Mechanisms of Transcriptional Precision in the *Drosophila* Embryo

by

Jacques Pierre Bothma

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Levine, Co-chair
Professor Susan Marqusee, Co-chair
Professor Nipam Patel
Associate Professor Jan Liphardt

Fall 2013
Mechanisms of Transcriptional Precision in the *Drosophila* Embryo

Copyright 2013
by
Jacques Pierre Bothma
Contemplating how a single cell can turn into the trillions of specialized cells that make a human being staggers the imagination. We still do not fully understand how the information in a genome is interpreted by a cell to orchestrate this incredible process. One thing that we do know is that much of the complexity we see in the natural world comes down to how essentially the same set of proteins are differentially deployed. One of the key places where this is controlled is at the level of transcription which is the first step in protein production. In this thesis we attempt to shed light on this process by looking at how transcription is regulated in the early Drosophila embryo with a focus on mechanisms of transcriptional precision. We developed imaging and segmentation techniques that allowed for the quantitative visualization of the transcriptional state of thousands of nuclei in the embryo. Using this approach we discovered the phenomenon of repression lag, whereby genes containing large introns are not only slow to be switched on (intron delay), but are also slow to be repressed. Many sequence-specific repressors have been implicated in early development, but the mechanisms by which they silence gene expression have remained elusive. We found that elongating Pol II complexes complete transcription after the onset of repression. As a result, moderately sized genes are fully silenced only after tens of minutes of repression. We propose that this "repression lag" imposes a severe constraint on the regulatory dynamics of embryonic patterning.

Having laid the foundations for using quantitative imaging in the early Drosophila embryo we next sought to understand the mechanisms underlying developmental timing, the temporal control of gene expression. Previous studies have provided considerable information about the spatial regulation of gene expression, but there is very little information regarding the temporal coordination of expression. Paused RNA Polymerase (pausd Pol II) is a pervasive feature of Drosophila embryos and mammalian stem cells, but its role in development is uncertain. We demonstrate that there is a spectrum of paused Pol II, which determines the "time to synchrony"—the time required to achieve coordinate gene expression across the
different cells of a tissue. To determine whether synchronous patterns of gene activation are significant in development, we manipulated the timing of snail expression, which controls the coordinated invagination of ~1000 mesoderm cells during gastrulation. Replacement of the strongly paused snail promoter with moderately paused or nonpaused promoters resulted in stochastic activation of snail expression and the progressive loss of mesoderm invagination. Computational modeling of the dorsal-ventral patterning network recapitulated these variable and bistable gastrulation profiles, and emphasized the importance of timing of gene activation in development. We concluded that paused Pol II and transcriptional synchrony are essential for coordinating cell behavior during morphogenesis.

These studies and others have helped launch a new approach to the well-established problem of differential gene expression in animal development. The quantitative imaging methods developed here have permitted the assessment of temporal dynamics of gene expression and the underlying mechanisms for coordinating gene expression across the different cells of a tissue. The next frontier will be to apply these methods to live embryos, thereby permitting an even deeper analysis of gene dynamics in development.
I dedicate this thesis and the work in it to Lawrence Alarcon. Without his unwavering support and encouragement most of it would never have seen the light of day.
Contents

List of Figures iv

1 Introduction 1
   1.1 Transcription Initiation and Noise ............................. 2
   1.2 Patterning of the Drosophila Embryo ............................ 3

2 Dynamics of Transcriptional Repression 8
   2.1 Chapter Summary .................................................. 8
   2.2 Introduction ..................................................... 8
   2.3 Results and Discussion ......................................... 10
   2.4 Experimental Procedures ....................................... 17

3 Pol II pausing and coordination of transcription 23
   3.1 Chapter Summary .................................................. 23
   3.2 Introduction ..................................................... 23
   3.3 Results and Discussion ......................................... 24
   3.4 Temporal Coordination of Dpp Target Genes ................. 25
   3.5 Minimal Promoter Sequences are Sufficient to Establish Paused Pol II .... 27
   3.6 Promoter-Associated Elements Influence Transcriptional Synchrony ........ 29
   3.7 A Spectrum of Synchrony ........................................ 31
   3.8 Transcriptional synchrony and rates of RNA synthesis .......... 31
   3.9 Model for the developmental timing of gene activation .......... 35
   3.10 Determining Promoter Strength .................................. 36
   3.11 Quantifying Initiation Dynamics and Determining t50 ........... 37
   3.12 Methods .......................................................... 38

4 Importance of transcriptional coordination in development 46
   4.1 Chapter Summary .................................................. 46
   4.2 Transcriptional synchrony is essential for coordinate invagination .......... 46
   4.3 Computational Models of Gastrulation Variability .................. 47
4.4 Dynamic control of the dorsal-ventral patterning network .......... 50
4.5 Spectrum of pausing and cell fate decisions ....................... 51
4.6 Methods ................................................. 51
4.7 Fly Genetics ............................................. 51
4.8 Modeling Evolution of Snail Protein ............................... 52
4.9 Model Behavior ........................................... 58
4.10 Sensitivity Analysis of Key Parameters ......................... 59

Bibliography 64
# List of Figures

1.1 The key steps in transcription initiation ............................................ 4  
1.2 Timetable of early *Drosophila* development ........................................ 6  
1.3 Schematic illustrating the Dorsal and Dpp gradients and some of their readouts. 7  

2.1 Schematic showing how the initiation of transcription and different schemes of repression affect the dynamics of full-length mRNA production ............................................ 10  
2.2 Time course of sog transcription from early cell cycle 13 to early cell cycle 14 . 11  
2.3 Time course of sog transcription during cc 14 ........................................ 13  
2.4 Repression of Delta and ASPP transcription in the presumptive mesoderm ... 14  
2.5 Visualizing repression of cno and sea .................................................. 15  
2.6 Quantification of nascent transcript detection efficiency for sog probes .......... 20  
2.7 Controls for detection of isolated 3' probe ........................................ 21  
2.8 List of primer sequences used to make mRNA probes relevant to this chapter ... 22  

3.1 BMP/Dpp target genes exhibit distinct coordination profiles ....................... 26  
3.2 The pur promoter is sufficient to delay tup expression. ............................ 28  
3.3 The minimal promoter mediates paused Pol II ......................................... 30  
3.4 A spectrum of synchrony ......................................................................... 32  
3.5 Temporal coordination with sog enhancer .............................................. 33  
3.6 Summary of the t50 values for all of the constructs used in this chapter ......... 34  
3.7 Minimal promoters are sufficient to perturb snail temporal coordination. .... 42  
3.8 Measuring and modeling mRNA levels .................................................. 43  
3.9 Determining t50 for the activation curves .............................................. 44  
3.10 Cumulative gamma fit parameters ......................................................... 45  

4.1 Stochastic expression of snai Results in gastrulation defects .................... 48  
4.2 Modeling gastrulation variability: The importance of coordination ............... 49  
4.3 Spatial Profile of the Dorsal Gradient. .................................................. 54  
4.4 Snail autoregulation. ............................................................................. 55  
4.5 Model behavior ...................................................................................... 61  
4.6 Model sensitivity analysis of key parameters .......................................... 62  
4.7 Sequences of primers ............................................................................. 63
Acknowledgments

First and foremost I need to thank my adviser Mike Levine. As an adviser Mike has been exceptional in many respects. He has been a vociferous advocate of my work, especially at times when it really mattered. Mike was consistently engaged in my research and made himself available to discuss science on a daily basis. During these discussions he constantly challenged me, and this was key for me to start thinking like a biologist and learn to differentiate between biological questions that were likely to lead to an incremental advance versus ones that may shed insights of broad reaching consequence. Through his example Mike also taught me how to be a better communicator. For all these things I will always be grateful to him, and be in his debt. These positive aspects also made indulging his notorious eccentricities bearable.

Joe Magliocco was an incredibly talented undergraduate that worked with me on the repression lag story. In a very short period of time Joe learned how to clone, code and image and made himself incredibly useful. Alistair Boettiger is the person who introduced me to the Levine lab and helped me orient myself in transription and the early drosophila embryo. I doubt I would have even joined the lab had he not been there and been as encouraging as he was. I also enjoyed working with Alistair on coding and image analysis and after he left the lab greatly missed the stimulating conversations we used to have about science. Mike Perry is one of the most talented experimentalists I have come across. His meticulous attention to detail and his quick thinking (which outpaces even his rapid pace of speech) makes him a great scientist. Mike was immensely helpful in terms of technical troubleshooting and thinking about the big picture. Much of the molecular biology that underpinned the work in this thesis are elaborations on tools he built and protocols he got up and running in the lab. Emilia Esposito has been an amazing collaborator. For the most part the things that Mike didn’t get working Emilia did. Emilia’s hands were the ones that performed much of the molecular biology that went into the work on pausing and synchrony and I can only describe her as a cloning wizard. Val´ erie Hilgers was immensely helpful when it came to questions about fly genetics and science in general and she always had insightful questions and comments on my work and I am grateful for this. Kevin Tsui was a technician in the lab for much of the time I was there and was fantastic in this role and he also did some cloning and a lot of injections for the synchrony and pausing work. Last but not least I want to acknowledge my collaborator Mounia Lagha for the the key role she played in the pausing and synchrony work. Mounia joined the lab shortly after I embarked on the project and her timing was perfect. Mounia’s ambition, focus and technical skill as a molecular biologists were of great value and played and important role in bringing the work to fruition. I look forward to hearing about the exciting discoveries I have no doubt she will make in her own lab. The Levine lab can be a tough environment and I need to acknowledge the following past and present members for bits of advice and encouragement over the years: Jessica Cande, Ben Haley, David Hendrix, Jessica Piel, Eileen Wagner, Emma Farley, Kasia Oktaba, Blair Gainous and Phil Abitua. I also wish to acknowledge our other collaborators
on the pausing and synchrony project namely Laura Stefanik, Jeffrey Johnston, Kai Chen, David S. Gilmour and Julia Zeitlinger. David Jun and Sam Ng provided great technical support for a number of the projects.

I need to acknowledge Nipam Patel for all his advice and guidance over the years. Nipam was always willing to talk about all things related to science from whether a particular step in an in situ protocol was absolutely necessary to what the best strategy was to image an embryo. His encyclopedic knowledge about development, protocols and imaging has been an invaluable resource. I also need to acknowledge Ehud Isacoff who was the director of the biophysics program when I joined for the seemingly minor but highly consequential role he played in my graduate career. Without Udi’s sage advice I would never have done rotations and been exposed to areas of research which I am now very passionate about but which were, at the time, very much out of my comfort zone. I also want to acknowledge Susan Marqusee for all the help and guidance she has provided me over the years.
Chapter 1

Introduction

All that exists in the natural world arises from information stored in nucleic acids [1]. These strings of As, Cs, Ts and Gs encode elaborate molecular recipes that turn a collection of simple chemicals into a cell, and a cell into an organized collection of 100 trillion cells that make a human being [2]. Physical interactions within and between different macromolecules on the nanoscale are what drive this process, and it is these interactions that lead to emergent behavior on length scales that span the next nine orders of magnitude. Our understanding of molecular physics has advanced greatly in the last century [3]. We have elucidated the relevant physical laws that dictate the properties of atoms and molecules and built quantitative frameworks that enable us to predict specific properties. We have also built computers that have vastly improved our ability to perform numeric calculations [4]. In principle it should be possible to implement this theoretical framework in a program that can predict how the different molecules in a cell interact and so calculate how simple chemicals turn into a cell, and how you organize trillions of cells into a person. Such a program would have immense and immediate benefits for human health and society at large.

Unfortunately such a program does not yet exist. Even thought we understand the relevant physics to produce it, the problem with implementing it is one of scale. As you increase the number of atoms the size of the calculation grows exponentially. A typical protein contains many thousands of atoms and a human cell contains \( \sim 10 \) billion protein molecules that interact with each other by diffusing in a sea of \( \sim 10 \) trillion water molecules [5]. With current computing technology the problem of modeling a cell and it’s contents is intractable and modeling a whole organism impossible. There are encouraging developments in the field of quantum computation that may make these kinds of calculations feasible in the coming decades [6, 7], but the area of practical quantum computing is still in it’s infancy. Hence the only tractable approach currently is to characterize and model a particular biomolecular process by making approximations and simplifications appropriate to the length and timescales of the process and then try to piece together the insights gleamed at the different levels to get a holistic sense of what’s going on.
In this thesis the goal was to do exactly that by focusing on the process of transcription. It is one of the most heavily regulated steps in going from the information in DNA to a phenotype [8]. It is also one where stochastic molecular scale fluctuations described so well by the physics we know can affect downstream cellular processes [9, 10, 11, 12, 13, 14]. There are typically only one or two DNA templates from which to transcribe mRNA in a cell and the process is regulated by a small number of proteins that diffuse and bind DNA. The number of mRNA molecules in a cell of a particular type are also typically relatively small number on the order of 100s and so are subject to large variance [15, 16]. This can lead to variability in the amount of mRNA in a given cell at a given time and variability in the amount of mRNA among different cells. In single celled organisms this variability from cell to cell may have limited consequences but in the development of multicellular organisms where patterning genes need to be deployed in a robust and coordinated manner across thousands of cells, this is unlikely to be the case. This is why this work was done in the context of the developing *Drosophila* embryo.

The genes studied in this thesis are those that pattern the dorsal ventral axis of the *Drosophila* embryo, i.e. those that specifying back from belly. Within just under an hour the foundations of the adult body plan are laid. The early patterning program in the *Drosophila* embryo is also very well studied and so provides the perfect context in which to ask questions about gene regulation [17]. *Drosophila* embryogenesis is incredibly rapid and so it is tantalizing to speculate that the system is optimized for deploying gene programs in a way that is running up against physical limits.

