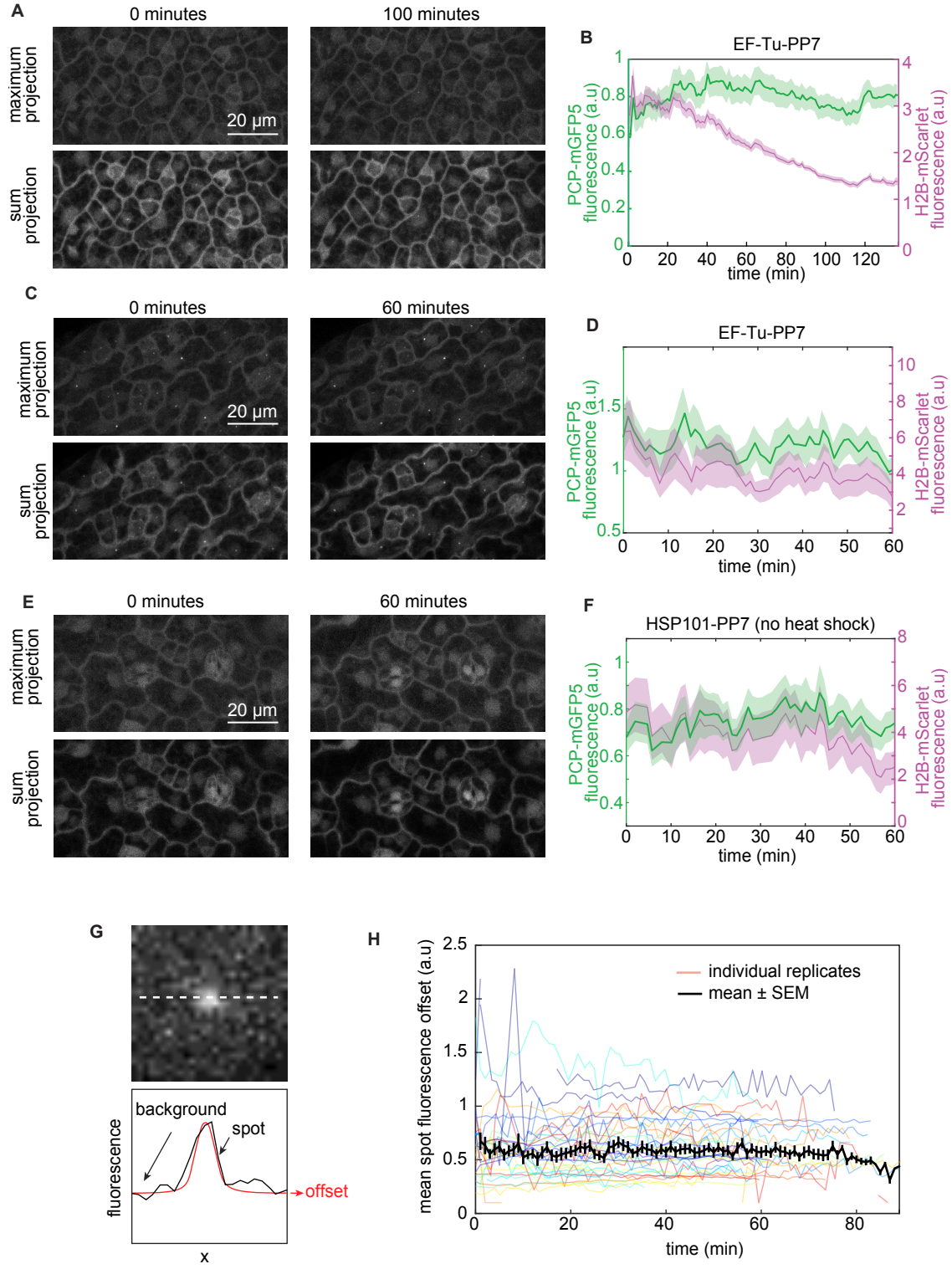

Figure 4.11: **Related to Figure 4.2B. Absence of GFP photobleaching during time lapse experiments.** ***See caption on next page.***

Figure 4.6: ***Continued from previous page*: Absence of GFP photobleaching during time lapse experiments. (A)** Snapshots of the PCP-GFP channel in leaves of an Arabidopsis plant carrying a constitutively expressed EF-Tu-PP7 reporter at the beginning of the experiment (left) and after 100 minutes of imaging (right). Two types of z-projections are shown: maximum projection (top) and sum projection (bottom). **(B)** Mean nuclear fluorescence in the GFP and the mScarlet channel in the movie shown in (A) ($n \approx 48$ nuclei per frame). See Materials and Methods: Image analysis: nucleus fluorescence for details on nuclear fluorescence measurements. **(C)** Same as (A) in an second EF-Tu-PP7 line. **(D)** Mean nuclear fluorescence in the GFP and mScarlet channels in the movie shown in (C) ($n \approx 26$ nuclei per frame). **(E)** Same as (A) in uninduced plant carrying HSP101-PP7. **(F)** Mean nuclear fluorescence in the GFP and mScarlet channels in the movie shown in (E) ($n \approx 29$ nuclei per frame). In (A)-(F) Nuclear PCP-GFP levels remain relatively stable, ruling out that photobleaching is affecting measurements of mRNA production. **(G)** Schematic showing how the spot fluorescence offset is calculated (for details see Materials and Methods: Spot fluorescence and tracking). On top, a maximum projection snapshot of a transcription spot. The dashed line indicates one of the dimensions along which fluorescence is calculated. At the bottom, the fluorescence profile along this line is used to fit a Gaussian curve (red). The baseline of the Gaussian corresponds to the spot fluorescence offset shown in (H). **(H)** Mean spot fluorescence offset over time in all the movies included in this study (colored lines) and mean spot fluorescence offset across all movies (black line). The background fluorescence, measured as spot offset, is stable over 60 minutes of continuous imaging, indicating that PCP-GFP is not being photobleached at an appreciable level. In (B), (D) and (F) the shaded areas correspond to the standard error of the mean over nuclei. In (H) the error bars correspond to the standard error across movies. In (A), (C) and (E) the same brightness and contrast setting were used to display the images corresponding to both time points.

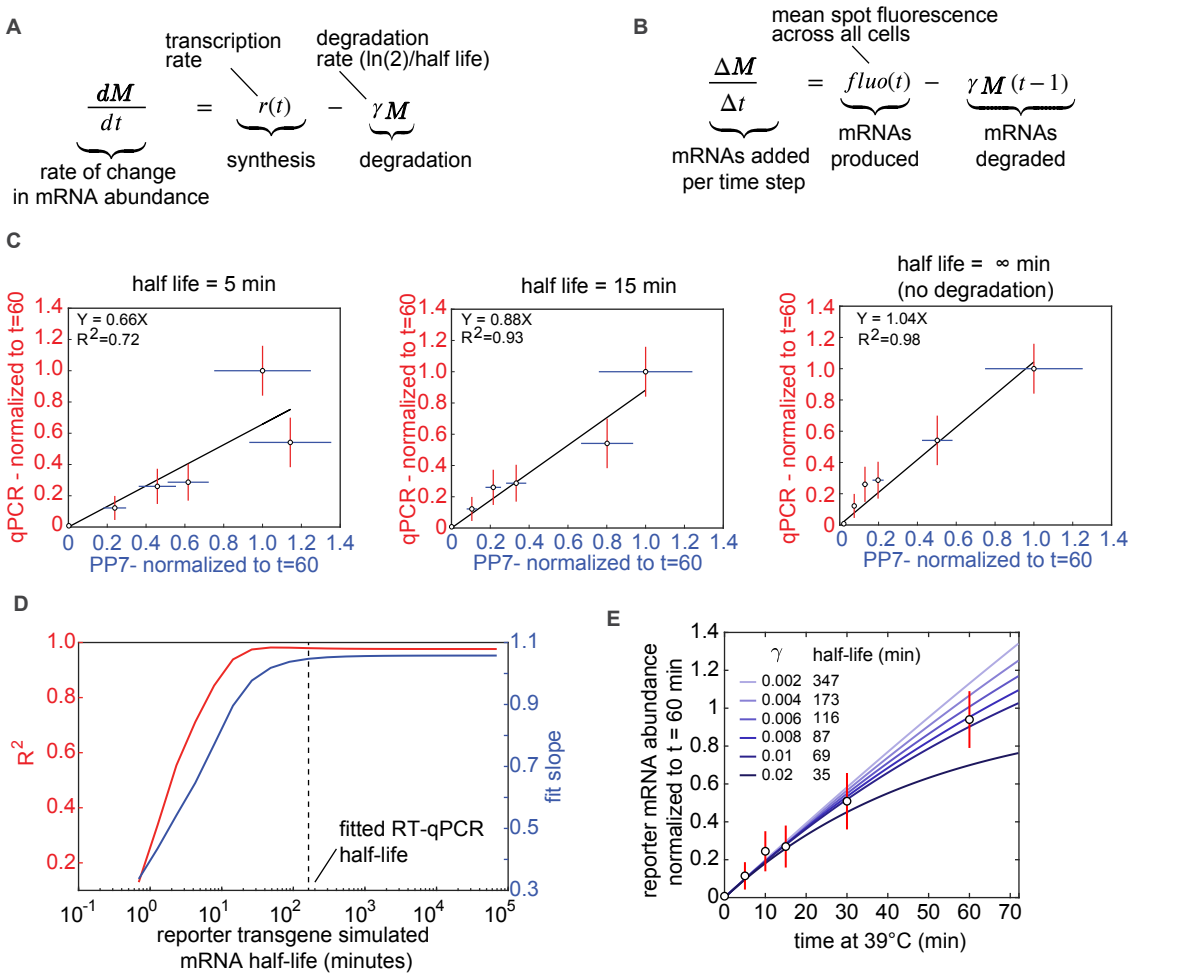

Figure 4.7: **Related to Figure 4.2B. Exploring the effect of the mRNA degradation rate on the validation of the PP7 system against RT-qPCR measurements. (A)** The rate of change in mRNA abundance is determined by a time-dependent rate of mRNA synthesis $r(t)$ and a constant mRNA degradation rate γ. **(B)** Discretized version of equation (A) used to obtain the accumulated mRNA based on spot fluorescence measurements. At each time point, the rate of synthesis is equal to the spot fluorescence while the number of mRNA molecules accumulated up to the previous time point are degraded at a simulated rate γ. Note that the mRNA half-life is defined as $\tau_{1/2} = \ln(2)/\gamma$. **(C)** Linear regression between the reporter mRNA abundance measured by RT-qPCR versus microscopy as in Figure 4.2C using the equation in (B) to incorporate mRNA degradation into the microscopy-based measurement. Because microscopy only reports on the synthesized, not the degraded, mRNA, we considered different, constant degradation rates and included this correction in the linear regression. ***Caption continues on next page***.

Figure 4.7: ***Continued from previous page*: Related to Figure 4.2B. Exploring the effect of the mRNA degradation rate on the validation of the PP7 system against RT-qPCR measurements. (D)** Fit parameters (R^2 and fit slope) as shown in (C) were calculated for a range of mRNA degradation rates expressed as half-lives. There is a good correlation and a constant slope between RT-qPCR and microscopy for half-lives longer that ~10 minutes. The dashed horizontal line indicates the fitted reporter mRNA half-life obtained in (C). **(E)** The reporter mRNA abundance measured by RT-qPCR was fitted to the mRNA accumulation model in (A) assuming a constant synthesis rate. mRNA accumulation according to RT-qPCR is almost linear on the timescales tested, resulting in a relatively long half-life. This half-life value is within the regime where there is a good correlation between PP7 fluorescence and qPCR (see vertical dashed line in (D)). For details about these calculations see Section 4.7.1.

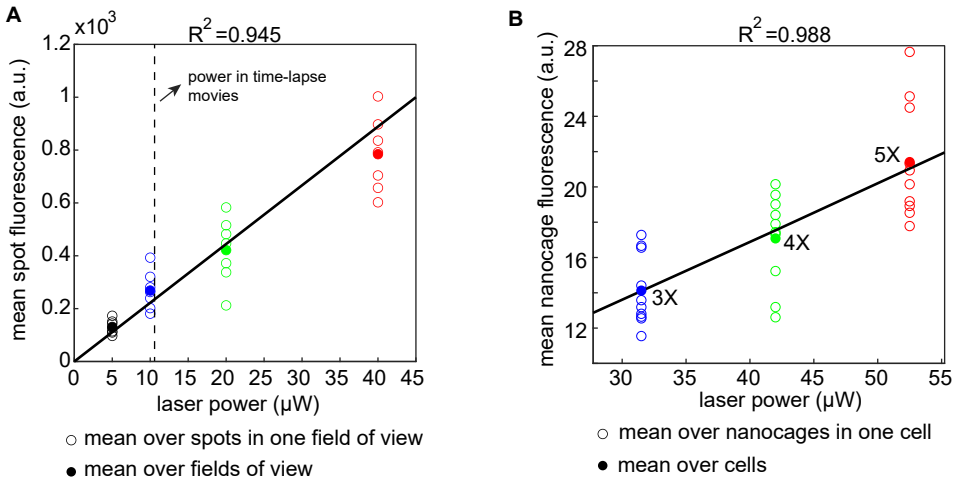

Figure 4.8: **Related to Figure 4.2: The fluorescence intensity of PP7 transcription spots and 60mer nanocages is linear with laser power intensity. (A)** Mean spot fluorescence of PP7 transcription spots driven by the constitutive EF-Tu promoter as a function of laser power intensity. Open circles correspond to the mean of all spots in a single snapshot in one field of view. Filled circles correspond to the mean taken over the mean of each snapshot. The vertical dashed line indicates the laser power used in time-lapse experiments. The solid black line corresponds to a linear fit to the data going through the origin, with R^2 = 0.945. **(B)** Mean fluorescence of 60mer GFP nanocages in tobacco cells as a function of laser power intensity. Open circles correspond to the mean nanocage fluorescence in one cell. Filled circles indicate the mean over the mean of each cell. The black solid line corresponds to a linear fit to the data going through the origin, with an with R^2 value of 0.988. Shown next to each mean value is how much stronger the laser power is compared to the power in time-lapse experiments (3, 4 or 5 times stronger).

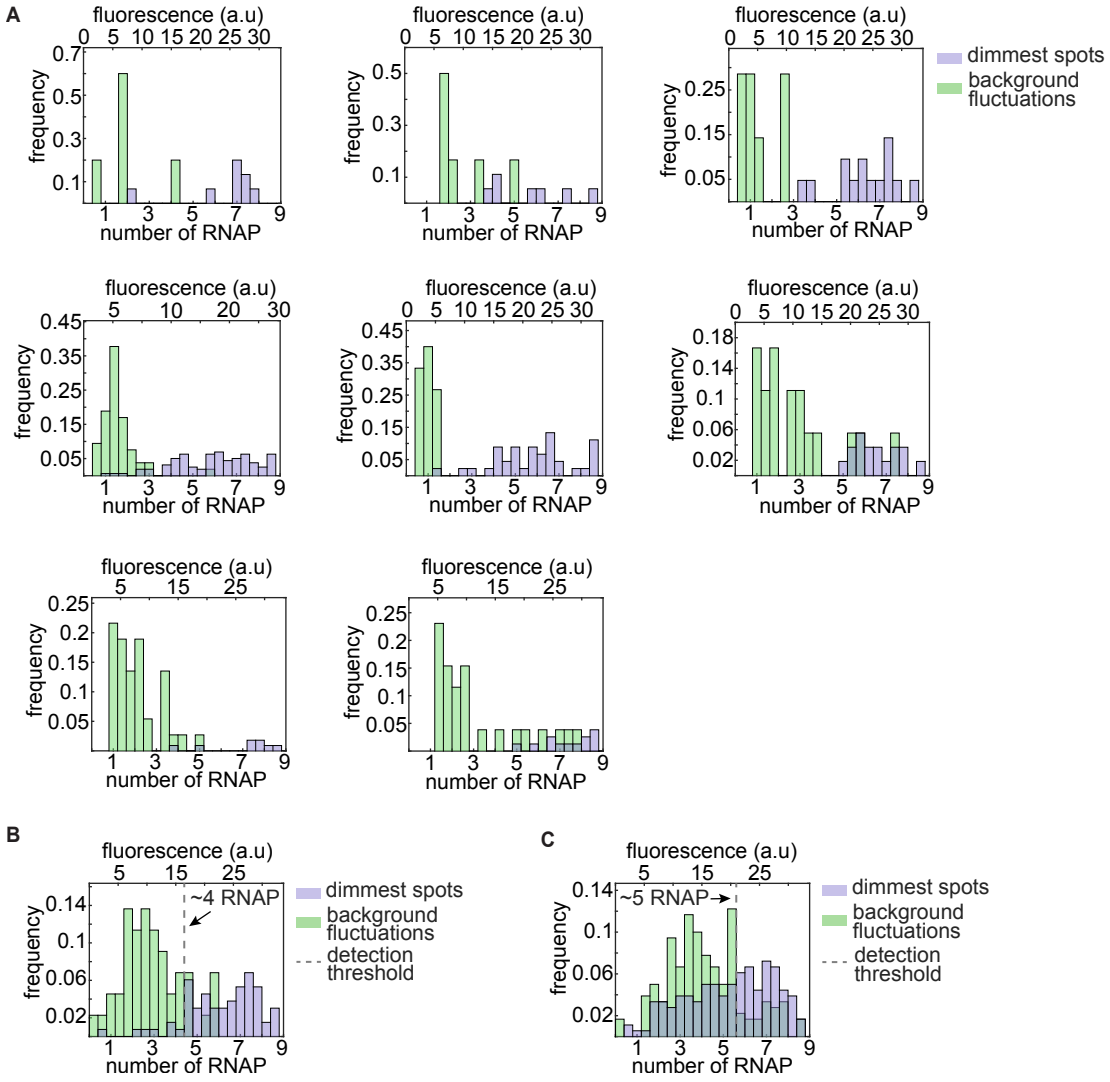

Figure 4.9: **Related to Figure 4.2E. Detection threshold analysis in individual HSP101-PP7 replicates and different reporters. (A)** Histograms of the calibrated number of transcribing RNAP molecules in the dimmest three frames of the weakest half of HSP101-PP7 fluorescence time traces (blue) and their associated fluorescence background fluctuations (green) as in Figure 4.2E. Each panel corresponds to an individual HSP101-PP7-1 replicate. **(B)** Same as (A) and Figure 4.2E where all the HsfA2-PP7-1 replicates were pooled together. **(C)** Same as (A) and Figure 4.2E where all the EF-Tu-PP7-1 replicates were pooled together. Note that, due to larger background fluctuations, the estimated detection threshold in (B) and (C) is larger than that of HSP101-PP7 shown in Figure 4.2E. This is likely due to a slightly higher PCP-GFP concentration in these lines.

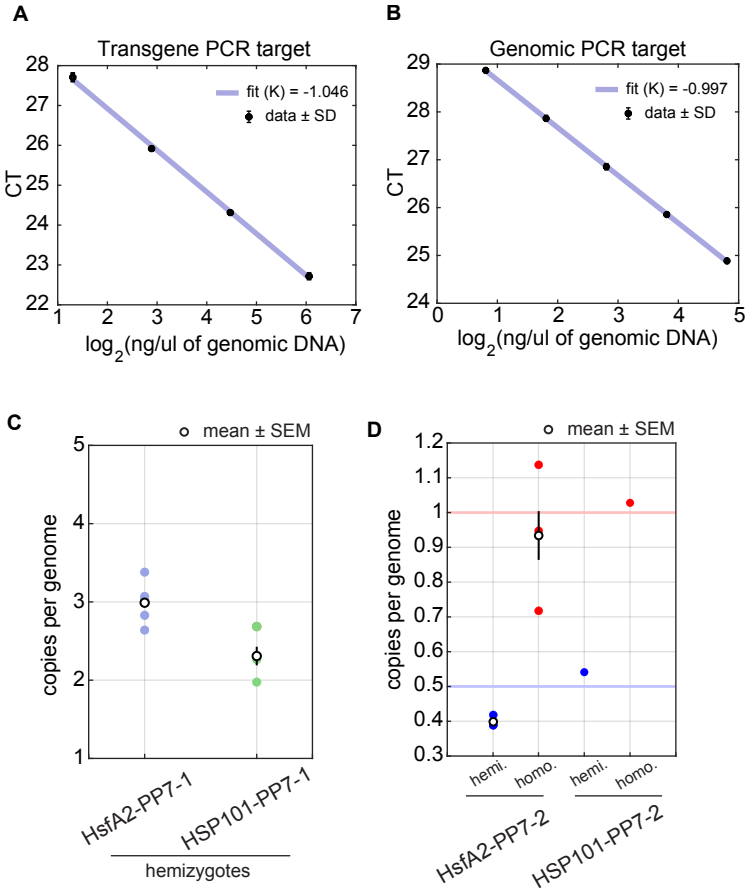

Figure 4.10: **Amplification efficiency of primer pairs and determination of the copy number of single insertion lines. (A)** qPCR results for serial dilutions of HSP101-PP7-2 Arabidopsis plants using primer pairs targeting the reporter transgene. **(B)** Same as (A) for a primer pair targeting a genomic location upstream of the Lhcb3 gene that we use to determine the CT value corresponding to one genomic copy. In (A) and (B), the slope of the linear fit corresponds to $K = 1/(1+log_2(\epsilon))$ where ϵ is the amplification efficiency. **(C)** Number of copies of the PP7 reporter transgene per genome copy in hemizygous individuals of HSP101-PP7-1 and HsfA2-PP7-1. **(D)** Number of copies of the PP7 reporter transgene per genome copy in two single insertion reporter lines in hemizygous and homozygous individuals. The horizontal blue line indicates the expected value for a single-copy hemizygous plant where the insertion locus contains a single copy of the transgene. The red horizontal line indicates the expected value for a plant homozygous for a single insertion where this insertion contains a single copy of the transgene.

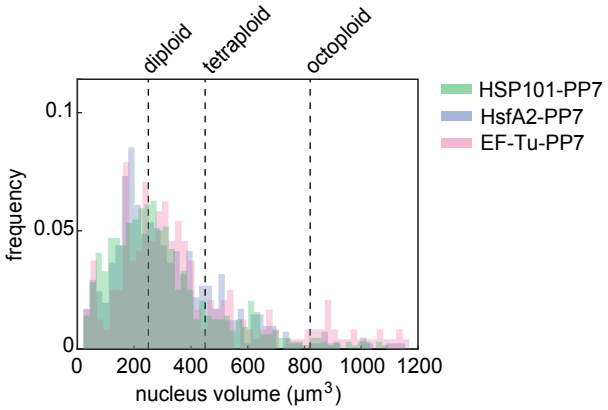

Figure 4.11: **Related to Figure 4.3. Nuclear volume distribution.** Histograms showing the volume of all nuclei in all the datasets included throughout this study. The nuclear volume was estimated by fitting maximum projections of the nuclear Histone-mScarlet channel to ellipsoids to obtain the mayor and minor axes for each nucleus. Shown on top are the mean nucleus volume of cells with different ploidy levels in Arabidopsis sepals according to Robinson2018.

