TOPOLOGICAL REGULATION OF ENHANCER-PROMOTER COMMUNICATION

Tyler Matthew Heist

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY
THE PROGRAM IN
QUANTITATIVE AND COMPUTATIONAL BIOLOGY

Adviser: Dr. Michael Levine

September 2019
ABSTRACT

The activation of gene expression by transcriptional enhancers remains a central mystery in molecular biology. One prevailing view has been that enhancers loop directly to the promoters of genes they regulate in a targeted fashion. However, recent studies have suggested that enhancer-promoter interactions occur with the context of topologically associating domains, opening up the possibility for potential enhancer-promoter and enhancer-enhancer crosstalk within these domains. Further, it is still unknown at what distances (both genomic and physical) enhancers can interact with their target promoters to elicit transcription.

In this dissertation, I employ quantitative live-imaging methods to investigate enhancer function in living Drosophila embryos. In Chapter 1, I introduce what is known about transcriptional regulation, the cis-regulatory elements that mediate it, and the current models for enhancer function in this process. In Chapter 2, I describe our strategy to visualize the process of transvection, whereby enhancers located on one homolog activate transcription on the other homolog. Interestingly, a shared enhancer was found to produce coordinated transcription from linked reporter genes in trans. In Chapter 3, I describe work done to explore transcriptional regulation at the endogenous fushi tarazu (ftz) and even-skipped (eve) loci. Evidence for enhancer-enhancer communication was found upon deletion of the eve stripe 1 enhancer. In Chapter 4, I investigate the distances that exist between genes co-regulated by a single enhancer during transcription. Using the transvection assay, large distances of at least 100-200nm were found to separate an enhancer from its target promoter.
Finally, in Chapter 5, I argue that these observations are all consistent with the occurrence of transcription “hubs”, which trap the transcriptional machineries mediating gene expression. These “hubs” therefore allow for action at a distance from transcriptional activators binding and acting at enhancers to the recruitment and release of Pol II at promoters. Together, these findings and resultant models bring novel insights to our understanding transcriptional regulation as well as highlight future directions for research.
ACKNOWLEDGEMENTS

In my time as a PhD student, I’ve had the pleasure of working in Mike Levine’s lab. While it’d be unfair to say that working for Mike didn’t come with its struggles (e.g., the inevitable onslaught of emails when Mike’s and my time in lab on Saturdays didn’t quite line up), Mike has been a great mentor and incredibly supportive no matter what direction I wanted to take my research. Mike introduced me to some of the most interesting questions in modern molecular biology and taught me what it meant to be truly passionate about the work you do.

I must also thank the members of the Levine lab, past and present, for their support, patience, and friendship. Particularly, I must thank Takashi Fukaya and Bomyi Lim, two postdocs in the lab who not only interested me enough to want to join the lab to work on transcription with their amazing research, but who also taught me nearly everything I know as a scientist today. Takashi is, in my opinion, the world’s greatest molecular biologist and taught me all the molecular biology and fly husbandry basics I needed. Bomyi was the image analysis expert in the lab and I’m very grateful that she passed on some of that expertise, even with the many instances of me asking her how some function worked line-by-line. I must also thank Nicholas Treen and Laurence Lemaire for their unending support in answering dumb experimental questions as well as reading and providing comments on my dissertation.

One of the greatest things about doing fly work at Princeton is the community built around it. I was lucky enough to have Stas Shvartsman and Paul Schedl, two Drosophila experts,
on my committee; their combined knowledge and support helped move my research in the right direction. I am also grateful for the great support from all the staff in LSI. Evangelos Gatzogiannis in particular provided endless instruction and support when I was learning how to correctly image with the lab’s confocal and really helped me better understand how to produce the best quality imaging data.

I have also been lucky enough to find several like-minded people in Princeton who I’m happy to call my friends. Whether it’s playing nerdy games, pub trivia at Ivy, hitting up B-dubs, or tea time in LSI, I’m glad that I have had such a supportive group of people to spend my free time with and cheer me up even when science might have gotten me down. I’m particularly thankful to Kurt Ristroph, who bravely volunteered to proofread this dissertation in its more early and vulnerable forms.

I am also grateful to my family for their love and support over the past four years while I’ve been pursuing my PhD. Even though their eyes might have glazed over any time I tried to explain what I work on, I still appreciated being asked how my flies were doing on nearly every phone call.
TABLE OF CONTENTS

Abstract …………………………………………………………………………………………… iii

Acknowledgements ………………………………………………………………………………… v

List of Figures ……………………………………………………………………………………… ix

1. Introduction ……………………………………………………………………………………… 1
   Gene regulation in development ……………………………………………………………… 2
   Core promoters ………………………………………………………………………………… 3
   Enhancers ……………………………………………………………………………………… 5
   Insulators ……………………………………………………………………………………… 6
   Emerging models of transcription in the era of genome topology …………………………… 8
   References ……………………………………………………………………………………… 12
   Figures …………………………………………………………………………………………… 19

2. Visualization of transvection in living Drosophila embryos ……………………………… 20
   Abstract ……………………………………………………………………………………… 21
   Introduction …………………………………………………………………………………… 21
   Results ……………………………………………………………………………………….. 23
   Discussion ……………………………………………………………………………………... 29
   Materials and Methods ……………………………………………………………………… 33
   References ……………………………………………………………………………………… 49
   Figures …………………………………………………………………………………………… 55

3. Temporal dynamics of pair-rule stripes in living Drosophila embryos ………………… 69
   Abstract ……………………………………………………………………………………… 70
   Significance …………………………………………………………………………………… 70
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>A summary schematic of the function of cis-regulatory elements implicated in transcriptional regulation</td>
<td>19</td>
</tr>
<tr>
<td>2-1</td>
<td>A pair of <em>gypsy</em> insulators mediates transvection</td>
<td>55</td>
</tr>
<tr>
<td>2-2</td>
<td>Dynamics of cis- and trans-activation of lacZ transcription</td>
<td>56</td>
</tr>
<tr>
<td>2-3</td>
<td>Orientation-dependent and -independent insulator pairing</td>
<td>57</td>
</tr>
<tr>
<td>2-4</td>
<td>Specificity of insulator pairing in early embryos</td>
<td>58</td>
</tr>
<tr>
<td>2-5</td>
<td><em>gypsy</em> insulators foster stable homolog pairing</td>
<td>59</td>
</tr>
<tr>
<td>2-6</td>
<td>3D tracking of allelic distances for unpaired and paired homologs</td>
<td>60</td>
</tr>
<tr>
<td>2-7</td>
<td>A shared enhancer can co-activate two reporter genes <em>in cis</em> and <em>trans</em></td>
<td>61</td>
</tr>
<tr>
<td>2-8</td>
<td>Transvection involves homolog pairing</td>
<td>62</td>
</tr>
<tr>
<td>2-9</td>
<td>3D tracking of allelic distances during transvection</td>
<td>63</td>
</tr>
<tr>
<td>2-10</td>
<td>Transvection depends on sequence homology</td>
<td>64</td>
</tr>
<tr>
<td>2-11</td>
<td>Coordinated transcription across homologous chromosomes</td>
<td>65</td>
</tr>
<tr>
<td>2-12</td>
<td>Measurement of coordination between MS2 and PP7 reporters under the control of separate enhancers</td>
<td>66</td>
</tr>
<tr>
<td>2-13</td>
<td>Promoter competition attenuates the timing and the level of transvection</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1</td>
<td>Dynamics of endogenous <em>ftz</em> expression pattern</td>
<td>87</td>
</tr>
<tr>
<td>3-2</td>
<td><em>In situ</em> images of <em>yw, ftz-MS2</em>, and <em>eve-MS2</em> embryos</td>
<td>88</td>
</tr>
<tr>
<td>3-3</td>
<td>Snapshots from an embryo expressing <em>ftz-MS2</em></td>
<td>89</td>
</tr>
<tr>
<td>3-4</td>
<td>Dynamics of endogenous <em>ftz</em> expression</td>
<td>90</td>
</tr>
</tbody>
</table>
Figure 3-5. Dynamics of endogenous *eve* expression pattern 91
Figure 3-6. Anterior shift of *eve* expression patterns 92
Figure 3-7. Dynamics of endogenous *eve* expression 93
Figure 3-8. Relative dynamics of *ftz* and *eve* 94
Figure 3-9. Relative dynamics of endogenous *ftz* and *eve* 95
Figure 3-10. Dynamics of *eve* in the absence of the stripe 1 enhancer 96
Figure 4-1. Large distances separate co-regulated genes *in trans* 111
Figure 4-2. Distances separating MS2/PP7 foci in an interleaved reporter gene 112
Figure 4-3. Distances separating independently regulated MS2 or PP7 reporter genes on separate alleles 113
Figure 4-4. Distal and proximal configurations of two *cis*-linked reporter genes yield unique distance distributions 114
Figure 4-5. Distance distributions of two *cis*-linked reporter genes affected by genomic location 115
Figure 4-6. Distance dynamics between co-regulated genes during transvection 116
Figure 4-7. Distance trajectories for transvecting nuclei 117
Figure 4-8. A model for the distances separating co-regulated genes in different configurations 118
Chapter 1

Introduction
GENE REGULATION IN DEVELOPMENT

Developing multicellular organisms must rapidly select fates from a single genome to give rise to different cell types. Each of these cell types has a unique function and structure imparted by the specific subset of proteins and mRNAs that each cell expresses. This specification occurs via transcriptional regulation initiated first by maternally-deposited factors to pattern the embryo.

This preeminent role of gene regulation in controlling the mechanisms of early development has long been appreciated (Davidson, 1986). Precise regulation of transcriptional activity gives rise to several well-characterized gene expression patterns in development, such as the even-skipped (eve) stripes setting up embryo segmentation in *Drosophila melanogaster* (Small, Blair, and Levine, 1992; Bothma et al., 2014) or the goosecoid (gsc) domain establishing Spemann’s organizer in *Xenopus laevis* (Cho et al., 1991). Much work has been done to identify and characterize the cis-regulatory DNA elements that mediate the regulation of these intricate expression domains. These efforts have led to an increased understanding of how certain DNA elements contribute to transcriptional dynamics in development as well as to general cellular behavior.

The three major cis-regulatory DNA elements that have been extensively studied in the context of transcription and are of key importance to this dissertation are core promoters, enhancers, and insulators (Figure 1-1). Each of these elements play an important role in ensuring precise spatiotemporal expression of mRNAs in organisms while also restricting
the number of incorrect transcriptional initiation events. The history and function of these cis-regulatory elements are further described below.

Recently, new technologies have allowed for the characterization of these DNA elements in unprecedented detail. For example, MS2 and PP7 stem loops and their associated fluorescently-tagged coat proteins now permit visualization of nascent transcription in live embryos, allowing for precise quantification of transcriptional dynamics, such as transcriptional bursting (Garcia et al., 2013; Bothma et al., 2014; Fukaya, Lim, and Levine, 2016). The emergence of the CRISPR/Cas9 system in enabling precise genome editing has also allowed for endogenous investigations into the function of these cis-regulatory elements (Li et al., 2014; Bassett and Liu, 2014). These advances have further elucidated our understanding of how these elements act together to achieve spatial and temporal regulation of transcriptional activity and will be used heavily in this dissertation.

**CORE PROMOTERS**

The core promoter of a gene is an approximately 100bp DNA sequence that determines the precise position of the transcriptional start site (TSS) as well as the direction of transcription (Juven-Gershon and Kadonaga, 2010). This role of core promoters is ubiquitously used across all of life, from bacteria and even viruses to humans. In eukaryotes, the core promoter functions via recruitment of RNA Polymerase II (Pol II) and the general transcription factors, which include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIF, in a protein assembly called the pre-initiation complex (PIC) in a specific and stepwise process (Sainsbury et al., 2015). Only once the PIC is formed can transcription proceed, leading to productive elongation of the transcript via processive movement of Pol
II along the gene body (Schilbach et al., 2017; Vos et al., 2018). These transitions are also well correlated with key post-translational modifications to the RPB1 subunit of the Pol II protein, largely in its C-terminal domain (Phatnani and Greenleaf, 2006; Hsin and Manley, 2012). Importantly, this process alone results in only weak basal transcription.