### 1.1 Transcription Initiation and Noise

Transcription of DNA into RNA is the first step in going from the information encoded in the genome into cellular effectors that have chemical and biological activity [8]. During this process 10s to 100s of proteins bound in different complexes are responsible for recruiting the mRNA Polymerase II enzyme (Pol II) to the gene of interest, getting it ready to be able to transcribe the DNA starting at the transcription start site (TSS) and then releasing the transcriptionally competent Pol II into the body of the gene to make the full length mRNA [8, 18]. This process can be broken up into 5 distinct steps (see Figure 1.1), each of which can in principle be regulated [18]. The first step consists of opening chromatin. Normally the DNA of eukaryotic genes that are not transcribed are locked up into a complex consisting of DNA and nucleosomes which prevent access of Pol II and other parts of the transcription machinery to the promoter of the gene. An activator can bind to the promoter proximal region (or in some cases to a distal enhancer) and recruit nucleosome remodellers to clear the promoter, priming it for the next step. The second step consists of the formation of the pre-initiation complex at the promoter of the gene. During this step the binding of different types of activators promotes the binding of the general transcription factors and
other cofactors that can recruit Pol II to the promoter region. The third step consists of the engagement of the polymerase with the DNA template. During this step DNA is unwound at the TSS and an open complex is formed. Step four consists of promoter escape where Pol II breaks contacts with factors responsible for recruiting it to the promoter and transcribes 20-50 bases downstream of the TSS producing an RNA molecule, and then it pauses. The final step is the release of the polymerase from the pausing step into productive elongation. This step is mediated by activators like P-TEFb after which the polymerase transcribes the length of the gene producing the full length mRNA capable of being translated into protein or functioning as a cellular effector in its own right. In multicellular organisms the activators of transcription often bind to enhancer sequences that are not promoter proximal and so they need to be brought to the proximity of the promoter through DNA looping to exert their effect\[19, 20\]. Studies from yeast and other model organisms provided evidence that the most heavily regulated step in transcription was the recruitment of Pol II to the promoter but genome wide studies of Pol II occupation in metazoan has recently challenged that view (see [21] and references there in)

Due to the stochastic nature of the steps involved in initiating transcription and the fact that there are typically only several templates from which it occurs the amount of mRNA produced in a cell can potentially be highly variable. This has been well documented in a number of organisms that include prokaryotes and eukaryotes [9, 10, 11, 12, 13, 14]. It has also been shown that since mRNA is the first step in the cascade of going from information in DNA to cellular effectors even small fluctuations in the number of molecules of mRNA in a cell can be amplified and lead to large variations in the concentration of specific proteins in the cell [13]. Moreover in the development of multicellular organisms many morphogenic events require the coordination of cell behavior across 1000s of cells and if they are out of sync bad things can happen. Although the variability in the amount of mRNA produced as a consequence of the inherently stochastic nature of transcription has been demonstrated development of multicellularity organisms proceed with incredible precision. Embryo after embryo is made reliably and reproducibly despite the potential variability. This begs the questions of which molecular mechanisms exist to ensure that the noise is not amplified and development proceeds robustly and reproducibly.

1.2 Patterning of the Drosophila Embryo

The Drosophila embryo starts life as an ellipsoid egg approximately 180 microns in diameter and 510 microns long [22]. Figure 1.2 shows the key steps during the early development of the Drosophila embryo. During the first two hours after fertilization the zygotic nucleus undergoes 13 rapid division cycles and these nuclei migrate to the periphery of the egg. For most of the time relevant to us the embryo is a syncytium which means there are no membranes separating nuclei and so proteins can freely diffuse from the neighbourhood of one nucleus
Figure 1.1: The key steps in transcription initiation. Step 1: chromatin opening. An untranscribed gene and its regulatory regions are entirely packaged as nucleosomes (green). An activator (orange oval) binds and facilitates the recruitment of nucleosome remodellers to clear the promoter of nucleosomes and opens it. Step 2: PIC formation. A second type of activator (yellow diamond) binds and promotes the binding of the general transcription factors (blue rectangle) and recruits coactivators (green hexagon), facilitating Pol II (red rocket) entry to the preinitiation complex (PIC). Step 3: initiation. DNA strands are separated (oval inside Pol II) at the transcription start site (TSS), and an open complex is formed. Step 4: promoter escape. Pol II breaks contact with promoter-bound factors and transcribes 20-50 bases downstream of the TSS and produces a short RNA (purple line) and pauses. This is partially mediated by SPT5 in Drosophila (pink pentagon) and the negative elongation factor (NELF) complex (purple circle). The Ser residues at position 5 (Ser 5) of the Pol II carboxy-terminal domain (CTD) tail are phosphorylated (red P) during this step. Step 5: escape from pausing. P-TEFb (blue triangle) is recruited by the activator and phosphorylates Ser 2 of the Pol II CTD repeats, SPT5 and the NELF subunits (blue Ps) which causes NELF to dissociate from the rest of the complex. Pol II then escapes from the pause and enters into productive elongation. Step 6: productive elongation. The Pol II elongation complex transcribes through the gene. (Schematic reproduced from [18])
to its neighbours. Most of the important patterning events of the embryo occurs during the 14th cell cycle (cc). The first major morphogenic movement is the process of gastrulation which occurs towards the end of cell cycle 14. The *Drosophila* embryo is unique in how rapid development occurs. Within just hours broad maternal gradients are turned into precise domains of gene expression. This necessitates very rapid changes in transcription programs.

Dorsal-ventral (DV) patterning of the *Drosophila* embryo is specified by the morphogen Dorsal, a sequence-specific transcription factor related to mammalian NF-κB [24, 25, 26]. The Dorsal protein is distributed in a nuclear gradient along the DV axis in the shape of a Gaussian distribution, with peak levels present in ventral most nuclei (see representation in Fig 1.3). The Dorsal gradient initiates DV patterning by regulating 50-60 target genes in a concentration-dependent fashion through enhancer sequences [27, 28]. In the dorsal most regions of the embryo the Dorsal gradient is shallow and so contains limited spatial information. In this regions of the embryo much of the spatial information is contained in the Dpp gradient, which is indirectly specified by the Dorsal gradient [29]. The Dpp gradient regulates a number of genes in a concentration-dependent fashion including pnr and tup which will be discussed in later sections.
Figure 1.2: Timetable of early *Drosophila* development. This figure is meant to orient the reader as to the events that occur in the early development of the *Drosophila* embryo which are relevant to this thesis. On the left are sketches of what the embryo looks like at the different stages and to the right is a timeline with important events annotated. This figure was drawn according to information in [23].
Figure 1.3: Schematic illustrating the Dorsal and Dpp gradients and some of their readouts. A The Dorsal nuclear gradient generates a series of distinct patterns of gene expression. The expression domains of some of these genes are depicted on a diagram representing a cross-section through an early embryo. The filled green circles represent high levels of nuclear Dorsal protein, and shaded green and yellow circles represent intermediate and low levels, respectively. (Schematic reproduced from [17]) B Dpp gradient activity. Diagram representing a cross-section through an early embryo showing the areas of Dpp activity, the more green shaded regions correspond to the highest activity which then tapers off towards the ventral regions of the embryo. (Schematic reproduced from [30])
Chapter 2

Dynamics of Transcriptional Repression

2.1 Chapter Summary

The development of the precellular Drosophila embryo is characterized by exceptionally rapid transitions in gene activity, with broadly distributed maternal regulatory gradients giving way to precise on/off patterns of gene expression within a one hour window, between 2 and 3 hrs after fertilization [31]. Transcriptional repression plays a pivotal role in this process, delineating sharp expression patterns (e.g., pair-rule stripes) within broad domains of gene activation. As many as 20 different sequence-specific repressors have been implicated in this process, yet the mechanisms by which they silence gene expression have remained elusive [32]. Here we report the development of a method for the quantitative visualization of transcriptional repression. We focus on the Snail repressor, which establishes the boundary between the presumptive mesoderm and neurogenic ectoderm [33]. We find that elongating Pol II complexes complete transcription after the onset of Snail repression. As a result, moderately sized genes (e.g., the 22 kb sog locus) are fully silenced only after tens of minutes of repression. We propose that this repression lag imposes a severe constraint on the regulatory dynamics of embryonic patterning, and further suggest that post-transcriptional regulators, like microRNAs, are required to inhibit unwanted transcripts produced during protracted periods of gene silencing. The results reported in this chapter were published in the following paper [34].

2.2 Introduction

Repression of transcription plays an important role in patterning a range of developmental systems [35]. In many developmental networks transcription factors mutually repress each other presumably to confer robustness to the system [36, 37]. Because of this understanding the detailed mechanisms responsible for how this occurs and affects the dynamic properties
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

of gene regulatory networks remains an important question. Repression might result from
the passive inhibition of upstream activators, such as the failure of the activators to mediate
looping to the core promoter. Alternatively, repressors might alter the chromatin state of
the promoter region, resulting in diminished access of the Pol II transcription complex [32,
38]. Such repression mechanisms might cause a lag in gene silencing due to the continued
elongation of Pol II complexes that were released from the promoter prior to the onset of
repression (Fig. 2.1B). As in the case of the delay in the production of mature mRNAs after
initiation (Fig. 2.1A), the lag in repression would be commensurate with the size of the
gene, with large genes taking longer to silence than small genes. This can take a significant
amount of time due to the surprisingly slow rate of Pol II elongation, just \( \sim 1 \text{ kb/min} \) [39].

Alternatively, elongating Pol II complexes might be arrested or released from the DNA
template due to changes in chromatin structure and/or attenuation of Pol II processivity.
Such mechanisms could lead to the immediate silencing of all genes regardless of size (see
Fig. 2.1C). Recent studies have documented rapid changes in the chromatin structure across
the entire length of genes, exceeding the rate of Pol II processivity [40]. Certain co-repressors
in the *Drosophila* embryo (e.g., Groucho) are thought to mediate repression by a spreading
mechanism that modifies chromatin over extensive regions [41]. Indeed, this type of mecha-
nism has been invoked to account for the repression of the pair-rule gene, even-skipped (eve),
by the gap repressor, Knirps (see below) [42]. The attenuation of Pol II elongation has been
implicated in a variety of processes. For example, Pol II attenuation has been documented
for the transcriptional repression of MYC [43]. Moreover, the activation of the HIV genome
is regulated by Pol II processivity [44].

In an effort to distinguish these potential mechanisms we visualized the repression dy-
namics of several Snail target genes as they are silenced in the presumptive mesoderm of
precellular embryos. The zinc finger Snail repressor is one of the most extensively studied
repressors in the *Drosophila* embryo. It has been implicated in a variety of developmental
and disease processes, including epithelial-mesenchyme transitions and tumorigenesis [33, 45,
46, 47, 48]. Snail typically binds to repressor sites located near upstream activation elements
within distal enhancers [49, 28]. Short gastrulation (sog) is transcriptionally repressed by
Snail and encodes an inhibitor of BMP/Dpp signaling that restricts peak Dpp signaling to
the dorsal midline of cellularizing embryos [50, 29, 51]. The sog locus is \( \sim 22 \text{ kb} \) in length
and contains three large introns, including a 5' intron that is \( \sim 10 \text{ kb} \) in length and a 3' intron that is \( \sim 5 \text{ kb} \) in length (see Fig. 2.1C). The use of separate intronic hybridization
probes permits independent detection of 5' (see Fig. 2.1F) and 3' (see Fig. 2.1G) sequences
within sog nascent transcripts (Fig. 2.1). Individual nuclei are then false colored according to
the probe combination they contain (see Fig. 2.1H) which we discuss in the following section.
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

Figure 2.1: Schematic showing how the initiation of transcription and different schemes of repression affect the dynamics of full-length mRNA production. A Gene models showing the differences in the distribution of polymerase on a short versus long gene and the amount of full-length mRNA produced some time after initiation. It illustrates that there is a significantly longer delay before Pol II complexes can reach the end of the 20 kb gene (20 mins) and produce productive transcripts, compared to the 2 kb gene (2 mins). B Gene models showing the differences in the distribution of polymerase on the two genes and the amount of full-length mRNA some time after transcriptional repression, assuming no new Pol II complexes are recruited to the gene after repression but that those on the gene finish elongating. It shows that there is a significantly longer delay before full-length mRNA production is repressed in the case of the long, compared to the short gene. C Similar to B except that elongating Pol II complexes on the template are arrested or have their processivity attenuated when the genes are repressed. This would result in a rapid cessation in the production of full-length mRNA for both the short and long genes (assumed elongation speed of Pol II is 1 kb/min throughout).

2.3 Results and Discussion

sog exhibits synchronous activation at the onset of cc13, \( \sim 2 \) hrs after fertilization (see [52]). There is a lag between the time when nascent transcripts are first detected with the 5’ probe
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

Figure 2.2: Time course of sog transcription from early cell cycle 13 to early cell cycle 14. A Lateral view of an embryo in the early stages of cc13. Most of the nuclei contain nascent transcripts. Only the 5′ (green) intronic probe is detected (see C). The nuclei are false colored according to the combination of probes they contain (see F-H). B Lateral view of an embryo midway through cc13. Most of the nuclei show in situ signal for the 5′ probe, and about half of these also show staining for the 3′ (red) probe. C Simplified gene model for the sog transcript showing the location to which the 5′ (green), sog1, and 3′ (red), sog3, intronic in situ probes hybridize. D Lateral view of an embryo in the late stages of cc13. Most of the cells express both the 5′ and 3′ probes. E Ventral view of an embryo in the early stages of cc14. Only isolated 5′ probe is detected. F Zoomed-in section of a cc14 embryo showing the expression of nascent transcript labeled by the 5′ probe in green. G The same section as in F, but with the 3′ probe labeled in red and the nuclear stain false colored green. H The same section shown in F and G, but after it has been processed with the segmentation algorithm. Nuclei that contain only isolated green and red nascent dots have been false colored green and red, respectively. Nuclei that contain a coincident red and green dot have been labeled in yellow.

and subsequently cross-hybridize with both the 5′ and 3′ intronic probes (Fig.2.1A,B). This lag is consistent with the established rates of Pol II elongation in flies, approximately 1.1-1.5
kb/min [39]. cc13 persists for \( \sim 20 \) min [53], and by the completion of this time window, most of the nuclei in ventral and lateral regions exhibit yellow staining, indicating the occurrence of multiple nascent transcripts containing 5' and 3' intronic sequences within each nucleus (Fig. 2.1D). There is little or no repression in ventral regions, presumably due to insufficient levels of the Snail repressor prior to cc14 [46, 54].

As shown previously, nascent transcripts are aborted during mitosis [55, 56]. Consequently, only the 5' hybridization probe detects sog nascent transcripts at the onset of cc14 (Fig. 2.1E). Moreover, a small number of nuclei (at the ventral midline) fail to exhibit nascent transcripts with either the 5' or 3' probe, suggesting repression by Snail. This repression becomes progressively more pronounced during cc14 (Fig. 2.3).

Within about 10 min of the first detection of sog nascent transcripts at the onset of cc14 (Fig. 2.3A,G), most of the nuclei exhibiting sog expression stain yellow, indicating expression of both 5' (green) and 3' (red) intronic sequences (Fig. 2.3B,H). During the next several minutes, progressively more nuclei exhibit only 3' (red) hybridization signals in ventral regions (Fig. 2.3C,I). This transition from yellow to red continues and culminates in a red flash where the majority of the ventral nuclei that contain nascent transcripts express only the 3' (red) probe (Fig. 2.3D). As cc14 continues there is a progressive loss of staining in the presumptive mesoderm (Fig. 2.3E), and eventually sog nascent transcripts are lost entirely in the presumptive mesoderm (Fig. 2.3F).

These results suggest that after its release from the promoter, Pol II continues to elongate along the length of the sog transcription unit, even as Snail actively represses its expression in the mesoderm. The red flash observed during mid-cc14 represents partially processed sog nascent transcripts that have lost the 5' intron (hence no green signals with the 5' hybridization probe) but retain 3' sequences (summarized in Fig. 2.3G-I). Previous studies are consistent with sequential processing of nascent transcripts, beginning with the removal of 5' intronic sequences and concluding with the removal of 3' introns [57]. As a control, two separate hybridization probes were used to label opposite ends of sog intron 1. As expected, there is no red flash since both hybridization signals are simultaneously lost when intron 1 is spliced (see Figs. 2.6; 2.7).