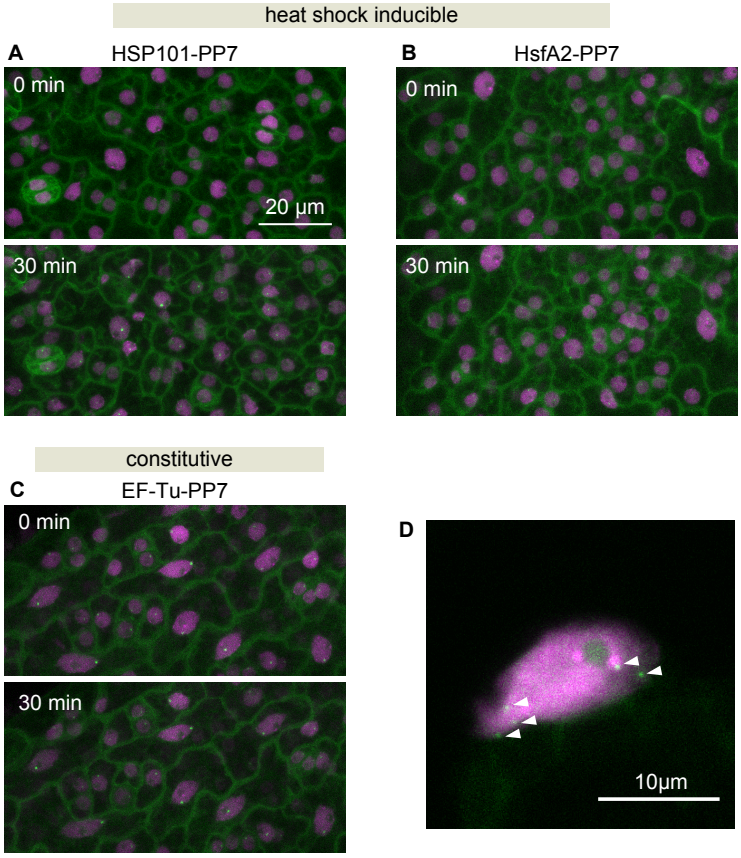

Figure 4.12: **Related to Figure 4.3B: Young diploid cells in hemizygous single insertion lines have a single spot per nucleus. Polyploid cells display multiple spots. (A)** Maximum projection snapshot of epidermis tissue near the base of the leaf from PCP-GFP Arabidopsis hemizygous for a single insertion of HSP101-PP7. On top, the sample at the beginning of a heat shock experiment. At the bottom, the same field of view after 30 minutes at 39°C. The PCP-GFP channel is shown in green, the Histone-mScarlet channel is shown in magenta. **(B)** Same as (B) but with HsfA2-PP7 instead of HSP101-PP7. **(C)** Plant hemizygous for a single insertion of a constitutively expressed EF-Tu-PP7 reporter at room temperature. Note that in (A)-(C), each nucleus has at most one transcription spot. **(D)** Polyploid Nucleus in a fully mature leaf from the plant in (B). White arrowheads indicate multiple transcription spots.

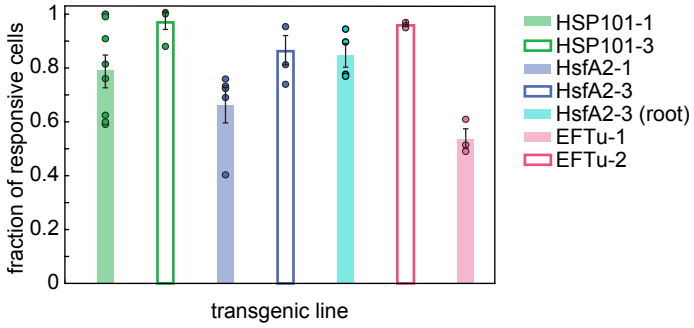

Figure 4.13: **Related to Figure 4.3A. Reproducibility of the fraction of responsive cells.** Mean fraction of transcriptionally responsive cells, defined as the number of nuclei that display reporter activity at least in one time point during the experiment divided by the total number of nuclei in the field of view (see Fig. 4.3A, bars on the right of each heat map). Circles represent single biological replicates (i.e movies). Error bars correspond to the standard error of the mean across 3 or more replicates.

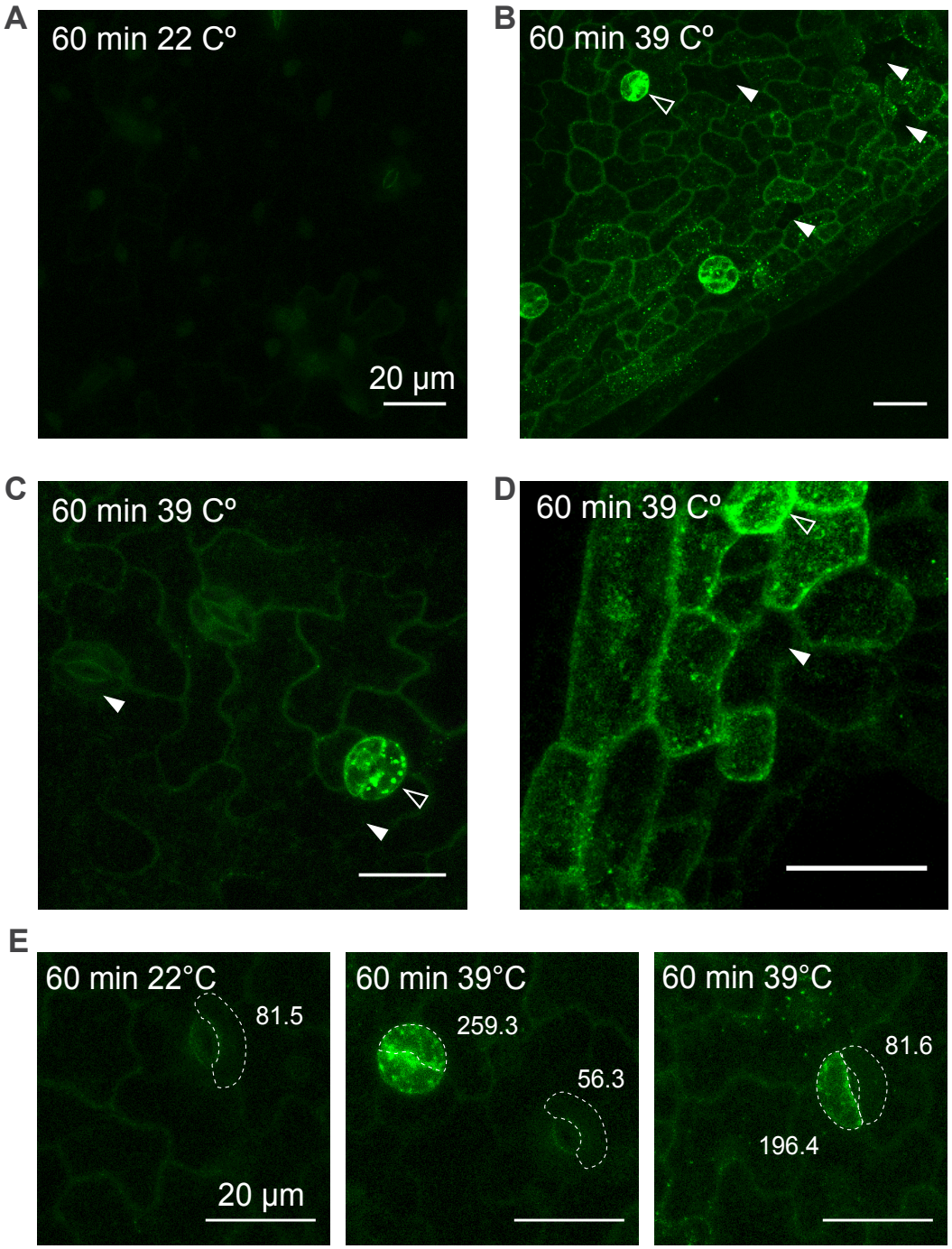

Figure 4.14: **Related to Figure 4.3: A rescue construct of HSP101-GFP reveals how refractory cells lead to substantial cell-to-cell heterogeneity in HSP101-GFP accumulation upon heat shock. (A-E)** Maximum fluorescence projections of leaf epidermis cells from hsp101 knockout mutant plants complemented with a transgene coding for a HSP101-GFP fusion driven by 734 bp of the endogenous HSP101 promoter [181]. ***Caption continues on next page***.

Figure 4.14: ***Continued from previous page*: Related to Figure 4.3: A rescue construct of HSP101-GFP reveals how refractory cells lead to substantial cell-to-cell heterogeneity in HSP101-GFP accumulation upon heat shock.** Detached leaves were treated with 39°C or 22°C for 60 minutes prior to imaging. **(A)** Untreated control. **(B-D)** Treated samples. White filled arrowheads indicate cells with negligible levels of GFP accumulation. Empty white arrowheads indicate cells with high levels of GFP accumulation. **(E)** Quantification of GFP fluorescence in treated and untreated cells. The dashed line highlights cells whose fluorescence was calculated. The numbers next to each cell correspond to the integrated GFP fluorescence of the volume of each cell highlighted.

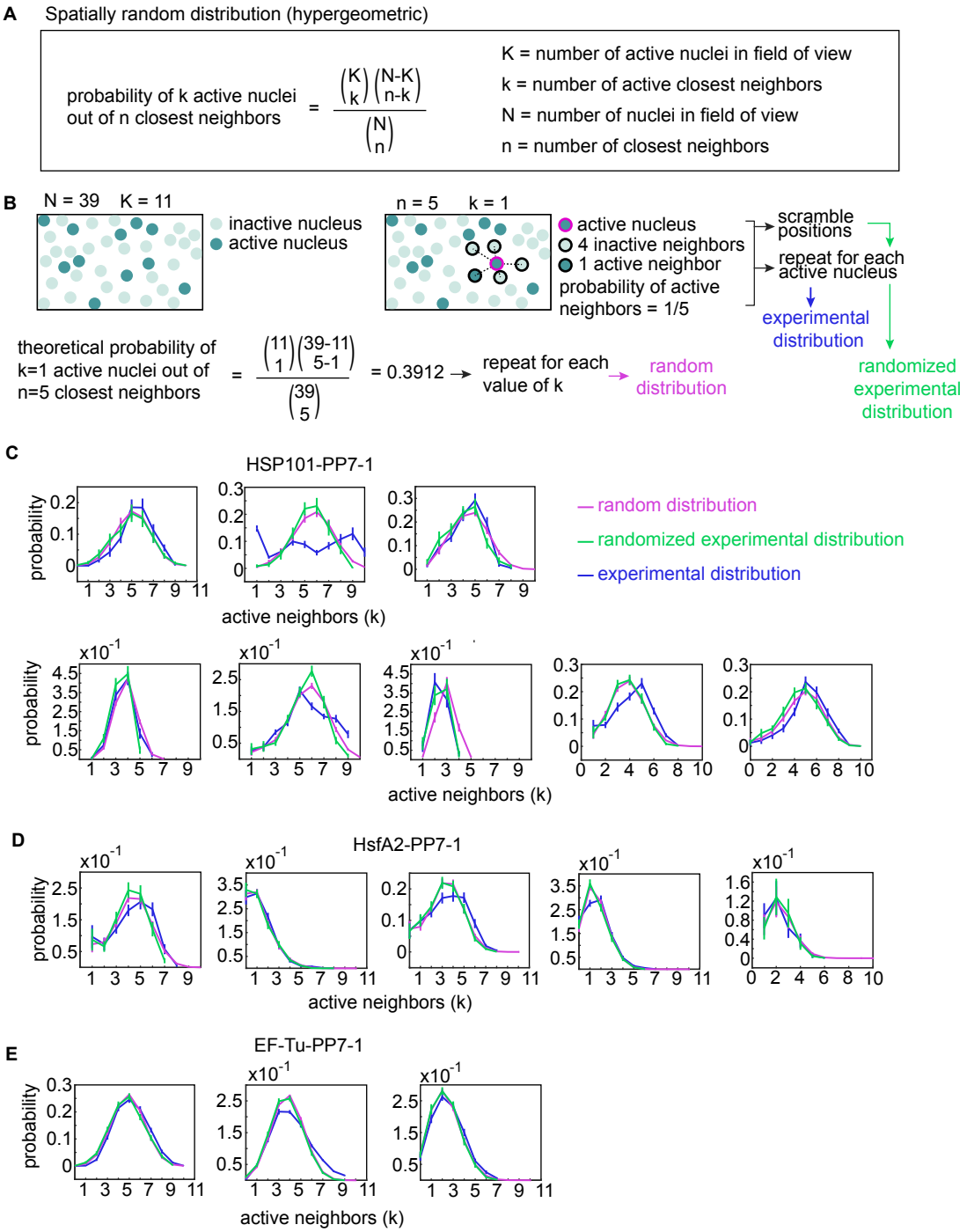

Figure 4.15: **Related to Figure 4.3: Transcriptionally active nuclei are randomly distributed in space.** ***See caption on next page.***

Figure 4.15: ***Continued from previous page*: Related to Figure 4.3: Transcriptionally active nuclei are randomly distributed in space. (A)** The hypergeometric distribution describes the probability of finding k successes in a sample of size n drawn randomly from a population of size N with K total successes. If nuclei containing transcription spots are randomly distributed in space, the hypergeometric distribution would capture the probability of a nucleus having k active nuclei among its n closest neighbors given K total active nuclei in a field of view containing N nuclei. **(B)** Schematic showing how the formula in (A) is applied to nuclei in a field of view. Nuclei with spots are represented by dark green circles. Light green circles represent nuclei without spots. For each transcribing nucleus (dark green circle with magenta border), we calculate the probability of finding another active nucleus among its closest neighbors (also denoted by a black border). An experimental probability distribution of active neighbors is then built by repeating this operation for all active nuclei. To build an experimental random distribution based on the data we randomize the positions of active nuclei and repeat this procedure. The random distribution can also be calculated analytically using the hypergeometric distribution in (A). **(C)** Probability distribution of the number of active neighbors (k) among the 10 closest neighbors (n) to each nucleus in the field of view of HSP101-PP7-1 replicates. Shown in magenta is the hypergeometric distribution (i.e., expectation if active nuclei are randomly distributed in space). In green is the distribution resulting from randomizing the position of actively transcribing nuclei. Actual experimental data is shown in blue. **(D)** Same as (C) for HsfA2-PP7-1 replicates. **(E)** Same as (C) for EF-Tu-PP7-1 replicates. Error bars in (C)-(E) correspond to the standard error over frames. The spatial distribution of active nuclei is close to that of the randomized data and similar to the theoretical random expectation. Thus, we conclude that there is no evidence for spatial structure in the transcriptional state of nuclei in the field of view.

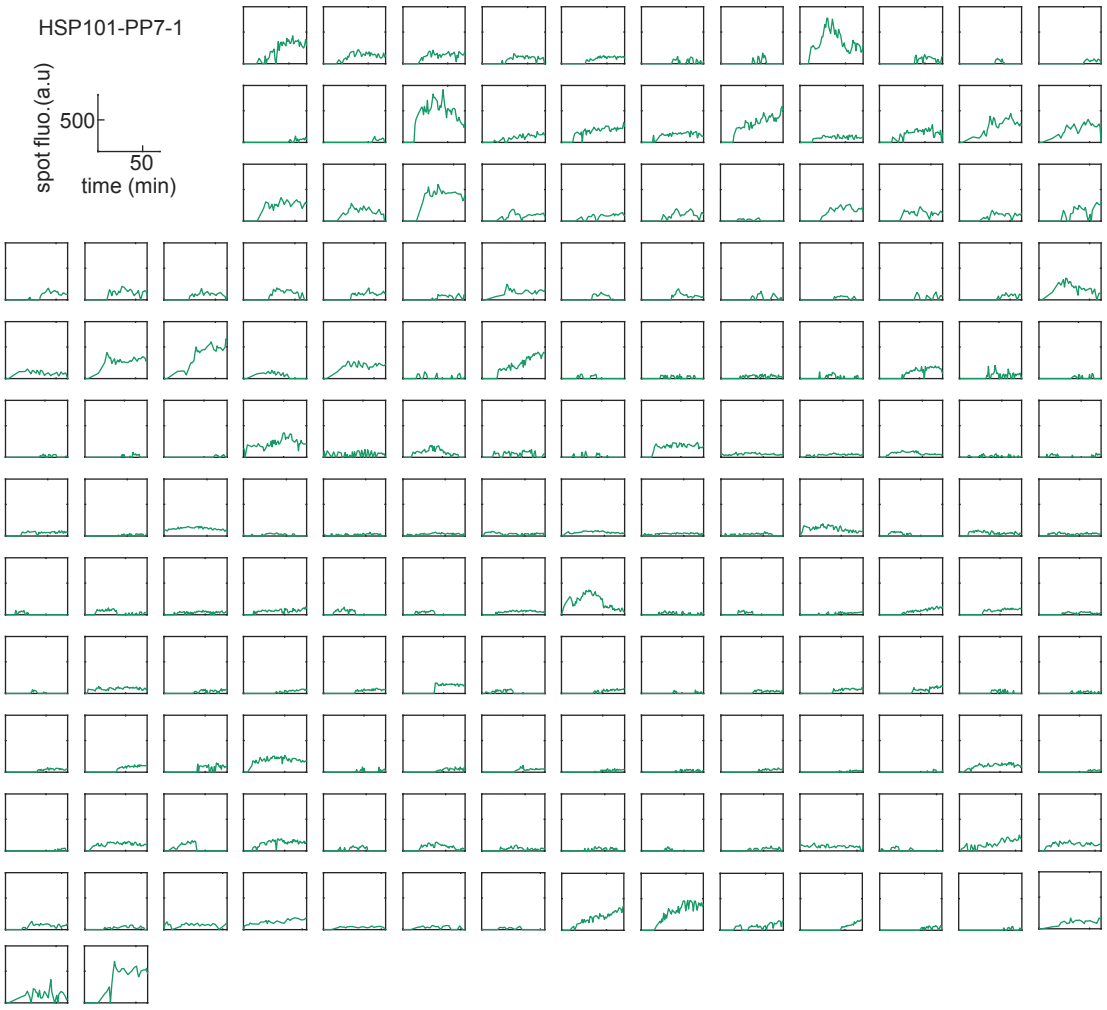

Figure 4.16: **Related to Figure 4.3: Behavior of single loci in HSP101-PP7-1.** Spot fluorescence time traces of individual loci in 8 replicates of HSP101-PP7-1 plants.

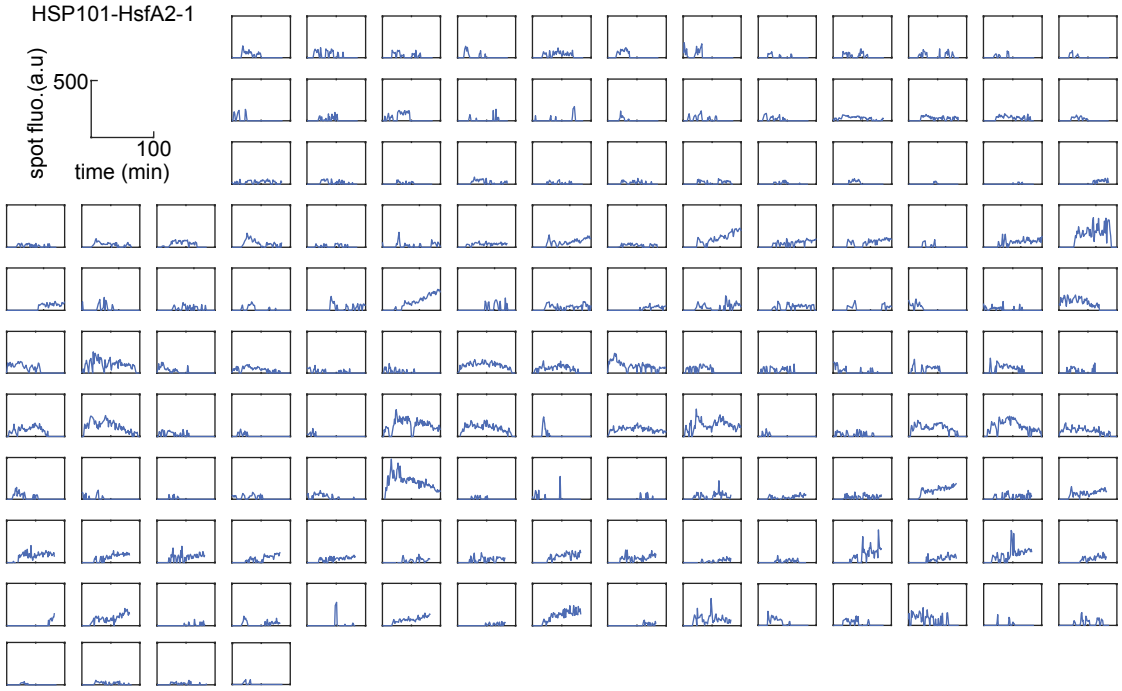

Figure 4.17: **Related to Figure 4.3: Behavior of single loci in HsfA2-PP7-1.** Spot fluorescence time traces of individual loci in 4 replicates of HsfA2-PP7-1 plants.

Figure 4.18: **Related to Figure 4.3: Behavior of single loci in EF-Tu-PP7-1.** Spot fluorescence time traces of individual loci in 3 replicates of EF-Tu-PP7-1 plants.

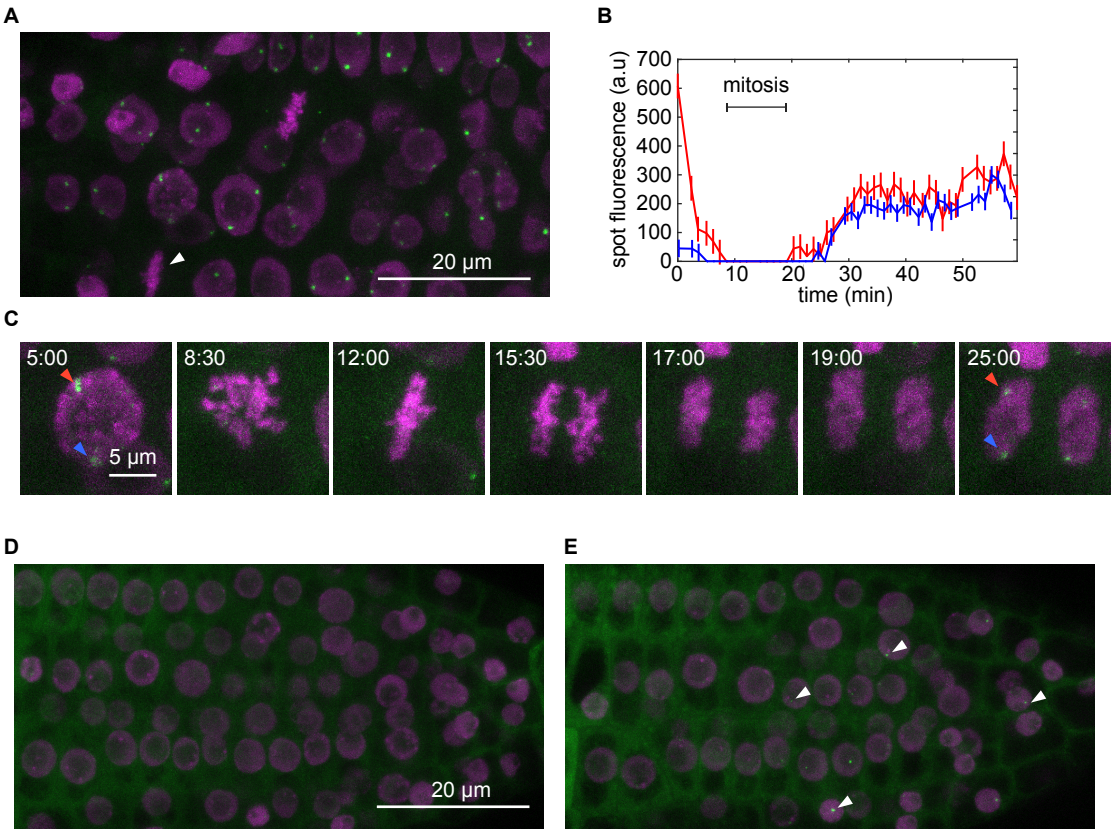

Figure 4.19: **Related to Figure 4.3A. Imaging transcription in Arabidopsis roots. (A)** Maximum projection snapshot of Arabidopsis root cells expressing H2B-mScarlet, PCP-GFP and EF-Tu-PP7. The white arrowhead indicates a cell undergoing mitosis. **(B)** Spot fluorescence before and after mitosis in the cell highlighted in (A). Each line corresponds to a different transcription spot. Error bars correspond to the uncertainity in spot fluorescence calculation as described in the Materials and Methods. **(C)** Snapshots of the cell undergoing mitosis in (A). Red and blue arrowheads indicate the spots whose fluorescence is shown in (B). **(D)** Maximum projection snapshot of Arabidopsis root cells expressing H2B-mScarlet, PCP-GFP and HsfA2-PP7 at room temperature. **(E)** Same sample as in (D) after 30 min under a 39°C heat shock treatment. white arrowheads indicate transcription spots.