The first hint that there was a conserved structure to the eukaryotic core promoter came from work done in the Hogness lab examining the *Drosophila* histone gene clusters (Lifton et al., 1978). They found that each of the histone genes had a common upstream AT-rich motif with the consensus sequence TATAAAT, which would later be called the TATA box (or Hogness box) and found to be a common feature of many eukaryotic promoters, situated 25-30bp away from the TSS. It would be later discovered that the TATA box is the direct binding target of the TATA binding protein, a subunit of TFIID (Starr and Hawley, 1991), and thus plays a crucial role in Pol II recruitment via PIC assembly. In addition to the TATA box, there are a number of other important conserved motifs common across promoters, including the downstream promoter element (DPE) and the initiator sequence (Butler and Kadonaga, 2002). However, it is important to note that not all core promoters contain these elements, despite being largely conserved across eukaryotes; these variations in the make-up and presence of these core sequences likely underlie the differences between transcriptional burst dynamics as well as between enhancer specificities seen across various promoters (Butler and Kadonaga, 2001; Zabidi et al., 2015; Tunnacliffe et al., 2018).
It has also been found that Pol II can be further regulated prior to productive elongation by stalling or pausing shortly after promoter escape (Adelman and Lis, 2012). This phenomenon was first discovered at the heat-shock gene locus, where paused Pol II would be rapidly released upon heat shock, allowing for an instant transcriptional response (Rougvie and Lis, 1988; Lis et al., 2000). It was later described that paused promoters allow for synchronous activation of multiple genes across the genome (Boettiger and Levine, 2009), which has important implications for tissue patterning and morphogenesis (Lagha et al., 2013). Genome-wide analyses also pointed out that differences in core consensus motifs likely contribute to whether or not Pol II will be regulated through promoter-proximal pausing (Chen et al., 2013).

**ENHANCERS**

Transcriptional enhancers are segments of DNA, usually 50bp-2kb in length, that are bound by transcription factors and act upon the core promoters of target genes (Levine, 2010). Transcription factors are proteins that contain both a DNA-binding domain that bind specific, short 6-12bp motifs as well as a transactivating domain that allows for the recruitment of specific transcriptional co-regulators (Spitz and Furlong, 2012). Enhancers were first discovered from the Simian virus 40 (SV40) by the Schaffner lab based upon their ability to increase the transcriptional output of a transgenic β-globin gene (Banerji, Rusconi, and Schaffner, 1981). An approximately 200bp sequence from SV40 was found capable of ‘enhancing’ transcription even when placed at a large distance (>1kb) from the core promoter and in an orientation-independent manner. These characteristics are still largely accepted as what defines a transcriptional enhancer (Schaffner, 2015).
Since their discovery, enhancers have been found to be a major factor in specifying domains of gene expression. Several studies have shown that changes in enhancer activity, both at an individual and population level, can lead to drastic phenotypic alterations, ranging from limb development to cancer risk (Lettice et al., 2002; Kvon et al., 2016; He et al., 2015). In developmental systems, enhancers have been largely characterized based on their ability to reproduce the expression patterns of particular genes in transgenic reporter assays (Levine, 2010; Kvon, 2015). Unlike core promoters, there has not been any evidence of conserved sequence motifs or structures across all enhancers; however, it has been found that enhancers are enriched for particular histone modifications, such as H3K4me1 and H3K27ac (Calo and Wysocka, 2013). It has also recently been estimated that the human genome contains nearly 400,000 enhancers, meaning that each human gene would be regulated by roughly 20 enhancers (ENCODE, 2012).

Enhancers have been found to be able to act at tremendous distances from their target promoters. For example, the ZRS enhancer of the Sonic hedgehog (Shh) gene, which patterns developing limbs, is located nearly 1 Mb away from the Shh promoter. In fact, it has been found that the median distance between enhancers and their target promoters is about 125kb in humans and approximately 10kb in Drosophila (Jin et al., 2013; Kvon et al., 2014). Despite these findings, it is still unclear how such long-range function via distal enhancers occurs.

**INSULATORS**

Insulator or boundary elements were first characterized in Drosophila near the heat shock locus by the Schedl lab based on their unique nucleoprotein structure as well as their ability
to restrict enhancers from acting on nearby promoters when situated between the two (Udvardy, Maine, and Schedl, 1985; Kellum and Schedl, 1991; Geyer and Corces, 1992). Since this early work describing the scs and scs’ insulator elements, insulators have been found to exhibit various different functions, from enhancer blocking (Geyer and Corces, 1992; West, Gaszner, and Felsenfeld, 2002) and chromatin organization (Ghirlando et al., 2012; Fujioka, Sun, and Jaynes, 2013) to topologically associating domain (TAD) formation (de Wit et al., 2015) and homolog pairing (Kravchenko et al., 2005; Fujioka et al., 2016), which are mediated by the aggregation of insulator-binding proteins. Insulator function is perhaps best understood in Drosophila (Negre et al., 2010; Maksimenko et al., 2015), where at least eight different insulator-binding proteins have been characterized. In mammals, however, canonical insulator function has only been observed with CTCF (Ong and Corces, 2014).

Insulators also play a key role in facilitating specific long-range enhancer-promoter interactions by bringing them into close physical proximity through insulator pairing, which ensures precise transcriptional regulation through stabilization of these interactions. Evidence for this function of insulators comes from knockdowns of CTCF protein as well as removal of CTCF sites in mouse Th2 cells, which results in an increase in cell-to-cell transcriptional noise (Ren et al., 2017). Further, Lupiáñez et al. (2015) have shown that disruptions in the boundaries of the TAD containing the Epha4 locus results in ectopic expression of surrounding genes, such as Pax3, Wnt6, and Ihh, in mice, yielding pups with limb deformations. This suggests that aberrations in insulator function likely underlie certain human diseases, and that the pathogenicity of these diseases can be caused by
dysregulation of enhancer-promoter communication. Recent work done by Fujioka et al. (2016) has laid out the architectural rules to insulator pairing using the homie and Nhomie insulators from the *eve* locus in *Drosophila*. These findings include the observation that insulator pairing in *cis* takes place via head-to-tail binding versus head-to-head binding in *trans* and, overall, provide a mechanistic model for how insulator binding occurs *in vivo* to allow physical interactions between enhancers and their promoters. These rules were recently tested when researchers were able to utilize a synthetic homie insulator element to drive expression of an enhancer-less lacZ reporter from the endogenous *eve* enhancers 142 kb away (Chen et al., 2018).

**EMERGING MODELS OF TRANSCRIPTION IN THE ERA OF GENOME TOPOLOGY**

As researchers have discovered more about the function of these cis-regulatory elements, there have been a number of models for transcriptional regulation (reviewed in Furlong and Levine, 2018). One of the perhaps most pervasive models over the past decade has been the looping model, whereby the enhancer loops stably and directly to the promoter to recruit Pol II and initiate transcription (Levine, Cattoglio, and Tjian, 2014). This model has gained increased traction as, through chromosomal conformation capture assays, such as Hi-C, it has recently become clear that transcriptional regulation takes place in a highly organized genomic context (Lieberman-Aiden et al., 2009). These assays have allowed for the discovery that the 3D structure of the genome is heavily structured into TADs that preferentially interact more often with genomic sequences found in the same TAD than with sequences found in other TADs (Dixon et al., 2012). In mammalian systems, CTCF
and cohesin have been found to be enriched at TAD boundaries and to play a functional role in TAD structure (Ong and Corces, 2014; Wutz et al., 2017).

It has been suggested that increased physical proximity allowed by TAD structure might contribute to enhancer-promoter contact by looping and subsequent transcriptional initiation. This is supported by studies from the Blobel lab where forcing loops between the LCR enhancer and its target B-globin promoter activates gene expression (Deng et al., 2012; Bartman et al., 2016). Recent work has also found that TADs act largely to restrict interactions between cis-regulatory elements found in neighboring, but distinct TADs. When cohesin is depleted from cells, novel enhancer-enhancer interactions happen at high frequency, further suggesting that TADs serve to restrict these sorts of contacts (Rao et al., 2017).

While Hi-C and other techniques have offered strong support for the looping model, there have also been several recent findings that have cast doubt on this model. The first was the realization that transcription occurs in discontinuous bursts (e.g. Raj et al., 2006; Bothma et al., 2014). It was also found that binding of certain TFs, such as Bicoid, at enhancers is incredibly transient, binding in some cases only on the scale of seconds (Mir et al., 2017; Donovan, 2018). A recent study using live-imaging methods in Drosophila embryos described how a single enhancer could coordinately and simultaneously activate two reporter genes (Fukaya, Lim, and Levine, 2016). Together, these findings suggest a level of dynamism to transcriptional regulation, particularly in Pol II recruitment, that isn’t explicit in the classical looping model.
These recent studies have instead led to the burgeoning idea of a transcriptional hub or transcriptional condensate (Hnisz et al., 2017; Furlong and Levine, 2018). This model differs from the looping model in that it proposes that transcription is carried out by large and dynamic pools of the transcription machinery formed between the enhancer and target promoters. While the looping model might suggest a tight physical distance between these cis-regulatory elements, the transcriptional hub model allows for much larger working distances between these elements thanks to the accumulation of Pol II and other activators. Further, the transcriptional hub model would support the dynamic behavior of Pol II recruitment that has been suggested by recent work as noted above. Much work has been done examining the viability of this model in explaining transcriptional regulation. For example, Pol II has been found to aggregate in dynamic clusters in living cells and move in these clusters to actively-transcribing loci (Cisse et al., 2013; Cho et al., 2016; Cho et al., 2018). Recent work has also described that particular TFs, such as Brd4 and Oct4, have the ability to phase-separate in vitro and appear to largely colocalize in vivo at active loci (Boija et al., 2018; Sabari et al., 2018). Together, these results provide strong evidence in favor of the transcriptional hub/condensate model.

However, there are still several outstanding questions in transcriptional regulation that, by addressing, will directly test some of the assumptions of the transcriptional hub model and give us a clearer picture of how cis-regulatory elements work. In this dissertation, I will address the following questions: How do enhancers contact their target promoters? What physical distances separate these cis-regulatory elements? What does this look like at an endogenous locus where a single promoter is regulated by multiple enhancers?
The early *Drosophila* embryo is an ideal experimental model system to address these questions. The phenomenon of transvection, the ability of an enhancer to interact in *trans* with a promoter on the homologous chromosome (Hopmann, Duncan, and Duncan, 1995; Southworth and Kennison, 2002; Ronshaugen and Levine, 2004; Mellert and Truman, 2012), in *Drosophila* provides a tractable method for understanding the mechanism behind long-range enhancer-promoter interactions. New technologies, such as the previously mentioned live imaging system with MS2/PP7 (Garcia et al., 2013; Bothma et al., 2014; Fukaya, Lim, and Levine, 2016) or the CRISPR/Cas9 system (Li et al., 2014; Bassett and Liu, 2014), have also been readily adapted by the *Drosophila* research community. These tools will allow me to study these cis-regulatory elements and how they contribute to the dynamics and precision of transcriptional regulation via transgenic assays as well as in their endogenous contexts.
REFERENCES


chromosomal domains. Cell 64, 941–950.
Kravchenko, E., Savitskaya, E., Kravchuk, O., Parshikov, A., Georgiev, P., and Savitsky,
M. (2005). Pairing between gypsy insulators facilitates the enhancer action in trans
enhancers in animals. Genomics 106, 185–192
Kvon, E. Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J.,
Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M., Schernhuber, K.,
Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of
Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J.,
(2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial
Biol 20, R754–R763.
Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward:


Figure 1-1. A summary schematic of the function of cis-regulatory elements implicated in transcriptional regulation.
Enhancers act at a distance on the promoters of target genes. Insulators interact with other insulators, participating in the formation of the 3D genomic organization.
Chapter 2

Visualization of transvection in living *Drosophila* embryos

Author Statement:

The work in Chapter 2 was previously published in *Molecular Cell* (Lim*, Heist*, et al., 2018). My contributions for this work included performing experiments, analyzing data, interpreting results, and writing the manuscript. I previously presented this work at the NHGRI Research Training and Career Development Annual Meeting (2017) and 59th Annual Drosophila Research Conference (2018).