There is a \( \sim 20 \) minute lag between the onset of repression at early cc14 (Fig. 2.3B) and the complete silencing of sog expression in the presumptive mesoderm during mid to late cc14 (Fig. 2.3F). To determine whether this repression lag is a common feature of Snail-mediated gene silencing, we examined additional target genes, including ASPP, Delta, canoe and scabrous (sca). ASPP (Fig. 2.4D,E) encodes a putative inhibitor of apoptosis [58], while Delta (Fig. 2.4A,B) encodes the canonical ligand that induces Notch signaling. All four of these genes exhibit repression lag as they are silenced in the presumptive mesoderm of cc14 embryos (Fig. 2.4; Fig. 2.5).
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

Figure 2.3: Time course of sog transcription during cc 14. A Embryo in the early stages of cc14 (older than Figure 2.2E). Most nuclei only express the 5’ probe, but a small number also express the 3’ probe. B Within about 10 min of the first detection of nascent sog transcripts at the onset of cc14, most of the nuclei exhibiting both 5’ and 3’ probe. C During the next several minutes, progressively more nuclei exhibit only 3’ probe in ventral regions. D This transition from yellow to red continues and culminates in a red flash where the majority of the ventral nuclei express only the 3’. E As cc14 continues, there is a progressive loss of staining in the presumptive mesoderm. F Eventually, nascent sog transcripts are lost almost entirely in the presumptive mesoderm before gastrulation. G Gene model depicting a gene like sog with multiple introns, where a 5’ probe recognizes the mRNA coded for by the first intron and a 3’ probe recognizes the mRNA coded for by the second intron. Initially, only the 5’ probe will hybridize to the nascent transcript. H After enough time has elapsed for some Pol II complexes to reach the second intron labeled by the 3’ probe, so both probes will hybridize and this will manifest as a yellow dot in a nucleus. I If repression inhibits new polymerases from initiating transcription but allows elongating polymerases to finish transcription, then after a time, only the 3’ probe will hybridize to nascent transcripts, because all of the intronic sequences containing the 5’ probe will have been spliced out and degraded. In situ, nuclei where this has occurred will have an isolated red dot at the site of nascent transcription. Anterior is to the left, showing ventral views.

With the notable exception of Delta, the genes examined in this study contain promoter-proximal paused Pol II, as do most developmental patterning genes active in the precellular
Figure 2.4: Repression of Delta and ASPP transcription in the presumptive mesoderm. A cc14 embryo showing staining for Delta. Both the 5' (green) and 3' (red) probes (see C) hybridize to the nascent transcripts in most of the nuclei. However, in the ventral regions, a number of nuclei only show the presence of isolated the 3' probe, consistent with repression. B Older embryo showing more nuclei expressing the 3' probe, consistent with the continuation of Snail-mediated repression. C Simplified gene model for the Dl transcript showing the location of the largest introns and the location of the mRNA sequences to which the 5' and 3' in intronic in situ probes hybridize. D cc14 embryo showing staining for ASPP. Both the 5' and 3' probes (see F) hybridize to the nascent transcripts in most of the nuclei. However, in the ventral regions, there are a large number of nuclei that only show the presence of the 3' probe, consistent with repression. E Older embryo showing most of the nuclei in the mesoderm without any staining but some isolated nuclei expressing the 3' probe, consistent with the continuation of Snail-mediated repression. F Simplified gene model for the ASPP transcript showing the location of the largest introns and the location of the mRNA sequences to which the 5' (green) and 3' (red) in intronic in situ probes hybridize.

embryo [59]. Moreover, results from whole-genome Pol II binding assays indicate that these genes maintain promoter-proximal paused Pol II in the presumptive mesoderm as they are actively repressed by Snail. These findings are consistent with the observation that the seg-
mentation gene, sloppy-paired-1, retains promoter-proximal paused Pol II even after being silenced by the ectopic expression of the Runt repressor [60]. Thus, the Snail repressor does not affect Pol II recruitment, but rather, inhibits the release of Pol II from the proximal promoter of paused genes. At every round of de novo transcription each Pol II complex at the pause site must receive an activation signal for its release into the transcription unit. We propose that the Snail repressor interferes with this signal, resulting in the retention of Pol II at the pause site.

It is currently unclear whether repression lag is a general feature of transcriptional silencing. A recent study suggests that the gap repressor, Knirps, reduces the processivity of Pol II complexes across the eve transcription unit [42]. Snail and Knirps might employ distinctive modes of transcriptional repression. Snail recruits the short range corepressor, CtBP [61], while Knirps recruits either CtBP or the long-range corepressor, Groucho [62]. When bound to certain cis-regulatory elements within the eve locus, Knirps recruits Groucho, which might propagate a repressive chromatin structure. In contrast, Snail-CtBP might interfere with the release of Pol II from the proximal promoter, as discussed above. There is a considerable difference in the lengths of the genes examined in the two studies. The eve transcription unit is only 1.5 kb in length, less than a tenth the size of sog. In fact, many patterning genes active in the early fly embryo contain small transcription units, just a few kb in length. Small transcription units offer dual advantages in rapid patterning processes: little or no lag in activation or repression.
All five Snail target genes examined in this study exhibit Pol II elongation after the onset of repression. The number of transcripts produced during repression lag depends on the Pol II density across the transcription unit at the onset of repression. Pol II binding assays (e.g., ChIP-chip, ChIP-Seq, and Gro-Seq) suggest that there are at least several Pol II complexes per kb [59]. This estimate is based on comparing the total amount of Pol II within these genes to that present at the promoter of the uninduced hsp 70 gene, for which there are accurate measurements. As a point of reference, the Pol II density on induced heat shock genes is one complex per 75-100 bp [63], which is comparable to the footprint size, 50 bp, of an elongating Pol II complex [64]. Thus, something like 50 (or more) sog transcripts may be produced in a diploid cell after the onset of Snail repression. This represents a significant fraction of the steady-state expression of a typical patterning gene (~200 transcripts per cell [15, 16]).

Repression lag could impinge on a number of patterning processes, such as Notch signaling. The specification of the ventral midline of the CNS depends on the activation of Notch signaling in the ventral-most regions of the neurogenic ectoderm [65]. Sca products somehow facilitate the activation of the Notch receptor [66], and repression lag could potentially disrupt this process by producing high steady state levels of Sca in the mesoderm where Notch is normally inactive. Similar arguments might apply to the unwanted accumulation of Delta products in the mesoderm. Perhaps microRNAs are required to inhibit these transcripts, and thereby facilitate localized activation of Notch signaling. Indeed, miR-1 is expressed in the presumptive mesoderm, at the right time and place to regulate Sca and/or Delta [67], and is known to be able to target Delta transcripts [68]. Repression lag is potentially quite severe for Hox genes, particularly Antp and Ubx, which contain large transcription units (100-75 kb) that could take over an hour to silence after the onset of repression. It is conceivable that miRNAs encoded by the miR-iab4 gene, which are known to target Antp and Ubx transcripts [69, 70], might inhibit post-repression transcripts.

The precellular Drosophila embryo possesses a number of inherently elegant features for the detailed visualization of differential gene activity in development. Indeed, such studies were among the first to highlight the importance of transcriptional repression in the delineation of precise on/off patterns of gene expression. Here we extend this rich tradition of visualization by providing the first dynamic view of gene silencing. The key feature of our method is the use of sequential 5’ and 3’ intronic probes to distinguish nascent transcripts produced by Pol II complexes shortly after their release from the promoter vs. mature Pol II elongation complexes that have already transcribed 5’ intronic sequences. We show that elongating Pol II complexes complete transcription after the onset of Snail repression and as a result, moderately sized genes are fully silenced only after a significant lag. We suggest that this repression lag represents a previously unrecognized constraint on the regulatory dynamics of the precellular embryo. In the next chapter we examine to what extent the timing and coordination of the initiation of transcription depends on properties encoded by the promoter.
2.4 Experimental Procedures

In situ hybridization

Yellow white embryos were developed at room temperature and were collected at 1-4 hours of development and were fixed as described in [71]. The in situ hybridization protocol was that used in [71] with minor modifications. No Proteinase K was used and signal was not amplified using Tyramide amplification. All probes were made with digoxigenin conjugated haptens and biotin-conjugated haptins, the primary antibodies used to detect probes were sheep anti-digoxigenin and mouse anti-biotin (Roche Applied Sciences, Invitorgen), followed by Alexa Fluor 555-donkey-anti-sheep and Alex Fluor 488-donkey-anti-mouse secondary antibodies (Invitrogen, Molecular Probes). The primers that were used to amplify the DNA fragments that were used to make the different RNA probes are listed in Figure 2.8.

Quantitative confocal imaging

Embryos were imaged on Carl Zeiss LSM 700 Laser Scanning microscope as 20-25 section z-stacks through the nuclear layer at 1/2 micron intervals, using a Plan-Apochromat 20x/0.8, WD=0.55 mm lens. Confocal images were taken at 2048x2048 resolution at 8 bit color depth. Independent z-sections spanning the nuclear layer of the blastoderm embryo were sampled at 1/2 m intervals to find any active transcripts in all nuclei. Nuclei were counter-stained with either DAPI (Invitrogen) or DRAQ5 (Biostatus) and sequentially imaged for AlexaFluor 488, and AlexaFluor 555. Embryos were staged by quantitative analysis of nuclear density and by developmental morphology. We examined embryos from the beginning of cell cycle 13 through to the onset of gastrulation.

Automated image analysis

We wrote an automated image segmentation program in MatlabR2008b (Mathworks) to identify and count all stained nuclei and detected probes. For the channel containing the DNA counter stain, Z-stacks were projected into two-dimensional images by selecting the maximum intensity pixel in each stack. For the channels that contained the signal from the in situ probes each image plane was filtered by a Laplacian of Gaussian and then the stack was max projected to a two dimensional image. The core of each individual nucleus was determined using the DNA counter-stain, processed with a Laplacian of Gaussian filter, which allows robust determination of nuclei using blob detection, size selection, and signal strength to inform classification. This was followed by an object dilation algorithm to create a com
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

putational mask in which all pixels in an embryo are assigned to a uniquely identified nucleus.

The script then determines the transcriptional activity of nuclei by identifying which in situ signals from nascent mRNA are present in the region assigned to different nuclei. True hybridization foci or dots are identified for both channels separately through a series of filtering steps and segmentation. The 2D projection is filtered by a Laplacian of Gaussian filter which is then segmented using a combination of a watershed algorithm (to split most dots that are joined) and intensity based thresholding. This allows the location of true hybridization dots to be determined in the two different channels. The dots are then grouped into two categories, isolated or paired, depending on whether dots are only present in one of the two channels, or whether dots are present within several pixels of each other in both channels. In the former case the dots are labeled to only consist of one probe, and in the latter the dots are treated as a single locus expressing both probes. To address possible misclassification due to uncertainty in determining the nuclear boundary, transcripts localized to pixels adjacent to the nuclear boundary may be automatically reassigned to the neighboring nucleus if the original parent nucleus already contains an interior localized probe. This exploits the fact that the hetero or homo allelic expression of the reporter should result in no more than one or two foci per nucleus, respectively. An iterative extension of this algorithm also insured reliable classification when several adjacent nuclei each had transcripts that localized to boundary pixels.

Segmentation Controls

To estimate the uncertainty in our quantitative measurements we first checked the accuracy of our nuclear segmentation against manual nuclei determination. For all embryos examined, small adjustments to the filter parameters readily allow almost all of nuclei to be accurately segmented. We used double staining for overlapping probes to quantify the accuracy of our transcript localization filter and nuclear assignment routine. This analysis indicates >96% accuracy in identifying of overlapping probes is typical (Fig. 2.6). Errors in detection could arise from a number of different sources including failure of the probe to react or failure of the segmentation to detect very weak signal. These figures also show that the majority of the mistakes that are made in misidentifying co localization occurs when the in situ signal is of low intensity. This provides a measure of the sensitivity limits of our assay to detect transcriptional state.

In order to validate that the reliable detection of isolated red probe we examined how the intensity in the red and green channels correlated for the different dots. Figure 2.7C shows a scatter plot of the intensity of dots in the green and red channels. This is from an embryo undergoing repression, as shown in Fig. 2.7B, which has been labeled with the sog1 probe in green and the sog3 probe in red. The scatter plot shows that almost 18% of the detected transcripts are isolated red probe and Fig. 2.7B shows these are localized to the
presumptive mesoderm. Moreover, the scatter plot also shows that these are not restricted
to low intensity part of the plot, and that there is no significant correlation between the
intensity of the isolated red dots and the intensity in the green channel. This trend also
holds when a different probe, sog2, is used in the first intron (See Fig. 2.7F; 2.7G). These
trends were also consistently observed when the color of the different probes were reversed,
i.e. sog3 was labeled with green and sog1 and sog2 labeled with red (data not shown), which
is consistent with findings shown in Fig. 2.6.

This is qualitatively different to what is observed when both red and green probes are
within the first intron. Fig. 2.7J shows an embryo that has sog1 labeled in green and sog2
in red. The intensity scatter plot shown in Fig. 2.7K shows that only a small fraction, 3%,
of the dots detected are isolated second probe. This is comparable to the error rate of the
assay as seen when around 2% of isolated red probe was detected in the controls where the
same probe was labeled in different colors (See Fig 2.6B-2.6D). The isolated red probes also
did not show the lack of dependence on the intensity of the green channels seen when the
probes were in different introns, indicating that many of them were actually in pairs but were
misclassified. These trends were also consistently observed when the color of the different
probes was reversed, i.e. sog2 was labeled with green and sog1 was labeled with red (data
not shown), which is consistent with findings shown in Fig. 2.6.
Figure 2.6: Quantification of nascent transcript detection efficiency for sog probes. A Simplified gene model for the sog transcript showing the location of the three biggest introns and the location of the three different intronic in situ probes, sog1, sog2 and sog3. B Ventral view of a cc 14 embryo stained for sog1 probe in both red and green channels where nuclei have been false colored depending on which probes are detected in the nuclei. The vast majority of nuclei are classified as expressing both probes. C Intensity correlation plot for the embryo in B. D Magnified section of imaging data that was processed to produce B. E Segmented and false colored version of D. F-I Same as B-E except the probe used was sog2. J-M Same as B-E except the probe used was sog3.
Figure 2.7: Controls for detection of isolated 3’ probe. A Simplified gene model for the sog transcript showing the location of the three biggest introns and the location of the different intronic in situ probes, sog1, sog2 and sog3. B Ventral view of a cc 14 embryo stained for sog1 probe in green and sog3 probe in red, where nuclei have been false colored depending on which probes are detected in the nuclei. The majority of lateral nuclei are classified as expressing both probes, and most ventral nuclei express isolated red probe. C Intensity correlation plot for the embryo shown in B. D Magnified section of imaging data that was processed to produce B. E. Segmented and false colored version of D. F-I Same as B-E except sog2 probe was used in green and sog 3 probe in red. J-M Same as B-E sog1 probe was used in green and sog2 probe in red.
CHAPTER 2. DYNAMICS OF TRANSCRIPTIONAL REPRESSION

Figure 2.8: List of primer sequences used to make mRNA probes relevant to this chapter