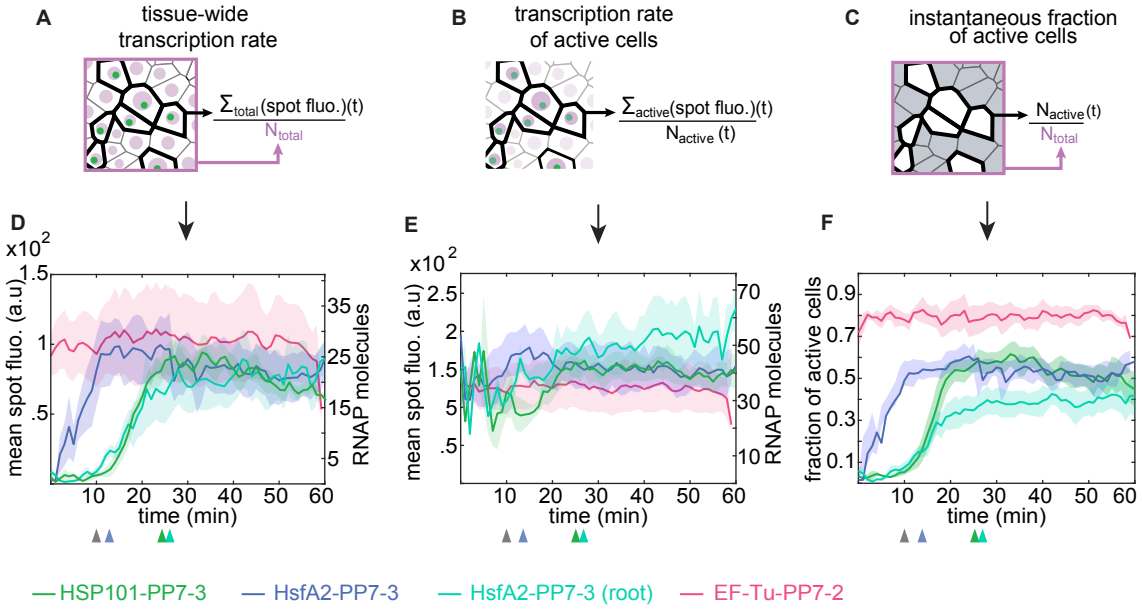

Figure 4.20: **Related to Figure 4.4: Experimental replicates using different independent transgenic lines of each promoter construct.** We repeated the experiments and analysis described in Figure 4.4 using different single-insertion transgenic lines carrying the same reporter constructs. **(A)** The tissue-wide transcription rate is calculated by adding the fluorescence of all spots in the field of view ($\sum_{total}$) in each frame and dividing by the total number of nuclei (N_{total}). **(B)** The transcription rate of active cells is calculated as in (A) except that the average is taken only over nuclei with spots in each frame ($N_{active}(t)$). **(C)** The instantaneous fraction of active nuclei corresponds to the number of nuclei exhibiting a spot in each frame divided by the total number of nuclei in the field of view. **(D)** Mean tissue-wide transcription rate in independent Arabidopsis transgenic lines carrying PP7 reporters driven by the promoters of HSP101, HsfA2 and EF-Tu as in Figure 4.4 inserted in different genomic locations. **(E)** Mean transcription rate of actively transcribing cells. **(F)** Mean fraction of active nuclei as a function of time. In (D-F) the shaded area corresponds to the standard error of the mean taken over 4, 4, 3 and 3 replicates for lines HSP101-3, HsfA2-3 (leaves), HsfA2-3 (roots) and EF-Tu-2, respectively. The arrowheads under each graph indicate the time points used to calculate the fold-change with respect to 10 minutes since the detection of the first spot (gray arrowhead). Because HsfA2-PP7-3 (blue) peaks near 10 minutes, 5 minutes were used for the fold change calculation of this dataset. These fold changes are shown in Figure 4.4H.

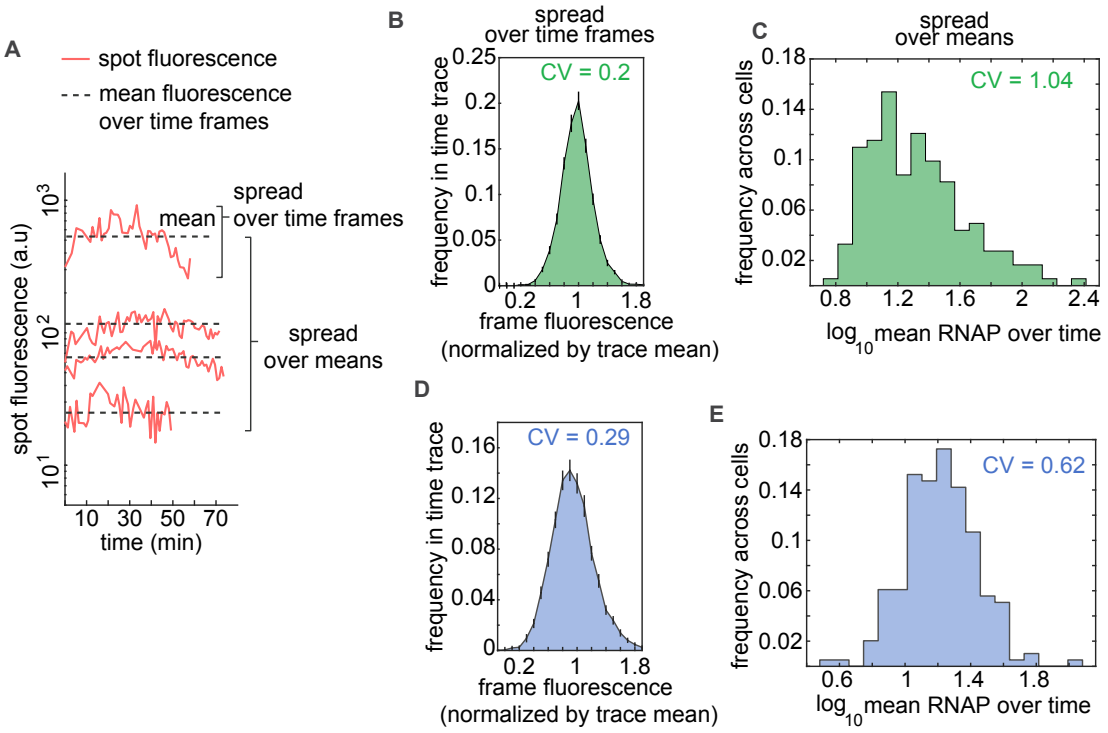

Figure 4.21: **Related to Figure 4.4: Spot fluorescence varies widely across cells but is relatively stable over time in individual cells. (A)** Representative spot fluorescence time traces in HSP101-PP7-1 replicates from Figure 4.3. Dashed lines correspond to the mean level of fluorescence of each trace taken over time. The spread of fluorescence values around this mean for each individual trace ("spread over time") informs about temporal fluctuations in transcriptional activity for each individual spot. The variability of mean fluorescence values across cells is captured by the "spread over means" and informs about cell-to-cell heterogeneity in activity. **(B)** Spread over time revealed by the distribution of frame fluorescence values normalized by the mean over time for each fluorescence trace pooled from all HSP101-PP7-1 replicates from Figure 4.3. The spread over time of fluorescence values of a given spot is very close to the mean, resulting in a coefficient of variation (CV=standard deviation/mean) of 0.2. **(C)** spread over means as reported by the distribution of mean fluorescence over time (see dashed lines in (A)) of all cells in HSP101-PP7-1 replicates. The average transcriptional activity varies widely across cells, with a coefficient of variation of 1.04. **(D,E)** Same as (B) and (C) for HsfA2-PP7-1 fluorescence traces pooled across replicates from Figure 4.4

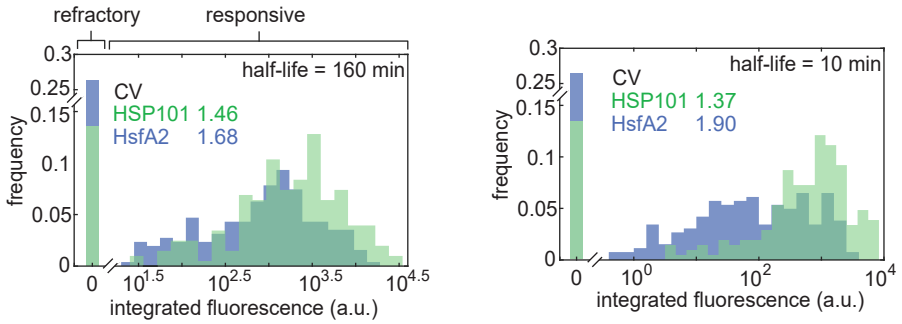

Figure 4.22: **Related to Figure 4.5B: Distribution of accumulated mRNA taking degradation into account.** Histograms showing the distribution of accumulated mRNA per cell in all pooled replicates of HSP101-PP7-1 and HsfA2-PP7-1 shown in Figure 4.3 as in Figure 4.5B. Two different mRNA half-lives were simulated, a realistic one of 160 minutes and very short one of 10 minutes. The value of 160 minutes was determined by fitting the RT-qPCR signal in Figure 4.7E. The calculation of accumulated mRNA based on spot fluorescence data is based on the assumptions described in Figure 4.10 and calculated as described in Section 4.7.1. The coefficients of variation (CV = standard deviation/mean) with a half-life of 160 minutes are virtually identical to those in Figure 4.5B obtained with an infinite half-life. The CV values are qualitatively similar even with an unrealistically short half-life of 10 minutes.

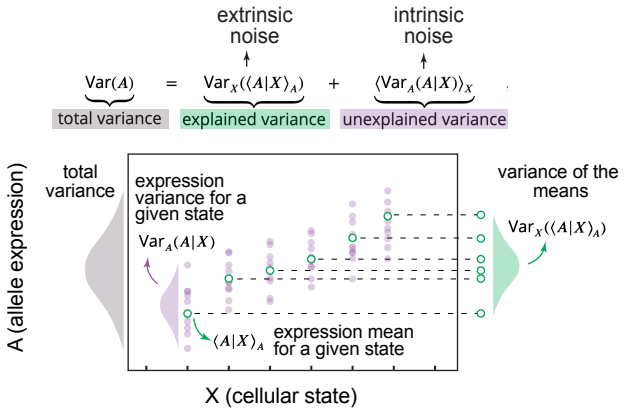

Figure 4.23: **Related to Figure 4.5 and calculations in Section 4.7.1: Visual explanation of the law of total variance.** Shown as a gray distribution on the left of the graph is the total variance in the expression of a gene (A) in a population of cells which varies depending on the cellular state (X). The total variance is composed of two types of variance, explained and unexplained, corresponding to extrinsic and intrinsic noise, respectively. As depicted by the green distribution to the right of the graph, subpopulations of cells belonging to different states will have different mean values of A since A depends on X. This variance is explained by the value of X being shared across cells within a subpopulation but different accross different subpopulation and is thus referred to as explained variance. On the other hand, cells in an identical state X can still have variable values of A (purple distribution). Since these cells share the same value of X, their variance is not explained by differences in cellular state. Thus, this intra-state variability is referred to as unexplained variance. .

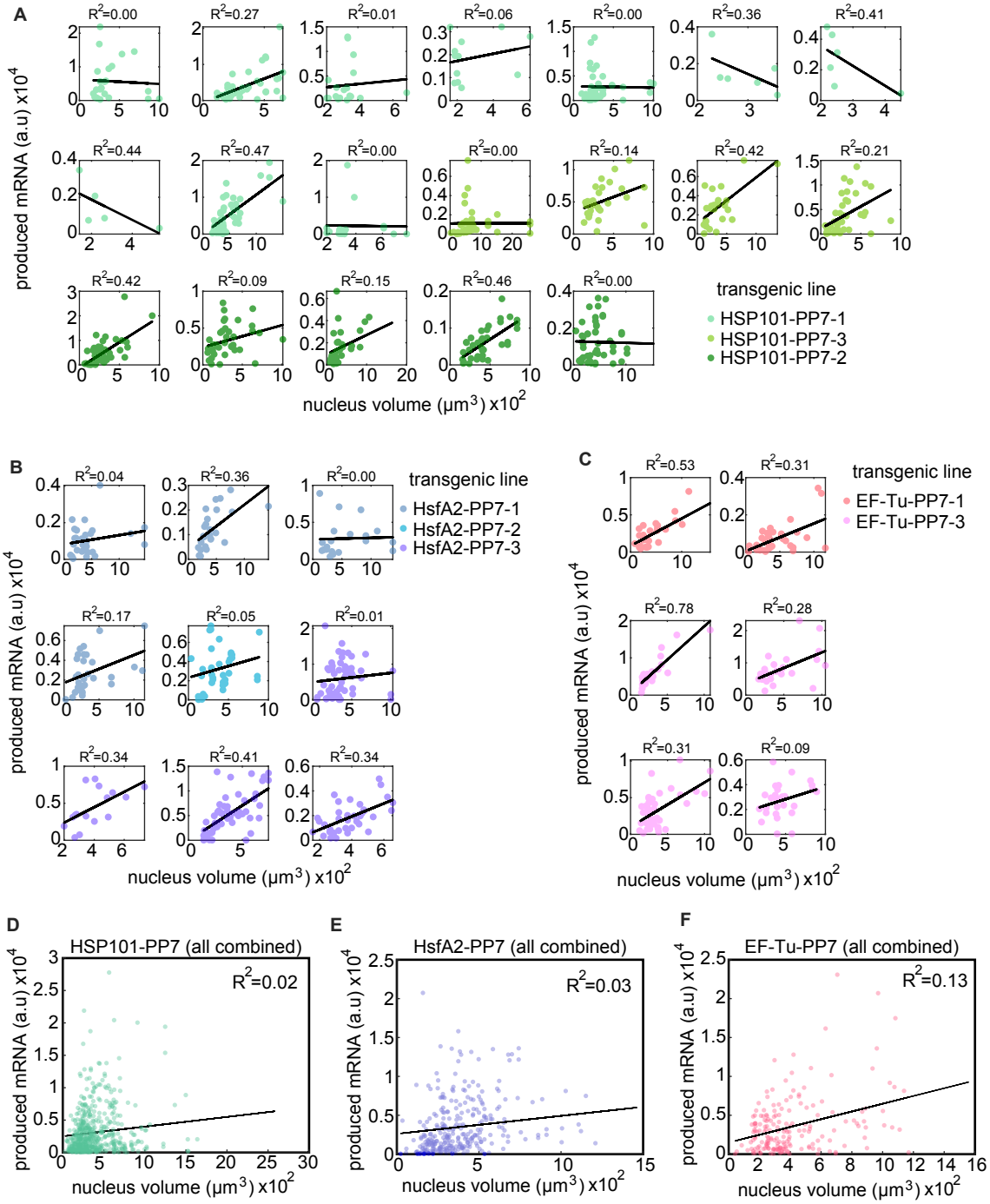

Figure 4.24: **Related to Figure 4.5: Nucleus volume is positively but only weakly correlated with transcriptional output. (A-C)** In each movie, nuclei were segmented at a single frame at ≈ 30 minutes based on Histone-mScarlet using the ImageJ Weka machine learning toolbox [10].

Figure 4.25: ***Continued from previous page*: Related to Figure 4.5: Nucleus volume is positively but only weakly correlated with transcriptional output.** To calculate their volume, nuclei were fitted to an ellipsoid based on the length of their mayor and minor axes. If a nucleus contained a transcription spot, its produced mRNA (calculated as integrated fluorescence over time) is plotted against its corresponding nuclear volume as a scatter plot. If a nucleus contained two transcription spots, as in the case of homozygous individuals, the integrated fluorescence of spots was averaged. Black lines on top of each scatter plot show the best fit to the data based on a linear model. The coefficient of determination (R^2) is shown on top of each plot. **(D-F)** Same as (A-C) except that nuclei from all replicates and transgenic lines were pooled together for each reporter construct.

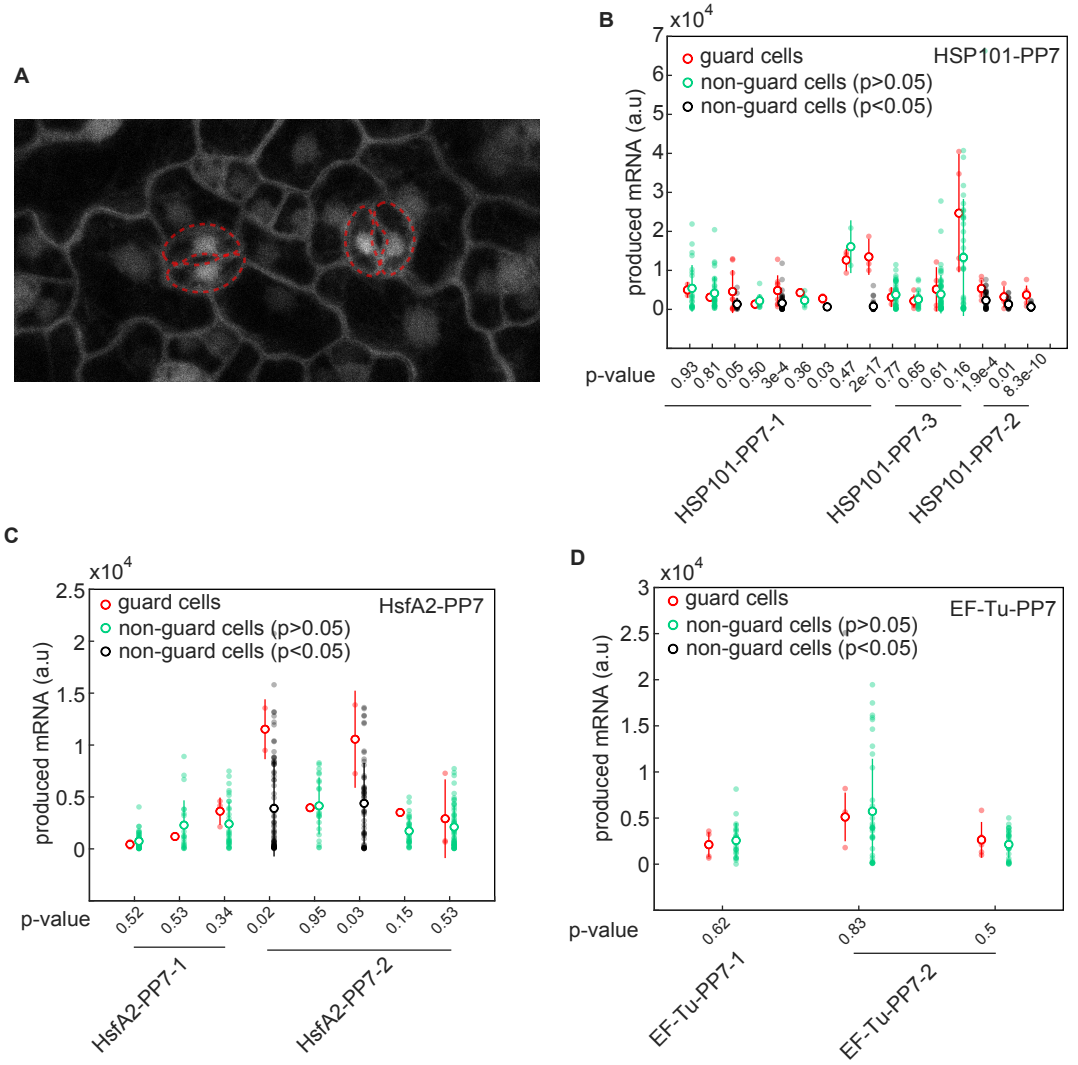

Figure 4.26: **Related to Figure 4.5: Guard cells do not consistently transcribe at different levels than the rest of cells. (A)** Arabidopsis epidermis cells expressing PCP-GFP. Dashed red lines highlight guard cells. **(B-D)** In each movie of each line presented in this study, the total mRNA produced by guard cells (red) was compared to that of non-guard cells. Filled circles correspond to individual cells, open circles correspond to their mean. A two-sided t-test was used to determine if guard cells are statistically different than the rest of cells. Non-guard cells are plotted in black if the test p-value is lower than 0.05 and in green otherwise, showing that guard cells do not transcribe at a different level in a consistent manner. Error bars represent the standard deviation. Only replicates in which guard cells were present are shown.