Reference:

ABSTRACT

How remote enhancers interact with appropriate target genes persists as a central mystery in gene regulation. Here we exploit the properties of transvection to explore enhancer-promoter communication between homologous chromosomes in living Drosophila embryos. We successfully visualized the activation of an MS2-tagged reporter gene by a defined developmental enhancer located in trans on the other homolog. This trans-homolog activation depends on insulator DNAs, which increase the stability—but not frequency—of homolog pairing. A pair of heterotypic insulators failed to mediate transvection, raising the possibility that insulator specificity underlies the formation of chromosomal loop domains. Moreover, we found that a shared enhancer co-activates separate PP7 and MS2 reporter genes in cis and in trans. Transvecting alleles weakly compete with one another, raising the possibility that they share a common pool of the transcription machinery. We propose that insulators establish stable trans-homolog associations, providing “hubs” for the accumulation of transcription assemblies.

INTRODUCTION

Enhancers are short segments of genomic DNA (typically ~100 bp to ~1 kb in length) that switch genes on and off in response to a variety of intrinsic and external signals. The human genome is thought to contain on the order of ~400,000 enhancers (ENCODE, 2012), an average of nearly 20 enhancers per protein coding gene. There is emerging evidence that sequence polymorphisms in enhancer DNAs represent a major source of population diversity and predilection to disease (e.g. Abraham et al., 2017; Indjeian et al., 2016). And yet, we do not understand how enhancers work over long distances to stimulate transcription at select target genes.
The phenomenon of transvection in *Drosophila* provides an intriguing model for analyzing enhancer-promoter interactions. Transvection was discovered by E.B. Lewis in 1954 in the course of his groundbreaking studies on the *Ultrabithorax* (*Ubx*) locus of the Bithorax complex (Lewis, 1954). He found that certain *Ubx* alleles are able to complement one another via transvection. Subsequent molecular studies suggested that this complementation is best explained by regulatory sequences on one homolog controlling the expression of the cognate transcription unit on the other homolog (Peifer and Bender, 1986). In this study, we sought to explore the molecular mechanisms underlying this regulatory communication between homologous chromosomes in living embryos.

Studies of transvection at *Ubx* and other genetic loci identified a number of properties that influence *trans*-homolog regulatory interactions, including chromosome pairing, promoter competition, and insulators (reviewed in Fukaya and Levine, 2017). Chromosomal inversions that disrupt pairing cause reductions in transvection at the *Ubx* locus (Lewis, 1954), although an unusual pairing-independent transvection process was documented at the *Abdominal-B* (*Abd-B*) locus (Hendrickson and Sakonju, 1995; Hopmann et al., 1995). Insulators positioned near transvecting alleles have been shown to augment the efficiency of *trans*-homolog interactions, possibly by facilitating stable chromosome pairing (Kravchenko et al., 2005). Finally, enhanced transvection was observed upon removal of nearby *cis*-linked promoters (Geyer et al., 1990; Martínez-Laborda et al., 1992), suggesting that enhancers on one homolog can more easily activate gene expression *in trans* on the other homolog when they are not impeded by proximal *cis*-linked promoters.
In this study, we visualized transvection in living Drosophila embryos. A well-defined developmental enhancer was placed in trans to a lacZ reporter gene containing a series of MS2 stem loops, permitting detection of nascent transcripts using an MCP-GFP fusion protein (Bertrand et al., 1998; Garcia et al., 2013; Lucas et al., 2013). Trans-activation of the MS2-lacZ reporter is observed in early embryos only when insulators are inserted into both alleles. Evidence is presented that insulators increase the stability, but not frequency, of homolog pairing. Different classes of insulators display distinct orientation requirements, and transvection is not observed with a pair of heterotypic insulators. We also show that a shared enhancer can co-activate a cis-linked PP7 reporter gene along with a trans-linked MS2 reporter gene. We propose that transvecting alleles form a trans-homolog “hub”, which serves as a scaffold for the accumulation of transcription complexes.

RESULTS

Insulators facilitate transvection

Site-directed integration was used to insert the well-defined snail (sna) shadow enhancer in trans to a lacZ reporter gene at the same location on homologous chromosomes (Figure 2-1A; top). The ~1.5-kb enhancer mediates stable expression in ~800-1000 cells comprising the presumptive mesoderm of early embryos (Figure 2-2A and B) (Dunipace et al., 2011; Perry et al., 2010). The lacZ reporter gene contains the minimal 100-bp even-skipped (eve) core promoter as well as 24 copies of the MS2 RNA stem loop sequence within the 5’ untranslated region (UTR). In situ hybridization assays were used to monitor expression of lacZ mRNAs. No significant expression was observed either in the absence of an insulator (Figure 2-1A), or upon insertion of a single copy of the minimal 432 bp
*gypsy* insulator (Figure 2-1B). However, expression was observed when *gypsy* insulators were placed on both homologs, upstream of the *sna* shadow enhancer and *MS2-lacZ* reporter gene, respectively (Figure 2-1C). Reporter transcripts were detected in about 3-4% of the cells comprising the *sna* expression pattern within the presumptive mesoderm (Figure 2-1D). Importantly, *trans*-activation of the MS2 reporter gene was observed with different core promoters, transcription units, and genetic organization (Figure 2-2C and D). Insulators were placed adjacent to the enhancer since this configuration is commonly seen for endogenous loci that are active in the early embryo, including *eve* and *fushi tarazu* (*ftz*) (e.g. Fujioka et al. 2013).

These observations are consistent with previous reports that homolog pairing is inefficient in early embryos as compared with later stages of the *Drosophila* life cycle (Gemkow et al., 1998; Hiraoka et al., 1993). This might be due to the slow rate of homolog pairing during the brief interphase periods that characterize the early embryo. In principle, *gypsy* insulators could facilitate *trans*-activation of *MS2-lacZ* by increasing the frequency or stability of homolog pairing. Previous genetic studies of transvection at the *yellow* locus suggest that insulators increase the efficiency of *trans*-homolog interactions (Kravchenko et al., 2005).

**Different classes of insulators**

To explore the contributions of insulators to transvection, we manipulated the orientation of *gypsy*, and examined the activities of additional insulators (Figure 2-3). The *gypsy* insulator contains a highly ordered structure, with each of the 12 Suppressor of Hairy-wing (Su(Hw)) binding sites spanning the insulator positioned in the same orientation (Spana et
Past studies have shown that insulators do not pair when placed in the same orientation *in cis* on the same chromosome (Kyrchanova et al., 2008). However, inverting one of the *gypsy* insulators did not diminish *trans*-activation of the *MS2-lacZ* reporter gene (Figure 2-3A and B), raising the possibility of distinct mechanisms for insulator interactions *in cis* and *in trans*.

We next examined the Homie insulator, which is located at the 3′ boundary of the endogenous *eve* locus (Fujioka et al., 2009; Fujioka et al., 2013). Homie is bound by Su(Hw), but unlike *gypsy*, it is also recognized by another insulator-binding protein CCCTC-binding factor (CTCF), which is highly conserved in vertebrates (Nègre et al., 2010). Homie insulators mediate a low rate of transvection when placed in the same orientation (Figure 2-3D and F), but not when positioned in opposing orientations (Figure 2-3E), similar to control embryos containing just a single Homie insulator (Figure 2-3C). This dependence on orientation is consistent with previous studies (e.g., Fujioka et al., 2016).

The *gypsy* and Homie insulators do not mediate transvection when placed *in trans* to one another (Figure 2-4A-C). This observation raises the possibility of separate classes of insulators: those that are bound by either Su(Hw) alone, or both Su(Hw) and CTCF. A third insulator, Fab-8, was also tested in this assay. It is bound by CTCF, but not Su(Hw) (Gerasimova et al., 2007). We examined a ~5 kb DNA fragment from the *Abd-B* transvection mediating region (tmr), which contains the Fab-8 insulator (Hopmann et al., 1995; Zhou et al., 1999). This tmr fragment also mediates *trans*-activation of the MS2
reporter gene, although expression is slightly delayed as compared with the timing observed for the gypsy and Homie insulators (Figure 2-4D-F). These observations raise the possibility that different classes of insulators mediate distinct kinetics of homolog interactions.

**Stabilization of homolog pairing**

It is possible that insulators facilitate transvection by increasing either the frequency or stability of homolog pairing. To address this question, we analyzed sna>lacZ alleles marked with either PP7 or MS2 in living embryos (Fukaya et al., 2016; Hocine et al., 2013) (Figure 2-5A). Most nuclei do not exhibit pairing, whether or not the two alleles contain gypsy insulators (Figure 2-5B-D, Figure 2-6A-D). Nuclei exhibiting pairing were divided into three classes based on the duration of association: 1-5 min, 5-10 min, and 10 min or more. Approximately 20% of these nuclei display pairing of sna>lacZ alleles whether there are two copies of gypsy or just one (control) (Figure 2-5H and I). Almost half of these nuclei correspond to the most stable class of pairing (10 min or more) with two copies of gypsy, but there is a two-fold reduction with a single copy. Thus, gypsy-gypsy interactions lead to a two-fold increase in stable pairing, and it is likely that this is important for transvection.

To determine how this frequency of stable pairing compares to endogenous loci, we examined the pairing of ftz alleles (Figure 2-6E and F). ftz is located within the Antennapedia complex, which exhibits a variety of transvection phenomena (e.g. Southworth and Kennison, 2002). Moreover, ftz contains well-defined SF1 and SF2 insulators (Belozerov et al., 2003; Li et al., 2015). MS2 and PP7 stem loops were inserted
into the 3’ UTR of each ftz allele using CRISPR-directed genome editing, and their activities were monitored in living embryos during nuclear cycle (nc) 14. There is an overall higher frequency of transient pairing as compared with the sna>lacZ transgenes (Figure 2-6E). However, stable pairing is seen in ~9% of all nuclei examined (Figure 2-6F), similar to the proportion of nuclei exhibiting stable pairing of the synthetic alleles. We therefore conclude that our transvection assay captures critical properties of the endogenous pairing process.

**Co-activation in trans**

An unresolved mystery of transvection is the relative levels of expression of the cis and trans transcription units when regulated by a shared enhancer. Does the enhancer on one homolog have equal access to target genes located on both homologs? To address this question, we created a co-transvection assay, whereby a single enhancer is challenged with both a cis-linked PP7 reporter gene as well as the MS2 reporter gene in trans (Figure 2-7A).

In most transvecting nuclei, the initial bursts of de novo transcription arise solely from the cis-linked PP7 reporter gene during the first 15-20 min following mitosis. Based on the kinetics of allele pairing (Figure 2-5E-G), we believe that this delay is due to a lag in pairing, which typically occurs 10-20 min after mitosis. Alternatively, the shared enhancer may prefer the cis-linked promoter due to its proximity (see Discussion). Following this delay, there is a high incidence of co-activation of the two reporter genes (Figure 2-7B and C, Figure 2-8A and B). During transvection, the two alleles exhibit tight association, although the signals do not completely coincide with one another (Figure 2-7C-E, Figure
2-8B-D, Figure 2-9). It is therefore possible that trans-activation of the MS2 reporter gene depends on proximity, rather than direct physical contact with the shared enhancer (see Discussion). Importantly, trans-activation was nearly abolished when the MS2 and PP7 alleles were inserted at non-homologous locations (Figure 2-10), suggesting that activation in trans depends on sequence homology of paired alleles.

Co-activation of the linked PP7 and MS2 reporter genes in cis and in trans is somewhat surprising since it suggests that they share “common resources” despite their location on different homologs. Further validation was obtained by statistical analysis, comparing the coordination of PP7 and MS2 reporter genes containing either a single shared enhancer (Figure 2-11A and B) or two separate enhancers (Figure 2-11C and D). There is a clear statistical trend: transvecting reporter genes sharing a common enhancer display significantly greater coordination in expression as compared with those containing separate enhancers (Mann-Whitney-U test, \( p < 1.5 \times 10^{-7} \); Figure 2-11E, Figure 2-12A-C). This is consistent with the idea that the two alleles share a local pool of the transcription machinery during transvection (see below).