<table>
<thead>
<tr>
<th>Sequences of Primers Used to Make All In Situ Probes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sog1F 5’-TGTTGCGTGGTGGTGGTGAAC-3’</td>
<td></td>
</tr>
<tr>
<td>Sog1R 5’-ACTGTGCTGGTGGTGGTGAAC-3’</td>
<td></td>
</tr>
<tr>
<td>Sog2F 5’-TCTCATCTCGGTTCTCTCATTG-3’</td>
<td></td>
</tr>
<tr>
<td>Sog2R 5’-TCGACGCACGCACTTAATCAACC-3’</td>
<td></td>
</tr>
<tr>
<td>Sog3F 5’-GAGGAGAGATGATGCTGGGCGATAAGG-3’</td>
<td></td>
</tr>
<tr>
<td>Sog3R 5’-TATCAAGTTGCGTTCGATTTTC-3’</td>
<td></td>
</tr>
<tr>
<td>ASPP1F 5’-ATTTCCACCTCCCTCCGAAG-3’</td>
<td></td>
</tr>
<tr>
<td>ASPP1R 5’-CGGAAAGCGAAGAACACAC-3’</td>
<td></td>
</tr>
<tr>
<td>ASPP2F 5’-AAGATAGCCACACAAAGGGAAG-3’</td>
<td></td>
</tr>
<tr>
<td>ASPP2R 5’-TGGAGCAAACGAAAGGGAAC-3’</td>
<td></td>
</tr>
<tr>
<td>Delta1F 5’-AACATCCAGATCAGAGACCC-3’</td>
<td></td>
</tr>
<tr>
<td>Delta1R 5’-ACGATTGAGGAGGTTGCGAAGG-3’</td>
<td></td>
</tr>
<tr>
<td>Delta2F 5’-ATCACCTGTCCACGACATC-3’</td>
<td></td>
</tr>
<tr>
<td>Delta2R 5’-GAGTCTTTGTAAGCCTTTCTCC-3’</td>
<td></td>
</tr>
<tr>
<td>Canoe1F 5’-CCTCTCCATTTCTCGTTGC-3’</td>
<td></td>
</tr>
<tr>
<td>Canoe1R 5’-TCCAATGTCGGGCTATTATG-3’</td>
<td></td>
</tr>
<tr>
<td>Canoe2F 5’-GCAATCTGGGCATAGCGAAG-3’</td>
<td></td>
</tr>
<tr>
<td>Canoe2R 5’-AAGTTGAATGGCTGGGAGG-3’</td>
<td></td>
</tr>
<tr>
<td>Sca1F 5’-GTCATATGTCTCGTGCAATTTTG-3’</td>
<td></td>
</tr>
<tr>
<td>Sca1R 5’-GGCCAGGTAATCTTTTAATTGCTACAT-3’</td>
<td></td>
</tr>
<tr>
<td>Sca2F 5’-CTCAGTCTAATGCTTATATCT-3’</td>
<td></td>
</tr>
<tr>
<td>Sca2R 5’-CTCAATGTAGCCCAAAATATGC-3’</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3

Pol II pausing and coordination of transcription

3.1 Chapter Summary

Morphogen gradients have been implicated in a variety of metazoan developmental processes, and much is known about how these gradients produce different spatial patterns of gene expression. However, there is considerably less information about the mechanisms underlying the temporal control of gene expression. Here we employ a quantitative imaging assay to measure the dynamics of gene activation during patterning in the Drosophila embryo. Genes containing paused RNA polymerase (Pol II) exhibit more synchronous and rapid induction of gene expression than those lacking Pol II, but not all paused promoters are equivalent. Using transgenes with different promoters and enhancers we demonstrate that a spectrum of paused Pol II determines the “time to synchrony” the time required to achieve coordinate gene expression across the cells of a tissue. This differential timing correlates with Pol II pausing stability, and mutations that destabilize paused Pol II cause a delay in gene expression. We conclude that the promoter is a prime determinant of developmental timing. The results of this chapter form the first part of the following publication on which I am co-first author [72].

3.2 Introduction

The early Drosophila embryo is the premier system for visualizing gene activity in animal development. In a period of just one hour, broadly distributed maternal determinants generate localized patterns of gene activity, including segmentation stripes of gene expression [73]. A variety of studies suggest that enhancers, typically 300-500 bp in length, are responsible for determining where and when developmental control genes are switched on and off [74]. With few exceptions, localized patterns of expression can be attributed to discrete enhancers
located upstream, downstream or within the gene of interest.

The enhancer is therefore seen as the key agent of differential gene activity in animal development [74]. Considerably less is known about the role of the promoter in the regulation of the spatial or temporal limits of gene expression, although they are known to control the rates of RNA synthesis [75]. In the simplest view, enhancers determine the limits of gene expression (where and when genes are active), while the promoter controls the levels of expression (e.g., how many transcripts are produced in a given unit of time).

The purpose of this study is to determine whether the promoter regions of developmental control genes can influence the timing or spatial limits of gene expression in the early Drosophila embryo. We were motivated by the recent finding that many developmental control genes contain paused RNA Polymerase (Pol II) prior to their activation during embryogenesis [76, 21]. The function of paused Pol II is uncertain, despite its apparent prevalence (~30% of all protein coding and noncoding genes) in both Drosophila embryos and mammalian stem cells [77].

The prototypic example of paused Pol II, Drosophila heat shock genes, underlies rapid induction of gene expression in response to stress [78]. There is also evidence that paused Pol II serves to keep promoters open by excluding or diminishing the occurrence of positioned nucleosomes that occlude the transcription start site in cultured cells [79]. Recent quantitative imaging methods suggest that paused Poll II influences synchronous induction of gene expression across the different cells of presumptive tissues in the early Drosophila embryo [52].

In the latter study, quantitative in situ hybridization assays were used to detect the first nascent transcripts encoded by different developmental control genes, within the first 10-20 min after the onset of expression in precellular embryos, ~2 hrs after fertilization. Genes were classified as synchronous if nascent transcripts were detected in over 50% of the nuclei that will eventually express a given gene, or stochastic if expressed in fewer than 50%. Most paused genes exhibited synchronous patterns of activation, while most nonpaused genes displayed stochastic expression. The evidence linking paused polymerase and synchrony was strictly correlative, and there is no evidence that these modes of activation are significant in development.

3.3 Results and Discussion

Previous studies suggested a correlation between paused Pol II and synchronous patterns of gene activation in the Drosophila embryo [52]. Moreover, computational analyses identified sequence elements that are associated with promoters containing paused Pol II, including
GAGA and pause button (PB) motifs [79, 80, 81, 82]. These observations raise the possibility that the core promoter might be sufficient to determine whether a gene is paused or not paused, and activated in a synchronous or stochastic fashion.

As a first step towards testing this possibility we examined the regulation of two Dpp (TGFβ) target genes, pannier (pnr; GATA4) and tailup (tup; Islet-1), transcription factors essential for the specification of a variety of dorsal tissues, including the heart [83]. These genes are co-activated in the dorsal ectoderm of 2-hour embryos [84], but nonetheless display opposite Pol II binding profiles. tup is strongly paused, while pnr lacks Pol II [85]. The use of quantitative imaging methods revealed differences in their activation profiles that were missed in previous studies, as discussed below (Figure 3.1).

3.4 Temporal Coordination of Dpp Target Genes

tup is activated by high levels of the Dpp gradient while pnr is triggered by low levels (Figure 3.1A-H) [84]. These distinctive spatial expression patterns depend on previously identified tup and pnr enhancers. Quantitative imaging methods reveal that they also exhibit dissimilar temporal profiles (Figure 3.1E-I).

It was previously shown that tup contains paused Pol II and is activated in a synchronous fashion, whereas, pnr lacks Pol II and exhibits stochastic expression [52]. We developed high-resolution confocal visualization and novel image segmentation methods to measure the time to synchrony, the degree of temporal coordination in gene activation during nuclear cleavage cycle (cc) 14, the one-hour interval preceding gastrulation (Figure 3.1A-H). The ~ 6000 cells comprising the pregastrula embryo are synchronized within the cell cycle, thereby permitting direct comparisons of transcriptional coordination. Quantitative FISH assays permit detection of nascent transcripts shortly after the onset of gene expression [34]. In this assay, activation is defined as the time it takes for 50% of the nuclei to express nascent transcripts (t50). Using a cumulative gamma distribution, we fit a curve to each experimental dataset (see Figures 3.1 and 3.9). t50 values are calculated by measuring the fraction of nuclei that express a given gene for each fitted activation profile. Pregastrula cc14 embryos are selected based on nuclear density and embryo morphology and then ordered relative to one another based on the fraction of the expression pattern containing nascent transcripts. The collections are designed to ensure that embryos are distributed in an unbiased way across the entirety of cc14. This approach allows us to measure the t50 values with an accuracy of ±5 min (see section 3.11, Figures 3.6 and 3.10).

The endogenous tup and pnr genes exhibit distinct t50 activation profiles: tup achieves t50 expression ~ 26 min after the onset of cc14, while pnr does not exhibit comparable expression for another 15 min (3.6). This represents a significant delay since the entire
Figure 3.1: BMP/Dpp target genes exhibit distinct coordination profiles A-H cc14 embryos hybridized with tup and pnr fluorescent (magenta) intronic probes for detecting nascent transcripts (nuclei stained with DAPI [blue]). Raw images for tup and pnr transcripts are shown in B and D, and the corresponding processed images are shown in (B') and (D'). Images shown in B and D are magnifications of bracketed regions in A and C. E-H tup (E and G) and pnr (F and H) expression during mid (E and F) and late (G and H) cc14. I and J Dynamics of gene expression during cc14 based on the fraction of nuclei containing nascent transcripts. I Endogenous tup expression (blue) reaches 50% of the complete pattern (t50, 26) 15 min earlier than does pnr (black) (t50, 41). J There is a delay in tup dynamics when the minimal promoter of a tup BAC transgene (tupY) is replaced by that of pnr (tupY-PnrPr) (see also Figure 3.2). The red curves represent the fitted curves (using a cumulative gamma distribution) to the data depicted in (I) and (J) (see section 3.11; Figures 3.9, 3.6 and 3.10). t50 values are determined from these fitted curves.

cc14 interphase extends for just 55 min (see below). To determine whether these divergent temporal expression profiles are due to enhancer or promoter sequences, we created a BAC
transgene encompassing the entire tup transcription unit and flanking regulatory DNAs that recapitulates the rapid and synchronous activation profile of the endogenous tup locus (Figure 3.1J, Figure 3.2). In these experiments the tup transcription unit was replaced with the yellow reporter gene to facilitate detection of nascent transcripts [47]. There is a slight delay in the t50 value of the BAC transgene (~32 min) as compared with the endogenous tup locus (~26 min) (Figure 3.6), which is likely due to the heterologous site of transgene insertion, a slower rate of yellow transcription, or the use of heterozygous embryos to measure expression of BAC transgenes (see Figure 3.1I, J and legends).

We next examined the activation profile obtained upon replacement of the paused tup promoter with the non-paused pnr promoter (Figure 3.1J, Figure 3.2). The modified BAC transgene is identical to the control, except for the substitution of just 200 bp centered around the +1 transcription start site of the pnr promoter. The modified transgene was inserted into the same chromosomal location as the control transgene, thereby permitting direct quantitative comparisons of their activation dynamics. Surprisingly, this 200 bp substitution within the large 60 kb BAC transgene is sufficient to convert the rapid and synchronous tup-yellow expression pattern into a slow and stochastic mode of activation (Figure 3.1J).

The modified transgene exhibits a t50 value of 53 min, which is considerably slower than the t50 values seen for the endogenous tup locus (26 min) or unmodified tup BAC transgene (32 min). It is somewhat slower than the t50 value seen for the endogenous pnr locus (41 min), although the differential timing of the tup vs. pnr promoters (t50), is similar for the endogenous loci and BAC transgenes, at 15 min and 21 min, respectively (Figure 3.6). These findings suggest that the pnr promoter, not enhancers, is the prime determinant of its slow and stochastic activation profile during development.

### 3.5 Minimal Promoter Sequences are Sufficient to Establish Paused Pol II

The preceding results suggest that minimal promoter sequences might be sufficient to determine whether a gene is activated in a synchronous or stochastic fashion. To determine whether they are also sufficient for determining the presence or absence of paused Pol II, we analyzed minigenes containing the pnr intronic enhancer (pnrE), tup promoter (tupPr), and yellow reporter gene (pnrE>tupPr/yellow). This minigene exhibits synchronous expression in the the dorsal ectoderm of wild-type embryos (see below), but is inactive in Toll10b mutants due to the absence of Dpp signaling [86]. Both the endogenous tup locus and the minigene nonetheless contain paused Pol II in these silent Toll10b embryos (Figure 3.3A,B); as expected, the endogenous pnr locus lacks paused Pol II [85]. Permanganate footprint assays identified hypersensitive thymidine residues at positions +48 and +51 nu-
Figure 3.2: The pur promoter is sufficient to delay tup expression. Processed images after in situ hybridization of transgenic embryos with intronic probes for the yellow reporter gene (BAC transgenes) (shown in yellow) or endogenous tup (shown in magenta) in the same embryo. A and A’ Reporter (A) and endogenous tup expression (A’) in a mid cc14 embryo containing one copy of the tup-Y BAC. B and B’ Reporter (B) and endogenous tup expression (B’) in an embryo just prior to the onset of gastrulation which contains one copy of the tup-Y control BAC. C and C’ Reporter C and endogenous tup expression C’ in a mid cc14 embryo that contains one copy of the tup-Y BAC where the tup promoter region has been replaced by an equivalent sequence from the pur promoter. Note that expression is significantly perturbed. D and D’ Same as in C and C’ except the embryo is older being at the onset of gastrulation.

cleotides downstream of the tup transcription start site in transgenic embryos (Figure 3.3C), strengthening the evidence that the stalled Pol II identified at the tup promoter represents promoter-proximal paused Pol II.
Thus, the 200 bp tup promoter region is sufficient for the establishment of paused Pol II (and synchronous expression, as shown below). It contains key signatures of paused promoters [79, 80, 81, 82], including 5' GAGA elements located ~100 bp upstream of the transcription start site, and PB motifs positioned +54 to +64 bp downstream of the start site, in the vicinity of the hypersensitive thymidine residues identified by permanganate protection assays (Figure 3.3C). We therefore conclude that minimal promoter sequences are sufficient to establish paused Pol II in vivo, in the *Drosophila* embryo. In principle, any gene can be artificially paused or de-paused by exchanging minimal promoter sequences. Such an approach may be relevant to the stem cell field since some of the key determinants of pluripotency (e.g., Nanog) exhibit stochastic expression among the different ICM cells of mouse embryos [87, 88].

3.6 Promoter-Associated Elements Influence Transcriptional Synchrony

To establish a sharper connection between pausing elements in the tup promoter and transcriptional synchrony, we expressed the pnrE>tupPr/yellow transgene in embryos containing diminished levels of the GAGA-binding protein, GAF (or Trl). Previous studies implicated GAGA and Trl in the stable association of paused Pol II within the proximal promoter of Hsp70 [89, 90, 82]. The tup promoter region contains GAGA elements located ~100 bp upstream of the transcription start site (Figure S3), and whole-genome assays confirm GAF/Trl binding to this region in the *Drosophila* embryo [64]. Reduced levels of Trl caused an ~20 min delay in the activation of the pnrE>tupPr/yellow transgene, as compared with wild-type embryos (Figure 3.3D), similar to the activation profile mediated by the nonpaused thisbe promoter (see below). An equivalent delay is observed with a truncated tup promoter lacking upstream GAGA elements, but retaining all core elements such as the INR (Figure 3.3D). These studies suggest a close correlation between Trl/GAGA and the temporal coordination of gene activation.