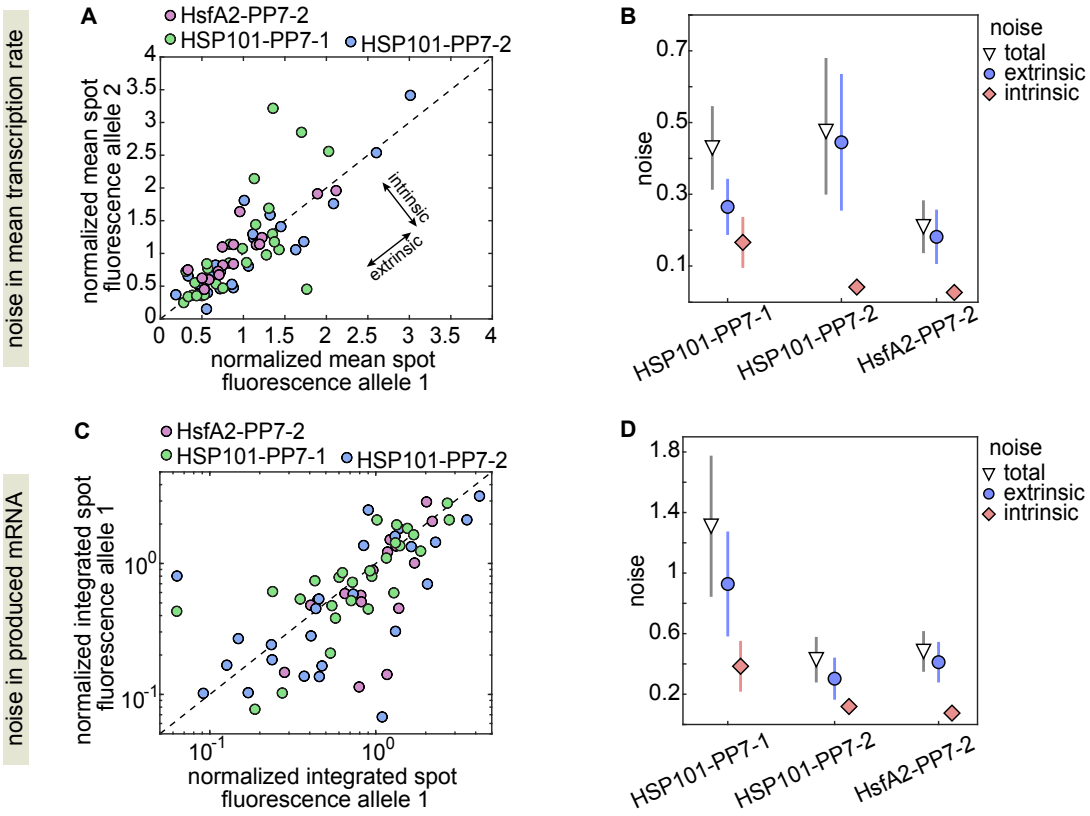

Figure 4.27: **Related to Figure 4.5: Extrinsic noise is larger than intrinsic noise among nuclei with two active alleles. (A)** Scatter plot showing the mean spot fluorescence over time for allele pairs belonging to the same nucleus in three different single-insertion lines homozygous for the PP7 reporter. **(B)** Decomposition of the total variability in (A) into its intrinsic and extrinsic components. **(C)** Scatter plot of integrated fluorescence over time in allele pairs belonging to the same nucleus in three different single-insertion reporter lines homozygous for the PP7 transgene (same as Figure 4.5E except that inactive alleles are not included). **(C)** Decomposition of the total noise in (C). In (A) and (C) values were normalized to the mean across all alleles in that line and the diagonal line shows y=x. Error bars in (B) and (C) correspond to the bootstrapped error.

4.7.4 Supplementary Videos

1. Video 1. **Constitutive reporter in tobacco.** Movie of tobacco cell expressing PCP-GFP and GAPC2-PP7. The scale bar is 10 μm.

2. Video 2. **Inducible reporter in tobacco.** Movie of tobacco cell expressing PCP-GFP and HSP70-PP7 under heat shock treatment starting at 10 min. The scale bar is 10 μm.

3. Video 3. **Inducible HSP101-PP7 reporter in Arabidopsis tissue.** Movie of leaf cells in Arabidopsis line stably transformed with PCP-GFP and HSP101-PP7 under heat shock treatment starting at 6 min. The scale bar is 10 μm.

4. Video 4. **Inducible HsfA2-PP7 reporter in Arabidopsis tissue.** Movie of leaf cells in Arabidopsis line stably transformed with PCP-GFP and HsfA2-PP7 under heat shock treatment starting at 8 min. The scale bar is 10 μm.

5. Video 5. **Constitutive reporter in Arabidopsis tissue.** Movie of leaf cells in Arabidopsis line stably transformed with PCP-GFP and EF-Tu-PP7. The scale bar is 10 μm.

6. Video 6. **Arabidopsis plant homozygous for an inducible reporter.** Movie of leaf cells in a homozygous Arabidopsis line stably transformed with PCP-GFP and HSP101-PP7 under a heat shock treatment starting at 0 min. The scale bar is 10 μm.

Chapter 5

Transcriptional dynamics of Phytochrome signaling during deetiolation in Arabidopsis

5.1 Abstract

Plants can sense the wavelength of light through photoreceptor proteins that transmit this information to the nucleus to regulate gene expression. The Phytochrome photereceptor senses the ratio between red and far red wavelengths of light and promotes the degradation of the PIF family of transcription factors upon activation by red light, leading to transcriptional changes genome-wide. The molecular aspects of Phytochrome B signaling have been extensively studied through genetic and biochemical approaches, particularly in the context of deetiolation, a dark-to-light transition that involves massive physiological changes. It is however currently unknown how this pathway regulates transcription in a wild type genetic background in a natural setting. In addition, the quantitative relationship between PIF nuclear dynamics and downstream gene expression has not been examined. Here, we used live cell imaging approaches to study the transcriptional induction of a PIF3-regulated gene during deetiolation.

5.2 Introduction

As sessile autotrophs, plant survival depends on connecting developmental transitions to changes in their environment. Seeds germinated in the dark (e.g underground) follow a developmental program known as skotomorphogenesis, where they grow heterotrophically and prioritize vertical growth to reach light. Upon light exposure, seedlings undergo a massive reprogramming at multiple levels of regulation, including changes in the mRNA abundance of thousands of genes. This process, termed deetiolation, reverses

skotomorphogenic growth and leads to the development of autotrophic structures such as chloroplasts and expanded cotyledons.

Multiple signaling pathways convey the quality of the light environment to the plant nucleus. Perhaps the best studied is the Phytochrome (Phy) pathway, which senses the ratio between far red and red light to trigger the degradation of members of Phytochrome Interacting Factor (PIF) family of transcription factors (reviewed by [157]). Briefly, in seedlings germinated in true darkness, the Phy photosensory protein exists in its inactive, Pr state and PIFs accumulate in the nucleus. Exposure to light covering the red parts of the spectrum converts Phytochromes into the active form Pfr. Pfr then binds to PIFs, eventually leading to the degradation of the complex [20, 192]. PIFs can act as transcriptional repressors or activators of thousands of targets [286] and their degradation leads to transcript level changes at a genomic scale [159].

The targets of the Phy-PIF signaling pathway and the mechanism of regulation by removal of PIFs have been studied using genetic approaches such as transcriptome comparisons between wild type and PIF knock-out mutants. However, to date it is not known how light affects the expression dynamics of PIF-regulated genes and what are the mechanisms by which PIF concentration regulates transcription of its targets. According to the Arabidopsis Information Resource website (TAIR), Phytochromes are the most studied Arabidopsis proteins. This solid knowledge of the molecular players and their biochemical transactions make the Phy-PIF pathway an ideal model to study transcriptional input-output functions in plants.

5.3 Results

5.3.1 Experimental setup

To measure the transcriptional activity of the RPT2 promoter, I adapted a system to fluorescently label nascent mRNAs described in detail in Chapter 4 (Fig. 5.1A) [155].

Here, the promoter of interest is used to drive a Luc-GUS fusion tagged on its 5' with 24 repeats of the PP7 stem loop sequence. When transcribed, these RNA stem loops bind with low nanomolar affinity to the PP7 phage coat protein (PCP) [46]. Fusing PCP to a fluorescent protein such as GFP results in the decoration of nascent RNAs with fluorophores. Because of the slow mobility of genomic loci and the accumulation of GFPs in the same diffraction-limited volume, transcribing alleles appear as fluorescent puncta [25]. To implement this system in Arabidopsis, I first generated multiple lines carrying PCP-GFP and kept several lines with moderate levels of fluorescence. Subsequently, one of these lines was transformed with a reporter construct driven by the RPT2 promoter (3329 bp upstream of the start codon) (Fig. 5.1B). Four RP2-PP7 lines were selected for further experiments based on the presence of spots and the 3/4 segregation ratio of Kanamycin resistance in the T2 generation (indicative of a single reporter insertion locus).

To measure the transcriptional activity of this reporter, plants were grown under 'true dark' conditions, following the protocol by Leivar and coworkers [158]. T2 seeds were plated in small petri dishes containing 0.5 strength MS agar. The dishes were placed at 4 ° C for 2-3 days to break seed dormancy and then exposed to white light for $\approx$ 6 hours to trigger germination. Immediately after, seeds were treated with far red light in a dark room for 5 minutes to completely inactivate the seed Phytochrome pool [158]. The plates were then wrapped in aluminum foil. Seeds were then left to germinate in the dark for $\approx$ 2 days prior to imaging.

As shown in Fig. 5.1D, swapping the dark period for a period under regular growth chamber conditions results in the presence of transcription spots at the beginning of the experiment. However, as expected, seedlings grown under true dark conditions are transcriptionally silent at the start of the experiment and gradually start activating the reporter (Fig. 5.1E).

To get a better sense of how reproducible this behavior is across different lines carrying random reporter insertions, we calculated a coarse and aggregated metric of reporter activity, the integral of the spot fluorescence over time for all spots in the field of view (Fig. 5.1F). This calculation can be thought of as the total amount of mRNA produced by the reporter locus at any given time (see Fig. 4.10) and is thus comparable to RT-qPCR data, assuming a negligible degradation rate (see Fig. 4.22).

5.3.2 The timescales of RPT2 activation in response to light

Next, I sought to measure the speed and kinetics of transcriptional induction in response to PIF3 removal. To correlate these dynamics to PIF3 abundance, I first measured the nuclear concentration of PIF3-GFP using microscopy in plants grown under exactly the same conditions as in PP7 experiments. As shown in Fig 5.2A, immediately after exposing plants to light, the nuclear distribution of PIF3 is relatively homogeneous although diffuse 'clusters' can be observed. These clusters become increasingly pronounced as the overall PIF3-GFP nuclear concentration drops exponentially to undetectable levels at $\approx$ 50 minutes (Fig 5.2B), similar to what has been previously reported by western blotting [3, 191].

Next, we determined the time at which the first spot is detected in multiple replicates of each of the four RPT2-PP7 transgenic lines we had established. Fig. 5.2C shows the time at which the first spot appeared in each individual replicate. To choose a single number for the start of transcription, I took the mean of the means of each replicate, resulting in 7.3 $\pm$ 1.3 minutes after light exposure. The lower bound of this estimation depends on the delay inherent to the PP7. According to experiments shown in Chapter 5, the detection threshold is 3 RNAP fully loaded with GFP. This means that to detect a spot, 3 RNAP have to traverse the 1.5 Kbp of the loops. For an elongation rate of 1 kbp per minute (a reasonable number in eukaryotes as reviewed by [9]) this takes approximately 1 minute, thus we have 5 minutes as the lower bound instead of 6 minutes.

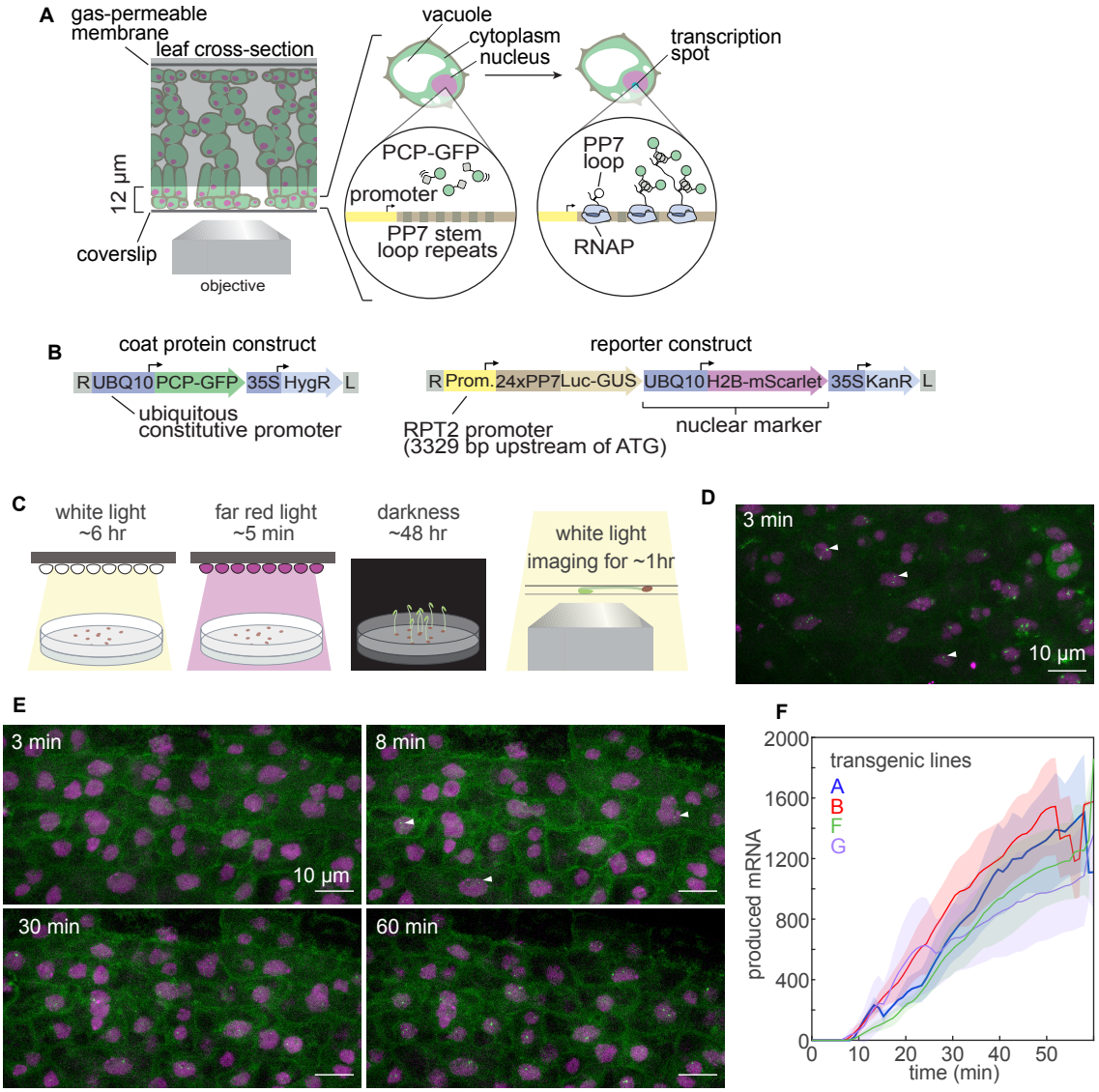

Figure 5.1: **Experimental setup**. **(A)** Schematic representation of the PP7 system to label nascent RNAs in Arabidopsis. **(B)** Constructs used in the experiments shown in this chapter. **(C)** The true dark protocol used to completely inactivate Phytochromes in Arabdiopsis seeds. **(**D) Snapshot of an Arabidopsis cotyledon expressing the constructs in (B) using a variation of the treatment in (C) where the 2-day dark period was replaced by two days under dark-light cycles. **(E)** Snapshots of cotyledons of a transgenic line transformed with the constructs in (B) and grown under true dark as in (C). Spots are not detectable immediately after light exposure but they rapidly appear. **(F)** mean total mRNA produced by the RPT2 reporter in at least three replicates of four independent transgenic lines measured as the integral of spot fluorescence over time across all nuclei in the field of view.

It is conceivable that these early detected spots correspond to outliers that happen to stochastically transcribe. In this case, the metric in Fig. 5.2C is not very representative of the behavior of the pathway across the plant. To determine if the response can be captured by a single unified process and to determine what the rate of this process is, I fit the number of alleles per frame to a simple equation,

$$\frac{dF}{dt} = r, \tag{5.1}$$

where F corresponds to the number of transcribing alleles divided by the number of nuclei and r corresponds to the rate this quantity increases with time under white light. A fit of this formula to one dataset is shown in Fig. 5.2D. In general, we found that this equation fits the data well, indicating that the first detected spots, rather than outliers, are part of a sustained induction behavior that can be captured by a single rate throughout the experiment. On average, $\approx$ 2% of the alleles that are active at 60 minutes turn on every minute under light (Fig. 5.2E).

5.3.3 Aspects of transcription regulated by PIF3

The temporal patterns of mRNA accumulation shown in Fig. 5.1F can arise from multiple single cell behaviors. We saw that removal of PIF3 leads to an increase in the number of actively transcribing alleles across the tissue. As we show in Chapter 5, this by itself is sufficient to create temporal expression patterns. To determine if PIF3 concentration can regulate the rate of transcription after the reporter turns on, I calculated the mean transcription rate across actively transcribing alleles (Fig. 5.2F). This analysis revealed that the average number of Polymerases transcribing the reporter does not depend on PIF3 concentration or time under white light. Instead, induction seems to operate like an all-or-nothing switch. Indeed, the mean transcription rate *per nucleus* (regardless of whether it had active reporter alleles or not) increases steadily over time (Fig. 5.2G), and these dynamics closely resemble those of the number of active alleles in the tissue (Fig. 5.2G). Thus, the PIF3 temporal dynamics bias the probability that the RPT2 reporter switches to an active state but do not affect its rate of transcription afterwards.

5.4 Discussion

The fact that transcription of a PIF3-repressed gene starts when the PIF3 nuclear concentration is still $\approx$ 90% of its original value is intriguing. It seems unlikely that temporal fluctuations in the order of 10% are sufficient to commit a gene to switch to an active state. Experiments in the Drosophila embryo have shown that nuclei are capable of distinguishing a 10% difference in transcription factor concentration but only if they are exposed to it for $\approx$ two hours [108]. It is possible that the rapid aggregation of PIF3 accelerates the drop in local concentration at PIF3 target genes, meaning that the curve

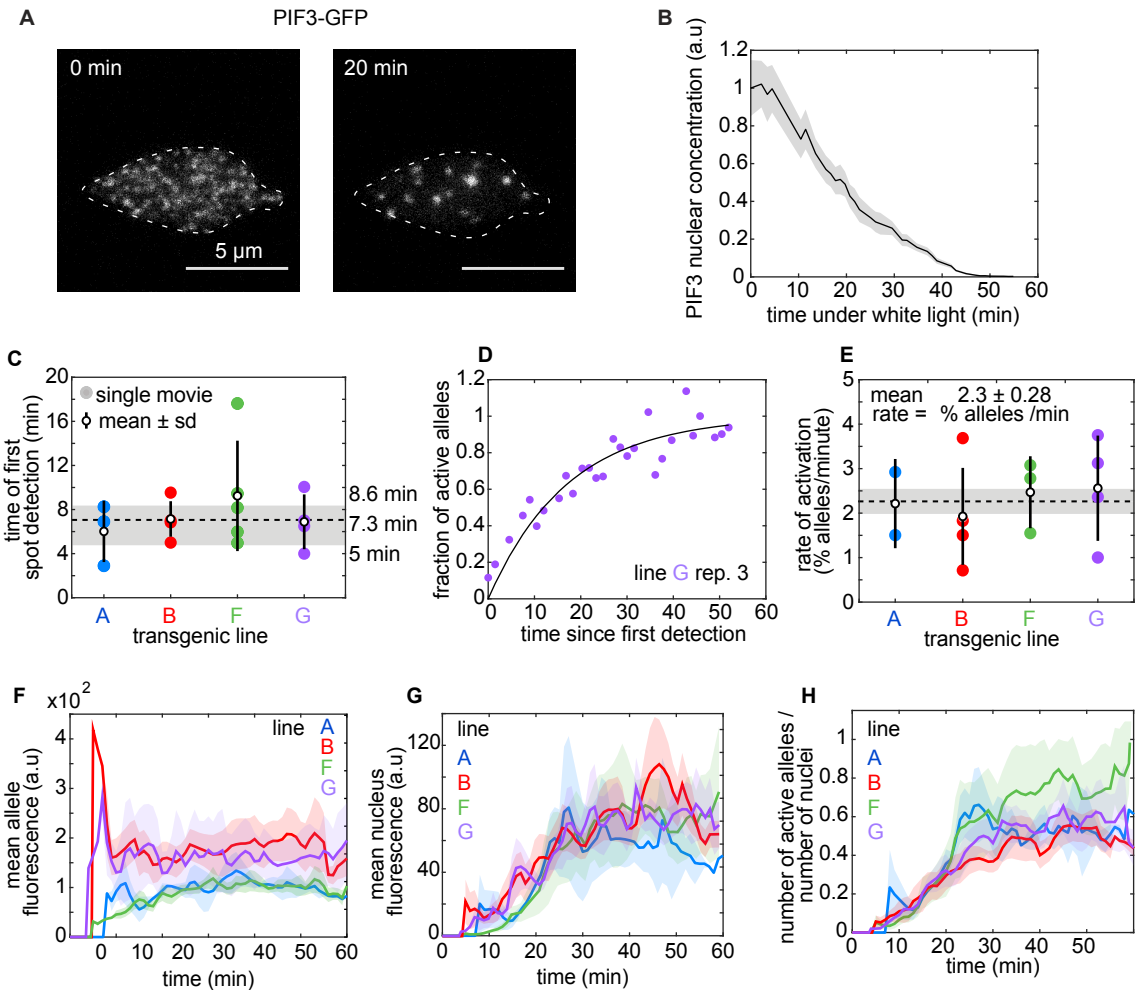

Figure 5.2: **The timescales of transcriptional activation in response to PIF3 degradation.** **(A)** Snapshots of Arabidopsis nuclei expressing PIF3-GFP. Seedlings were grown under true dark and exposed to white light at time = 0. **(B)** Mean PIF3 nuclear concentration over time following a dark-to-light transition. **(C)** Time at which the first RPT2-PP7 spot was detected in the field of view across multiple replicates and transgenic lines. The dashed line corresponds to the mean and the shaded area represents the confidence interval. **(D)** Number of active alleles divided by the number of nuclei in the field of view in an RPT2-PP7 cotyledon from line G. The solid black line is a fit to an irreversible one step process. **(E)** Mean rate of transcriptional engagement of the RPT2-PP7 reporter across replicates and transgenic lines measured as the fitted rate in (D). $\approx$ 2% of alleles that have turned on after 60 minutes turn on every minute following light exposure. **(F)** Mean transcription rate of active alleles measured as the sum of spot fluorescence across all spots divided by the number of spots. ***Caption continues on next page***.

Figure 5.2: ***Continued from previous page:*** **The timescales of transcriptional activation in response to PIF3 degradation**. **(G)** Mean transcription rate per nucleus measured as the sum of spot fluorescence across all spots divided by the number of nuclei. **(H)** Mean number of actively transcribing alleles normalized by the number of nuclei in the field of view. In (C) and (E), error bars correspond to the standard deviation. In (F)-(H) shaded areas correspond to the standard error of the mean across replicates (movies)

shown in Fig. 5.2B does not represent the PIF3 concentration being experienced by the RPT2 promoter. In addition, there could be a positive feedback mechanism operating at the promoter level that makes the response to PIF3 concentration extremely sharp and/or irreversible.

The type of regulation by which the temporal dynamics of RPT2 expression are dictated by PIF3 are remarkably similar to the scenario we described for the heat shock response in Arabidopsis leaves in Chapter 4. Alleles can exist in two 'states', active or inactive, and PIF3 concentration biases the likelihood of switching to the active state. Once locked into a state, the rate of transcription does not respond to the instantaneous PIF3 nuclear concentration anymore.