**Allele competition during transvection**

To explore this idea of “shared resources” during transvection, we measured the timing and amplitudes of MS2 transcription in the absence (Figure 2-13A) or presence (Figure 2-13C) of a linked PP7 reporter gene. There is a significant reduction in the amplitude of MS2 transcription in the presence of PP7 (Mann-Whitney-U test, \( p < 0.001 \); Figure 2-13B and D, Figure 2-12D). Moreover, there is a delay in the timing of MS2 transcription within a subset of transvecting nuclei (Figure 2-13E and F). As discussed earlier, most
transvecting nuclei display a delay in the onset of trans-activation, possibly due to a lag in allele pairing (Figure 2-5) or preferential activation of the cis-linked reporter gene. However, in the absence of the cis-linked PP7 reporter, ~10% of transvecting nuclei show precocious activation of MS2 within the first 10 min following mitosis (Figure 2-13E and F). Such precocious expression is never observed in the presence of the PP7 reporter gene, again suggesting competition in the expression of PP7 and MS2 reporter genes during transvection. Thus, we conclude that the timing and quality, but not frequency, of transvection is affected by the presence of a cis-linked PP7 reporter. We believe that co-activation of the PP7 and MS2 reporter genes during transvection, along with weak competition between the two alleles, suggests that they share a common conglomeration of transcription factors (Figure 2-14).

**DISCUSSION**

The most striking finding of this study is the co-activation of linked PP7 and MS2 reporter genes in cis and trans during transvection. Promoter competition has been observed in a variety of systems including both invertebrates and vertebrates (Choi and Engel, 1988; Foley and Engel, 1992; Fukaya et al., 2016; Ohtsuki et al., 1998). Enhancers tend to prefer nearby promoters as opposed to those positioned in more remote locations. We therefore expected preferential activation of the cis-linked PP7 reporter gene as compared with the MS2 gene located in trans on the other homolog. Instead, once transvection commences, the shared enhancer seems to have equal access to both reporter genes. The only preference for the cis-linked PP7 is seen during the initial periods of interphase following mitosis (Figure 2-7). It is possible that this delay in trans-activation is due to preferential enhancer-promoter interactions with the proximal cis-linked reporter gene. A nonexclusive
alternative possibility is that the delay is due to the time it takes for the two alleles to become physically associated following mitosis. We prefer the latter explanation since the analysis of $sna>MS2$-$lacZ$ and $sna>PP7$-$lacZ$ hetero-allelic embryos reveals a delay in association during the first 10-20 min of nuclear cycle 14 (Figure 2-5E-G).

Transvection is not observed during the timeframe of our analysis (nc 14) without the insertion of insulators into both alleles (Figure 2-1). As a result, we were able to use transvection as an assay for examining insulator function. Most previous studies have centered on the role of insulators in creating chromosomal loop domains in cis, such as topologically associating domains (TADs). Insulators that are separated by large distances along a chromosome (~50 kb to 3 Mb) are thought to come into close physical proximity (Dekker and Mirny, 2016; Dixon et al., 2016). This pairing of insulators depends on their relative orientations (Guo et al., 2015; Rao et al., 2014). When in the convergent orientation in cis, they are able to form a chromosomal loop domain, but not when placed in the same orientation. We do not see this orientation dependence for the gypsy insulator in trans, although the Homie insulator does exhibit such a requirement (Figure 2-3). Moreover, we found that a pair of heterotypic insulators were unable to mediate transvection, e.g., there is no trans-activation of the MS2 reporter gene when one allele is tagged with the Su(Hw)-dependent gypsy insulator and the other with Su(Hw)/CTCF-dependent Homie insulator (Figure 2-4). This observation raises the possibility of insulator specificity in the creation of topological domains both in cis and in trans.
We found that insulators do not significantly increase the frequency of allelic pairing, but instead increase the stability of pairing, as suggested previously (Kravchenko et al., 2005). It is possible that somatic pairing of homologous chromosomes is mediated by DNA sequence homology, as seen for pairing during meiosis (Tsai and McKee, 2011). Support for this view stems from the analysis of trans-chromosomal interactions when the gypsy insulator, sna shadow enhancer and PP7 reporter gene are placed in non-homologous locations in the genome (Figure 2-10). Although trans-activation is exceedingly rare, we have detected a few nuclei that display co-expression of MS2 and PP7, and the resulting transcriptional bursts are comparable to those seen in pairing-dependent transvection. One interpretation of these findings is that sequence homology increases the probability of transient associations between alleles, while insulators stabilize the association. In the case of non-homologous sites, associations between the enhancer and MS2 reporter are rare, but, when they occur, the insulators stabilize the association for trans-activation of the MS2 reporter. There are instances of non-homologous interactions in vertebrates (Ling et al., 2006; Lomvardas et al., 2006), and we suggest that these depend on compatible insulators located near interacting loci. It is also important to note that some enhancers can act in trans independently of insulators (e.g. Bateman et al., 2012; Blick et al., 2016; Mellert and Truman, 2012), giving rise to the possibility that there might be other DNA elements that can mediate transvection.

We propose that insulator-insulator interactions create a trans-homolog topological domain, comparable to those formed in cis (Fukaya et al., 2016). This association places the shared enhancer in close proximity with both the cis-linked PP7 reporter gene and the
trans-linked MS2 reporter (Figure 2-14). It is possible that paired insulators create a trap or “hub” for condensates of the transcriptional machinery. The formation of this transcriptional hub might be driven by the coalescence of Pol II complexes and transcription factors, accumulated at promoters and enhancers respectively (Hnisz et al., 2017). As implied by a recent live-imaging study (Gu et al., 2018), the formation of such hubs might influence the sub-diffusive mobility of enhancers. The release of Pol II from the hub could contribute to transcriptional bursting (Bothma et al., 2014), as well as co-activation of the PP7 and MS2 reporter genes in cis (Fukaya et al., 2016) and in trans.

This view of transcriptional control is quite distinct from the classical models of sequential recruitment of individual Pol II complexes to their target promoters. It was recently proposed that transcriptional bursts are the result of periodic phase transitions of activator proteins and the Pol II machinery at active loci (Hnisz et al., 2017). These transitions are thought to arise from protein-protein interactions via intrinsically disordered domains within transcription factors, co-activators and Pol II subunits (Gemayel et al., 2015; Hnisz et al., 2017). Indeed, recent super-resolution studies suggest highly dynamic clustering of Pol II complexes and transcription factors at active foci (Cisse et al., 2013; Mir et al., 2017; Tsai et al., 2017). The proposed transcription hubs also challenge traditional models of targeted enhancer-promoter loops. The formation of extended transcriptional condensates consisting of multiple activators and Pol II complexes might require approximate proximity rather than direct, physical association of enhancers with their target promoters.
MATERIALS AND METHODS

Experimental model and subject details

In all experiments, we studied *Drosophila melanogaster* embryos at nuclear cycle 14. The following fly lines were used in this study: *nos>MCP-GFP, His2Av-mRFP* (this study), *nos>MCP-GFP, nos>mCherry-PCP, His2Av-eBFP2* (this study), *sna shadow enhancer* (this study), *gypsy-sna shadow enhancer* (this study), *sna shadow enhancer-gypsy* (this study), *sna shadow enhancer-evePr-MS2-lacZ* (this study), *gypsy-sna shadow enhancer-evePr-MS2-lacZ* (this study), *gypsy-evePr-MS2-lacZ* (this study), *gypsy (inverted)-evePr-MS2-lacZ* (this study), *Homie-evePr-MS2-lacZ* (this study), *Homie (inverted)-evePr-MS2-lacZ* (this study), *Homie-sna shadow enhancer* (this study), *tmr-sna shadow enhancer* (this study), *tmr-evePr-MS2-lacZ* (this study), *gypsy-sna shadow enhancer-evePr-PP7-lacZ* (this study), *gypsy-snaPr-MS2-yellow* (this study), *fushi tarazu-MS2* (this study), *fushi tarazu-PP7* (this study), *y¹ w¹¹¹⁸; PBac{y+-attP-3B} VK00033* (Bloomington Drosophila Stock Center #9750), *y¹ w¹¹¹⁸; PBac{y+-attP-3B} VK00031* (Bloomington Drosophila Stock Center #9748).

*fushi tarazu-MS2* and *fushi tarazu-PP7* were generated using CRISPR/Cas9 based insertion of 24x MS2 RNA stem loop or 24x PP7 RNA stem loop into the 3’ UTR of endogenous *fushi tarazu*. In brief, ~1-kb DNA fragment of 5’ and 3’ homology arm sequences were PCR amplified from the genomic DNA, and inserted into the pBS-MS2-loxP-dsRed-loxP and pBS-PP7-loxP-dsRed-loxP donor plasmids (see below). These plasmids were co-injected with the pCFD3 gRNA expression plasmid to *nos-Cas9/CyO* embryos (Ren et al., 2013).
**Method Details**

**Site specific transgenesis by phiC31 system**

All reporter plasmids were integrated into a unique landing site on the third chromosome using strain 9750 (Bloomington *Drosophila* Stock Center). In Figure 2-10, *PP7-lacZ* reporter plasmid was integrated into a unique landing site on the third chromosome using strain 9748 (Bloomington *Drosophila* Stock Center). Microinjection was performed as described (Ringrose, 2009). Zero to 1-hour embryos were collected and dechorionated with bleach. Aligned embryos were dried with silica gel for 12 min and covered with Halocarbon oil 27 (Sigma). Subsequently, microinjection was performed using Picospritzer III (Parker) and Narishige M-152 Micromanipulator (Narishige). Injection mixture contains 500 ng/μl plasmid DNA, 5 mM KCl, 0.1 mM phosphate buffer, pH 6.8. *mini-White* marker was used for subsequent screening.

**Genome editing by CRISPR/Cas9**

pCFD3 gRNA expression plasmid and pBS-dsRed donor plasmid were co-injected to *nos-Cas9/Cyo* embryos (Ren et al., 2013). Microinjection was performed as described (Ringrose, 2009). Injection mixture contains 500 ng/μl pCFD3 gRNA expression plasmid, 500 ng/μl pBS-dsRed donor plasmid, 5 mM KCl, 0.1 mM phosphate buffer, pH 6.8. *3xP3-dsRed* marker was used for subsequent screening.

**Fly strains**

*MCP-GFP, mCherry-PCP, His2Av-eBFP2*

The *nanos>SV40NLS-mCherry-PCP, His2Av-eBFP2* expression plasmid (Fukaya et al., 2017) was integrated into a unique landing site on the third chromosome using strain 9750 (Bloomington *Drosophila* Stock Center) to obtain maternal expression of the mCherry-
PCP and His2Av-eBFP2 fusion proteins. Maternal expression of the MCP-GFP fusion protein was obtained using a transgenic strain carrying a nanos>\textit{MCP-GFP} transgene that was integrated into the third chromosome by P-element mediated transformation (Garcia et al., 2013). These were mated to create the fly line \textit{mCherry-PCP, MCP-GFP, His2Av-eBFP2} in order to obtain co-expression of MCP-GFP, mCherry-PCP and His2Av-eBFP2 fusion proteins.

\textit{MCP-GFP, His2Av-mRFP}

The \textit{nanos} > MCP-GFP, His2Av-mRFP expression plasmid was integrated into a unique landing site on the third chromosome using strain 9750 (Bloomington \textit{Drosophila} Stock Center) to obtain maternal expression of the MCP-GFP and His2Av-mRFP fusion proteins.