The stability of paused Pol II also depends on negative elongation factors, such as NELF and Spt5, which bind nascent transcripts shortly after the onset of transcription [79, 91]. There is an ~30 min delay in the activation profile of the pnrE>tupPr/yellow transgene in embryos containing reduced levels of NelfE and Spt5 (Figures 3.3D and 3.6). Thus, the preceding findings suggest a close correlation between minimal promoter sequences, paused Pol II and the time to synchrony in the *Drosophila* embryo.
Chapter 3. Pol II pausing and coordination of transcription

Figure 3.3: The minimal promoter mediates paused Pol II. A Pol II Chip-seq reads of the pnr/tup transgene in a tissue where it is silent. B Pol II ChIP followed by qPCR showing enrichment at the tupPr/yellow junction. y ORF, yellow open reading frame. Error bars represent SD. C Permanganate footprinting reveals a promoter-proximal transcription bubble in mutant embryos where the tupPr/yellow transgene is silent. D Reduced levels of maternal Trl (turquoise) or NelfE/Spt5 (pink) cause a delay in the expression profile of the pnrE > tupPr transgene. A similar effect is observed with a truncated version of the tupPr lacking the upstream GAGA sites.
3.7 A Spectrum of Synchrony

Whole-genome Pol II ChIP-Seq assays suggest that genes might not be simply paused or nonpaused, and activated in a strictly synchronous or stochastic fashion. Instead, there are different levels of Pol II in the promoter regions of genes previously identified as stalled or not stalled (Figure 3.4F; [85]). Normalized levels of paused Pol II were measured in vivo in dorsal-ventral patterning mutants containing a single embryonic tissue, in which the gene in question is silent (Figure 3.4F). For example, tup is not expressed in Toll10b mutant embryos since they contain only mesoderm due to the transformation of ectoderm into mesoderm. Conversely, snail is not expressed in gd7 mutant embryos, which display the reciprocal transformation of mesoderm into ectoderm. The tup and snail (sna) promoters contain significantly more Pol II sequence reads than sog in silent mutant embryos, even though all three genes were classified as stalled or paused in previous studies [85]. Similarly, thisbe contains more Pol II than pnr, even though both genes were classified as nonstalled. Tup is consistently seen to contain the highest levels of promoter-proximal Pol II read counts in a variety of tissues [92].

To investigate the significance of these different levels of Pol II, we analyzed the expression of a series of minigenes containing the pnr enhancer (pnrE) and six different promoter sequences encompassing a spectrum of paused Pol II. Remarkably, the activation profiles of these minigenes mirror the levels of Pol II binding (Figure 3.4E and 3.6). The tup promoter contains the highest levels of Pol II and exhibits a t50 value of just \( \sim 15 \) min. This is followed by progressively slower profiles for sna (t50 = 24 min), hsp70 (28 min), and sog (38 min), which contain successively lower levels of Pol II.

Finally, the promoter regions of the nonpaused genes ths and pnr exhibit the slowest activation dynamics, although ths is somewhat faster (t50 = 55 min) than pnr (t50 = 74 min). A similar correlation between the levels of paused Pol II and the time to synchrony was seen for minigenes containing the sog intronic enhancer, which mediates activation in the neurogenic ectoderm (Figures 3.5A-D, 3.6), and for the distal snail enhancer in the mesoderm (see below).

3.8 Transcriptional synchrony and rates of RNA synthesis

The preceding findings demonstrate that the same enhancer can produce a spectrum of activation profiles in the ectoderm of early embryos. To determine the feasibility of manipulating the timing of gene expression in the presumptive mesoderm, we placed the distal (shadow) snail enhancer (snaE) [93, 47] upstream of the sna, sog, ths, and pnr promoters and yellow reporter gene (Figure 3.7A-C). We observed similar relative t50 values as those obtained
Figure 3.4: A spectrum of synchrony. The pnrE was placed upstream of the tup, sna, Hsp70, sog, ths, and pnr promoters (Pr) (see diagram in upper left). A-D Examples of transgenic embryos stained with a yellow intronic probe at the midpoint of cc14. The tupPr mediates synchronous expression in the dorsal ectoderm A, whereas the pnrPr mediates stochastic expression D. E Temporal coordination profiles during cc14. The tupPr provides the rapid coordination profile, whereas the pnr promoter exhibits the slowest coordination while sog and ths give intermediate. F Relative amounts of Pol II at the promoter regions of inactive genes. For actively expressed genes, we denote them as expressed.

with the pnr enhancer (Figures 3.4 and 3.6). The snail promoter mediates a $t_{50}$ value of 3 minutes, while the more weakly paused sog promoter exhibits a 19 min delay in the $t_{50}$ profile (Figure 3.7, 3.6). As expected, the nonpaused ths and pnr promoters mediate even slower activation profiles ($t_{50} = 40$ and 42 min, respectively). The snaE enhancer mediates more rapid onset of expression in cc14 than the pnrE enhancer, probably due to the earlier availability of activators (e.g., Dorsal and Twist) in the mesoderm as compared with the dorsal ectoderm (e.g., pSmad). Nonetheless, after upstream activators initiate expression, the detailed temporal dynamics ($t_{50}$ activation) are determined by the different promoter sequences.
Figure 3.5: Temporal coordination with sog enhancer. A–C Processed images showing ventral views of embryos after fluorescent in situ hybridization using a yellow intronic probe showing nascent transcripts from a sog enhancer-tup promoter A transgenic embryo or from a sog enhancer-sna promoter embryo B and a sog enhancer-pnr promoter C examined at cc13 (the activity of the sog intronic enhancer starts at the onset of cc13 and during this time period the gene is not repressed by Snail). D Quantification of the dynamics of de novo transcription in these sog enhancer transgenic lines.

The sna, sog, and ths promoters provide a nice spectrum of activation during cc14 (t50 values of 3, 22, and 40 min, respectively), and seem ideally suited for manipulating the synchrony of snail expression in the presumptive mesoderm. Our choice of snail stems from the short lag time, less than 90 min, between the onset of transcription and morphogenesis - the coordinate invagination of the ventral mesoderm during gastrulation (see below). However, the accurate interpretation of any changes in gastrulation arising from the use of heterologous promoters requires an understanding of the relationship between t50 activation profiles and the levels of gene expression.

We expected promoters mediating slow synchrony profiles (e.g., ths) to produce weaker expression than those mediating rapid synchrony (e.g., sna promoter). Single molecule in situ hybridization assays (smFISH) [16] were employed to measure the number of yellow mRNAs produced by different yellow minigenes, snaE>pnrP/yellow, snaE>thsP/yellow, and snaE>snaP/yellow (Figure 3.7E). As expected, the slow minigenes produce lower levels of yellow mRNAs than the fast genes. The pnr, ths, and sna promoters produce 30 ± 10, 60
Figure 3.6: Summary of the t50 values for all of the constructs used in this chapter. t50 corresponds to the time it takes for an embryo to show nascent transcription in 50% of the pattern. t50 is an estimated time based on the measured activated kinetics of many embryos (see section 3.11 for further details).

<table>
<thead>
<tr>
<th>Promoter</th>
<th>t50 Time (min)</th>
<th>Uncertainty from Simulations (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PnrP Endo</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td>TupP Endo</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Tup-Y BAC</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>Tup-Y-PnrPr BAC</td>
<td>53</td>
<td>2</td>
</tr>
<tr>
<td>PnrE-TupPr NelfE/SptS</td>
<td>39</td>
<td>4</td>
</tr>
<tr>
<td>PnrE-TupPrShort</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>PnrE-TupPr Trl</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>PnrE-TupPr</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>PnrE-SnaPr</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>PnrE-Hsp70Pr</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>PnrE-SogPr</td>
<td>38</td>
<td>4</td>
</tr>
<tr>
<td>PnrE-ThsPr</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>PnrE-PnrPr</td>
<td>74</td>
<td>4</td>
</tr>
<tr>
<td>SnaE-SnaPr</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>SnaE-SogPr</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>SnaE-ThsPr</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>SnaE-PnrPr</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>SogE-TupPr</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>SogE-SnaPr</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>SogE-PnrPr</td>
<td>35</td>
<td>7</td>
</tr>
</tbody>
</table>

± 20 and 100 ± 30 mRNAs/cell, respectively, in the mesoderm prior to invagination.

Modeling methods were used to estimate promoter strength based on activation kinetics (see Section 3.10). The different levels of yellow mRNAs produced by the ths and sna promoters can be attributed to their respective t50 activation profiles. The snaE>thsP/yellow minigene is expressed at lower levels than snaE>snaP/yellow due to its slower synchrony profile. However, once activated in a given cell, the ths promoter appears to mediate a similar rate of RNA synthesis as the sna promoter (see Section 3.10). Similarly, quantitative measurements suggest that the sog promoter mediates a similar rate of expression as the snail and ths promoters once activated (see below). In contrast, the low levels of yellow mRNAs produced by the pnr promoter probably result from the combination of a slow synchrony...
profile and a lower rate of RNA synthesis. Thus, we focused on the use of the ths and sog promoters to examine the consequences of desynchronizing the onset of snail expression in the next chapter.

3.9 Model for the developmental timing of gene activation

Recent studies in S2 cells suggest that developmentally regulated genes tend to contain either paused Pol II or inhibitory nucleosomes [79]. RNAi-mediated depletion of NELF led to reduced levels of paused Pol II and a concomitant increase in promoter-positioned nucleosomes. These studies prompted the proposal that paused Pol II might render genes poised for activation by excluding the formation of inhibitory nucleosomes at the core promoter.

It is possible that nonpaused genes mediate slow activation dynamics due to cell-cell variation in the eviction of inhibitory nucleosomes at the core promoter. If occupied by an inhibitory nucleosome, a distal enhancer will not be able to stimulate transcription as it engages the promoter. Either the enhancer must await repositioning or dynamic turnover of inhibitory nucleosomes to allow recruitment of Pol II. Either way, this process might be inherently stochastic, resulting in cell to cell variations in the onset of transcription.

In principle, this model can account for the spectrum of activation profiles seen for genes containing different levels of paused Pol II. A gene containing high levels, such as tup, is more likely to contain Pol II than an inhibitory nucleosome in a given cell at a given time as compared with genes containing little or no paused Pol II (e.g., ths and pnr, respectively). Consequently, upon induction, strongly paused genes exhibit synchronous patterns of activation since most of the promoters in the different cells of a tissue contain Pol II. In contrast, genes containing little or no paused Pol II are more likely to contain an inhibitory nucleosome in a given cell at a given time, resulting in variable delays in the onset of gene expression. Thus, the ratio of poised and inhibited states might determine the time to synchrony.

In this chapter we have shown using minigenes, that there exits a tight correlation between pausing and coordination of transcription. This however doe not address the question of whether this is important in the context of development. We explore this question in the next chapter.
3.10 Determining Promoter Strength

We determine the relative strength of the pnr, ths and sna promoters by combining our data, i.e. measured levels of the yellow reporter mRNA and promoter activation kinetics with a model describing how the amount of mRNA in a cell changes with time. This model has two parameters: the rate at which the promoter produces mRNA once activated, i.e. the promoter strength, and the degradation rate of this mRNA. Our measurements alone do not allow us to independently specify both parameters. However by combining the known sna promoter strength [16] with our data, we can determine the half-life of the yellow reporter mRNA. Fortunately, the reporter mRNA is the same for all constructs. Using this method (detailed below, Figure 3.8) we determine the relative promoter strength of the different promoters is 1, 0.93 ± 0.15 and 0.47 ± 0.10 for sna, ths and pnr, respectively.

The model we use assumes that once a locus starts transcribing mRNA it produces mRNA at a constant rate, which depends on the intrinsic strength of the promoter, $\alpha$. The fraction of loci in the core region that are transcribing at a given time is approximated by the measured activation curve defined by $A(t)$. The history of promoter activity, i.e. how recently the promoter entered an actively transcribing state in a given nucleus can have a significant impact on the amount of mRNA present before steady state is reached. The other parameter is the degradation rate of the mRNA, $\lambda$. Under these assumptions it is possible to formulate an ordinary differential equation (ode) that describes how the average number of mRNAs per cell, $N(t)$, change as a function of time, $t$:

$$\frac{dN(t)}{dt} = \alpha \times A(t) - \lambda \times N(t) \quad (3.1)$$

Early estimates of promoter strength in *Drosophila* vary from about 1 mRNA/minute per template for histone genes [94] to 10 mRNAs/minute per template in the case of the heat shock genes [95]. The promoter strength for the sna promoter has recently been measured to be 6 mRNAs/minute [16].

The half-life for mRNA in the *Drosophila* embryos has also been measured for ftz to be between 6-10 minutes [96]. The half-life of snail mRNA has been measured to be 15 minutes [16]. Many mRNAs expressed in the early embryo are likely to have similarly half-lives around 5-15 minutes since mRNA expression patterns change on this timescale and mRNA accumulation has been shown to closely follow that of the transcription. Figure S5D shows the calculated number of yellow mRNAs at the onset of gastrulation for the sna promoter construct and how this varies as a function of promoter strength and mRNA half-life. By using the known sna promoter strength (6 mRNAs/minute), the measured number of yellow mRNAs (100 ± 30 mRNAs/cell) for this construct and the calculated number of yellow mRNAs from the model we can back out an estimate for the yellow mRNA half-life of 6 ±
1 minutes.

By using the promoter activation curve measured for the construct with the ths promoter, we can estimate the number of yellow mRNAs at the onset of gastrulation and how this varies as a function of promoter strength and mRNA half-life (Figure 3.8). Since the mRNA half-life is 6 ± 1 minutes and the number of yellow mRNAs is 60 ± 20, we can determine the strength of the ths promoter. This yields a value for the ths promoter, which is 5.6 ± 1 mRNAs/min and so relative to the strength of the snail promoter 0.93 ± 0.15. This shows that the ths and sna promoters have very similar promoter strengths in spite of showing different initiation kinetics.

We can similarly calculate the promoter strength for the pnr promoter construct (Figure 3.8) (mRNA half-life of 6 ± 1 minutes; number of yellow mRNAs which is 30 ± 10). This yields a value for the pnr promoter, which is 2.8 ± 0.6 and so relative to the strength of the snail promoter 0.47 ± 0.10. This shows that the ths and pnr promoters have very different promoter strengths, in spite of showing very similar initiation kinetics.

3.11 Quantifying Initiation Dynamics and Determining t50

To quantify the initiation dynamics, we combined the calculated fraction on for all the ~50-100 embryos for the different constructs tested. The embryo collections were designed to ensure that embryos are distributed in an unbiased way across the entirety of the one hour window of cc14. Under the assumption that the fraction of the core pattern active increases monotonically with time (which was verified by examining the fraction of pattern on at different stages of membrane invagination), we can order these embryos from youngest to oldest based on the fraction of pattern that is active. This means that we know the relative age of each embryo but dont know the absolute age. To determine the kinetics, we can assume that the timing between embryos is uniform at the cost of introducing uncertainty in the activation kinetics. We measure large numbers of embryos and so we sample the period of interest densely and as a result, the uncertainty introduced with this method is relatively small. We can explicitly calculate the uncertainty introduced using our model to describe the timing of gene activation. The uncertainty in time introduced with this method is less than 5 minutes in most cases (see Figure 3.6, Figure 3.9 and Figure 3.10), which is sufficient for our purposes and significantly better than what can be obtained by monitoring membrane invagination.