Note again that these observations are based on the *average* PIF3 nuclear concentration, which might not represent the *local* concentration at its targets. The very obvious aggregation of PIF3 is indicative that this might be the case. In the future, it would be interesting to image PIF3 and RPT2-PP7 simultaneously in the same plant to use the position of transcription spots as a guide to measure the local PIF3 levels. The ParB-*parS* system could be used to label the reporter locus independent of PP7 detection. Interestingly, alleles of PhyB have been described that trigger either PIF3 degradation or aggregation [201]. These mutants could be used to shed light on the regulatory role of nuclear aggregates to inactivate transcriptional repressors like PIF3.

Chapter 6

Overcoming technical limitations in live imaging experiments

6.1 Abstract

Over the course of the research presented in the previous chapters, we encountered a number of technical constraints related to detection of molecules in living cells in their tissue context. In this chapter, I describe the use of dimerization-dependent fluorescent proteins to improve detection of nascent RNAs. To follow up on the potential biological consequences of our results, it will be necessary to go beyond transcription. Here, I show preliminary results to simultaneously image nascent RNAs and proteins in plants. We also discuss the use of ParB DNA labeling in plants and flies and an improvement on optogenetic approaches to study the Dorsal transcription factor.

6.2 Introduction

Most of the experimental results presented in previous chapters revolve around the problem of detecting molecules such as RNAs, DNA, and proteins in living cells. In this context, any extra level of resolution is not only poised to improve the quality of our measurements, it also opens the door to uncovering novel biological phenomena. Detecting inactive or weakly active alleles is particularly relevant in the case of plants because many genes transcribe at relatively low rates, and cells can carry an unknown number of alleles. Part of this chapter is dedicated to describing preliminary results that improve technology to image transcription. Transcriptional regulation is the first layer going from genotype to phenotype. Answering how the quantitative aspects of gene expression impact phenotypic changes requires going beyond RNA to track changes in protein abundance. In this chapter, I describe preliminary efforts to simultaneously image transcription and the protein product of transgenes in plants. On the theoretical front, we advance models that incorporate temporal dynamics. A straightforward way

of testing these models is through dynamical perturbations. This was the motivation to develop the optogenetic tools presented in Chapter 3. Here, I also present preliminary results aimed towards establishing an optogenetic experiment that can dialog with the theoretical models we introduced in Chapter 4.

6.3 Results

6.3.1 Lowering detection thresholds using dimerization-dependent fluorescent proteins

Detecting single transcribing RNA Polymerases in plants would open the door to performing single molecule experiments in living cells within their tissue context. Single polymerase resolution has been achieved in the past in mammalian cancer cell lines using the MCP/MS2 system by tagging a reporter with 128 copies of the MS2 stem loop [253]. Compared to the more standard method of using just 24 stem loop repeats, this approach reduces the detection threshold by increasing the signal while presumably leaving the background and the noise levels untouched (Fig. 6.1A). In many applications it is undesirable to introduce large sequences in the gene of interest and, in addition, alternative methods to decrease the fluorescence detection threshold could have a use outside labeling nascent RNA. As shown in Fig. 6.1A, distinguishing a signal depends on its magnitude being larger than the fluctuations of the background (or background "noise") and on it being larger that the magnitude of the background itself. In principle, by reducing the background and/or its fluctuations it should be possible to improve detection. A promising approach is using fluorescent proteins that only fluoresce upon binding to the molecule being detected. This was the logic behind the methods developed by [273] and [204] where two RNA binding proteins are fused to two fragments of a fluorescent protein which come together when the RNA binding domains recognize the target RNA. However, these methods are based on split fluorescent proteins, which have to undergo maturation after dimerization, a process that can take several minutes. This is much longer than the timescales of RNA Polymerase activity. In contrast, dimerization-dependent fluorescent proteins (ddFPs) contain mature fluorophores that are non-fluorescent in their monomeric state. Dimerization of the ddFP partners ddFPA and ddFPB results in appearance of fluorescence in a matter of seconds [65]. We speculated that by fusing ddFPA to PCP and ddFPB to MCP, it would be possible to reconstitute a fluorescent ddFP dimer by recruiting MCP and PCP to an RNA through the MS2 and PP7 stem loop, respectively (Fig. 6.1B). As a proof of principle, we tested this approach by transiently expressing MCP-ddFPA, PCP-ddFPB and a reporter gene tagged with 48 copies of a sequence containing MS2 and PP7 loops driven by a strong constitutive promoter (Fig. 6.1C). This preliminary experiment shows that, compared to the signal obtained from PCP-GFP binding to 24 PP7 repeats, the ddFP approach outperforms it in terms of signal-to-background and signal-to-noise ratios (Fig. 6.1D). Intramolecular ddFPs were used in the past to create

biosensors based on protein conformation [65]. A promising application of ddFPs is the detection of proximity between proteins. For example, as shown in Fig. 6.1E, ddFPB fused to Histone 3 forms a fluorescent dimer only in the presence of a specific modification to the Histone 3 tail which is recognized by a nanobody fused to ddFPA.

6.3.2 Methods to image transcription and the protein product of transgenes in plant cells to follow the flow of information along the central dogma and across cells.

Transcription is the first of multiple regulatory steps in gene expression that lead to a phenotype. In this flow of information from DNA to phenotype, protein concentration and localization are much closer to gene function than transcription and can also be measured in single live cells using fluorescent proteins and dyes. Protein activity and post-translational modifications are important layers of regulation but are comparatively much harder to measure at the single cell level. The extent of cellular heterogeneity in transcription presented in the previous chapters prompts a number of questions. First, how does this variability impact cell to cell heterogeneity in downstream levels of gene expression? Are there buffering mechanisms between transcription and protein accumulation? Can cells communicate with each other to ensure coordinated levels of protein concentration? What is the impact of heterogeneity at different levels of gene expression in a phenotype such as stress tolerance? To answer these questions it is necessary to measure the transcriptional activity and the protein product of the gene of interest. This can be achieved by placing MS2 or PP7 loops in the 3'UTR with the risk of interfering with mRNA function. Another approach is placing the stem loops within endogenous or synthetic introns. The disadvantage of this strategy is that any level of co-transcriptional splicing will reduce the signal in ways that are not always easy to predict. To demonstrate the feasibility of this approach in plants, the PP7 loops were placed in the first intron of a transgene coding for full-length Arabidopsis HSP70. A C terminal mCherry fusion allows imaging the HSP70 protein while a second transgene coding for PCP-GFP enables visualization of nascent HSP70 mRNAs (Fig. 6.2A). It should be noted that mCherry is a very slowly maturing fluorophore, causing this reporter to underestimate the amount of HSP70 present. Despite the caveats of the intronic approach, it offers the advantage that a PP7/MS2-containing intron placed within the coding sequence of a fluorescent protein constitutes a single self-contained tag to report on transcription and protein accumulation. Such a tag could be inserted in the 5' or 3' of any gene of interest. I created this tag by introducing a potato intron within the coding sequence of mGFPmut2, the fastest maturing fluorescent protein [15]. As shown in Fig. 6.2B, coexpressing this tag and PCP-mCherry results in the appearance of mCherry transcription spots and GFP in the same cell. Even fast maturing fluorescent proteins do not report on the instantaneous protein concentration. This is particularly problematic for proteins with fast production and degradation rates. To go around this problem, [39]

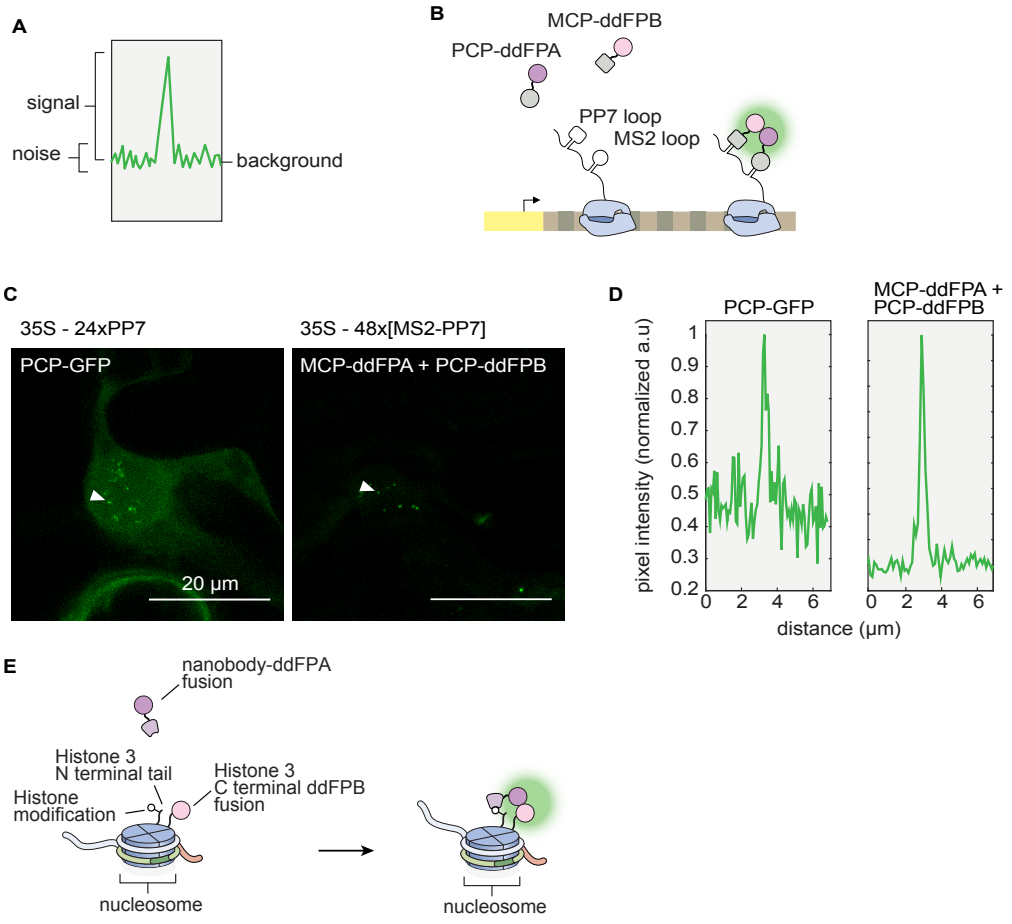

Figure 6.1: **Improving fluorescence detection with dimerization-dependent fluorescent proteins.(A)** Cartoon of the pixel intensity profile of a diffraction-limited fluorescent spot. The signal can be understood as the difference between the peak intensity and the background level. The noise is associated with the magnitude of the background fluctuations. **(B)** Schematic representation of the ddFP approach to label nascent RNAs. The RNA is tagged with repeats of interlaced MS2 and PP7 loops. The quenched ddFP binding partners ddFPA and ddFPB are brought to close proximity by binding of MCP and PCP to their corresponding loops. Dimerization of ddFPA and ddFPB restores fluorophore fluorescence in a timescale of seconds. **(C)** Snapshots of transcription of a 35S reporter detected via PCP-GFP or ddFPs. **(D)** Pixel intensity profile of the spots indicated by white arrowheads in (C). **(E)** The ddFP approach could be used to detect protein proximity in real time with minimal background signal. For example, a nanobody binding a specific modification in the tail of Histone 3 could be used to restore fluorescence to a ddFP monomer fused to its C terminus.

developed the LlamaTag approach. Here, the protein of interest is fused to a nanobody raised against a fluorescent protein such as mCherry or GFP. Genetically provided mature fluorescent proteins are then non-covalently bound to the protein of interest (Fig. 6.2C, top). To test LlamaTags in plants, I coexpressed in Tobacco a nanobody-NLS fusion and free GFP. In Tobacco cells expressing free GFP only, the protein freely diffuses in the cytoplasm and into the nucleus. Coexpression of the nanobody-NLS transgene sequesters GFP in the nucleus, demonstrating that LlamaTags can be used to track proteins in plants (Fig. 6.2D). In the future, these technologies will be useful to study non-cell autonomous signaling in plant stress responses. For example, after exposure to environmental inputs including high light, UV radiation and nitrogen sources, Arabidopsis leaves induce the expression of HY5 [92]. It has been reported that the HY5 protein can diffuse between cells through plasmodesmata and that this movement allows inter-organ communication of high light stress [139] and energy source levels [50]. It is possible that HY5 movement at much shorter scales can help buffer noise at the transcriptional level to promote coordinated tissue-level responses. Another interesting example of inter-cellular communication is the high light inducible expression of ZAT10 and ZAT12 transcription factors in tissues that do not experience direct stress exposure [226, 105]. ZAT10 and ZAT12 have been implicated in photoprotection [131, 187], raising the possibility that this cell-to-cell mobile signal can play a role in coordinating the high light stress response at the tissue level. To pursue these ideas, it will be necessary to track transcription and accumulation of the protein product of these genes and later determine the level of stress protection at the single cell level (Fig. 6.2F). Multiple stress treatments such as excess light generate reactive oxygen species, which can lead to proteotoxic stress by modifying protein residues. Heat shock also generates proteotoxicity by promoting protein misfolding. Interestingly, it is possible to measure the stability of the proteome using dyies that recognize misfolded proteins [173]. Such assay could be performed after live imaging to determine the physiological consequences of single cell gene expression behaviors.

6.3.3 Avenues to improve optogenetic Dorsal experiments

As shown in Chapter 2, using a new approach based on nanobodies we were able to generate a functional version of the Dorsal transcription factor in fruit flies (Fig. 2.5. However, when we tested the effect of optogenetic perturbations on the expression of a minimal enhancer containing a single Dorsal binding site, we found no differences compared to the unperturbed control. This should not be surprising since in the range of Dorsal concentration reached by this fly line the activity of this enhancer is barely modulated. Thus, to establish a viable optogenetic experiment there are two possible but not mutually exclusive strategies. First, using an enhancer with a sharper response to Dorsal concentration could improve the sensitivity of transcription to the optogenetic perturbation. One such enhancer is the proximal element from the *twist* gene (TwiPE). The fraction of refractory nuclei for this enhancer goes from 0 to $\approx$0.8 in the

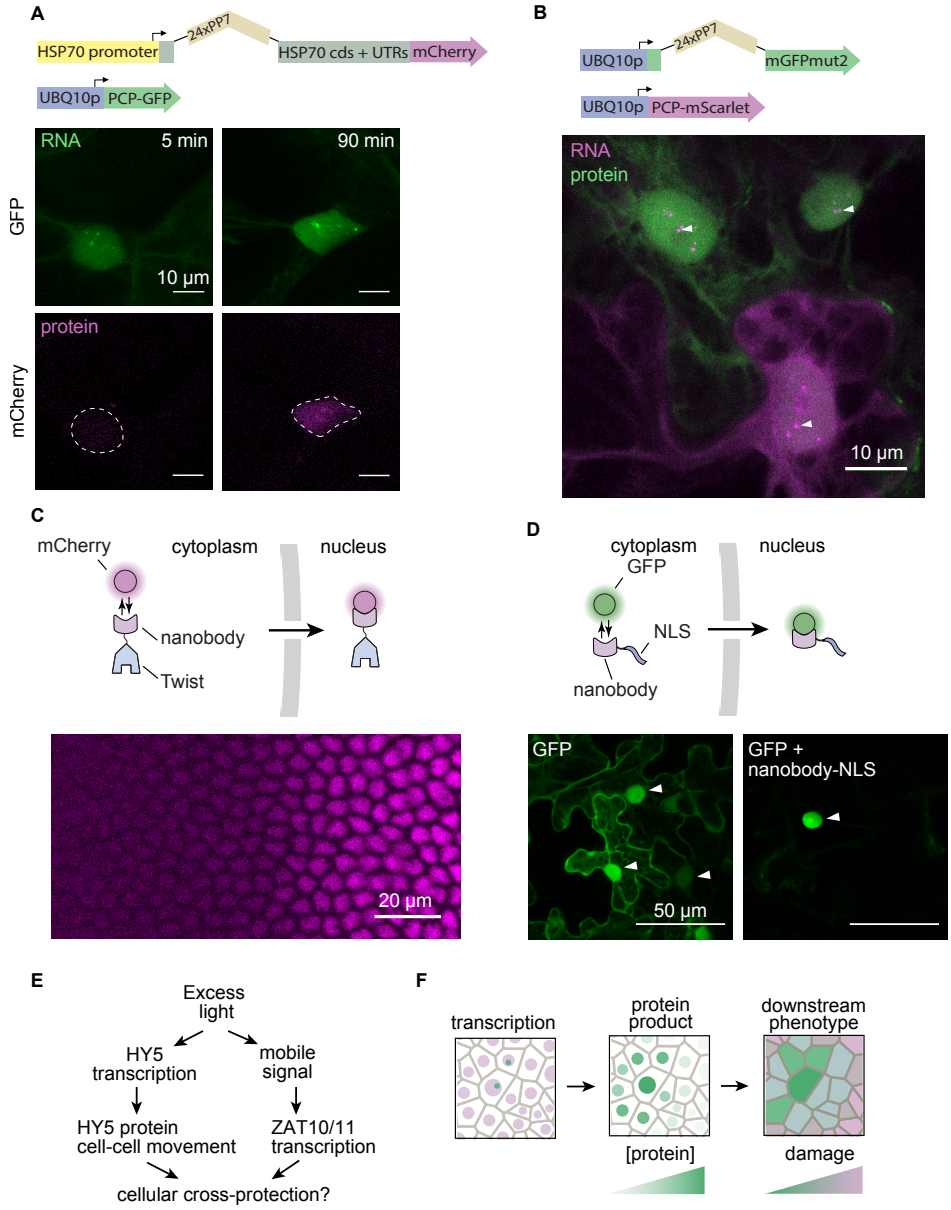

Figure 6.2: **Methods to measure transcription and translation products simultaneously in live plant cells and their potential applications**. **(A)** Imaging transcription and the protein product of Arabidopsis HSP70. Top: the PP7 loops were placed in the first intron of HSP70 and mCherry was fused to its C terminus. PCP-GFP was used to label nascent transgene RNA. Bottom: Tobacco cells expressing these constructs during a heat shock treatment. **(B)** Placing the PP7 loops within a synthetic intron in mGFPmut2 allows imaging transcription with PCP-mScarlet (magenta spots highlighted by white arrowheads) and the protein product. ***Caption continues on next page***

Figure 6.2: ***Continued from previous page*: Methods to measure transcription and translation products simultaneously in live plant cells and their potential applications.(C)** Top: LlamaTag approach to imaging the Twist protein in fruit fly embryos. An anti-mCherry nanobody was fused to the C terminus of the *twist* gene. Coexpressing mCherry allows fluorescently labeling the Twist protein. Bottom: Snapshot of the Twist gradient. **(D)** Top: Testing LlamaTags in plants. Bottom: Tobacco cells expressing free GFP with or without coexpressing a nanobody-NLS fusion. **(E)** Examples of non-cell autonomous signaling in the transcriptional response to excess light in Arabidopsis. **(F)** To study the consequences of the pathways in (E) at the cellular level, it will be necessary to distinguish cells that transcribe the gene from cells in which the protein is present and then to correlate this to a phenotypic outcome.

range of concentrations in which the previously tested enhancer hovers around ≈0.3 (Fig. 6.3A). This enhancer is thought to respond exclusively to Dorsal as its spatial input [138]. However, because it has two Dorsal binding sites placed within ≈ 400 bp of extra DNA, it does not easily lend itself to be treated with the same theoretical tools developed for the minimal enhancers in Chapter 3. The second strategy involves increasing the concentration of optogenetic Dorsal. Because the Llama Shepherd approach can be applied to any genotype, in practice, this entails increasing the dosage of Dorsal-Venus. To achieve this, we created a transgene carrying the Dorsal coding sequence fused to Venus driven by a strong maternally expressed promoter (Fig. 6.3B). A preliminary test of this transgene showed that it drives Dorsal expression at levels comparable to a CRISPR Venus knock-in (data not shown).

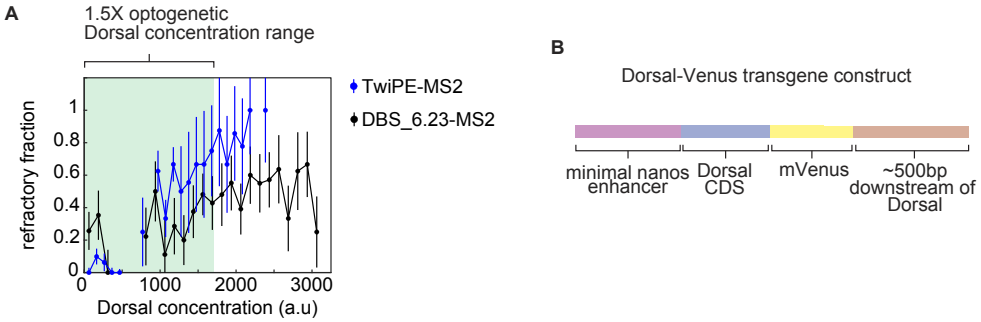

Figure 6.3: **Improvements to the optogenetic Dorsal experiment.(A)** Comparison of the fraction of refractory cells between the minimal DBS-6.23 enhancer and a reporter driven by the Twist proximal element (TwiPE). In the range of concentration observed in the optogenetic 1.5X Dorsal Llama Shepherd, DBS-6.23 is barely modulated while TwiPE goes from 0 to ≈ 0.8. **(B)** A Dorsal-Venus transgene driven by a strong enhancer should allow further increasing the Dorsal concentration and thus increasing the dynamic range of optogenetic Dorsal.

6.3.4 Using ParB labeling of DNA to improve measurements of transcription factor inputs and transcriptional outputs.

The nucleus is not a well-mixed uniform environment. Chromatin forms stable spatial domains and proteins such as transcription factors and RNA Polymerase tend to preferentially occupy discrete volumes within the nucleus. The presence of phase-separated nuclear compartments has long been recognized in the case of the nucleolus, but recent evidence supports a role for phase separation in transcription as well. Despite this mounting evidence, efforts to correlate transcription factor inputs to transcriptional outputs tend to use the average nuclear concentration. Recently, there have been efforts to use the position of the gene locus to correlate the local transcription factor concentration to transcriptional activity. However, these approaches depend on the detection of transcription using live imaging and as a consequence they cannot detect the protein input when transcription is not detected [67, 85]. These shortcomings could be overcome by labeling DNA instead of nascent RNA. Labeling of DNA using the bacterial ParB-*parS* system was introduced in Chapter 1 and Chapter 3. Briefly, hundred of ParB proteins polymerize on the surface of the *parS* site, fusing ParB to a fluorescent protein results in the appearance of bright diffraction-limited spots at the position of the locus (Fig. 6.4A). Tracking inactive loci is particularly important in the case of polyploid cells where, out of several alleles, only a few might be active at a time. In addition to accounting for allele-specific activity, ParB labeling could be used to unequivocally determine cellular ploidy in live imaging experiments (Fig. 6.4B). The relevance of local protein concentration is exemplified by the nuclear distribution of PIF3 as revealed by a GFP fusion (Fig. 6.4C, left). PIF3, like other proteins involved in light signaling forms dynamic "nuclear speckles" also known as "photobodies" in response to light [223]. The Dorsal transcription factor does not form such distinctive nuclear foci as PIF3 but it is far from homogeneously distributed in the nucleus. Optogenetic export of nuclear Dorsal-Venus reveals the presence of an aggregated fraction (Fig. 6.4C, right). Three channels are necessary to measure transcriptional activity and local transcription factor concentration independent of detecting transcription. One channel is dedicated to the detection of the locus through ParB, while the other two channels detect the protein and nascent RNA at the ParB spot position (Fig. 6.4D), respectively. To implement this approach, we created a fly line carrying a Dorsal-Venus CRISPR knock-in, MCP-mCherry and ParB-GFP. The emission spectra of GFP Venus are very close, but it should still be possible to separate them with linear unmixing.