\textit{In situ hybridization}

Embryos were dechorionated and fixed in fixation buffer (0.5x PBS, 25 mM EGTA, 4% formaldehyde and 50% Heptane) for 20 min at room temperature. Antisense RNA probes labeled with digoxigenin (DIG RNA Labeling Mix 10 × conc, Roche) and biotin (Biotin RNA Labeling Mix 10 × conc, Roche) were used to detect \textit{lacZ} and \textit{snail} RNAs, respectively. Hybridization was performed at 55°C overnight in hybridization buffer (50% formamide, 5x SSC, 50 μg/ml Heparin, 100 μg/ml salmon sperm DNA, 0.1% Tween-20). Subsequently, embryos were washed with hybridization buffer at 55°C and incubated with Western Blocking Buffer (Roche) at room temperature for one hour. Then, embryos were incubated with sheep anti-digoxigenin (Roche) and mouse anti-biotin primary antibodies (Invitrogen) at 4°C for overnight, followed by incubation with Alexa Fluor 488 donkey anti-sheep (Invitrogen) and Alexa Flour 555 goat anti-mouse (Invitrogen) fluorescent secondary antibodies at room temperature for two hours. DNA was stained with Hoechst
33342 (Thermo Fisher Scientific), and embryos were mounted in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). Imaging was performed on a Zeiss LSM 880 confocal microscope. Plan-Achromat 20x / 0.8 N.A. objective was used. Images were captured in 16 bit.

**Dual-color MS2/PP7 Live imaging**

*MC-P-GFP, mCherry-PCP, His2Av-eBFP2* virgins were mated with homozygous males carrying *PP7-lacZ* reporter genes. Resulting *trans*-heterozygote virgins were collected and mated with homozygous males carrying MS2 reporter genes. The resulting embryos were dechorinated and mounted between a semipermeable membrane (In Vitro Systems & Services) and a coverslip (18 mm x 18 mm), and embedded in Halocarbon oil 27 (Sigma). Embryos were imaged using a Zeiss LSM 880 at room temperature. Plan-Achromat 40x / 1.3N.A. oil immersion objective was used. At each time point, a stack of 21 images separated by 0.5 μm was acquired and the final time resolution is 13 sec. Images were captured in 16 bit. For each cross, more than three biological replicates were taken using the same setting of the microscope. The same laser power and microscope setting were consistently used in the same experiments.

**Single-color MS2 Live imaging**

*MC-P-GFP, His2Av-mRFP* virgins were mated with homozygous males carrying the MS2 reporter gene. Resulting *trans*-heterozygote virgins were collected and mated with homozygous males carrying *snail* shadow enhancer. In Figure 2-2A and B, heterozygote *MC-P-GFP, His2Av-mRFP* virgins were mated with homozygous males carrying *snail shadow enhancer-evePr-MS2-lacZ* reporter gene. The resulting embryos were dechorinated and mounted between a semipermeable membrane (In Vitro Systems &
Embryos were imaged using a Zeiss LSM 880 at room temperature. Plan-Apochromat 40x / 1.3N.A. oil immersion objective was used. At each time point, a stack of 21 images separated by 0.5 μm was acquired and the final time resolution is 13 sec. Images were captured in 16 bit. For each cross, three biological replicates were taken using the same setting of the microscope. The same laser power and microscope setting were used throughout this study.

Plasmids

*pbphi-snail shadow enhancer*

A DNA fragment containing *snail* shadow enhancer was amplified from the genomic DNA using primers (5´-GGG GGA AGC TTG CAT TGA GGT GTT TTG TTG G-3´) and (5´-CCC CCG CTA GCT AAA TTC CGA TTT TTC TTG T -3´), and digested with HindIII and NheI. The resulting fragment was inserted between the HindIII and NheI sites in pbphi-multi cloning site (Fukaya et al., 2016). Subsequently, a DNA fragment containing *Burkholderia cenocepacia ParSc3* sequence (Saad et al., 2014) was inserted into the unique Ndel site of the plasmid. ParSc3 was not used in this study.

*pbphi-gypsy-snail shadow enhancer*

A DNA fragment containing *gypsy* insulator was amplified from the genomic DNA using primers (5´-GGG GGG CGG CCT GGC CAC GTA ATA AGT GTG CG-3´) and (5´-CCC CCG GAT CCG TTG TTG GTT GTC GGC ACA CCA C-3´), and digested with NOTI and BamHI. The resulting fragment was inserted between the NOTI and BglIII sites in pbphi-snail shadow enhancer.
pbphi-snail shadow enhancer-gypsy

A DNA fragment containing gypsy insulator was amplified from the genomic DNA using primers (5’-GGG GGG CTA GCC TGG CCA CGT AAT AAG TGT GCG -3’) and (5’-CCC CCT CTA GAG TTG TTG GTT GGC ACA CCA C-3’), and digested with NheI and XbaI. The resulting fragment was inserted between the NheI and XbaI sites in pbphi-snail shadow enhancer. ParSc3 sequence is depleted from this plasmid.

pbphi-lacZ-αTub 3’UTR

A DNA fragment containing α-Tubulin at 84B 3’UTR was amplified from the genomic DNA using primers (5’-GGG GGC TCG AGG CGT CAC GCC ACT TCA ACG C -3´) and (5’-CCC CCT CTA GAA AGC TTG AGC TTC GCA TGG TTT TGC C -3´), and digested with XhoI and XbaI. The resulting fragment was inserted between the XhoI and XbaI sites in pbphi-multi cloning site. Subsequently, a DNA fragment containing lacZ was amplified using primers (5’-GGG GGA GAT CTA TGC AGA ACT GGG AGA CGA C -3´) and (5’-CCC CCC TCG AGT TAT TTT TGA CAC CAG ACC A -3´), and digested with BglII and XhoI. The resulting fragment was inserted between the BglII and XhoI sites in the plasmid.

pbphi-evePr-MS2-lacZ-αTub 3’UTR

A DNA fragment containing eve core promoter was amplified from the genomic DNA using primers (5’-GGG GGC GGC CGC ACT AGT TTT GCC TGC AGA AGA CGA C -3´) and (5’-CCC CCA GAT CTA ACG AAG GCA GTT AGT TGT T-3´), and digested with NOTI and BglII. The resulting fragment was inserted between the NOTI and BglII sites in pbphi-lacZ-αTub 3’UTR. Subsequently, a DNA fragment containing 24x MS2 stem loops was purified from pCR4-24xMS2SL-stable (Bertrand et al., 1998) by
digesting with BamHI and BglII. The resulting fragment was inserted into the unique BglII site in the plasmid.

\textit{pbphi-gypsy-evePr-MS2-lacZ-\alphaTub 3´UTR}

A DNA fragment containing \textit{gypsy} insulator was amplified from the genomic DNA using primers (5´-GGG GGG CGG CCG CCT GGC CAC ATA AGT GTG-3´) and (5´-CCC CCG CGG CCG CGT TGT TGG TTG GCA CAC CAC-3´), and digested with NOTI. The resulting fragment was inserted into the unique NOTI site in the \textit{pbphi-evePr-MS2-lacZ-\alphaTub 3´UTR}. By sequencing resulting plasmids, orientation of \textit{gypsy} insulator was determined.

\textit{pbphi-gypsy-snail shadow enhancer-evePr-MS2-lacZ-\alphaTub 3´UTR}

A DNA fragment containing \textit{gypsy} insulator and \textit{snail} shadow enhancer was purified from \textit{pbphi-gypsy-snail shadow enhancer} by digesting with NOTI and NheI. The resulting fragment was inserted between the NOTI and SpeI sites in the \textit{pbphi-evePr-MS2-lacZ-\alphaTub 3´UTR}.

\textit{pbphi-snail shadow enhancer-evePr-MS2-lacZ-\alphaTub 3´UTR}

A DNA fragment containing \textit{snail} shadow enhancer was purified from \textit{pbphi-snail shadow enhancer} by digesting with NOTI and NheI. The resulting fragment was inserted between the NOTI and SpeI sites in the \textit{pbphi-evePr-MS2-lacZ-\alphaTub 3´UTR}.

\textit{pbphi-evePr-PP7-lacZ-\alphaTub 3´UTR}

A DNA fragment containing \textit{eve} core promoter was amplified from the genomic DNA using primers (5´-GGG GGG CGG CCG CGC TAG CTT TGC CTG CAG AGC GCA GCG-3´) and (5´-CCC CCA GAT CTA ACG AAG GCA GTT AGT TGT T-3´), and digested with NOTI and BglIII. The resulting fragment was inserted between the NOTI and
BglII sites in pbphi-lacZ-αTub 3´UTR. Subsequently, a DNA fragment containing 24x PP7 stem loops was purified from pBS-24x PP7 (Fukaya et al., 2016) by digesting with BamHI and BglII, and the resulting fragment was inserted into the unique BglII site in the plasmid.

**pbphi-gypsy-snail shadow enhancer-evePr-PP7-lacZ-αTub 3´UTR**

A DNA fragment containing gypsy insulator and snail shadow enhancer was purified from pbphi-gypsy-snail shadow enhancer by digesting with NOTI and NheI. The resulting fragment was inserted between the NOTI and NheI sites in pbphi-evePr-PP7-lacZ-αTub 3´UTR.

**pbphi-gypsy-snaPr-MS2-yellow-αTub 3´UTR**

A DNA fragment containing partial sequence of lacZ was amplified using primers (5´-GGG GGC TGC AGG TAT GCA GAA CTG GGA GAC GAC-3´) and (5´-CCC CCC TGC AGC GAA ACC GAC ATC GCA GGC TT-3´), and digested with PstI. The resulting fragment was inserted into the unique PstI site of the snaPr-MS2-yellow plasmid (Ferraro et al., 2016) to discriminate from the endogenous yellow gene. Subsequently, a DNA fragment containing gypsy insulator was amplified from the genomic DNA using primers (5´-GGG GGG CGG CGG CCT GGC CAC GTA ATA AGT GTG CG-3´) and (5´-CCC CCC TCG AGG TTG TTG GTT GGC ACA CCA C-3´), and digested with NotI and XhoI. The resulting fragment was inserted between the NotI and XhoI sites of the plasmid.

**pbphi-Homie-snail shadow enhancer**

A DNA fragment containing Homie insulator was amplified from the genomic DNA using primers (5´-GGG GGG CGG CCG CCT GGC CAC GTA ATA AGT GTG CG-3´) and (5´-CCC CCC TCG AGG TTG TTG GTT GGC ACA CCA C-3´), and digested with NOTI
and BglII. The resulting fragment was inserted between the NOTI and BglII sites in pbphi-snail shadow enhancer.

\textit{pbphi-Homie-evePr-MS2-lacZ-\alpha Tub 3’UTR}

A DNA fragment containing Homie insulator was amplified from the genomic DNA using primers (5’-GGG GGG CGG CCG CAA TAC TAA AAA GTT TTT ACG AGC-3’) and (5’-CCC CCG CGG CCG CGA TTA CAC GCT GCG ATG GTT-3’), and digested with NOTI. The resulting fragment was inserted into the unique NOTI site in pbphi-evePr-MS2-lacZ-\alpha Tub 3’UTR. By sequencing resulting plasmids, orientation of Homie insulator was determined.

\textit{pbphi-tmr-snail shadow enhancer}

A DNA fragment containing \textit{Abd-B} transvection mediating region was amplified from the genomic DNA using primers (5’-GGG GGG CGG CCG CGA ATT CGT CTT CTA GCT ACC TGC CC-3’) and (5’-CCC CCA GAT CTC GTG TGG AAT TCT GTT CTG TCA CAA-3’), and digested with NOTI and BglII. The resulting fragment was inserted between the NOTI and BglII sites in pbphi-snail shadow enhancer.
*pbphi-nanos>MCP-GFP*

A DNA fragment containing MCP was amplified from pUC18-Pnos-MCP-GFP-αTub 3’UTR plasmid (Garcia et al., 2013) using primers (5’-GGG GGC TCG AGA TGG CTT CTA ACT TTA CTC A-3’) and (5’-CCC CCA AGC TTG TAG ATG CCG GAG TTT GCT G-3’), and digested with XhoI and HindIII. The resulting fragment was inserted between the XhoI and HindIII sites in pbphi-nanos promoter-αTubulin 3’UTR expression vector (Fukaya et al., 2016). Subsequently, a DNA fragment containing GFP was amplified using primers (5’-GGG GGG CTA GCA TGG TGA GCA AGG GCG AGG A-3’) and (5’-CCC CCG GAT CCT TAC TTG TAC AGC TCG TCC A-3’) and digested with NheI and BamHI, and the resulting fragment was inserted between the NheI and BamHI sites in the plasmid.