In order to better understand the process of gene activation and be able to compare activation kinetics across different constructs and promoters, we needed to develop a model
to characterize the time to gene activation. From what we have measured, it is clear that the time it takes for a particular nucleus to start transcribing a gene is a random variable. Exactly how pausing will affect the stochastic molecular processes that set the timescale for the activation of gene expression are not understood. Hence, it would be very difficult to derive a distribution with few free parameters from a minimal mechanistic model. Following the principle of maximum entropy, we decided that the most appropriate distribution to use to model the time to activation would be the gamma distribution [97].

The gamma distribution is widely used to model stochastic waiting times in a range of other contexts and has only two free parameters [98, 97]. If the time to activation in a given nucleus is described by a gamma distribution, then the fraction of the pattern on at a given time is given by the gamma cumulative distribution function (GCDF). We were encouraged to find that even with just 2 free parameters, the GCDF was able to fit the activation curves we measured for the myriad of different promoter and enhancer combinations very well (See Figure 3.9A and 3.9B). Based on these fitted curves (which took the entire dataset into account), we determined a parameter to characterize how quickly genes were activated. We calculated what we refer to as the t50, which is the expected time at which half of the nuclei in the core of a pattern are activated (See Figure 3.9A and 3.9B). With this model, we can computationally choose random embryos and then construct kinetic curves by ordering them (Figure 3.9C). Figure 3.9C shows that even with multiple different sampling, the generated activation curve is very similar to the original data and the fitted curve. We can go through this process of simulating activation profiles many hundreds of times and fit a GCDF to each of the computational curves generated this way (See Figure 3.10 for fit parameters). By comparing the different curves to each other, we can determine the uncertainty in the timing associated with a particular fraction of the pattern on and t50 specifically (Figure 3.9D). Since the simulations explicitly take into account the number of embryos used and the profile of the activation for each different experiment, we can obtain an uncertainty value unique to each set of data. This analysis shows that for almost all of the constructs this is less than 5 minutes (Figure 3.6).

3.12 Methods

Recombineering, Cloning and Transgenesis

Plasmid constructs containing various enhancers (pannier enhancer, sog enhancer and snail shadow enhancer) and promoters (pnr, tup, sna, sog, ths, Hsp70) were built using the pbPHi backbone vector [99, 47]. Primers used for construct building and recombineering are listed in Figure 4.7. All plasmids and BACs were integrated in the same landing site on chromosome 3 (VK33). The reverse transcription has been performed on RNA collected from 2-4h yw embryos.
Permanganate genomic footprinting
Permanganate footprinting on embryos was carried as described in [100], by adapting the volumes to small amount of DNA. LM-PCR reactions were performed as previously described [101]. All reactions started with 100ng of piperedine-cleaved DNA. The primers used are listed in Figure 4.7. Naked DNA stands for genomic DNA. The time of permanganate treatment is indicated in the Figure, 0, 30 and 60.

Pol II ChIP
Chromatin immunoprecipitations (ChIP) from approximately 1g of Toll10b, Pnre-tupPr-Y; gd7, TwiBac-PurPr-Y or gd7, TwiBac-Y mutant embryos were performed as described in [85]. A monoclonal antibody recognizing both the phosphorylated and the non-phosphorylated form of Pol II was used (CTD4H8, Millipore). Sequencing libraries were prepared from 10ng of immunoprecipitated DNA and 50ng input DNA following the Illuminas instructions. The primers used for ChIP-qPCR are listed in Figure 4.7.

ChIP-Seq Analysis
Paired-end sequenced reads for both the RNA Pol II ChIP and the whole-cell extract input control were aligned to the Drosophila reference genome (UCSC version dm3) with the addition of a pseudo-chromosome containing the sequence of the synthetic promoter and yellow gene from the BAC insert. Reads were uniquely aligned using Bowtie version 0.12.7 with the following parameters: -X 300 k 1 m 1 l 51 n 3. Successfully aligned paired reads were merged into single fragments and used to calculate read-count-normalized coverage over the entire genome.

Fluorescent In Situ Hybridization and Immuno-staining
Yellow white and transgenic embryos were developed at room temperature and were collected at 1-4 hours of development and were fixed as described in [71]. We used the in situ hybridization protocol from [71] with minor modifications. No Proteinase K was used and signal was not amplified using Tyramide amplification. All probes were made with either digoxigenin, biotin or dinitrophenyl conjugated haptens. The primary antibodies used to detect probes were sheep anti-digoxigenin, mouse anti-biotin and rabbit anti-dinitrophenyl (Roche Applied Sciences, Invitrogen). These were labeled with Alexa dyes using Alexa Fluor 555-donkey-anti-sheep, Alexa Fluor 488 donkey-anti-mouse and Alexa Fluor 647 donkey-anti-rabbit secondary antibodies (Invitrogen, Molecular Probes). Nuclei were counter-stained with either DAPI (Invitrogen) or DRAQ5 (Biostatus). All the embryos used for the initiation curves were co-stained with a probe labeling sog nascent transcripts. Sog has a distinctive expression pattern that allows unambiguous identification of the dorso-ventral orientation of
embryos during cc13 and 14. The primers that were used to amplify the DNA fragments that were used to make the different RNA probes are listed in Figure 4.7.

Confocal Imaging

Embryos were imaged on a Carl Zeiss LSM 700 Laser Scanning microscope, equipped with a motorized stage. For embryos used to examine activation kinetics 20-25 section z-stacks through the nuclear layer at 1/2 micron intervals were taken using a Plan-Apochromat 20x/0.8, WD=0.55 mm lens. They were taken at 2048x2048 resolution with 8 bit color depth. Embryos were roughly staged by quantitative analysis of nuclear density and by developmental morphology. We examined embryos from the beginning of cell cycle 13 through to the onset of cephalic furrow formation. For the initiation curves acquired during cc13 and cc14 all the embryos on a slide that were in the appropriate orientation to see the expression pattern of the enhancer of interest were imaged (i.e. dorsal view for the pur enhancer constructs and ventral view for the snail enhancer constructs). The orientation was judged by examining the expression pattern of sog nascent mRNA which was placed in a different channel to that of the transgene.

For the images acquired for the mRNA counting 50 section z-stacks through the outer layer of cells at 1/3 micron intervals were taken using a Plan-Apochromat 63x/1.40 oil lens. They were taken at 2048x2048 resolution at 16 bit color depth.

Image Analysis for Determining Fraction Active Nuclei

We wrote an automated image segmentation program in MatlabR2011 (MathWorks) to identify and count all stained nuclei and detect foci of nascent transcription. This was then used to determine what fraction of nuclei in a region of gene expression was active for many different embryos.

For the image stack containing the DNA counter stain, Z-stacks were projected into two-dimensional images by selecting the maximum intensity pixel in each stack. For the channels that contained the signal from the in situ probes each image plane was filtered by a Laplacian of Gaussian and then the stack was max projected to a two dimensional image. The core of each individual nucleus was determined using the DNA counter-stain, processed with a Laplacian of Gaussian filter, which allows robust determination of nuclei using blob detection, size selection, and signal strength to inform classification. This was followed by an object dilation algorithm to create a computational mask in which all pixels in an embryo are assigned to a uniquely identified nucleus.

The script then determines the transcriptional activity of nuclei by identifying which in situ signals from nascent mRNA are present in the region assigned to different nuclei. True hybridization foci or dots are identified through a series of filtering steps and segmentation.
The 2D projection is filtered by a Laplacian of Gaussian filter which is then segmented using a combination of a watershed algorithm (to split dots that are joined) and intensity based thresholding.

To address possible misclassification due to uncertainty in determining the nuclear boundary, transcripts localized to pixels adjacent to the nuclear boundary may be automatically reassigned to the neighboring nucleus if the original parent nucleus already contains an interior localized probe. This exploits the fact that the hetero or homo allelic expression of the reporter should result in no more than one or two foci per nucleus, respectively. An iterative extension of this algorithm also insured reliable classification when several adjacent nuclei each had transcripts that localized to boundary pixels.

To determine the fraction of inducible nuclei that are actively expressing a particular gene/enhancer, the subset of cells that make up the expression pattern needs to be defined. The enhancers used in this study are heavily studied and therefore the exact locations where they drive expression are well characterized. This allows the software user to manually select the region of expression based on the percentage of the embryo and spatial profile of other genes that serve as fiduciary markers. For our analysis we focus on a subset of the expression pattern that we define as the core region of expression. The core region of expression satisfies the following two criteria. First, nuclei are activated with no significant spatial bias in this region. Second, the expression pattern displays no spatial dynamics across cc14. For example the expression of snail is refined at the poles of the embryo during cc14 and so nuclei within 15%EL of the poles were not included in the core region for analysis of constructs with the snail enhancer. We specifically focused on the core of the pattern to avoid convoluting the dynamics of changes in the expression pattern of the gene with that of gene activation. The core region consists of many hundreds of nuclei and using this we can calculate the fraction of core pattern on. This is done for many embryos and forms the basis for determining the kinetics of activation described in later sections.
Figure 3.7: Minimal promoters are sufficient to perturb snail temporal coordination. The distal snaE was placed upstream of the snaPr (A), sogPr (B), and thsPr (C) promoters attached to the yellow reporter gene (see diagram in upper left). A-C Processed images after FISH using a yellow intronic probe. D Temporal coordination profiles during cc14. E High-resolution confocal image of yellow mRNAs encoded by the snaE > snaPr/yellow minigene. Arrowheads point to individual cytoplasmic mRNAs; arrow indicates nascent transcripts. F Bar graph showing the estimated promoter strength from the pnr, ths, and sna promoters just prior to gastrulation (see section 3.10 and Figure 3.8). Error bars represent SD. a.u., arbitrary units. G-I False-colored nuclei showing the presence of nascent transcripts for sna in the rescue BAC constructs containing the ths (G), sog (H), and sna (I) promoter.
Figure 3.8: Measuring and modeling mRNA levels. A Graph illustrating the activation curves measured for a given construct. B Ordinary Differential Equation (ODE) used to estimate the average number of mRNA molecules per cell as a function of time. C Curves for different means of mRNA counts obtained for the snaiE-snaPromoter-yellow minigene. These curves are obtained by using A and B. D Knowing that the promoter strength for snail is about 6 mRNA/min [16] and that the mean number of yellow full-length mRNA at the end of cc14 is about 100 molecules per nuclei, we determine the half-life of our reporter. E Knowing the number of mRNA molecules and the half-life of the yellow reporter, we determine the promoter strength for pnr and ths.
Figure 3.9: Determining t50 for the activation curves. A Graph showing the ordered fraction of the core pattern that exhibits nascent transcripts for different embryos. The total time of the relevant part of cc 14 is 60 min and so the axis represents this. Three different transgenic lines are represented: pnrE < snaPr (red), pnrE < sogPr (green) and pnrE < thsPr (blue). For each measured curve, we fit a cumulative gamma distribution (see supplementary methods for details), shown in black. From the fitted cumulative gamma function, we determine the t50 value, defined by the time it takes for an embryo to show nascent transcripts in 50% of the core pattern. B Similarly to A, this graph shows the measured and fitted activation curves to determine the t50 values for transgenic embryos with the snail enhancer. C Simulated activation curves plotted with actual curve to illustrate that for large numbers of embryos they compare favorably (see supplementary methods for details). D Graph illustrating how the t50 uncertainties are obtained. The uncertainty is the standard deviation in the difference between the measured t50 and the estimated t50 for hundreds of simulations. All t50s and their estimated uncertainties are summarized in Figure 3.6 (see also Figure 3.10.)
## Cumulative Gamma Function Fit Parameters

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Gamma Shape Parameter (a)</th>
<th>Unitless Gamma Scale Parameter (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PnrPr Endo</td>
<td>68.41</td>
<td>9.97E-03</td>
</tr>
<tr>
<td>TupPr Endo</td>
<td>147.35</td>
<td>2.94E-03</td>
</tr>
<tr>
<td>Tup-Y BAC</td>
<td>13.06</td>
<td>4.17E-02</td>
</tr>
<tr>
<td>Tup-Y-PnrPr BAC</td>
<td>17.05</td>
<td>5.26E-02</td>
</tr>
<tr>
<td>PnrE-TupPr NelfE/Spt5</td>
<td>15.48</td>
<td>4.25E-02</td>
</tr>
<tr>
<td>PnrE-TupPrShort</td>
<td>6.47</td>
<td>8.49E-02</td>
</tr>
<tr>
<td>PnrE-TupPr Trl</td>
<td>6.38</td>
<td>8.19E-02</td>
</tr>
<tr>
<td>PnrE-TupPr</td>
<td>3.62</td>
<td>7.55E-02</td>
</tr>
<tr>
<td>PnrE-SnaPr</td>
<td>3.10</td>
<td>1.46E-01</td>
</tr>
<tr>
<td>PnrE-Hsp70Pr</td>
<td>3.66</td>
<td>1.40E-01</td>
</tr>
<tr>
<td>PnrE-SogPr</td>
<td>6.18</td>
<td>1.07E-01</td>
</tr>
<tr>
<td>PnrE-ThsPr</td>
<td>7.14</td>
<td>1.33E-01</td>
</tr>
<tr>
<td>PnrE-PnrPr</td>
<td>7.24</td>
<td>1.79E-01</td>
</tr>
<tr>
<td>SnaE-SnaPr</td>
<td>0.35</td>
<td>5.27E-01</td>
</tr>
<tr>
<td>SnaE-SogPr</td>
<td>0.61</td>
<td>1.14E+00</td>
</tr>
<tr>
<td>SnaE-ThsPr</td>
<td>1.92</td>
<td>4.24E-01</td>
</tr>
<tr>
<td>SnaE-PnrPr</td>
<td>2.85</td>
<td>2.77E-01</td>
</tr>
<tr>
<td>SogE-TupPr</td>
<td>1.54</td>
<td>6.14E-02</td>
</tr>
<tr>
<td>SogE-SnaPr</td>
<td>1.68</td>
<td>5.35E-02</td>
</tr>
<tr>
<td>SogE-PnrPr</td>
<td>2.40</td>
<td>1.13E+00</td>
</tr>
</tbody>
</table>

**Figure 3.10:** Cumulative gamma fit parameters
Chapter 4

Importance of transcriptional coordination in development

4.1 Chapter Summary

To determine whether the time to synchrony is important in development, we manipulated the coordinate expression of snail (sna) (Slug/Sna2 in vertebrates), a major determinant of epithelial-mesenchyme transitions (EMT) in animal development. There is also evidence that snail may be directly or indirectly involved in regulating its own expression, which would potentially make it sensitive to small differences in coordination between cells. In Drosophila, snail is expressed in ~1000 cells comprising the presumptive mesoderm [33, 102]. These cells undergo coordinated invagination during gastrulation, within 90 min of the onset of snail expression [103, 104]. To determine whether synchronous activation of snail expression is essential for coordinated invagination of the mesoderm, we replaced the native snail promoter with those from moderately paused (short gastrulation; Chordin) or nonpaused (thisbe; Fgf8) genes. These heterologous promoters result in less synchronous patterns of snail activation, and a progressive reduction in mesoderm invagination during gastrulation. Computational modeling of the dorsal-ventral patterning network recapitulates these variable and bistable gastrulation profiles and emphasizes the importance of timing of gene activation in development. We therefore conclude that paused Pol II and transcriptional synchrony are essential for coordinating cell behavior during morphogenesis.