6.4 Discussion

The preliminary results presented in this Chapter offer concrete, viable ways of surmounting the experimental limitations we encountered throughout the course of this research. They offer a way to give more precise answers to the questions addressed

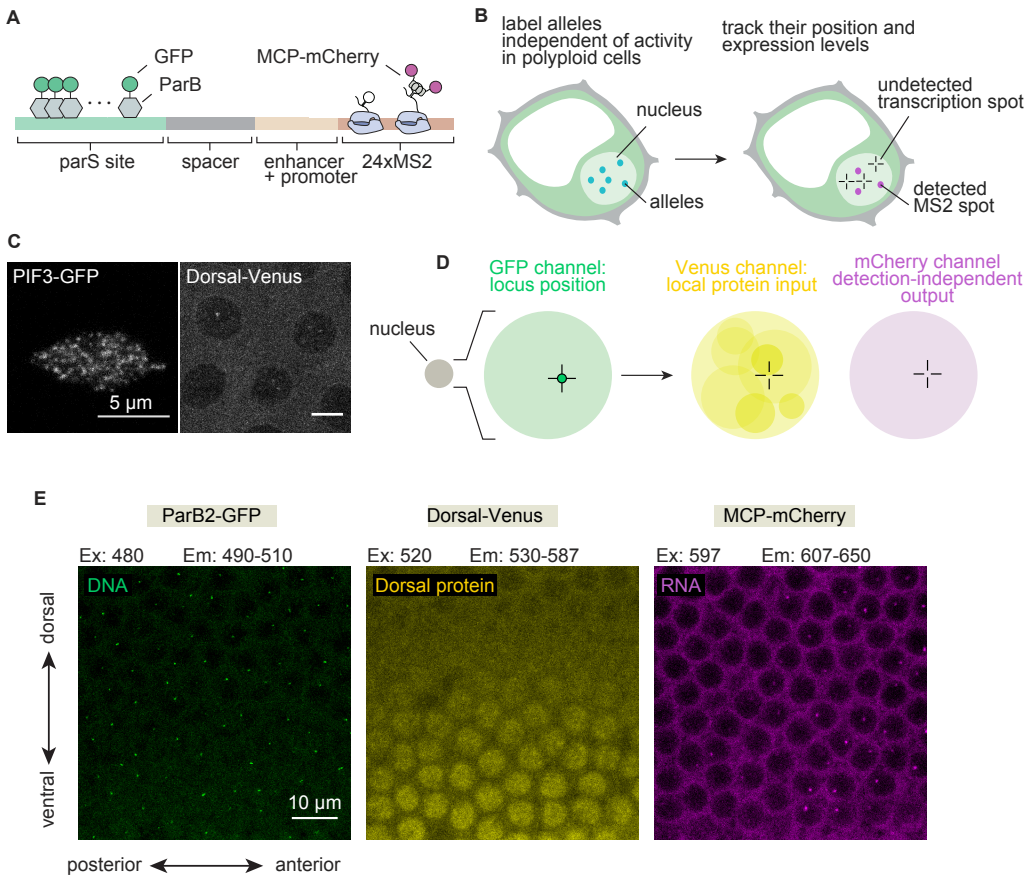

Figure 6.4: **Future applications of ParB DNA labeling.(A)** Schematic of the ParB system to label DNA simultaneously with RNA imaging via MS2 or PP7. **(B)** Left: Cartoon of a polyploid plant cell carrying six transgene alleles labeled with ParB-GFP. Right: Cartoon of the same cell where the transcriptional activity from these alleles is detected with MS2 or PP7. The black cross hairs represent the position of undetected alleles based on the ParB signal. **(C)** Images of nuclei where a transcription factor was fluorescently labeled showing non-homogeneous distribution of this protein. PIF3-GFP was imaged in Arabidopsis cotyledons, Dorsal-Venus is expressed in fruit fly embryos. **(D)** Schematic of a three channel experiment where one channel is dedicated to locating the gene of interest using ParB. This positional information is then used to measure the transcription factor signal nearby the gene. The transcriptional activity is measured by calculating the fluorescence of the MCP or PCP channel in the position of the ParB spot.

throughout the previous chapters and open the door to studying previously inaccessible phenomena.

6.5 Materials and Methods

The plasmids used in the experiments shown in each panel throughout this chapter can be found in Table 6.1. The methods associated with the experiments presented in this chapter can be found in previous chapters as follows:

- **Figure 6.1**: Agroinfilitration and imaging setup is described in detail in Chapter 4.
- **Figure 6.2**: Agroinfilitration and imaging setup used in Figure 6.2A, B and D is described in detail in Chapter 4. The experiment in Figure 6.2A, including the heat shock treatment is described in detail in Chapter 4, Figure 4.7.

6.6 Biological Material

Plasmids		
Name (hyperlinked to Benchling)	**Description**	**code**
UMddGA	Ubiquitous expression of MCP-ddGFPA in plants	
UPddGB	Ubiquitous expression of PCP-ddGFPB1 in plants	
UPddGB3	Ubiquitous expression of PCP-ddGFPB3 in plants	
UMddGA-UPddGB	Ubiquitous expression of MCP-ddGFPA and PCP-ddGFPB in plants used in Figure 6.1	
UMddGA-UPddGBv2	Ubiquitous expression of MCP-ddGFPA and PCP-ddGFPB3 in plants	
48xMP-35S	reporter construct driven by the 35S promoter and tagged with 48 repeats of PP7 and MS2. Used in Figure 6.1	
HSP70-pp7i-mChUPG	Arabidopsis *HSP70* fused to mCherry and tagged with PP7v5 in its first intron. Expression of PCP-GFP. Used in Figure 6.2A	
UmGFPmut2pp7i	Ubiquitous expression of mGFPmut2 carrying an intron with the PP7v5 loops in it. Used in Figure 6.2B	
UPmS	Ubiquitous expression of PCP-mScarlet fusion. Used in Figure 6.2B	
TwCJB10dsRed	Donor construct for CRISPR C terminal fusion of the JB10 mCherry nanobody to the Drosophila *twist* gene. Used in Figure 6.2C	
TwgRNA3	synthetic guide RNA to knock in DNA to the C terminus of Drosophila. Used in Figure 6.2C *twist*	

Table 6.1: List of plasmids used in this chapter. *Continues on next page*

Plasmids		
Name (hyperlinked to Benchling)	**Description**	**code**
UsfG	Ubiquitous expression of sfGFP in plants. Used in Figure 6.2D	
UJB3nls	Ubiquitous expression of a JB3 GFP nanobody carrying a nuclear localization signal. Used in Figure 6.2D	
pIB-1Dg-MS2v5-LacZ-tub3UTR	Minimal MS2 reporter carrying a single strong Dorsal binding site. Used in Figure 6.3	
pIB-TwiPEv2-MS2v5-LacZ-tub3UTR	MS2 reporter driven by the Twist proximal element. Used in Figure 6.3	
pBPhi-His-miRFP670-gyp-2xeNos-Dl-Venus-gyp-eNosx2-MCP-mCherry	Maternal expression of Dorsal-Venus, Histone-iRFP and MCP-mCherry. Used in Figure 6.3	

Table 6.1: *Continued from previous page*: List of plasmids used in this chapter.

Fly lines	
genotype	**Function**
yw;1Dg(11);+	minimal Dorsal synthetic enhancer MS2 reporter in Figure 6.3
yw;TwiPEv5(7);+	MS2 reporter driven by the *twist* proximal element in Figure 6.3
yw;2xIntB2-hbP2P-MS2x1-lacZ (2)/CyO;+	MS2 reporter driven by the *hunchback* P2 enhancer and promoter taged with 2 ParB2 binding sites. See Figure 6.4E
yw;Dorsal-Venus,MCP-mCherry;Dorsal-Venus,His-iRFP,MCP-mCherry	2X Dorsal-Venus used to generate 1.5x Dorsal-Venus Llama Shepherd transheterozygote in Figure 6.3
yw;Dorsal-Venus,MCP-mCherry;JB3-LEXY	1X Dorsal-Venus Llama Shepherd used to generate 1.5x Dorsal-Venus Llama Shepherd transheterozygote in Figure 6.3
yw;Dorsal-Venus,MCP-mCherry;Dorsal-Venus,[Dorsal-Venus,MCP-mCherry,His-iRFP]	3x Dorsal-Venus used to generate used to generate 2.5x Dorsal-Venus Llama Shepherd transheterozygote in Figure 6.3
yw;Dorsal-Venus,ParB2-GFP,MCP-mCherry	imaging of DNA, nascent RNA and Dorsal simultaneously in Figure 6.4

Table 6.2: List of fly lines used in this chapter.

Bibliography

[1] Katie Abley, James C.W. Locke, and H. M.Ottoline Leyser. "Developmental mechanisms underlying variable, invariant and plastic phenotypes". In: *Annals of Botany* 117.5 (2016), pp. 733–748. ISSN: 10958290. DOI: 10.1093/aob/mcw016.

[2] Matthew Akamatsu et al. "Principles of self-organization and load adaptation by the actin cytoskeleton during clathrin-mediated endocytosis". In: *eLife* 9 (2020), pp. 1–40. ISSN: 2050084X. DOI: 10.7554/eLife.49840.

[3] Bassem Al-Sady et al. "Photoactivated Phytochrome Induces Rapid PIF3 Phosphorylation Prior to Proteasome-Mediated Degradation". In: *Molecular Cell* 23.3 (2006), pp. 439–446. ISSN: 10972765. DOI: 10.1016/j.molcel.2006.06.011.

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

[5] Henrik Albert et al. "Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome". In: *The Plant Journal* 7.4 (1995), pp. 649–659. ISSN: 1365313X. DOI: 10.1046/j.1365-313X.1995.7040649.x.

[6] José M. Alvarez et al. "Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade". In: *Nature Communications* 11.1 (2020), pp. 1–13. ISSN: 20411723. DOI: 10.1038/s41467-020-14979-6.

[7] Andrew Angel et al. "A Polycomb-based switch underlying quantitative epigenetic memory". In: *Nature* 476.7358 (2011), pp. 105–109. ISSN: 00280836. DOI: 10.1038/nature10241.

[8] Ilka Schultheiß Araújo et al. "Stochastic gene expression in Arabidopsis thaliana ". In: *Nature Communications* 8.1 (2017), p. 2132. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02285-7. URL: https://doi.org/10.1038/s41467-017-02285-7.

[9] M. Behfar Ardehali and John T. Lis. "Tracking rates of transcription and splicing in vivo". In: *Nature Structural and Molecular Biology* 16.11 (2009), pp. 1123–1124. ISSN: 15459993. DOI: 10.1038/nsmb1109-1123.

[10] Ignacio Arganda-Carreras et al. “Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification”. In: *Bioinformatics* 33.15 (2017), pp. 2424–2426. ISSN: 14602059. DOI: 10.1093/bioinformatics/btx180.

[11] Žiga Avsec et al. *Deep learning at base-resolution reveals motif syntax of the cis-regulatory code*. 2019, pp. 0–62. ISBN: 3550445335504.

[12] Keren Bahar Halpern et al. “Bursty gene expression in the intact mammalian liver”. In: *Molecular Cell* 58.1 (2015), pp. 147–156. ISSN: 10974164. DOI: 10.1016/j.molcel.2015.01.027. URL: http://dx.doi.org/10.1016/j.molcel.2015.01.027.

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

[14] Timothy L Bailey et al. “MEME : discovering and analyzing DNA and protein sequence motifs”. In: 34 (2006), pp. 369–373. DOI: 10.1093/nar/gkl198.

[15] Enrique Balleza, J Mark Kim, and Philippe Cluzel. “Systematic characterization of maturation time of fluorescent proteins in living cells”. In: 15.1 (2018). DOI: 10.1038/NMETH.4509.

[16] Roopa Banerjee et al. “The signaling state of Arabidopsis cryptochrome 2 contains flavin semiquinone”. In: *Journal of Biological Chemistry* 282.20 (2007), pp. 14916–14922. ISSN: 00219258. DOI: 10.1074/jbc.M700616200.

[17] C J Bashor et al. “Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies”. In: 8287.April (2019), pp. 1–11.

[18] Jack R. Bateman, Anne M. Lee, and C. Ting Wu. “Site-specific transformation of Drosophila via ϕC31 integrase-mediated cassette exchange”. In: *Genetics* 173.2 (2006), pp. 769–777. ISSN: 00166731. DOI: 10.1534/genetics.106.056945.

[19] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. “Control of Transcript Variability in Single Mammalian Cells”. In: *Cell* 163.7 (2015), pp. 1596–1610. ISSN: 10974172. DOI: 10.1016/j.cell.2015.11.018. URL: http://dx.doi.org/10.1016/j.cell.2015.11.018.

[20] Diana Bauer et al. “Constitutive photomorphogenesis 1 and multiple photoreceptors control degradation of phytochrome interacting factor 3, a transcription factor required for light signaling in arabidopsis”. In: *Plant Cell* 16.6 (2004), pp. 1433–1445. ISSN: 10404651. DOI: 10.1105/tpc.021568.

[21] Nathan M Belliveau et al. “Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria”. In: (2018). DOI: 10.1073/pnas.1722055115.

[22] S. Benzer. “Induced synthesis of enzymes in bacteria analyzed at the cellular level”. In: *BBA - Biochimica et Biophysica Acta* 11.C (1953), pp. 383–395. ISSN: 00063002. DOI: 10.1016/0006-3002(53)90057-2.

[23] Casey M Bergman et al. “Databases and ontologies Drosophila DNase I footprint database : a systematic genome annotation of transcription factor binding sites in the fruitfly , Drosophila melanogaster”. In: 21.8 (2005), pp. 1747–1749. DOI: 10.1093/bioinformatics/bti173.

[24] Augusto Berrocal et al. “Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene”. In: *bioRxiv (in press at eLife)* (2020), p. 335901. DOI: 10.1101/335901.

[25] Edouard Bertrand. “Localization of ASH1 mRNA Particles in Living Yeast”. In: *Molecular Cell* 2.4 (2004), pp. 437–445. ISSN: 10972765. DOI: 10.1016/s1097-2765(00)80143-4.

[26] Edouard Bertrand et al. “Localization of ASH1 mRNA Particles in Living Yeast”. In: *Molecular Cell* 2.4 (1998), 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.

[27] William Bialek and Thomas Gregor. “Action at a distance in transcriptional regulation”. In: (2019), pp. 1–5. arXiv: arXiv:1912.08579v1.

[28] Daphne S Bindels et al. “mScarlet : a bright monomeric red fluorescent protein for cellular imaging”. In: november (2016). DOI: 10.1038/nmeth.4074.

[29] L. Bintu et al. “Transcriptional regulation by the numbers: applications”. In: *Curr Opin Genet Dev* 15.2 (2005), pp. 125–35.

[30] L. Bintu et al. “Transcriptional regulation by the numbers: models”. In: *Curr Opin Genet Dev* 15.2 (2005), pp. 116–24.

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

[32] Kenneth D. Birnbaum. “Power in Numbers: Single-Cell RNA-Seq Strategies to Dissect Complex Tissues”. In: *Annual Review of Genetics* 52.1 (2018), pp. 203–221. ISSN: 0066-4197. DOI: 10.1146/annurev-genet-120417-031247.

[33] Marta Bjornson et al. “The transcriptional landscape of Arabidopsis thaliana pattern-triggered immunity”. In: *Nature Plants* (2021). DOI: 10.1038/s41477-021-00874-5.

[34] Elizabeth M. Blackwood and James T. Kadonaga. “Going the distance: A current view of enhancer action”. In: *Science* 281.5373 (1998), pp. 60–63. ISSN: 00368075. DOI: 10.1126/science.281.5373.60.

[35] Adam J Bogdanove and Daniel F Voytas. “TAL effectors: customizable proteins for DNA targeting.” In: *Science (New York, N.Y.)* 333.6051 (2011), pp. 1843–6. ISSN: 1095-9203. DOI: 10.1126/science.1204094. URL: http://www.ncbi.nlm.nih.gov/pubmed/21960622.

[36] Ann Boija et al. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains”. In: *Cell* 175.7 (2018), 1842–1855.e16. ISSN: 10974172. DOI: 10.1016/j.cell.2018.10.042. URL: https://doi.org/10.1016/j.cell.2018.10.042.

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

[38] Jacques P. Bothma et al. “Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo”. In: *eLife* 4.AUGUST2015 (2015). ISSN: 2050084X. DOI: 10.7554/eLife.07956.001.

[39] Jacques P. Bothma et al. “LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos”. In: *Cell* 173.7 (2018), 1810–1822.e16. ISSN: 10974172. DOI: 10.1016/j.cell.2018.03.069. URL: http://dx.doi.org/10.1016/j.cell.2018.03.069.

[40] K. J. Bradford and A. J. Trewavas. “Sensitivity thresholds and variable time scales in plant hormone action”. In: *Plant Physiology* 105.4 (1994), pp. 1029–1036. ISSN: 00320889. DOI: 10.1104/pp.105.4.1029.

[41] Christopher Buccitelli and Matthias Selbach. “mRNAs, proteins and the emerging principles of gene expression control”. In: *Nature Reviews Genetics* 21.10 (2020), pp. 630–644. ISSN: 14710064. DOI: 10.1038/s41576-020-0258-4. URL: http://dx.doi.org/10.1038/s41576-020-0258-4.

[42] L. Cai, N. Friedman, and X. S. Xie. “Stochastic protein expression in individual cells at the single molecule level”. In: *Nature* 440.7082 (2006), pp. 358–62. ISSN: 1476-4687 (Electronic).

[43] Enrico Cannavò et al. “Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks”. In: *Current Biology* 26.1 (2016), pp. 38–51. ISSN: 09609822. DOI: 10.1016/j.cub.2015.11.034.

[44] Stefano Ceolin et al. “A sensitive mNeonGreen reporter system to measure transcriptional dynamics in Drosophila development”. In: *Communications Biology* 3.1 (2020), pp. 1–12. ISSN: 23993642. DOI: 10.1038/s42003-020-01375-5. URL: http://dx.doi.org/10.1038/s42003-020-01375-5.

[45] Tomáš Čermák et al. “High-frequency, precise modification of the tomato genome”. In: *Genome Biology* 16.1 (2015). ISSN: 1474760X. DOI: 10.1186/s13059-015-0796-9.

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

[47] Yee-yung Charng et al. “A Heat-Inducible Transcription Factor , HsfA2 , Is Required for Extension of Acquired Thermotolerance”. In: 143.January (2007), pp. 251–262. DOI: 10.1104/pp.106.091322.

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

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

[50] Xiangbin Chen et al. “Shoot-to-Root Mobile Transcription Factor HY5 Coordinates Plant Carbon and Nitrogen Acquisition”. In: *Current Biology* 26.5 (2016), pp. 640–646. ISSN: 09609822. DOI: 10.1016/j.cub.2015.12.066. URL: http://dx.doi.org/10.1016/j.cub.2015.12.066.

[51] Eun Kyung Cho and Young Ju Choi. “A nuclear-localized HSP70 confers thermoprotective activity and drought-stress tolerance on plants”. In: *Biotechnology Letters* 31.4 (2009), pp. 597–606. ISSN: 01415492. DOI: 10.1007/s10529-008-9880-5.

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

[53] Adam M Corrigan et al. “A continuum model of transcriptional bursting”. In: *eLife* 5 (2016), pp. 1–38. DOI: 10.7554/elife.13051.

[54] Sandra Cortijo and James C.W. Locke. “Does Gene Expression Noise Play a Functional Role in Plants?” In: *Trends in Plant Science* (2020), pp. 1–11. ISSN: 13601385. DOI: 10.1016/j.tplants.2020.04.017. URL: https://doi.org/10.1016/j.tplants.2020.04.017.

[55] Sandra Cortijo et al. “Transcriptional Regulation of the Ambient Temperature Response by H2A.Z Nucleosomes and HSF1 Transcription Factors in Arabidopsis”. In: *Molecular Plant* 10.10 (2017), pp. 1258–1273. ISSN: 17529867. DOI: 10.1016/j.molp.2017.08.014. URL: https://doi.org/10.1016/j.molp.2017.08.014.

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

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

[58] Peter A Crisp et al. *Rapid Recovery Gene Downregulation During Excess-light Stress and Recovery in Arabidopsis*. 2017. ISBN: 0000000322683. DOI: 10.1105/tpc.16.00828.

[59] Justin Crocker, Garth R Ilsley, and David L Stern. "Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors." In: *Nature genetics* 48.3 (2016), pp. 292–8. ISSN: 1546-1718. DOI: 10.1038/ng.3509. URL: http://www.ncbi.nlm.nih.gov/pubmed/26854918.

[60] Tomasz Czechowski et al. "Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis". In: *Plant Physiology* 139.September (2005), pp. 5–17. ISSN: 00320889. DOI: 10.1104/pp.105.063743.1. arXiv: 1409.2752. URL: http://arxiv.org/abs/1409.2752{\%}5Cnhttp://www.arxiv.org/pdf/1409.2752.pdf.

[61] Nathalie Daigle and Jan Ellenberg. "λN-GFP: An RNA reporter system for live-cell imaging". In: *Nature Methods* 4.8 (2007), pp. 633–636. ISSN: 15487091. DOI: 10.1038/nmeth1065.

[62] Xavier Darzacq et al. "In vivo dynamics of RNA polymerase II transcription". In: 14.9 (2007), pp. 796–806. DOI: 10.1038/nsmb1280.

[63] Sulagna Das et al. "A transgenic mouse for imaging activity-dependent dynamics of endogenous arc mRNA in live neurons". In: *Science Advances* 4.6 (2018), pp. 1–14. ISSN: 23752548. DOI: 10.1126/sciadv.aar3448.

[64] M. Delarue et al. "mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding". In: *Cell* 174.2 (2018), 338–349.e20. ISSN: 10974172. DOI: 10.1016/j.cell.2018.05.042.

[65] Yidan Ding et al. "Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange". In: *Nature Methods* 12.3 (2015), pp. 195–198. ISSN: 15487105. DOI: 10.1038/nmeth.3261.

[66] Dong Yul Sung, E. Vierling, and C. L. Guy. "Comprehensive expression profile analysis of the Arabidopsis hsp70 gene family". In: *Plant Physiology* 126.2 (2001), pp. 789–800. ISSN: 00320889. DOI: 10.1104/pp.126.2.789.

[67] Benjamin T Donovan et al. "Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting". In: *EMBO J* 38.12 (2019), e100809. ISSN: 0261-4189.

[68] Wolfgang Driever, Gudrun Thoma, and Christiane Nüsslein-Volhard. "Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen". In: *Nature* 340.6232 (1989), pp. 363–367. ISSN: 00280836. DOI: 10.1038/340363a0.

[69] Jeremy Dufourt et al. “Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs”. In: *Nature Communications* 9.1 (2018), pp. 1–13. ISSN: 20411723. DOI: 10.1038/s41467-018-07613-z. URL: http://dx.doi.org/10.1038/s41467-018-07613-z.