*pbphi-His2Av-mRFP*

A DNA fragment containing mRFP was amplified using primers (5’-GGG GGA AGC TTG GCG GAT CAG GCT CGG GAT CAT CGA TGG CCT CCT CCG AGG ACG T-3’) and (5’-CCC CCG CTA GCT TAG GCG CCG GTG GAG TGG C-3’), and digested with HindIII and NheI. The resulting fragment was inserted between the HindIII and NheI sites in the pbphi-His2Av (Fukaya et al., 2017).

*pbphi-nanos>MCP-GFP, His2Av-mRFP*

A DNA fragment containing His2Av-mRFP was purified from pbphi-His2Av-mRFP by digesting with XbaI. The resulting fragment was inserted into the unique XbaI site in the pbphi-nanos>MCP-GFP.
**pCFD3-dU6-fitz gRNA**

Two DNA oligos (5´-GTC GAG TTA CTC TCT TCC CCA GAG-3´) and (5´-AAA CCT CTG GGG AAG AGA GTA ACT-3´) were annealed and inserted into the pCFD3-dU6:3gRNA vector (addgene # 49410) using BbsI sites.

**pBS-loxP**

Two DNA oligos (5´-CCA AGG ATC CAT AAC TTC GTA TAA TGT ATG CTA TAC GAA GTT ATC ATA TGC GAC TAG TGG TC-3´) and (5´-GAC CAC TAG TCG CAT ATG ATA ACT TCG TAT AGC ATA CAT TAT ACG AAG TTA TGG ATC CTT GG-3´) were annealed and inserted into the pBlueScript vector using BamHI and SpeI restriction sites.

**pBS-MS2-loxP-dsRed-loxP**

A DNA fragment containing 24x MS2 stem loops was purified from pCR4-24xMS2SL-stable (Bertrand et al., 1998) by digesting with BamHI and BglII. The resulting fragment was inserted into the unique BamHI site in pBS-loxP plasmid. Subsequently, a DNA fragment containing dsRed and loxP site was amplified from pHD-DsRed-attP (addgene #51019) using primers (5´-GGG GGC ATA TGG GAT CTA ATT CAA TTA GAG A-3´) and (5´-TCT TTA CTA GTA CCG GTT AAG ATA CAT TGA TGA G-3´), and digested with NdeI and SpeI. The resulting fragment was inserted between the NdeI and SpeI sites of the plasmid.

**pBS-PP7-loxP-dsRed-loxP**

A DNA fragment containing 24x PP7 stem loops was purified from pBlueScript-24x PP7 (Fukaya et al., 2016) by digesting with BamHI and BglII. The resulting fragment was inserted into the unique BamHI site in pBS-loxP plasmid. Subsequently, a DNA fragment
containing dsRed and loxP site was amplified from pHD-DsRed-attP (addgene #51019) using primers (5´-GGG GGC ATA TGG GAT CTA ATT CAA TTA GAG A-3´) and (5´- TCT TTA CTA GTA CCG GTT AAG ATA CAT TGA TGA G-3´), and digested with NdeI and SpeI. The resulting fragment was inserted between the NdeI and SpeI sites of the plasmid.

*pBS-ftz5´-MS2-loxP-dsRed-loxP-ftz3´*

A DNA fragment containing 3´ homology arm of ftz was amplified from genomic DNA using primers (5´-GGG GGA CTA GTG AGC GGA ACC GAA AGC CGT ACC GCC-3´) and (5´-CCC CCG CGG CCG CGC CCC GAA GCC AGA CCA AAA TGT GA-3´), and digested with SpeI and NotI. The resulting fragment was inserted between the SpeI and NotI sites of pBS-loxP-MS2-dsRed-loxP. Subsequently, a DNA fragment containing 5´ homology arm of ftz was amplified from genomic DNA using primers (5´-GGG GGG GTA CCC GTC GAG CAG GTG AAG AAG GCT CCC-3´) and (5´-CCC CCC TGC AGT GGG GAA GAG AGT AAC TGA GCA TC-3´), and digested with KpnI and PstI. The resulting fragment was inserted between the KpnI and PstI sites of the plasmid.

*pBS-ftz5´-PP7-loxP-dsRed-loxP-ftz3´*

A DNA fragment containing 3´ homology arm of ftz was amplified from genomic DNA using primers (5´-GGG GGA CTA GTG AGC GGA ACC GAA AGC CGT ACC GCC-3´) and (5´-CCC CCG CGG CCG CGC CCC GAA GCC AGA CCA AAA TGT GA-3´), and digested with SpeI and NotI. The resulting fragment was inserted between the SpeI and NotI sites of pBS-PP7-loxP-dsRed-loxP. Subsequently, a DNA fragment containing 5´ homology arm of ftz was amplified from genomic DNA using primers (5´-GGG GGG GTA CCC GTC GAG CAG GTG AAG AAG GCT CCC-3´) and (5´-CCC CCC TGC AGT GGG GAA GAG AGT AAC TGA GCA TC-3´), and digested with KpnI and PstI. The resulting fragment was inserted between the KpnI and PstI sites of the plasmid.
GAA GAG AGT AAC TGA GCA TC-3′), and digested with KpnI and PstI. The resulting fragment was inserted between the KpnI and PstI sites of the plasmid.

**Quantification and Statistical Analysis**

All the image processing methods and analysis were implemented in MATLAB (R2016b, MathWorks).

**Nuclei segmentation and tracking**

For each time point, maximum projections were obtained for all 21 z-sections per image. His2Av-mRFP and His2Av-eBFP2 were used to segment nuclei for single-color imaging and dual-color imaging, respectively. Nuclei-labeled channel images were pre-processed with Gaussian filtering, top-hap filtering, and adaptive histogram equalization, in order to enhance the signal-to-noise contrast. Processed images were converted into binary images using a threshold value obtained from Otsu’s method. The number and the position of separate components within a frame was obtained, where each component serves as a mask for individual nuclei. Since nuclei hardly move past 5 min into nc 14, nuclei tracking was obtained by finding the component with minimal movement across these frames (5 min into nc 14 to the onset of gastrulation).

**Recording MS2 and PP7 signals**

Maximum projections of raw images were used to record fluorescent intensities. Using nuclei segmentation files as a mask for each nucleus, fluorescence intensities within each nucleus was extracted. After subtracting the background nuclear signal, the signal of MS2 and PP7 transcription foci was determined by taking an average of the top three pixels with the highest fluorescence intensity within each nucleus.
3D tracking of MS2 and PP7 alleles

The entire 21 z-stack images were used to track the three-dimensional position of the active MS2 and PP7 transcription foci within a nucleus. For each nucleus, x, y, and z coordinates of the pixel with the maximum MS2 or PP7 signal was determined respectively using raw image files. The MS2 and PP7 foci coordinates were then transformed with respect to the center of each corresponding nucleus. The resulting x, y, and z resolution is 0.277 µm, 0.277 µm, and 0.5 µm per pixel, respectively. The distance between MS2 and PP7 alleles was measured only when the MS2-lacZ and PP7-lacZ are both transcriptionally active. MS2 and PP7 alleles were considered to be paired when the average distance between the two alleles was less than 1 µm (Figure 2-5H, Figure 2-6E and Figure 2-10B). In Figure 2-5I, Figure 2-6F and Figure 2-10B, pairing duration was determined by measuring the time during which the allelic distance was less than 1.5 µm.

Analysis of coordinated transcription

For each of the MS2 and PP7 trajectories from a given nucleus, the Pearson correlation coefficient was calculated. In the transvection assay, the correlation coefficients within the time window between the first and the last time point of active MS2 transcription were measured. When MS2 and PP7 reporter genes were individually placed under the control of separate enhancers, the correlation coefficients were measured using the MS2 and PP7 trajectories during the period of 5 min into nc 14 to the onset of gastrulation. To compare the histograms between two data sets, the distribution of correlation coefficients was determined by normalizing to the number of total analyzed nuclei. In Figure 2-11E and Figure 2-12C, the statistical significance of the difference between two distributions of correlation coefficients was examined using a two-sided Mann-Whitney-U test.
Measurement of cumulative fraction of active nuclei

The time of activation was determined as the first time frame wherein MS2 fluorescence intensity exceeded a threshold value. To determine the threshold value, maximum fluorescence intensity throughout the entirety of nc 14 was measured for each embryo. Subsequently, 15% of the maximum fluorescence intensity was calculated and averaged among replicates. This value was used as the threshold for each genotype. Nuclei were considered as active if MS2 signal exceeded the threshold for longer than 2 min. The presumptive mesoderm region was defined as a domain containing ~17-18 cell widths centered to the ventral most nuclei, which corresponds to endogenous sna expression domain. The very first cells that undergo gastrulation were determined as the ventral most nuclei. The same measurement methods were used in Figure 2-1, 2-3, 2-13 and Figure 2-2, 2-4. In all these figures, error bars represent the standard error of the mean.

False coloring of active nuclei

Active nuclei were determined by possessing MS2 signal above a threshold value as described in the previous section. The set of nuclei that were active at any point during the ~50 min period of nc 14 was then identified per embryo. Using their segmentation mask, these active nuclei were then colored and layered over the raw His2Av-eBFP2 image at the time point of ~30 min into nc 14.

Characterization of transcriptional bursting

The amplitude was measured by taking the peaks of each transcriptional burst after smoothing with the local regression (LOESS) method. A transcriptional burst was defined as a change in fluorescence intensity where the local dynamic range (i.e., distance between peak and trough) was as least twice the dynamic range between the global and local minima
(i.e., distance between baseline and trough). When a nucleus had above-baseline transcriptional activity at the onset (5 min into nc14) or end (onset of gastrulation) of the analysis, that burst was excluded, due to uncertainty in the detection of its associated amplitude. A two-sided Mann-Whitney-U test was performed to determine the statistical significance of the difference between two distributions of burst amplitude shown in Figure 2-12D.

**Data and Software Availability**

Original imaging data of *in situ* hybridization assay has been deposited to Mendeley Data (http://dx.doi.org/10.17632/3ff644zs6s.1).
REFERENCES


**FIGURES**

(A) snail shadow enhancer is not sufficient to activate lacZ transcription in trans. Scale bar represents 50 µm.

(B) A single gypsy insulator cannot mediate trans-activation of the lacZ reporter.

(C) Insertion of gypsy insulators on both alleles facilitates trans-activation of lacZ transcription.

(D) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. Trans-activation of MS2-lacZ reporter gene was visualized with MCP-GFP fusion protein in living embryos. Gypsy insulators are positioned in the same orientation as shown in (C). A total of 1151 (1119 inactive, 32 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.

---

**Figure 2-1. A pair of gypsy insulators mediates transvection.**

(A-C) Double fluorescent in situ hybridizations of lacZ (green) and endogenous snail (red). The lacZ reporter gene contains the 100-bp minimal eve promoter and 24x MS2 RNA stem loops within the 5’ UTR. Embryos at late nc 14 are shown. Images were rotated to align embryos (anterior to the left and posterior to the right).

(A) snail shadow enhancer is not sufficient to activate lacZ transcription in trans. Scale bar represents 50 µm.

(B) A single gypsy insulator cannot mediate trans-activation of the lacZ reporter.

(C) Insertion of gypsy insulators on both alleles facilitates trans-activation of lacZ transcription.

(D) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. Trans-activation of MS2-lacZ reporter gene was visualized with MCP-GFP fusion protein in living embryos. Gypsy insulators are positioned in the same orientation as shown in (C). A total of 1151 (1119 inactive, 32 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-2. Dynamics of cis- and trans-activation of lacZ transcription.

(A) The MS2-lacZ reporter gene is linked to the sna shadow enhancer in cis. Cumulative fraction of cis-activating nuclei in the presumptive mesoderm cells. Activation of MS2-lacZ reporter gene was visualized with MCP-GFP fusion protein in living embryos. A total of 1056 (47 inactive, 1009 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.

(B) Schematic representation of the transvection assay. MS2-yellow reporter contains the 100-bp minimal sna core promoter and 24x MS2 RNA stem loop at the 5’ UTR. The reporter gene was modified by insertion of a 1-kb sequence derived from lacZ into the second exon of the yellow transcription unit to discriminate from the endogenous yellow locus. gypsy insulators were placed downstream of the sna shadow enhancer (allele 1), and upstream of the yellow reporter gene (allele 2).