4.2 Transcriptional synchrony is essential for coordinate invagination

A 25 kb snail BAC transgene encompassing the snail transcription unit, proximal enhancer, and neighboring Tim17B2 locus (which harbors the distal snail shadow enhancer) was shown to be sufficient to rescue the gastrulation defects of sna-/sna- mutant embryos [93, 47]. How-
ever, there is evidence that the proximal enhancer might attenuate snail expression by impeding access of the distal enhancer to the sna promoter [93] (data not shown). Consequently, we removed this enhancer in order to obtain a more direct assessment of the contributions of the different promoters in coordinating mesoderm invagination. It is important to note that the distal enhancer is sufficient for complete rescue of the gastrulation defects of sna-/sna-mutant embryos and the development of fully viable adult flies [93].

We employed recombineering methods to create a series of snail BAC transgenes that contain either sog or ths promoter sequences in place of the native sna promoter (replacement of 100-110 bp). The three BAC transgenes (native sna promoter, sog promoter, or ths promoter) exhibit distinctive patterns of activation during the onset of cc14 (Figure 3.7G-I), concomitant with the levels of paused Pol II and the t50 synchrony values seen for the snail minigenes (Figure 3.7A-D).

There is a tight correlation between these activation profiles and the extent to which mesoderm invagination is rescued in sna-/sna- embryos (Figure 4.1). Thus, the native transgene containing the strongly paused sna promoter mediates a coordinated ventral furrow and robust invagination of the mesoderm (Figure 4.1A-D). In contrast, the transgene containing the moderately paused sog promoter produces truncated furrows (Figure 4.1E-H) (n=18/22), but occasionally induces nearly complete furrows approaching those seen in wild-type embryos (Figure 4.1G,G’) (n=4/22). Finally, the nonpaused ths promoter produces highly variable phenotypes, ranging from the complete absence of invagination (n=11/16), to erratic pockets of ingressing cells (n=3/16) (Figure 4.1I,J), and rarely, extended grooves of invaginating cells (n=2/16) (Figure 4.1K,K’). Snail mRNAs and protein are detected only in the invaginating cells of partially rescued embryos exhibiting truncated furrows or isolated pockets of ingestion (Figure 4.1D,H,L). These studies suggest that the time to synchrony is a critical determinant of coordinate cell behavior in development (see below).

4.3 Computational Models of Gastrulation Variability

We constructed a mathematical model (Figure 4.2A–C) to explain the highly variable gastrulation phenotypes seen for the sna BAC transgenes containing the sog and ths promoters. Bistability of snail expression is often observed in the anterior third of the embryo encompassing ~ 300 of the ~ 1,000 cells comprising the ventral furrow (e.g., Figure 4.1E’,F’,K’). This model makes use of the wealth of knowledge about the transcription networks governing the dorsal-ventral patterning of the Drosophila embryo (reviewed by [105]). It also draws on recent dynamic imaging of the Dorsal nuclear gradient [106, 107, 108, 109], which revealed a slight narrowing of the gradient in anterior regions. Our model also invokes snail autoregulation, which is suggested by the rapid loss of snail transcripts [110] and yellow transcripts from a sna BAC transgene (Figure 4.4) in sna-/sna- null embryos. We explored different
mechanisms of autoregulation and obtained the most faithful results with an indirect model, whereby Snail represses a localized ectodermal repressor via a double negative feedback loop.
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

(see below and Figure 4.4).

Figure 4.2: Modeling gastrulation variability: The importance of coordination.

A Mesodermal region of a DAPI-stained embryo to show the segmentation process of the nuclei. The panel below is a schematic illustrating the neighbors \((j)\) of a given mesodermal nucleus \((i)\). We allow for nearest neighbor diffusion, where the \(i\) nucleus is diffusively coupled to its nearest neighbors that share a boundary \((j, 1:6\) in this case). B Simplified mathematical model for Snail dynamic expression in a given nucleus \((i)\). The key parameters are the timing of sna activation in the particular nuclei, the concentration of the neurogenic repressor \((Rep)\), and the concentration of activators like Dorsal \((k_1)\), number of nearest neighbors \((NN)\), and the strength of the diffusive coupling between nuclei D. C Activation curves computationally obtained for three different promoters: sna, sog, and ths. D to F Results of computational simulations when sna temporal coordination is affected; t50 values are indicated.

Computational simulations consistently produce uniform ventral furrows when the onset of sna expression is rapid and uniform \((t50 = 0-4\) min after the onset of cc14\) (Figure 4.2C,D), as seen for the native snail promoter (Figure 3.7I). However, slightly less coordinated patterns of activation \((t50 = 4-7\) min\), e.g., sog promoter (Figure 3.7H), produce highly variable ventral furrows (Figure 4.2E). As the coordination is further reduced \((t50 = 7-12\) min\) (Figure 4.2F), most simulations show a complete loss of the furrow, although a small fraction of simulations produce half furrows in the posterior half of the embryo, as
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

seen for the ths promoter (Figure 3.7G).

The bistable, all or none invagination of the anterior mesoderm can be explained by the combination of reduced levels of the Dorsal gradient, variable activation of Snail expression and delayed synthesis of critical threshold levels of the Snail repressor (Figure 4.2E,F). The key insight from these simulations is that short-range diffusion among neighboring nuclei can produce sufficient levels of Snail repressor to rescue small but not large patches of snail-expressing nuclei. The decision to maintain or repress snail expression occurs during a very tight time window, 10-20 min after the onset of gastrulation. The key parameter underlying bistability is the time to synchrony since similar results are obtained when computer simulations are performed with a range of Hill coefficients for sna regulation and different diffusion rates for the Snail protein (see Figure 4.6).

4.4 Dynamic control of the dorsal-ventral patterning network

The gene regulatory network underlying the spatial control of dorsal-ventral patterning has been extensively studied (reviewed by [105]). Considerably less is known about the temporal dynamics of this process. Indeed, developmental timing has only recently become a critical focus of study, even in well-defined systems such as the patterning of the vertebrate neural tube [36]. Here we have shown that perturbing coordinate activation of the snail expression pattern leads to various invagination defects during gastrulation. Computational modeling (Figure 4.2) highlights the importance of timing in producing these defects. Delayed and asynchronous patterns of activation uncouples Snail from the other components of the dorsal-ventral patterning network, resulting in variable gaps and bistability of the ventral furrow, particularly in the anterior mesoderm. Indeed, this uncoupling results in the expression of high levels of the Dorsal and Twist activators in regions that fail to invaginate due to the delay in Snail expression (e.g., Figure 4.1 H,L). This uncoupling of sna expression from its activators provides a vivid illustration of the importance of temporal dynamics in the control of complex developmental processes. A static gene network based on a simple Dorsal gradient affinity model does not appear to be sufficient to capture the intricacies of mesoderm morphogenesis.

The dorsal-ventral patterning network amplifies small changes in the levels of the dynamic Dorsal gradient to produce all or none patterns of sna expression. We believe that the key agent of this all or none, bistable expression of Snail is the anti-repression of competitive ectodermal repressors [110]. This indirect mechanism of Snail autoregulation may be the basis for producing the unusually sharp border of Snail expression at the boundary between the mesoderm and neurogenic ectoderm. This border determines whether cells become fully
committed to EMT at gastrulation. Delays in coordinate sna expression are amplified by the dorsal-ventral patterning network to produce bistable gaps in the ventral furrow, particularly in anterior regions where there are slightly diminished levels of Dorsal nuclear transport.

4.5 Spectrum of pausing and cell fate decisions

Our results indicate that the continuum of Pol II pausing seen for different promoters leads to a continuum of temporal coordination in gene activation, spanning from highly stochastic to synchronous. As discussed above, synchronous activation of transcription is essential for coordinating mesoderm invagination, however the stochastic regulation of gene expression is sometimes used to provide flexibility in cell fate specification within a tissue [111, 112]. For example, stochastic specification mechanisms underly fate decisions in the Drosophila eye and human immune system [113, 112], whereby cells must adopt alternate fates to achieve a distribution of distinct functions. For example, there is a 70:30 distribution of alternative ommatidial identities in the eyes of higher diptera that has been conserved for $\sim 120$ MYA [112]. Similarly, flexibility in the behavior of B lymphocytes is important for immune regulation [113]. Modulating the levels of paused Pol II could help tune the proportion of cells that adopt different fates through such stochastic specification mechanisms.

In summary, we have presented evidence that the promoter is a key agent for coordinating gene expression in the different cells of an embryonic tissue. Minimal promoter sequences are sufficient to establish paused Pol II and mediate synchronous patterns of gene expression. There is a tight correlation among the levels of paused Pol II, the time to synchrony, and the coordination of mesoderm invagination. We therefore propose that promoters ensure exquisite control of the complex cellular processes underlying morphogenesis.

4.6 Methods

4.7 Fly Genetics

The genetic procedure used for the sna BAC rescue experiments was performed as described previously by [47]. The snail deficiency used in this study also removes some neighboring genes (B1 3078). In addition to genotyping the embryos with a lacZ from a labeled balancer (hb-lacZ), we use a probe for the escargo (esg) gene, which next to snail and is expressed during cc14 in the ventral ectoderm. The snail deficiency we used also deletes this escargo gene, but the latter is not present in the 25kb snail BAC used in this study.
4.8 Modeling Evolution of Snail Protein

In order to understand the phenotypes we observed when we changed promoters we developed a model to describe the evolution of the Snail protein expression pattern in the early embryo. We sought to simplify the system as much as possible while keeping it realistic enough to incorporate the wealth of quantitative information available in the early Drosophila embryo. For example, the spatial profile and dynamic behavior of the Dorsal gradient that specifies Snail has been well characterized recently [109, 106, 107]. Estimates are also available for the timescales of diffusion, protein synthesis and protein half-lives and molecular binding constants [114, 115, 116, 117, 118, 119, 120]. Our goal was to see what would happen when you perturb the timing and coordination of Snail transcriptional activation in the context of a model based on what is known about the early embryo and some reasonable estimates for parameters that are not known. We do not claim that our model is completely comprehensive, rather it serves as a tool to judge how differences in kinetics of coordination may lead to changes in the Snail expression pattern and then coordination of morphogenesis.

We represent the embryo as a two dimensional grid of nuclei. The nuclear grid used for the modeling was obtained by segmenting a fluorescent confocal stack of a cc14 embryo (Figure 4.2). This yields a pseudohexagonal grid consisting of about 2000 nuclei where the location and connectivity of all the nuclei are known. In our model each of these nuclei are considered a unit. The evolution of the concentration of Snail in an individual nucleus is described by a differential equation. This differential equation has terms that describe synthesis rates that depend on the regulatory network, first order decay and diffusion between neighboring nuclei (see below).

The different kinetics of activation for the different promoters is included in the evolution of the model as follows. We assume that there is a delay between the start of cc14 and when each nucleus can start producing protein. This delay is modeled as a random variable that follows a gamma distribution and the parameters that define this distribution are different for different promoters. In the model we assume that to a good approximation there exists a linear correlation between the amount of snail mRNA and protein. This has been illustrated in the literature [121] and may be due to the short half-lives of both mRNA and protein in the early drosophila embryo [116]. We use the t50 value defined earlier to characterize the different timing of activation for the different constructs. All other parameters are kept the same. How the Snail concentration changes with time in each nucleus can then be obtained by numerically solving the differential equation given some initial conditions. The next section describes the relevant components of the regulatory network.
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

Regulatory Network

Dorsal

The key morphogen that patterns the dorso-ventral axis of the drosophila embryo is the transcription factor Dorsal [122]. The nuclear concentration of the Dorsal protein forms a gradient that is high in ventral nuclei and low in dorsal regions. The spatial profile of this gradient has been well characterized by a number of groups and has been shown to follow a Gaussian function [109, 108, 106, 107, 123]. Through a combination of modeling and experiments it has also been shown that the Dorsal gradient is dynamic. At a given anterior-posterior position the shape and width of the gradient is static but the amplitude increases with time [109, 108, 107]. It has also been shown that the width of the gradient changes significantly along the anterior posterior axis as shown in Figure 4.3 [107]. There has been some disagreement over the exact width of the gradient [109, 108, 106, 107, 123], but this is largely a result of different groups measuring the profile of the gradient at different anterior-posterior positions along the embryo in earlier studies. Due to the careful characterization of the spatial and temporal profile of the gradient we can build both of these features into our model to predict the evolution of the Snail expression pattern.

Snail

Through a combination of classic enhancer mutagenesis experiments it was shown that Dorsal acts as an activator that specifies the domain of Snail expression through a standard affinity threshold model [125]. It activates transcription in concert with another transcriptional activator Twist whose expression pattern is also defined by Dorsal. Twist plays a permissive role in the specification of Snail and because it follows Dorsal we have not explicitly included it in our model. Instead we assume that considering Dorsal as the major activator captures most of the important aspects of activation. Hence according to what is known about the specification of Snail so long as the Dorsal and Twist gradients are intact the domain of Snail specification should remain unchanged.

There has however been circumstantial evidence that Snail may play a role in defining its own expression pattern [110, 126]. Embryos that are homozygous for the strongest snail loss of function allele show normal patterns of snail mRNA during cc13 but that mRNA is lost in the presumptive mesoderm during cc14. The Snail[18] allele produces mRNA and protein but the lesion was not characterized [110]. To better understand the role that Snail is playing in shaping its own mRNA expression we looked at reporter expression driven by snail enhancers in embryos where there is no snail protein. We used a 25kb reporter BAC which contains all of the known snail regulatory sequences relevant for early expression where the snail coding region has been replaced with that of the yellow reporter [47]. We looked at the expression of this reporter BAC in embryos that were homozygous for a snail deficiency.
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

Figure 4.3: Spatial Profile of the Dorsal Gradient. A Relative width of the Dorsal gradient as measured at different positions along the embryo (taken directly from [106]). B Mean number of nuclei in a 1.5% EL slice of the embryo running from anterior to posterior for 42 embryos downloaded from the Berkeley Drosophila Group Database [124]. C By combining the data shown in A with that in B one can determine the width the Dorsal gradient in terms of number of cells shown in which is used in the simulations.

Using this approach we saw clear evidence for snail auto regulation during the course of cc14 (Fig 4.4). At the onset of cc 14 the expression of yellow nascent dots looks to that of a control embryo where there is snail present. However at the end of cc14 the reporter expression is severely compromised and absent through most of the presumptive mesoderm. This is despite the fact that both Dorsal and Twist expression are normal in their embryos.