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

[71] Susan Duncan et al. “A method for detecting single mRNA molecules in Arabidopsis thaliana”. In: (2016), pp. 1–10. DOI: 10.1186/s13007-016-0114-x.

[72] Elizabeth Eck et al. “Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility”. In: *eLife* 9 (2020), pp. 1–99. ISSN: 2050084X. DOI: 10.7554/eLife.56429.

[73] Klaus Eichenberg et al. “Arabidopsis phytochromes C and E have different spectral characteristics from those of phytochromes A and B”. In: *FEBS Letters* 470.2 (2000), pp. 107–112. ISSN: 00145793. DOI: 10.1016/S0014-5793(00)01301-6.

[74] Imane El Meouche, Yik Siu, and Mary J. Dunlop. “Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells”. In: *Scientific Reports* 6.January (2016), pp. 1–9. ISSN: 20452322. DOI: 10.1038/srep19538.

[75] Avigdor Eldar and Michael B Elowitz. “Functional roles for noise in genetic circuits.” In: *Nature* 467.7312 (2010), pp. 167–173. ISSN: 0028-0836. DOI: 10.1038/nature09326. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1038/nature09326.

[76] Michael B. Elowitz et al. “Stochastic gene expression in a single cell”. In: *Science* 297.5584 (2002), pp. 1183–1186. ISSN: 00368075. DOI: 10.1126/science.1070919.

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

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

[79] Julia Falo-Sanjuan et al. “Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling”. In: *Developmental Cell* 50.4 (2019), 411–425.e8. ISSN: 18781551. DOI: 10.1016/j.devcel.2019.07.002. URL: https://doi.org/10.1016/j.devcel.2019.07.002.

[80] Christine Faulkner. “Plasmodesmata and the symplast”. In: *Current Biology* 28.24 (2018), R1374–R1378. ISSN: 09609822. DOI: 10.1016/j.cub.2018.11.004. URL: http://dx.doi.org/10.1016/j.cub.2018.11.004.

[81] Fernán Federici et al. “Integrated genetic and computation methods for in planta cytometry.” In: *Nature methods* 9.5 (2012), pp. 483–5. ISSN: 1548-7105. DOI: 10.1038/nmeth.1940. URL: http://dx.doi.org/10.1038/nmeth.1940.

[82] Teresa Ferraro et al. “New methods to image transcription in living fly embryos: The insights so far, and the prospects”. In: *Wiley Interdisciplinary Reviews: Developmental Biology* 5.3 (2016), pp. 296–310. ISSN: 17597692. DOI: 10.1002/wdev.221.

[83] Steve Fiering, Emma Whitelaw, and David I.K. Martin. “To be or not to be active: The stochastic nature of enhancer action”. In: *BioEssays* 22.4 (2000), pp. 381–387. ISSN: 02659247. DOI: 10.1002/(SICI)1521-1878(200004)22:4<381::AID-BIES8>3.0.CO;2-E.

[84] Robert Foreman and Roy Wollman. “Mammalian gene expression variability is explained by underlying cell state”. In: *Molecular Systems Biology* 16.2 (2020), pp. 1–13. ISSN: 1744-4292. DOI: 10.15252/msb.20199146.

[85] Linda S Forero-quintero et al. “Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene”. In: *bioRxiv* (2020), p. 2020.04.03.024414. DOI: 10.1101/2020.04.03.024414.

[86] Peter C. Fridy et al. “A robust pipeline for rapid production of versatile nanobody repertoires”. In: *Nature Methods* 11.12 (2014), pp. 1253–1260. ISSN: 15487105. DOI: 10.1038/nmeth.3170.

[87] Audrey Qiuyan Fu and Lior Pachter. “Estimating intrinsic and extrinsic noise from single-cell gene expression measurements”. In: *Statistical Applications in Genetics and Molecular Biology* 15.6 (2016), pp. 447–471. ISSN: 15446115. DOI: 10.1515/sagmb-2016-0002. arXiv: 1601.03334.

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

[89] T. Fukaya, B. Lim, and M. Levine. “Rapid Rates of Pol II Elongation in the Drosophila Embryo”. In: *Curr Biol* 27.9 (2017), pp. 1387–1391. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: 10.1016/j.cub.2017.03.069.

[90] Timothy Fuqua et al. “Dense and pleiotropic regulatory information in a developmental enhancer”. In: *Nature* 587.7833 (2020), pp. 235–239. ISSN: 14764687. DOI: 10.1038/s41586-020-2816-5. URL: http://dx.doi.org/10.1038/s41586-020-2816-5.

[91] David W. Galbraith, Kristi R. Harkins, and Steven Knapp. “Systemic endopolyploidy in Arabidopsis thaliana”. In: *Plant Physiology* 96.3 (1991), pp. 985–989. ISSN: 00320889. DOI: 10.1104/pp.96.3.985.

[92] Sreeramaiah N. Gangappa and Javier F. Botto. "The Multifaceted Roles of HY5 in Plant Growth and Development". In: *Molecular Plant* 9.10 (2016), pp. 1353–1365. ISSN: 17529867. DOI: 10.1016/j.molp.2016.07.002. URL: http://dx.doi.org/10.1016/j.molp.2016.07.002.

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

[94] H. G. Garcia and R. Phillips. "Quantitative dissection of the simple repression input-output function". In: *Proc Natl Acad Sci U S A* 108.29 (2011), pp. 12173–8. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 1015616108[pii]10.1073/pnas.1015616108.

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

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

[97] Hernan G Garcia and Rob Phillips. "Quantitative dissection of the simple repression input – output function". In: 2011 (2011), pp. 1–6. DOI: 10.1073/pnas.1015616108.

[98] Hernan G Garcia et al. "Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning". In: *Current Biology* 23.21 (2013), pp. 2140–2145. ISSN: 09609822. DOI: 10.1016/j.cub.2013.08.054. URL: http://dx.doi.org/10.1016/j.cub.2013.08.054.

[99] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. "Construction of a genetic toggle switch in Escherichia coli". In: *Nature* 403.6767 (2000), pp. 339–342. ISSN: 00280836. DOI: 10.1038/35002131.

[100] Thomas Germier et al. "Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement". In: *Biophysical Journal* 113.7 (2017), pp. 1383–1394. ISSN: 15420086. DOI: 10.1016/j.bpj.2017.08.014.

[101] Agnieszka A. Gil et al. "Optogenetic control of protein binding using light-switchable nanobodies". In: *Nature Communications* 11.1 (2020), pp. 1–12. ISSN: 20411723. DOI: 10.1038/s41467-020-17836-8. URL: http://dx.doi.org/10.1038/s41467-020-17836-8.

[102] Stephen S. Gisselbrecht et al. "Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos". In: *Nature Methods* 10.8 (2013), pp. 774–780. ISSN: 15487091. DOI: 10.1038/nmeth.2558.

[103] Ido Golding. "Single-Cell Studies of Phage λ: Hidden Treasures under Occam's Rug". In: *Annual Review of Virology* 3 (2016), pp. 453–472. ISSN: 23270578. DOI: 10.1146/annurev-virology-110615-042127.

[104] Ido Golding et al. "Real-time kinetics of gene activity in individual bacteria". In: *Cell* 123.6 (2005), pp. 1025–1036. ISSN: 00928674. DOI: 10.1016/j.cell.2005.09.031.

[105] A. Gordon et al. "Single-cell quantification of molecules and rates using open-source microscope-based cytometry". In: *Nat Methods* 4.2 (2007), pp. 175–81.

[106] Peter D. Gould et al. "Coordination of robust single cell rhythms in the Arabidopsis circadian clock via spatial waves of gene expression". In: *eLife* 7 (2018), pp. 1–20. ISSN: 2050084X. DOI: 10.7554/eLife.31700.

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

[108] T. Gregor et al. "Probing the limits to positional information". In: *Cell* 130.1 (2007), pp. 153–64. ISSN: 0092-8674 (Print).

[109] Thomas Gregor, Hernan G Garcia, and Shawn C Little. "The embryo as a laboratory: quantifying transcription in Drosophila." In: *Trends in genetics : TIG* 30.8 (2014), pp. 1–12. ISSN: 0168-9525. DOI: 10.1016/j.tig.2014.06.002. URL: http://www.ncbi.nlm.nih.gov/pubmed/25005921.

[110] Thomas Gregor et al. "Probing the Limits to Positional Information". In: *Cell* 130.1 (2007), pp. 153–164. ISSN: 00928674. DOI: 10.1016/j.cell.2007.05.025. arXiv: NIHMS150003.

[111] Thomas Gregor et al. "Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient". In: *Cell* 130.1 (2007), pp. 141–152. ISSN: 00928674. DOI: 10.1016/j.cell.2007.05.026. arXiv: NIHMS150003.

[112] Oliver Grimm and Eric Wieschaus. "The Bicoid gradient is shaped independently of nuclei". In: *Development* 137.17 (2010), pp. 2857–2862. ISSN: 09501991. DOI: 10.1242/dev.052589.

[113] Guido Grossmann et al. "Green light for quantitative live-cell imaging in plants". In: *Journal of Cell Science* 131.2 (2018). ISSN: 14779137. DOI: 10.1242/jcs.209270.

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

[115] M. Hafner et al. "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP". In: *Cell* 141.1 (2010), pp. 129–41. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: S0092-8674(10)00245-X[pii]10.1016/j.cell.2010.03.009.

[116] Peter Hajdukiewicz, Zora Svab, and Pal Maliga. “The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation”. In: *Plant Molecular Biology* 25.6 (1994), pp. 989–994. ISSN: 01674412. DOI: 10.1007/BF00014672.

[117] Shigeki Hamada et al. “The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells.” In: *The Plant cell* 15.10 (2003), pp. 2253–2264. ISSN: 1040-4651. DOI: 10.1105/tpc.013466.

[118] Danielle C Hamm et al. “A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster”. In: (2017), pp. 1–22.

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

[120] Samuel J. Harrison et al. “A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation”. In: *Plant Methods* 2.1 (2006), pp. 1–7. ISSN: 17464811. DOI: 10.1186/1746-4811-2-19.

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

[122] Gerald Z. Hertz and Gary D. Stormo. “Identifying DNA and protein patterns with statistically significant alignments of multiple sequences”. In: *Bioinformatics* 15.7-8 (1999), pp. 563–577. ISSN: 13674803. DOI: 10.1093/bioinformatics/15.7.563.

[123] Andreas Hilfinger and Johan Paulsson. “Separating intrinsic from extrinsic fluctuations in dynamic biological systems”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.29 (2011), pp. 12167–12172. ISSN: 00278424. DOI: 10.1073/pnas.1018832108.

[124] D. Hnisz et al. “A Phase Separation Model for Transcriptional Control”. In: *Cell* 169.1 (2017), pp. 13–23. ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2017.02.007.

[125] S. Hocine et al. “Single-molecule analysis of gene expression using two-color RNA labeling in live yeast”. In: *Nat Methods* (2012). ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/nmeth.2305.

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

[127] Yang Hsia et al. “Design of a hyperstable 60-subunit protein icosahedron”. In: *Nature* (2016), pp. 1–12. ISSN: 0028-0836. DOI: 10.1038/nature18010. URL: http://dx.doi.org/10.1038/nature18010.

[128] Heng-Cheng Hu, Ya-Yun Wang, and Yi-Fang Tsay. "AtCIPK8, a CBL-interacting protein kinase, regulates the low-affinity phase of the primary nitrate response." In: *The Plant journal : for cell and molecular biology* 57.2 (2009), pp. 264–78. ISSN: 1365-313X. DOI: 10.1111/j.1365-313X.2008.03685.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/18798873.

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

[130] Robert Ietswaart et al. "Cell-Size-Dependent Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-to-Cell Expression Variation". In: *Cell Systems* 4.6 (2017), 622–635.e9. ISSN: 24054720. DOI: 10.1016/j.cels.2017.05.010.

[131] A. Iida et al. "A zinc finger protein RHL41 mediates the light acclimatization response in Arabidopsis". In: *Plant Journal* 24.2 (2000), pp. 191–203. ISSN: 09607412. DOI: 10.1046/j.1365-313X.2000.00864.x.

[132] Itaru Imayoshi et al. "Oscillatory control of factors determining multipotency and fate in mouse neural progenitors." In: *Science* 342.6163 (2013), pp. 1203–1208. ISSN: 1095-9203. DOI: 10.1126/science.1242366. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed{\&}id=24179156{\&}retmode=ref{\&}cmd=prlinks{\%}5Cnpapers2://publication/doi/10.1126/science.1242366{\%}5Cnpapers2://publication/uuid/66719717-F812-4175-838D-932B54A6BFFA.

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

[134] Ryu Iwatate et al. "Covalent Self-labeling of Tagged Proteins with Chemical Fluorescent Dyes in BY-2 Cells and Arabidopsis Seedlings". In: *The Plant Cell* (2020), tpc.00439.2020. ISSN: 1040-4651. DOI: 10.1105/tpc.20.00439.

[135] J. Jaeger et al. "Dynamic control of positional information in the early Drosophila embryo". In: *Nature* 430.6997 (2004), pp. 368–71. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature02678.

[136] Ken Jean-Baptiste et al. "Developmental and conditional dynamics of gene expression in single root cells of A. thaliana". In: *bioRxiv* (2018), p. 448514. DOI: 10.1101/448514. URL: https://www.biorxiv.org/content/10.1101/448514v1.

[137] J. Jiang et al. "Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo". In: *EMBO Journal* 11.8 (1992), pp. 3147–3154. ISSN: 02614189. DOI: 10.1002/j.1460-2075.1992.tb05387.x.

[138] Jin Jiang and La Jolla. "The dorsal morphogen gradient regulates the mesoderm determinant twist in early Drosophila embryos". In: (1991), pp. 1881–1891.

[139] Xiaochun Jiang et al. *Light-induced HY5 Functions as a Systemic Signal to Coordinate the Photoprotective Response to Light Fluctuation*. Vol. 184. 2. 2020, pp. 1181–1193. ISBN: 0086571869. DOI: 10.1104/pp.20.00294.

[140] Richard Jorgensen, Christine Snyder, and Jonathan D.G. Jones. "T-DNA is organized predominantly in inverted repeat structures in plants transformed with Agrobacterium tumefaciens C58 derivatives". In: *MGG Molecular & General Genetics* 207.2-3 (1987), pp. 471–477. ISSN: 00263925. DOI: 10.1007/BF00331617.

[141] Florian Jupe et al. "The complex architecture and epigenomic impact of plant T-DNA insertions". In: *PLoS Genetics* 15.1 (2019), pp. 1–25. ISSN: 15537404. DOI: 10.1371/journal.pgen.1007819.

[142] Noa Katz et al. "An in Vivo Binding Assay for RNA-Binding Proteins Based on Repression of a Reporter Gene". In: *ACS Synthetic Biology* 7.12 (2018), pp. 2765–2774. ISSN: 21615063. DOI: 10.1021/acssynbio.8b00378.

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

[144] Jeffery M Kimbrough et al. "Kimbrough et al 2004, The Fast and Transient Transcriptional Network of.pdf". In: 136.September (2004), pp. 2790–2805. DOI: 10.1104/pp.104.044594.a.

[145] Justin B. Kinney and David M. McCandlish. "Massively Parallel Assays and Quantitative Sequence-Function Relationships". In: *Annual Review of Genomics and Human Genetics* 20 (2019), pp. 99–127. ISSN: 1545293X. DOI: 10.1146/annurev-genom-083118-014845.

[146] Axel Kirchhofer et al. "Modulation of protein properties in living cells using nanobodies". In: *Nature Structural and Molecular Biology* 17.1 (2010), pp. 133–139. ISSN: 15459985. DOI: 10.1038/nsmb.1727.

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

[148] M. S. Ko, H. Nakauchi, and N. Takahashi. "The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates." In: *The EMBO Journal* 9.9 (1990), pp. 2835–2842. ISSN: 0261-4189. DOI: 10.1002/j.1460-2075.1990.tb07472.x.

[149] Minoru S. H. Ko. "Induction mechanism of a single molecule: stochastic or deterministic?" In: *BioEssays* 14.5 (1992), pp. 341–346.

[150] Shinya Koizumi et al. "Functional analysis of the Hikeshi-like protein and its interaction with HSP70 in Arabidopsis". In: *Biochemical and Biophysical Research Communications* 450.1 (2014), pp. 396–400. ISSN: 10902104. DOI: 10.1016/j.bbrc.2014.05.128. URL: http://dx.doi.org/10.1016/j.bbrc.2014.05.128.

[151] Hannes Kollist et al. “Rapid Responses to Abiotic Stress: Priming the Landscape for the Signal Transduction Network”. In: *Trends in Plant Science* 24.1 (2019), pp. 25–37. ISSN: 13601385. DOI: 10.1016/j.tplants.2018.10.003. URL: https://doi.org/10.1016/j.tplants.2018.10.003.

[152] Gabriel Krouk et al. “Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate”. In: *Genome Biology* 11.12 (2010), R123. ISSN: 14747596. DOI: 10.1186/gb-2010-11-12-r123. URL: http://genomebiology.com/content/11/12/R123.

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

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

[155] Daniel R. Larson et al. “Real-time observation of transcription initiation and elongation on an endogenous yeast gene”. In: *Science* 332.6028 (2011), pp. 475–478. ISSN: 00368075. DOI: 10.1126/science.1202142.

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

[157] Martina Legris, Yetkin Çaka Ince, and Christian Fankhauser. “Molecular mechanisms underlying phytochrome-controlled morphogenesis in plants”. In: *Nature Communications* 10.1 (2019). ISSN: 20411723. DOI: 10.1038/s41467-019-13045-0. URL: http://dx.doi.org/10.1038/s41467-019-13045-0.

[158] Pablo Leivar et al. “Multiple Phytochrome-Interacting bHLH Transcription Factors Repress Premature Seedling Photomorphogenesis in Darkness”. In: *Current Biology* 18.23 (2008), pp. 1815–1823. ISSN: 09609822. DOI: 10.1016/j.cub.2008.10.058. URL: http://dx.doi.org/10.1016/j.cub.2008.10.058.

[159] Pablo Leivar et al. “Definition of early transcriptional circuitry involved in light-induced reversal of PIF-imposed repression of photomorphogenesis in young Arabidopsis seedlings.” In: *The Plant cell* 21.11 (2009), pp. 3535–53. ISSN: 1532-298X. DOI: 10.1105/tpc.109.070672. URL: http://www.ncbi.nlm.nih.gov/pubmed/19920208{\%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2798328.

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

[161] C. Li et al. “Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation”. In: *Cell Syst* 6.4 (2018), 409–423 e11. ISSN: 2405-4712 (Print) 2405-4712 (Linking). DOI: 10.1016/j.cels.2018.01.012.

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

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

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

[165] Zhen Li et al. “Gene duplicability of core genes is highly consistent across all angiosperms”. In: *Plant Cell* 28.2 (2015), pp. 326–344. ISSN: 1532298X. DOI: 10.1105/tpc.15.00877.

[166] B. Lim et al. “Temporal dynamics of pair-rule stripes in living Drosophila embryos”. In: *Proc Natl Acad Sci U S A* 115.33 (2018), pp. 8376–8381. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1810430115.

[167] Bomyi Lim et al. “Visualization of Transvection in Living Drosophila Embryos”. In: *Molecular Cell* 70.2 (2018), 287–296.e6. ISSN: 10974164. DOI: 10.1016/j.molcel.2018.02.029. arXiv: 0411586 [cond-mat].

[168] Shawn C. Little, Mikhail Tikhonov, and Thomas Gregor. “Precise developmental gene expression arises from globally stochastic transcriptional activity”. In: *Cell* 154.4 (2013), pp. 789–800. ISSN: 00928674. DOI: 10.1016/j.cell.2013.07.025. URL: http://dx.doi.org/10.1016/j.cell.2013.07.025.

[169] Hongtao Liu et al. “The action mechanisms of plant cryptochromes”. In: *Trends in Plant Science* 16.12 (2011), pp. 684–691. ISSN: 13601385. DOI: 10.1016/j.tplants.2011.09.002. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.tplants.2011.09.002.

[170] Min Liu, Jiafu Zhu, and Zhicheng Dong. “Immediate transcriptional responses of Arabidopsis leaves to heat shock”. In: *Journal of Integrative Plant Biology* 00.00 (2020). ISSN: 17447909. DOI: 10.1111/jipb.12990.

[171] Tsung Li Liu et al. “Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms”. In: *Science* 360.6386 (2018). ISSN: 10959203. DOI: 10.1126/science.aaq1392.

[172] Kenneth J. Livak and Thomas D. Schmittgen. "Analysis of relative gene expression data using real-time quantitative PCR and the 2-$\Delta\Delta$CT method". In: *Methods* 25.4 (2001), pp. 402–408. ISSN: 10462023. DOI: 10.1006/meth.2001.1262. arXiv: 1003.3921v1.

[173] Ernesto Llamas et al. "The intrinsic chaperone network of Arabidopsis stem cells confers protection against proteotoxic stress". In: *bioRxiv* (2021), pp. 10.1101/2021.01.19.427268 URL: https://doi.org/10.1101/2021.01.19.427268.

[174] Tanguy Lucas et al. "Live imaging of bicoid-dependent transcription in Drosophila embryos." In: *Current Biology* 23.21 (2013), 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.

[175] Cristiana Lungu et al. "Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites". In: *Nature Communications* 8.1 (2017). ISSN: 20411723. DOI: 10.1038/s41467-017-00457-z. URL: http://dx.doi.org/10.1038/s41467-017-00457-z.

[176] H. Maamar, A. Raj, and D. Dubnau. "Noise in gene expression determines cell fate in Bacillus subtilis". In: *Science* 317.5837 (2007), pp. 526–9. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 1140818[pii]10.1126/science.1140818.

[177] Narendra Maheshri and Erin K. O'Shea. "Living with noisy genes: How cells function reliably with inherent variability in gene expression". In: *Annual Review of Biophysics and Biomolecular Structure* 36 (2007), pp. 413–434. ISSN: 10568700. DOI: 10.1146/annurev.biophys.36.040306.132705.

[178] Samuel Marguerat and Jürg Bähler. "Coordinating genome expression with cell size". In: *Trends in Genetics* 28.11 (2012), pp. 560–565. ISSN: 01689525. DOI: 10.1016/j.tig.2012.07.003.

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

[180] José L. McFaline-Figueroa, Cole Trapnell, and Josh T. Cuperus. "The promise of single-cell genomics in plants". In: *Current Opinion in Plant Biology* 54 (2020), pp. 114–121. ISSN: 13695266. DOI: 10.1016/j.pbi.2020.04.002.

[181] Fionn McLoughlin et al. "Class I and II small heat shock proteins together with HSP101 protect protein translation factors during heat stress". In: *Plant Physiology* 172.2 (2016), pp. 1221–1236. ISSN: 15322548. DOI: 10.1104/pp.16.00536.

[182] J. E. Melaragno, B. Mehrotra, and A. W. Coleman. "Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis". In: *Plant Cell* 5.11 (1993), pp. 1661–1668. ISSN: 10404651. DOI: 10.1105/tpc.5.11.1661.