(D) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. A total of 850 (842 inactive, 8 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-3. Orientation-dependent and -independent insulator pairing.
(A) Insertion of gypsy insulators in the opposite orientation can facilitate trans-activation of the lacZ reporter. Scale bar represents 50 µm.
(B) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. gypsy insulators are positioned in the opposite orientation as shown in (A). A total of 916 (879 inactive, 37 active) nuclei from three independent embryos were analyzed. Error bars represent the ± standard error of the mean of three independent embryos.
(C) A single Homie insulator cannot mediate trans-activation of the lacZ reporter. Scale bar represents 50 µm.
(D) Homie insulators on both homologs positioned in the same orientation facilitates trans-activation of the lacZ reporter.
(E) A pair of Homie insulators positioned in the opposite orientation fails to mediate trans-activation of the lacZ reporter.
(F) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. Homie insulators are positioned in the same orientation as shown in (D). A total of 980 (958 inactive, 22 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-4. Specificity of insulator pairing in early embryos.
(A-E) Double fluorescent in situ hybridizations of lacZ (green) and endogenous snail (red). Embryos at late nc 14 are shown. Images were rotated to align embryos (anterior to the left and posterior to the right).
(A) A pair of gypsy insulators on both homologs facilitates trans-activation of lacZ transcription. Scale bar represents 50 µm.
(B, C) gypsy insulator cannot pair with Homie insulator.
(D) Single Fab-8 insulator (tmr) cannot mediate trans-activation of lacZ transcription. Scale bar represents 50 µm.
(E) A pair of Fab-8 insulator (tmr) on both homologs mediates trans-activation of lacZ transcription. tmr sequences were positioned in the same orientation.
(F) Cumulative fraction of transvecting nuclei in the presumptive mesoderm cells. Fab-8 insulators are positioned in the same orientation as shown in (E). A total of 1038 (1017 inactive, 21 active) nuclei from three independent embryos were analyzed. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-5. *gypsy* insulators foster stable homolog pairing.

(A) Schematic representation of the *lacZ* reporter genes containing the 100-bp minimal *eve* promoter, the 1.5-kb *sna* shadow enhancer, the 432-bp *gypsy* insulator and 24x MS2 or 24x PP7 RNA stem loops within the 5’ UTR.

(B) 3D tracking of MS2-*lacZ* (green) and PP7-*lacZ* (red) transcription foci in a nucleus that exhibits allelic separation throughout nc 14. An ellipsoid with 4 µm short-axis radius and 7 µm long-axis radius was plotted as a proxy for the shape of the nucleus at mid nc 14. Location of MS2 and PP7 alleles are indicated as spheres with 0.5 µm radii to represent uncertainty in the spatial resolution in the z-axis.

(C, D) Cumulative traces of MS2-*lacZ* (green) and PP7-*lacZ* (red) transcription foci during early (5 min to 29.9 min; C) and late nc 14 (30.1 min to 54.6 min; D) in the nucleus shown in (B).

(E) 3D tracking of MS2-*lacZ* (green) and PP7-*lacZ* (red) transcription foci in a nucleus that undergoes stable homolog pairing at the midpoint of nc 14. Note the association of the two alleles during later stages of nc 14.

(F, G) Cumulative traces of MS2-*lacZ* (green) and PP7-*lacZ* (red) transcription foci during early phases of nc 14 when the two alleles are unpaired (5.0 min to 29.5 min; F) and later stages when paired (29.7 min to 54.6 min; G) in the nucleus shown in (E).

(H, I) The pairing efficiency across nuclei (H) and the distribution of the duration of homolog pairing in those nuclei exhibiting pairing (I). The most salient impact of two copies of *gypsy* is the doubling in the frequency of the most stably paired alleles (from ~4% in controls to ~9% with a pair of *gypsy* insulators). A total of 1135 and 936 nuclei from three independent embryos were analyzed for single *gypsy* and two *gypsy*, respectively. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-6. 3D tracking of allelic distances for unpaired and paired homologs.

(A) Trajectory of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity in the nucleus shown in Figure 2-5B-D.

(B) 3D distances between MS2 and PP7 alleles in the nucleus shown in (A) and Figure 2-5B-D.

(C) Trajectory of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity in the nucleus shown in Figure 2-5E-G.

(D) 3D distances between MS2 and PP7 alleles in the nucleus shown in (C) and Figure 2-5E-G.

(E, F) The pairing efficiency across nuclei (E) and the distribution of the duration of homolog pairing in those nuclei exhibiting pairing (F). A total of 728 nuclei from three independent embryos were analyzed for endogenous ftz. Nuclei within mature ftz expression domain were used for the analysis. The domain was defined as the composite of nuclei that exhibit transcriptional activity after 35 min into nc 14. Error bars represent ± the standard error of the mean of three independent embryos.
Figure 2-7. A shared enhancer can co-activate two reporter genes in cis and trans.

(A) Schematic representation of the transvection assay. The lacZ reporter gene contains the 100-bp minimal eve promoter and 24x MS2 or 24x PP7 RNA stem loops within the 5’ UTR. The 1.5-kb sna shadow enhancer was placed upstream of the PP7-lacZ reporter gene. The 432-bp minimal gypsy insulators were placed in both alleles to facilitate transvection.

(B) Representative trajectories of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity in a nucleus that displays transvection. Another example is shown in Figure 2-8.

(C) Snapshots of the maximum projected images of a representative transvecting nucleus within a living embryo. MS2 and PP7 activities were visualized with MCP-GFP (green) and mCherry-PCP (red). Nuclei were visualized with His2Av-eBFP2 (blue). Scale bar indicates 5 μm. The insets are magnifications of the transcription foci within the central nucleus. Images were processed to remove background fluorescence and enhance signals. Signal trajectories from the raw image are shown in (B). Minutes into nc 14 are indicated above each panel.

(D) 3D tracking of MS2-lacZ (green) and PP7-lacZ (red) transcription foci in the transvecting nucleus shown in (B, C).

(E) Cumulative traces of MS2-lacZ (green) and PP7-lacZ (red) transcription foci during transvection in the nucleus shown in (B-D). The plot spans from 24.5 min to 39.9 min. There is stable association of the two alleles during transvection.
Figure 2-8. Transvection involves homolog pairing.
(A) A representative trajectory of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity in a nucleus that displays transvection.
(B) Snapshots of the maximum projected images of a representative transvecting nucleus from a living embryo. MS2 and PP7 activities were visualized with MCP-GFP (green) and mCherry-PCP (red). Nuclei were visualized with His2Av-eBFP2. The insets are magnifications of the transcription foci in the center nucleus. Images were processed to remove background fluorescence and enhance signals. Signal trajectories from the raw image are shown in (A). Scale bar indicates 5 µm.
(C) 3D tracking of MS2-lacZ (green) and PP7-lacZ (red) transcription foci in the transvecting nucleus shown in (A, B).
(D) Cumulative traces of MS2-lacZ (green) and PP7-lacZ (red) transcription foci during transvection in the nucleus shown in (A-C). The plot spans from 25.8 min to 55.3 min into nc 14.
Figure 2-9. 3D tracking of allelic distances during transvection.

(A-B) 3D distances between MS2 and PP7 alleles in the transvecting nuclei shown in Figure 2-7 (A) and Figure 2-8 (B).
Figure 2-10. Transvection depends on sequence homology.

(A) The MS2 and PP7 transgenes were inserted at non-homologous locations. The MS2-lacZ was inserted at VK00033 landing site (Venken et al., 2006) and the PP7-lacZ was inserted at VK00031 landing site (Venken et al., 2006).

(B) The distribution of the duration of homolog pairing between non-homologous genomic locations. A total of 964 nuclei from three independent embryos were analyzed.

(C) False-coloring of transvecting nuclei in the embryo containing MS2 and PP7 transgenes in homologous locations. The image is oriented with anterior to the left and ventral view facing up.

(D) False-coloring of transvecting nuclei in the embryo containing MS2 and PP7 transgenes in non-homologous locations. The image is oriented with anterior to the left and ventral view facing up.
Figure 2-11. Coordinated transcription across homologous chromosomes.
(A) Schematic representation of the MS2-lacZ and PP7-lacZ reporter genes containing gypsy insulators. Only the PP7-lacZ reporter gene contains the sna shadow enhancer.
(B) Representative trajectory of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity. These reporters were placed under the control of a shared enhancer as shown in (A).
(C) Schematic representation of the MS2-lacZ and PP7-lacZ reporter genes containing gypsy insulators and separate sna shadow enhancers.
(D) Representative trajectory of MS2-lacZ (green) and PP7-lacZ (red) transcriptional activity. These reporter genes were placed under the control of separate enhancers as shown in (C).
(E) The distribution of correlation coefficients between MS2-lacZ and PP7-lacZ transcription activities during transvection (red) or under the control of separate enhancers (blue). A total of 99 and 883 nuclei from four and three independent embryos were analyzed, respectively. Both histograms were plotted with a bin width of 0.1. These two distributions were found to be significantly different ($p < 1.5 \times 10^{-7}$, Mann-Whitney-U test).
Figure 2-12. Measurement of coordination between MS2 and PP7 reporters under the control of separate enhancers.

(A) Schematic representation of the MS2-lacZ and PP7-lacZ reporter genes containing the 100-bp minimal eve promoter and the 1.5-kb sna shadow enhancer. Both alleles contain a gypsy insulator.

(B) gypsy insulator was depleted from the MS2-lacZ allele.

(C) The distribution of correlation coefficients for MS2-lacZ and PP7-lacZ reporter genes containing the sna shadow enhancer. Both alleles contain a gypsy insulator (two gypsy; blue) or only the PP7 allele contains the gypsy insulator (single gypsy; gray). A total of 883 and 1019 nuclei from three independent embryos were analyzed, respectively. The plot of two gypsy is identical to the plot shown in Figure 2-11E (blue). Bin width is 0.1. These two distributions were not found to be significantly different ($p > 0.05$, Mann-Whitney-U test).

(D) Boxplot showing the distribution of MS2 burst amplitudes. 95 and 97 bursts from three independent embryos were analyzed for – cis-linked PP7-lacZ and + cis-linked PP7-lacZ, respectively. The box indicates the lower (25%) and upper (75%) quantile and the solid line indicates the median. Whiskers extend to the 10th and 90th percentile of each distribution. Asterisk indicates $p < 0.001$ (Mann-Whitney-U test).
Figure 2-13. Promoter competition attenuates the timing and the level of transvection.

(A) Schematic representation of the transvection assay without a cis-linked PP7-lacZ reporter gene. MS2-lacZ reporter gene and snα shadow enhancer are located on separate alleles. gypsy insulators were placed in both alleles to facilitate transvection. Single-color live imaging was performed to visualize MS2-lacZ reporter gene.

(B) Representative trajectory of MS2-lacZ transcriptional activity in the absence of cis-linked PP7-lacZ reporter gene. The first trans-activation can occur within first 10 min of nc 14.

(C) Schematic representation of the transvection assay with a cis-linked PP7-lacZ reporter gene.

(D) Representative trajectory of MS2-lacZ transcriptional activity in the presence of cis-linked PP7-lacZ reporter gene. There is a delay in the first MS2 burst, and the overall amplitudes of the bursts are lower as compared with the nucleus lacking the cis-linked PP7 reporter gene (B).

(E) Cumulative fraction of nuclei expressing MS2-lacZ reporter gene in the presence (blue) or absence (green) of a cis-linked PP7-lacZ. A total of 1110 (1077 inactive, 33 active) and 1151 (1119 inactive, 32 active) nuclei from three independent embryos were analyzed, respectively. Error bars represent ± the standard error of the mean of three biological replicates. Plot shown as – cis-linked PP7-lacZ is identical to Figure 2-1D.