The key role that Snail plays in specifying its own expression pattern is also clear from what we observe in the rescue constructs (Fig. 4.1). We observe the complete loss of snail protein and mRNA in variable regions of the presumptive mesoderm. In the rescue constructs the spatial pattern of snail expression is changed. This is quite different to what
we observe for the pattern defined by the reporter genes with the snail enhancer (Fig. 3.7). Here nuclei across the whole presumptive mesoderm showed expression. The constructs with different promoters showed different kinetics of filling out the domain but we see no evidence for any spatial difference in the expression pattern.

**Mutual Repression**

The previous experimental evidence indicates that Snail protein is either directly or indirectly involved in regulating its own expression in a positive way. When snail protein expression is compromised the snail expression pattern is compromised. Snail has been shown to be a potent transcriptional repressor and there is no evidence for it being able to act as a transcriptional activator. However we attempted to model the feedback as direct by enabling snail to activate its own transcription but the simulations from this model were unable to reproduce the phenotypes we observed. Hence it is most likely that snail’s positive regulation on itself is mediated indirectly, by having it repress a transcriptional repressor that can repress snail. There is evidence for mutual repression between snail homologs and other ectodermal repressors [127]. From the domain of snail expression and the timing of the loss of snail it is most likely that the repressor is of an ectodermal origin and so would be activated by lower threshold concentrations of the Dorsal protein. By invoking such a mutually repressive interaction we were able to explain the dependence of snail transcription
on its own expression and by building it into the model we could reproduce the observed patches of snail expression.

**Equations and Parameters**

In the following section we list all the equations that were used in building the model of the DV gene network to look at the evolution of snail protein. This formalizes the concepts that were discussed in the earlier sections. Not all the parameters that the equations depend on have been measured and so some needed to be estimated from values that are available for similar processes. Moreover the approach with the modeling was not to fully constrain the model based on our observations but rather to see how the model would respond to difference in the coordination of transcription with some reasonable estimates of parameters that are not know. Equation 4.1 and 4.2 describe the temporal evolution of the snail ($Sna$) and repressor ($Rep$) concentration in the $i$th nucleus which has a particular position along the anterior-posterior axis ($y_i$) and dorso-ventral axis ($x_i$). All simulations were evolved forward for a period of 60 minutes representing the time from the onset of cc14 to when the snail pattern stabilizes and the first signs of gastrulation appear.

\[
\frac{d[Sna_i]}{dt} = \frac{k_{1i}(x_i, y_i, t)}{[Rep_i/K_R]^{n_1} + 1} - \lambda_1[Sna_i] + D(t) \sum_{j=1}^{NN} \frac{[Sna_j - Sna_i]}{NN}
\]

(4.1)

\[
\frac{d[Rep_i]}{dt} = \frac{k_{2i}(x_i, y_i, t)}{[Sna_i/K_S]^{n_2} + 1} - \lambda_2[Rep_i] + D(t) \sum_{j=1}^{NN} \frac{[Rep_j - Rep_i]}{NN}
\]

(4.2)

\[
D(t) = \begin{cases} 
D_0 \left[1 - \frac{t}{30}\right], & t \leq 30; \\
\frac{2D_0}{3} \left[1 - \frac{t-30}{30}\right], & 30 < t \leq 60. 
\end{cases}
\]

(4.3)

$NN$ is the number of nearest neighbor nuclei. $D_0$ is the diffusion constant between nuclei and was set to be equal to $3/\text{min}$ (unitless because its diffusion between nuclei) this number was based on the geometry of nuclei and the value of $\sim 1 \mu\text{m}^2/\text{s}$ which is measured for the Bcd protein in the early embryo [114, 115]. The strength of diffusion between nuclei was decreased with time during cc14 according to the dynamics of the slow and fast phases of membrane invagination according to equation 4.3. The half-lives of transcription factors in the drosophila embryo have been estimated to be between 5 to 30 min [116]. The snail protein domain refines rapidly at the poles when it is repressed during cc14 and so we reasoned that the half-life would be close to the lower end of this range. A half-life of 7 minutes was used in our simulations for the half-life of both snail and the repressor. We assumed binding affinities ($KR$ and $KS$) on the order of $\sim 10 \text{nM}$ based on the typical range seen for transcription factors in the early embryo ([116], and references therein). We speculate that there is likely to be some degree of cooperativity in binding and so chose values for $n_1$ and $n_2$ to be equal to 5. The role of Dorsal and differing initiation kinetics comes in through the
synthesis terms and so these are explicitly defined and explained in the next section.

Equation 4.4 describes the snail protein synthesis term in the ith nucleus and how it implicitly depends on space and time. Much of the implicit dependence on space and time comes through the spatial and temporal dependence of the Dorsal gradient itself. The absolute concentration of Dorsal has not been measured but because we know the relative value where the snail boundary occurs when the gradient stabilizes we can estimate the relative KDs value to be equal to 0.5 [107]. The dependence of the dorsal gradient on space and time will be discussed below. Some degree of cooperativity was assumed and so $n_3$ was chosen to be 5.

$$k_{1i}(x_i, y_i, t) = \frac{r_s}{K_{DS}^{Dorsal(x_i, y_i, t)} + 1} \times H(t - TA_i)$$ (4.4)

$$TA_i \sim \Gamma(a, b) \equiv Gamma(a, b)$$ (4.5)

The different promoter kinetics is included in the second half of the equation 4.4 and through equation 4.5. The term representing dorsal mediated activation is multiplied by a Heaviside step function that effectively allows protein only to be produced after a given time $TA_i$. This time is a random variable that follows a gamma distribution characterized by the variables a and b and is the source of the variability seen in our simulations. This is how we incorporate the delay in the onset of transcription into the model that looks at the evolution of the snail protein. For our analysis we fixed a to be equal to one and examined how the behavior of the system changed as b was increased which caused a corresponding increase in the value of $t_50$. We assumed a synthesis rate ($r_s$) of $\sim 1 \text{ nM/min}$ which is well within the realms of what has been estimated for the early embryo ([116], and references therein).

The synthesis term for the repressor is similarly defined in equation 4.6 except it doesn’t have timing of activation component (we assume is comes on quickly). Due to the fact that we consider it to be a neurogenic repressor the sensitivity to dorsal must be significantly higher than snail and so we set the relative KDs value to be equal to 0.1. The synthesis rate was set to ($r_r$) 0.3 nM/min. Some degree of cooperativity was assumed and so $n_4$ was chosen to be 5.

$$k_{2i}(x_i, y_i, t) = \frac{r_r}{K_{Dr}^{Dorsal(x_i, y_i, t)} + 1}$$ (4.6)

The final part of the model takes the spatial and temporal profile of the dorsal gradient into account and is shown in equation 4.7. The dynamics of the gradient were imaged live recently [109] and this revealed that the values for the parameters B, A and $\tau_D$ are 0.2, 0.8 and $\sim 5 \text{ min}$, respectively. How the profile of the Dorsal gradient changes as a function of anterior-posterior position was taken from measured data Fig 4.3, [107]. Nuclei in the core snail domain were initialized with $\sim 1 \text{ nM}$ of snail protein that was normally distributed with
a variance of 0.1 nM. Reducing this variance did not lead to significant differences in the simulation outcomes.

\[
Dorsal(x_i, y_i, t) = B + A \left(1 - \exp \left(\frac{t}{\tau_D}\right)\right) \times \exp \left(- \frac{x_i^2}{2\sigma(y_i)^2}\right)
\]  

(4.7)

4.9 Model Behavior

When we examined the behavior of the model with the parameter values listed above, we observed four qualitatively distinct classes of behavior as the t50 or time to synchrony was varied (see Figure 4.5). One of the main features that distinguish the four classes is the degree of variability across simulations. This variability is a consequence of the inbuilt random nature of the activation time of individual cells, which depending on the degree of coordination is either suppressed or enhanced as the system evolves.

Class 1 occurs for rapid t50s from 0 to 4 min, class 2 for moderate t50s from 4 to 7 min, class 3 for slow t50s from 7 to 12 min and class 4 for t50s greater than 12 min. When the time to activation is rapid, the behavior of the Snail expression pattern was consistent for many different simulation runs which represent individual embryos. All the simulation runs finished with embryos where the Snail domain specified was completely intact and encompassing the entire presumptive mesoderm (See Figure 4.5A & Figure 4.5A'). Class 1 which includes t50 values from 0 to 4 mins most closely resembles what we observe for rescues with the snail promoter where the domain of Snail expression consistently encompassed the entire mesoderm.

Class 2 is characterized by the onset of variability in the spatial distribution of Snail expression. As t50 is increased, a small proportion of simulations start showing gaps in the snail expression pattern in the anterior region of the embryo (See Figure 4.5B & Figure 4.5B'). As t50 is further increased the fraction of embryos showing these gaps grows and simulation runs start to appear where Snail is completely lost in the anterior third of the Drosophila embryo (See Figure 4.5C & Figure 4.5C'). The variability seen is also sufficiently large that rarely there are simulations that converge to a complete loss of snail expression. Class 2 encompasses t50 values from 4 to 7 min and is most similar to what we observe for rescues with the sog promoter. Class 3 is characterized by the consistent absence of Snail expression in the anterior third of the embryo. In this class simulations that show absolutely no snail expression become much more common and when there the snail domain is restricted to the posterior regions of the embryo (See Figure 4.5D & Figure 4.5D'). Class 3 encompasses t50 values from 7 to 12 min and is most similar to what we observe for the rescue constructs with the ths promoter. Class 4 is characterized by the complete absence of a stable region of snail expression. As t50 is increased beyond 12 mins out of hundreds of simulation runs none produce stable regions of snail expression.
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

4.10 Sensitivity Analysis of Key Parameters

The behavior of the model depends on a number of parameters but for simplicity we focus on describing the sensitivity to several of the key parameters, namely the diffusion strength, degree of cooperativity and protein half-life. In examining the behavior of the model as these parameters were varied, we saw a similar trend: as the parameters are changed the quantitative behavior of the model changes as a function of different activation kinetics, but the qualitative behavior of the model is robust, i.e. the 4 regimes described previously occur for different ranges of t50 values but they nonetheless persist (Figure 4.5 & Figure 4.6).

One of the key parameters in the model is the degree of cooperativity (n). In the simulations we chose a value for the strength of cooperativity of 5 (Figure 4.5). The qualitative behavior of the model is robust when the degree of cooperativity falls within the range of 3 to 7. When the degree of cooperativity is reduced, the onset in variable Snail expression patterns is shifted to larger values of t50. Specifically when the degree of cooperativity is set to 3, the onset of variable Snail expression patterns occurs at a t50 value of 9 min as compared to 4 min for a degree of cooperativity of 5. This shift is illustrated by comparing Figure 4.6A, which shows an embryo histogram when t50 is 5.4 min and n=3, and Figure 4.5B, which shows an embryo histogram when t50 is 5.4 min and n=5. For the simulation with the higher cooperativity and equivalent t50, the Snail expression is variable while for the simulation with lower cooperativity there is no significant variability in Snail expression. Figure 4.6B shows that as the cooperativity is increased, the onset of variable expression occurs for smaller values of t50. When comparing Figure 4.6B with Figure 4.5B, one can see that the Snail pattern is far more variable at the same t50 value when the degree of cooperativity is increased from 5 to 7.

The strength of diffusion plays a key role in the model of Snail evolution. As the strength of diffusion is varied the t50 value at which variability in Snail expression starts to occur shifts, but the qualitative behavior remains the same. Figure 4.5B, 4.6C and 4.6D show that when the strength of diffusion (D) is reduced from 3 to 1, the essentially invariant Snail expression profiles persist for larger values of t50 but Snail expression becomes variable as t50 is further increased. Figure 4.5B, 4.6E and 4.6F show that when the strength of diffusion is increased from 3 to 6 the Snail expression profiles become variable at smaller values of t50. While the Snail profiles are essentially invariant for a t50 value of 1.7 min when this is increased to 5.4 min, the profiles become highly variable with most simulations failing to have stable Snail expression which is quite different to what is observed when the diffusion strength is 3 for the same t50 value as shown in Figure 4.5B.

The degradation rate of the Snail protein, which is inversely related to the protein half-life, plays an important role in setting the timescale of model evolution. In the main modeling section we chose a degradation rate of $\lambda_s$ of 0.1/min. Figure 4.5A, 4.6G and 4.6H show that as one increases the degradation rate of $\lambda_s$ to 0.15/min the range of t50 values for which stable snail expression is obtained shifts to smaller t50 values. Specifically, even though
the Snail expression domain is essentially invariant for a t50 value of 1.4 min, it becomes
significantly more variable as t50 is increased to 2.8 min. Figure 4.5A shows that when λₖ is
equal to 0.1/min, the Snail expression pattern is invariant even when t50 is equal to 3.4 min.
Figure 4.6I and 4.6J show that as one decreases the degradation rate of λₖ to 0.05/min, the
range of t50 values for which stable Snail expression is obtained shifts to larger t50 values.
To have a comparable degree of variability in the Snail expression pattern for a s value of
0.15/min at a t50 value of 2.8 min the t50 value needs to be increased to 13.9 min for a λₖ
value of 0.05/min.
Figure 4.5: Model behavior. For each shown t50 value, 200 independent simulations were conducted and the model behavior analyzed. The false colored embryos show how frequently a particular nucleus had snail expression at the end of the simulation (Dark red to dark blue, more frequent to less frequent). These histograms show how many embryos had a particular number of nuclei expressing Snail. A and A’ When t50 is only 3.4 min all the simulations converge to the same point, where both the number and spatial domain of Snail expression is the same for all runs. B and B’ As the t50 is increased to 5.4 min, one starts to see significant variability in the spatial profile and number of nuclei expressing Snail. C and C’ For a t50 value of 6.9 min, the variability is so large that some embryos show a full domain of Snail expression while others lose it completely. D and D’ As t50 is increased even further, the Snail pattern starts to break up significantly and often disappears entirely.
CHAPTER 4. IMPORTANCE OF TRANSCRIPTIONAL COORDINATION IN DEVELOPMENT

Figure 4.6: Model sensitivity analysis of key parameters. As in Figure 4.5, 200 independent simulations were conducted and the model behavior analyzed. The false colored embryos show how frequently a particular nucleus had Snail expression at the end of the simulation (Dark red to dark blue, more frequent to less frequent). The parameter values that are not given on each panel were assigned values that were used in the default simulation as shown in Figure 4.5. A and B Embryo histograms showing how changing the degree of cooperativity (n), affects model behavior for a given t50 value, (Figure 4.5 shows an equivalent simulation for n = 5). C and D Embryo histograms showing how reducing the diffusion constant D from 3 to 1 changes model behavior, (Figure 4.5 shows simulations for a diffusion constant value of 3). E and F Embryo histograms showing how increasing the diffusion constant from 3 to 6 affects model behavior (Figure 4.5 shows simulations for a diffusion constant value of 3). G and H Embryo histograms showing how increasing the degradation rate of the Snail protein (λ) from 0.1/min to 0.15/min changes model behavior (Figure 4.5 shows simulations for a degradation rate of 0.1/min). I and J Embryo histograms showing how decreasing the degradation rate of the Snail protein from 0.1/min to 0.05/min changes model behavior, (Figure 4.5 shows simulations for a degradation rate of 0.1/min).
Figure 4.7: Sequences of primers
Bibliography


BIBLIOGRAPHY