[183] Heather M Meyer et al. “Fluctuations of the transcription factor ATML1 generate the pattern of giant cells in the Arabidopsis sepal”. In: *eLife* 6 (2017), pp. 1–41. ISSN: 2050-084X. DOI: 10.7554/elife.19131.

[184] Daisuke Miki et al. “CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation”. In: *Nature Communications* 9.1 (2018), pp. 1–9. ISSN: 20411723. DOI: 10.1038/s41467-018-04416-0.

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

[186] Ron Mittler, Andrija Finka, and Pierre Goloubinoff. “How do plants feel the heat?” In: *Trends in Biochemical Sciences* 37.3 (2012), pp. 118–125. ISSN: 09680004. DOI: 10.1016/j.tibs.2011.11.007. URL: http://dx.doi.org/10.1016/j.tibs.2011.11.007.

[187] Ron Mittler et al. “Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress”. In: *FEBS Letters* 580.28-29 (2006), pp. 6537–6542. ISSN: 00145793. DOI: 10.1016/j.febslet.2006.11.002.

[188] M Moore, Mo Vogel, and Kj Dietz. “The acclimation response to high light is initiated within seconds as indicated by upregulation of AP2/ERF transcription factor network in Arabidopsis thaliana.” In: *Plant signaling & behavior* 9.10 (2014), p. 976479. ISSN: 1559-2324. DOI: 10.4161/15592324.2014.976479. URL: http://www.ncbi.nlm.nih.gov/pubmed/25482793.

[189] Mahdi Moradpour, Siti Nor, and Akmar Abdulah. “CRISPR / dCas9 platforms in plants : strategies and applications beyond genome editing”. In: (2020), pp. 32–44. DOI: 10.1111/pbi.13232.

[190] Brian Munsky, Gregor Neuert, and Alexander Van Oudenaarden. “Using gene expression noise to understand gene regulation”. In: *Science* 336.6078 (2012), pp. 183–187. ISSN: 10959203. DOI: 10.1126/science.1216379.

[191] Weimin Ni et al. “Multisite light-induced phosphorylation of the transcription factor PIF3 Is necessary for both its rapid degradation and concomitant negative feedback modulation of photoreceptor phyB levels in Arabidopsis”. In: *Plant Cell* 25.7 (2013), pp. 2679–2698. ISSN: 10404651. DOI: 10.1105/tpc.113.112342.

[192] Weimin Ni et al. “A mutually assured destruction mechanism attenuates light signaling in Arabidopsis”. In: *Science* 344.6188 (2014), pp. 1160–1164. ISSN: 10959203. DOI: 10.1126/science.1250778.

[193] Damien Nicolas, Nick E. Phillips, and Felix Naef. “What shapes eukaryotic transcriptional bursting?” In: *Molecular BioSystems* 13.7 (2017), pp. 1280–1290. ISSN: 17422051. DOI: 10.1039/c7mb00154a. URL: http://dx.doi.org/10.1039/C7MB00154A.

[194] Dominik Niopek et al. “Optogenetic control of nuclear protein export.” In: *Nature communications* 7 (2016), p. 10624. ISSN: 2041-1723. DOI: 10.1038/ncomms10624. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4748110{\&}tool=pmcentrez{\&}rendertype=abstract.

[195] A. Novick and M. Weiner. “Enzyme Induction As an All-or-None Phenomenon”. In: *Proceedings of the National Academy of Sciences* 43.7 (1957), pp. 553–566. ISSN: 0027-8424. DOI: 10.1073/pnas.43.7.553.

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

[197] Naohiko Ohama et al. “The Transcriptional Cascade in the Heat Stress Response of Arabidopsis Is Strictly Regulated at the Level of Transcription Factor Expression”. In: 28.January (2016), pp. 181–201. DOI: 10.1105/tpc.15.00435.

[198] Ronan C. O’Malley et al. “Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape”. In: *Cell* 165.5 (2016), pp. 1280–1292. ISSN: 10974172. DOI: 10.1016/j.cell.2016.04.038. URL: http://dx.doi.org/10.1016/j.cell.2016.04.038.

[199] Olivia Padovan-Merhar et al. “Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms”. In: *Molecular Cell* 58.2 (2015), pp. 339–352. ISSN: 10974164. DOI: 10.1016/j.molcel.2015.03.005. URL: http://dx.doi.org/10.1016/j.molcel.2015.03.005.

[200] Aikaterini Papagianni et al. “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 (2018), pp. 1807–1812. ISSN: 10916490. DOI: 10.1073/pnas.1713930115.

[201] Eunae Park, Yeojae Kim, and Giltsu Choi. “Phytochrome B requires PIF degradation and sequestration to induce light responses across a wide range of light conditions”. In: *Plant Cell* 30.6 (2018), pp. 1277–1292. ISSN: 1532298X. DOI: 10.1105/tpc.17.00913.

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

[203] Soon Ju Park et al. “Optimization of crop productivity in tomato using induced mutations in the florigen pathway.” In: *Nature genetics* 46.12 (2014), pp. 1337–42. ISSN: 1546-1718. DOI: 10.1038/ng.3131. URL: http://dx.doi.org/10.1038/ng.3131.

[204] Sung Young Park, Hyungseok C. Moon, and Hye Yoon Park. "Live-cell imaging of single mRNA dynamics using split superfolder green fluorescent proteins with minimal background". In: *Rna* 26.1 (2020), pp. 101–109. ISSN: 14699001. DOI: 10.1261/rna.067835.118.

[205] Michael W Perry et al. "Report Shadow Enhancers Foster Robustness of Drosophila Gastrulation". In: *Current Biology* 20.17 (2010), pp. 1562–1567. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.07.043. URL: http://dx.doi.org/10.1016/j.cub.2010.07.043.

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

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

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

[209] Mark Ptashne. *A genetic switch: Gene control and phage lambda*. United States, 1986. URL: https://www.osti.gov/biblio/5413898.

[210] Qi Qiu et al. "Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq". In: *Nature Methods* 17.10 (2020), pp. 991–1001. ISSN: 15487105. DOI: 10.1038/s41592-020-0935-4. URL: http://dx.doi.org/10.1038/s41592-020-0935-4.

[211] Christine Queitsch et al. "Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis". In: *Plant Cell* 12.4 (2000), pp. 479–492. ISSN: 10404651. DOI: 10.1105/tpc.12.4.479.

[212] Arjun Raj and Alexander van Oudenaarden. "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences". In: *Cell* 135.2 (2008), pp. 216–226. ISSN: 00928674. DOI: 10.1016/j.cell.2008.09.050.

[213] Arjun Raj et al. "Stochastic mRNA Synthesis in Mammalian Cells". In: 4.10 (2006). DOI: 10.1371/journal.pbio.0040309.

[214] Jonathan M. Raser and Erin K. O'Shea. "Control of stochasticity in eukaryotic gene expression". In: *Science* 304.5678 (2004), pp. 1811–1814. ISSN: 00368075. DOI: 10.1126/science.1098641.

[215] Gregory T. Reeves et al. "Dorsal-Ventral Gene Expression in the Drosophila Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient". In: *Developmental Cell* 22.3 (2012), 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.

[216] Gretchen A. Rice, Michael J. Chamberlin, and Caroline M. Kane. “Contacts between mammalian RNA polymerase II and the template DNA in a ternary elongation complex”. In: *Nucleic Acids Research* 21.1 (1993), pp. 113–118. ISSN: 03051048. DOI: 10.1093/nar/21.1.113.

[217] Dana O. Robinson et al. “Ploidy and size at multiple scales in the arabidopsis sepal[open]”. In: *Plant Cell* 30.10 (2018), pp. 2308–2329. ISSN: 1532298X. DOI: 10.1105/tpc.18.00344.

[218] Joseph Rodriguez and Daniel R. Larson. “Transcription in Living Cells: Molecular Mechanisms of Bursting”. In: *Annual Review of Biochemistry* 89 (2020), pp. 189–212. ISSN: 15454509. DOI: 10.1146/annurev-biochem-011520-105250.

[219] Joseph Rodriguez et al. “Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity”. In: *Cell* 176.1-2 (2019), 213–226.e18. ISSN: 10974172. DOI: 10.1016/j.cell.2018.11.026. URL: https://doi.org/10.1016/j.cell.2018.11.026.

[220] Samuel G Rodriques et al. “RNA timestamps identify the age of single molecules in RNA sequencing”. In: *Nature Biotechnology* (2020). ISSN: 15461696. DOI: 10.1038/s41587-020-0704-z. URL: http://dx.doi.org/10.1038/s41587-020-0704-z.

[221] Daniel Rodrı et al. “Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing Engineering Quantitative Trait Variation”. In: (2017), pp. 470–480.

[222] Adrienne HK Roeder. “Use it or average it: stochasticity in plant development”. In: *Current Opinion in Plant Biology* 41 (2018), pp. 8–15. ISSN: 13695266. DOI: 10.1016/j.pbi.2017.07.010. URL: http://dx.doi.org/10.1016/j.pbi.2017.07.010.

[223] James Ronald and Seth J. Davis. “Focusing on the nuclear and subnuclear dynamics of light and circadian signalling”. In: *Plant Cell and Environment* 42.10 (2019), pp. 2871–2884. ISSN: 13653040. DOI: 10.1111/pce.13634.

[224] Nitzan Rosenfeld et al. “Gene Regulation at the Single-Cell Level”. In: 307.March (2005).

[225] IAN L. ROSS, CATHERINE M. BROWNE, and DAVID A. HUME. “Transcription of individual genes in eukaryotic cells occurs randomly and infrequently”. In: *Immunology and Cell Biology* 72.2 (1994), pp. 177–185. ISSN: 14401711. DOI: 10.1111/j.1440-1711.1994.tb03774.x.

[226] Jan Bart Rossel et al. “Systemic and Intracellular Responses to Photooxidative Stress in Arabidopsis”. In: 19.December (2007), pp. 4091–4110. DOI: 10.1105/tpc.106.045898.

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

[228] Keisuke Sako et al. “Optogenetic Control of Nodal Signaling Reveals a Temporal Pattern of Nodal Signaling Regulating Cell Fate Specification during Gastrulation”. In: *Cell Reports* 16.3 (2016), pp. 866–877. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.06.036.

[229] M. Salomon et al. “Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin”. In: *Biochemistry* 39.31 (2000), pp. 9401–9410. ISSN: 00062960. DOI: 10.1021/bi000585+.

[230] Jeremy E Sandler and Angelike Stathopoulos. “Stepwise Progression of Embryonic Patterning”. In: 32.7 (2016), pp. 432–443.

[231] Rupinder Sayal et al. “Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early Drosophila embryo”. In: *eLife* 5.MAY2016 (2016), pp. 1–25. ISSN: 2050084X. DOI: 10.7554/eLife.08445.

[232] Johannes Schindelin et al. “Fiji: An open-source platform for biological-image analysis”. In: *Nature Methods* 9.7 (2012), pp. 676–682. ISSN: 15487091. DOI: 10.1038/nmeth.2019.

[233] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. In: *Nature Methods* 9.7 (2012), pp. 671–675. ISSN: 15487091. DOI: 10.1038/nmeth.2089.

[234] Clarissa Scholes and Angela H Depace. “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: (2017), pp. 97–108. DOI: 10.1016/j.cels.2016.11.012.

[235] Johannes Schönberger, Ulrich Z. Hammes, and Thomas Dresselhaus. “In vivo visualization of RNA in plants cells using the λn 22 system and a GATEWAY-compatible vector series for candidate RNAs”. In: *Plant Journal* 71.1 (2012), pp. 173–181. ISSN: 09607412. DOI: 10.1111/j.1365-313X.2012.04923.x.

[236] Eran Segal et al. “Predicting expression patterns from regulatory sequence in Drosophila segmentation”. In: 451.January (2008). DOI: 10.1038/nature06496.

[237] Sydney M Shaffer et al. “reprogramming as a mode of cancer drug resistance”. In: *Nature Publishing Group* 546.7658 (2017), pp. 431–435. ISSN: 0028-0836. DOI: 10.1038/nature22794. URL: http://dx.doi.org/10.1038/nature22794.

[238] Jayasha Shandilya and Stefan G.E. Roberts. “The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling”. In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819.5 (2012), pp. 391–400. ISSN: 18749399. DOI: 10.1016/j.bbagrm.2012.01.010. URL: http://dx.doi.org/10.1016/j.bbagrm.2012.01.010.

[239] Marc S. Sherman et al. “Cell-to-Cell Variability in the Propensity to Transcribe Explains Correlated Fluctuations in Gene Expression”. In: *Cell Systems* 1.5 (2015), pp. 315–325. ISSN: 24054712. DOI: 10.1016/j.cels.2015.10.011. URL: http://www.sciencedirect.com/science/article/pii/S2405471215001854.

[240] Antony W Shermoen and Patrick H O Farrell. “y //”. In: 67 (1991), pp. 303–310.

[241] Sae Shimizu-Sato et al. “A light-switchable gene promoter system”. In: *Nature Biotechnology* 20.10 (2002), pp. 1041–1044. ISSN: 10870156. DOI: 10.1038/nbt734.

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

[243] Samuel O. Skinner et al. “Single-cell analysis of transcription kinetics across the cell cycle”. In: *eLife* 5.JANUARY2016 (2016). ISSN: 2050084X. DOI: 10.7554/eLife.12175.001.

[244] Sebastian Soyk, Matthias Benoit, and Zachary B Lippman. “New Horizons for Dissecting Epistasis in Crop Quantitative Trait Variation”. In: (2020), pp. 1–21.

[245] L Carine Stapel, Christoph Zechner, and Nadine L Vastenhouw. “Uniform gene expression in embryos is achieved by temporal averaging of transcription noise”. In: (2017), pp. 1–6. DOI: 10.1101/gad.302935.117.naarden.

[246] Timothy J. Stasevich et al. “Regulation of RNA polymerase II activation by histone acetylation in single living cells”. In: *Nature* 516.7530 (2014), pp. 272–275. ISSN: 14764687. DOI: 10.1038/nature13714. URL: http://dx.doi.org/10.1038/nature13714.

[247] Fabio Stossi et al. “Estrogen-induced transcription at individual alleles is independent of receptor level and active conformation but can be modulated by coactivators activity”. In: *Nucleic Acids Research* 48.4 (2020), pp. 1800–1810. ISSN: 13624962. DOI: 10.1093/nar/gkz1172.

[248] Paolo Struffi and David N. Arnosti. “Functional interaction between the Drosophila knirps short range transcriptional repressor and RPD3 histone deacetylase”. In: *Journal of Biological Chemistry* 280.49 (2005), pp. 40757–40765. ISSN: 00219258. DOI: 10.1074/jbc.M506819200.

[249] Akiko Sugio et al. “The cytosolic protein response as a subcomponent of the wider heat shock response in arabidopsis”. In: *Plant Cell* 21.2 (2009), pp. 642–654. ISSN: 10404651. DOI: 10.1105/tpc.108.062596.

[250] Nobuhiro Suzuki et al. “Ultra-fast alterations in mRNA levels uncover multiple players in light stress acclimation in plants”. In: (2015), pp. 760–772. DOI: 10.1111/tpj.13039.

[251] P Szymanski and M Levine. “Multiple modes of dorsal-bHLH transcriptional synergy in the Drosophila embryo.” In: *The EMBO Journal* 14.10 (1995), 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.

[252] Marvin E. Tanenbaum et al. “A protein-tagging system for signal amplification in gene expression and fluorescence imaging”. In: *Cell* 159.3 (2014), pp. 635–646. ISSN: 10974172. DOI: 10.1016/j.cell.2014.09.039. URL: http://dx.doi.org/10.1016/j.cell.2014.09.039.

[253] Katjana Tantale et al. “A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting”. In: *Nature Communications* 7.1 (2016), p. 12248. ISSN: 20411723. DOI: 10.1038/ncomms12248. URL: http://www.nature.com/articles/ncomms12248.

[254] Mallorie Taylor-Teeples, Mily Ron, and Siobhan M. Brady. “Novel biological insights revealed from cell type-specific expression profiling”. In: *Current Opinion in Plant Biology* 14.5 (2011), pp. 601–607. ISSN: 13695266. DOI: 10.1016/j.pbi.2011.05.007. URL: http://dx.doi.org/10.1016/j.pbi.2011.05.007.

[255] The Arabidopsis Genome Iniative. “Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.” In: *Nature* 408.December 2000 (2000), pp. 796–815. ISSN: 1573-5109. DOI: 10.1134/S1022795411020074.

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

[257] Jared E. Toettcher, Orion D. Weiner, and Wendell A. Lim. “Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module”. In: *Cell* 155.6 (2013), pp. 1422–1434. ISSN: 00928674. DOI: 10.1016/j.cell.2013.11.004. URL: http://dx.doi.org/10.1016/j.cell.2013.11.004.

[258] Silvia Tornaletti, Daniel Reines, and Philip C. Hanawalt. “Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA”. In: *Journal of Biological Chemistry* 274.34 (1999), pp. 24124–24130. ISSN: 00219258. DOI: 10.1074/jbc.274.34.24124.

[259] A. TREWAVAS. “How do plant growth substances work? II”. In: *Plant, Cell & Environment* 14.1 (1991), pp. 1–12. ISSN: 13653040. DOI: 10.1111/j.1365-3040.1991.tb01366.x.

[260] Edward Tunnacliffe and Jonathan R. Chubb. "What Is a Transcriptional Burst?" In: *Trends in Genetics* 36.4 (2020), pp. 288–297. ISSN: 13624555. DOI: 10.1016/j.tig.2020.01.003. URL: https://doi.org/10.1016/j.tig.2020.01.003.

[261] Gina M. Turco et al. "Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation". In: *Cell Reports* 28.2 (2019), 342–351.e4. ISSN: 22111247. DOI: 10.1016/j.celrep.2019.06.041.

[262] Evelina Tutucci et al. "Imaging mRNA In Vivo, from Birth to Death". In: *Annual Review of Biophysics* 47.1 (2018), pp. 85–106. ISSN: 1936-122X. DOI: 10.1146/annurev-biophys-070317-033037.

[263] Josh Tycko et al. "Advancing towards a global mammalian gene regulation model through single-cell analysis and synthetic biology". In: *Current Opinion in Biomedical Engineering* 4 (2017), pp. 174–193. ISSN: 24684511. DOI: 10.1016/j.cobme.2017.10.011. URL: https://doi.org/10.1016/j.cobme.2017.10.011.

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

[265] Mark C Walters et al. "Enhancers increase the probability but not the level of gene expression". In: *Gene Expression* 92.July (1995), p. 8.

[266] Mengyu Wang et al. "Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality". In: *Nature Microbiology* 4.12 (2019), pp. 2118–2127. ISSN: 20585276. DOI: 10.1038/s41564-019-0553-z.

[267] Qintao Wang et al. "Genome editing of model oleaginous microalgae Nannochloropsis spp . by CRISPR / Cas9". In: (2016), pp. 1071–1081. DOI: 10.1111/tpj.13307.

[268] Rachel Waymack et al. "Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic". In: *eLife* 9 (2020), pp. 1–57. ISSN: 2050084X. DOI: 10.7554/ELIFE.59351.

[269] Per K.I. Wilhelmsson et al. "Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae". In: *Genome Biology and Evolution* 9.12 (2017), pp. 3384–3397. ISSN: 17596653. DOI: 10.1093/gbe/evx258.

[270] Maxwell Z Wilson et al. "Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control Article Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control". In: (2017), pp. 1–13.

[271] Debbie Winter et al. "An "electronic fluorescent pictograph" Browser for exploring and analyzing large-scale biological data sets". In: *PLoS ONE* 2.8 (2007), pp. 1–12. ISSN: 19326203. DOI: 10.1371/journal.pone.0000718.

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

[273] Bin Wu, Jiahao Chen, and Robert H. Singer. “Background free imaging of single mRNAs in live cells using split fluorescent proteins”. In: *Scientific Reports* 4 (2014), pp. 11–13. ISSN: 20452322. DOI: 10.1038/srep03615.

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

[275] Zhe Wu et al. “Quantitative regulation of FLC via coordinated transcriptional initiation and elongation”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.1 (2016), pp. 218–223. ISSN: 10916490. DOI: 10.1073/pnas.1518369112.

[276] Z. Wunderlich and L. A. Mirny. “Different gene regulation strategies revealed by analysis of binding motifs”. In: *Trends Genet* 25.10 (2009), pp. 434–40. ISSN: 0168-9525 (Print) 0168-9525 (Linking). DOI: 10.1016/j.tig.2009.08.003.

[277] Bin Xu et al. “Magic numbers in polymer phase separation - The importance of being rigid”. In: *arXiv* 2020 (2019), pp. 4–11. ISSN: 23318422. DOI: 10.1038/s41467-020-15395-6. arXiv: 1901.09352. URL: http://dx.doi.org/10.1038/s41467-020-15395-6.

[278] Xiaosa Xu et al. “Illuminating Progress in Phytochrome-Mediated Light Signaling Pathways”. In: *Trends in Plant Science* 20.10 (2015), 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.

[279] Sora Yang et al. “Contribution of RNA polymerase concentration variation to protein expression noise”. In: *Nature Communications* 5 (2014), pp. 1–9. ISSN: 20411723. DOI: 10.1038/ncomms5761.

[280] Takumi Yoshida et al. “Arabidopsis HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression”. In: *Molecular Genetics and Genomics* 286.5-6 (2011), pp. 321–332. ISSN: 16174615. DOI: 10.1007/s00438-011-0647-7.

[281] Daseuli Yu et al. “Optogenetic activation of intracellular antibodies for direct modulation of endogenous proteins”. In: *Nature Methods* 16.11 (2019), pp. 1095–1100. ISSN: 15487105. DOI: 10.1038/s41592-019-0592-7. URL: http://dx.doi.org/10.1038/s41592-019-0592-7.

[282] Sara I. Zandalinas et al. “Identification and characterization of a core set of ROS wave-associated transcripts involved in the systemic acquired acclimation response of Arabidopsis to excess light”. In: *Plant Journal* 98.1 (2019), pp. 126–141. ISSN: 1365313X. DOI: 10.1111/tpj.14205.

[283] Sara I. Zandalinas et al. “Signal transduction networks during stress combination”. In: *Journal of Experimental Botany* 71.5 (2020), pp. 1734–1741. ISSN: 14602431. DOI: 10.1093/jxb/erz486.

[284] Daniel Zenklusen, Daniel R. Larson, and Robert H. Singer. “Single-RNA counting reveals alternative modes of gene expression in yeast”. In: *Nature Structural and Molecular Biology* 15.12 (2008), pp. 1263–1271. ISSN: 15459993. DOI: 10.1038/nsmb.1514.

[285] Fengli Zhang and Anne E. Simon. “A novel procedure for the localization of viral RNAs in protoplasts and whole plants”. In: *Plant Journal* 35.5 (2003), pp. 665–673. ISSN: 09607412. DOI: 10.1046/j.1365-313X.2003.01837.x.

[286] Yu Zhang et al. “A Quartet of PIF bHLH Factors Provides a Transcriptionally Centered Signaling Hub That Regulates Seedling Morphogenesis through Differential Expression-Patterning of Shared Target Genes in Arabidopsis”. In: *PLoS Genetics* 9.1 (2013), pp. 11–13. ISSN: 15537390. DOI: 10.1371/journal.pgen.1003244.