(F) A close-up of the delay in trans-activation during the onset into nc 14 (5-15min) from the plot shown in (E). Error bars represent ± the standard error of the mean of three biological replicates.
Figure 2-14. A model for gene activation via formation of a transcription hub.
A pair of insulators establishes a stable $trans$-homolog association. Subsequently, the transcription "hub" is assembled near the site of transcription by trapping transcription factors (yellow and orange ovals) and co-activators (pink oval) at enhancers as well as Pol II complexes (blue ovals) at promoters, leading to coordinated transcription activity of both reporter genes $in cis$ and $trans$. 
Chapter 3

Temporal dynamics of pair-rule stripes
in living *Drosophila* embryos

Author Statement:
The work in Chapter 3 was previously published in *Proceedings of the National Academy of Sciences of the United States of America* (Lim et al., 2018). My contributions for this work included performing experiments, interpreting results, and editing the final manuscript.

Reference:
ABSTRACT

Traditional studies of gene regulation in the *Drosophila* embryo centered primarily on the analysis of fixed tissues. These methods provided considerable insight into the spatial control of gene activity, such as the borders of *eve* stripe 2, but yielded only limited information about temporal dynamics. The advent of quantitative live-imaging and genome-editing methods permits the detailed examination of the temporal control of endogenous gene activity. Here, we present evidence that the pair-rule genes *fushi tarazu* (*ftz*) and *even-skipped* (*eve*) undergo dynamic shifts in gene expression. We observe sequential anterior shifting of the stripes along the anterior to posterior axis, with stripe 1 exhibiting movement before stripe 2 and the more posterior stripes. Conversely, posterior stripes shift over greater distances (two or three nuclei) than anterior stripes (one or two nuclei). Shifting of the *ftz* and *eve* stripes are slightly offset, with *ftz* moving faster than *eve*. This observation is consistent with previous genetic studies, suggesting that *eve* is epistatic to *ftz*. The precision of pair-rule temporal dynamics might depend on enhancer–enhancer interactions within the *eve* locus, since removal of the endogenous *eve* stripe 1 enhancer via CRISPR/Cas9 genome editing led to precocious and expanded expression of *eve* stripe 2. These observations raise the possibility of an added layer of complexity in the positional information encoded by the segmentation gene regulatory network.

SIGNIFICANCE

Classical studies of gene activity in development focused on spatial limits due to the use of fixed tissues for analysis. The advent of live-imaging methods provides an opportunity to examine the temporal control of gene expression. Here, we used quantitative live-imaging methods to visualize dynamic shifts in the endogenous expression of *eve* and *ftz*
pair-rule stripes in the early *Drosophila* embryo. We suggest that these temporal dynamics add to the complexity encoded by the segmentation gene network.

**INTRODUCTION**

The segmentation of the *Drosophila* embryo represents one of the most extensively examined gene regulatory networks in animal development. Maternal gradients of Bicoid, Hunchback, and Caudal establish sequential, overlapping patterns of gap gene expression, which, in turn, delineate pair-rule stripes that subdivide the embryo into a repeating series of body segments (1–3). The classical view of this system invokes threshold responses of pair-rule stripe enhancers to static gradients of maternal and gap gene regulatory factors (4–6). However, quantitative imaging studies have revealed anterior shifts in gap gene expression, particularly in posterior regions of cellularizing embryos (7–9). These shifts are thought to enrich the positional information encoded by this system through dynamic changes in the distribution of regulatory gradients over time. We sought to explore how these dynamic changes in gap gene expression would be read out at the level of the pair-rule genes that they regulate.

We analyzed the temporal dynamics of the pair-rule genes *fushi tarazu* (*ftz*) and *even-skipped* (*eve*), which are responsible for specifying the even and odd parasegments, respectively (10). MS2 and PP7 RNA stem loops were inserted into the 3′ UTRs of the endogenous loci using CRISPR/Cas9-mediated genome editing (11–14). Visualization of the expression patterns in living embryos reveals highly dynamic shifts in each of the stripes, including those located in anterior regions of the embryo. These shifts are particularly striking for *ftz*, in that the initial sites of expression correspond to the interstripe
regions of the mature stripes. Simultaneous visualization of \textit{ftz} and \textit{eve} expression suggests rapid shifting of the \textit{ftz} stripes, followed by a slight lag in refinement of the \textit{eve} stripes. We propose that the differential rates of \textit{ftz} and \textit{eve} temporal dynamics contribute to the positional information encoded by the segmentation network. Moreover, we provide evidence that enhancer–enhancer interactions within the \textit{eve} locus may be important for the precision of expression.

\textbf{RESULTS AND DISCUSSION}

\textit{ftz} is regulated by four separate enhancers located upstream and downstream of the transcription unit (15, 16). Three of the enhancers initiate stripes 3+6, 2+7, and 1+5. There is no stripe 4 enhancer, and thus its expression depends on the classical “Zebra” enhancer, which directs all seven expression stripes at later stages following initiation by the individual stripe enhancers (3, 17). We created separate \textit{ftz} alleles containing 24 copies of either MS2 or PP7 RNA stem loops inserted into the 3′ UTR of the endogenous locus (Figure 3-1A). Fluorescent \textit{in situ} hybridization (FISH) of \textit{ftz-MS2} revealed expression patterns comparable to those seen in wild-type \textit{yw} embryos (Figure 3-2). Moreover, homozygotes of these alleles were fully viable, fertile, and phenotypically indistinguishable from wild-type flies. This suggests that the stem loops do not significantly affect either transcription or translation.

Twelve embryos expressing either \textit{ftz-MS2} or \textit{ftz-PP7} were analyzed, and similar results were obtained for both alleles. Nuclei were visualized with His2AV-mRFP or His2AV-eBFP2, allowing us to trace transcriptional activity from individual nuclei during nuclear cycle (nc) 14. There was a clear anterior shift in the \textit{ftz} expression pattern for all stripes.
during a 15–30-min interval of nc 14 (Figure 3-1B and C, Figure 3-3). Strikingly, while each ftz stripe was formed at different time points in nc 14, the initial ftz transcripts were always first detected in the interstripe regions of the mature stripes (Figure 3-1B and C). Comparison of ftz-MS2 mRNA accumulation along the anteroposterior (AP) axis between early and late nc 14 revealed clear anterior shifts in stripe positions (Figure 3-4). Moreover, there was an anterior-to-posterior wave in the temporal shifts whereby stripe 1 shifted earlier than stripe 2, followed by the more posterior stripes (Figure 3-1C). These sequential shifts are evocative of the wavefront of gene expression seen during somitogenesis in vertebrates (18, 19). Following the shifts, there was a refinement in each of the stripes during the latter periods of nc 14, such that the anterior boundary of each stripe was fixed and the posterior nuclei gradually lost transcriptional activity (Figure 3-1B).

eve is regulated by five separate enhancers located upstream and downstream of the transcription unit (20, 21). As for ftz, we created separate eve alleles tagged in the 3′ UTR with either MS2 or PP7 (Figure 3-5A), both of which are fully viable as homozygotes. Furthermore, eve-MS2 expression was found to be similar to wild-type expression on FISH (Figure 3-2). A total of 17 embryos carrying the eve-MS2 or eve-PP7 allele were imaged. Strong transcriptional activity was observed near the stripe 1 domain, followed by the formation of more posterior stripes. Anterior shifting of each eve stripe was seen, with the posterior stripes displaying more significant movement than stripes 1, 2, and 3 (Figure 3-5, Figure 3-5, Figure 3-6). Nonetheless, there was an anterior-to-posterior wave in dynamics, similar to that seen for ftz (Figure 3-5C and Figure 3-6C).
To determine the relative dynamics of the *ftz* and *eve* expression patterns, we examined embryos containing both *ftz-PP7* and *eve-MS2* alleles (Figure 3-8). Dual-color live imaging revealed a slightly offset timing of expression dynamics. For example, *eve* and *ftz* transcripts were initially expressed in neighboring domains. During the first 20 min of nc 14, there was a slight overlap between *eve* and *ftz* stripes, whereby the anterior border of each *ftz* stripe was tightly juxtaposed to the posterior edge of the corresponding *eve* stripe (e.g., 20 min; Figure 3-8). The two sets of stripes did not adopt their defining complementary and evenly spaced expression patterns until late phases of cellularization (e.g., 40 min; Figure 3-8). Because overlap occurred only at the posterior boundary of *eve* and the anterior boundary of *ftz*, the simplest interpretation is that *ftz* stripes shift faster than *eve*. This idea is consistent with previous genetic studies suggesting dominance of *eve* in the pair-rule gene hierarchy, with *eve* exerting a strong repressive effect on *ftz* expression (22, 23).

To determine the relative rates of *ftz* and *eve* temporal dynamics, we conducted various quantitative analyses. We measured the fraction of active nuclei for *ftz* and *eve* within the mature expression limits. In general, *eve* nascent transcripts persisted longer within the *ftz* expression domains than do *ftz* transcripts within the *eve* domains (Figure 3-9). For example, *ftz* transcripts were rapidly lost in the *eve 2* domain as *eve* transcripts are being activated. On the other hand, *eve* transcripts were detected in the *ftz 2* domain even after the onset of *ftz* transcription (Figure 3-9, Top). These results support the idea that *ftz* stripes shift faster than *eve*. This might be due to asymmetric repression of *ftz* by Eve (22, 23).
Previous studies have documented the spatial precision of individual pair-rule stripes, such as *eve* stripe 2 (24–26). Our study suggests that there are similar constraints on the temporal dynamics. The anterior stripes shift first and cover a distance of just one or two nuclei, while the posterior stripes shift last and move over greater distances of two or three nuclei. Stripe 1 displays a broad expression profile at the onset of nc 14, with expression extending into the future limits of *eve* stripe 2 (Figure 3-5). Stripe 2 begins to form only after stripe 1 begins to refine and retreat from the stripe 2 territory. Thus, it is possible that stripe 1 exerts an “organizing” activity in the patterning of the early embryo. To explore this idea, we removed the stripe 1 enhancer, which is located downstream of the *eve* transcription unit, by CRISPR/Cas9 genome editing (Figure 3-10A). The stripe 1 deficiency was created in the context of the MS2-tagged allele. Trans-heterozygotes containing this MS2 allele along with the unmodified PP7-tagged allele were visualized using two-color imaging with MCP-GFP and mCherry-PCP fusion proteins (Figure 3-10).

As expected, very little MS2 signal was detected in the vicinity of stripe 1 during the initial phases of nc 14 due to the removal of the stripe 1 enhancer. Only the PP7 allele was expressed in this territory, while the MS2 allele was activated in the stripe 1 domain at later stages due to the late *eve* enhancer (Figure 3-10B) (27). However, *eve* stripe 2 expression from the deficiency allele (green) appeared earlier and extended more anteriorly into the interstripe region compared with the wild-type stripe 2 pattern (red). This broader pattern persisted for 10–15 min before refining into a more or less normal stripe 2 pattern (Figure 3-10B and C). These results suggest that *eve* stripe 2 takes over as the dominant stripe on removal of the stripe 1 enhancer. It will be interesting to see whether removal of both the
stripe 1 and stripe 2 enhancers would result in dominance of stripe 3 expression. Dominance of anterior stripes is consistent with the anterior-to-posterior coordination in the stripe shifts seen for both *eve* and *ftz*. It is also possible that this coordination depends on enhancer–enhancer interactions within the *eve* locus.

In summary, we have presented evidence for dynamic shifts of *ftz* and *eve* stripes. These shifts are slightly offset, with the *ftz* stripes shifting and refining faster than the corresponding *eve* stripes. These temporal dynamics have the potential to augment the positional information encoded by the segmentation regulatory hierarchy.
MATERIALS AND METHODS

Fly Strains

*eve-MS2* and *eve-PP7* were generated using CRISPR/Cas9-based insertion of 24× MS2 RNA stem loop or 24× PP7 RNA stem loop into the 3‘ UTR of endogenous *eve*. In brief, approximately 1-kb DNA fragments of 5‘ and 3‘ homology arm sequences were PCR-amplified from the genomic DNA and then inserted into the pBS-MS2-loxP-dsRed-loxP and pBS-PP7-loxP-dsRed-loxP donor plasmids (28). These plasmids were coinjected with the pCFD3 gRNA expression plasmid to nos-Cas9 embryos (29). *ftz-MS2* and *ftz-PP7* have been described previously (28).

Genome Editing by CRISPR/Cas9

For insertion of MS2 and PP7 stem loops into the 3‘ UTR of *eve* locus, pCFD3 gRNA expression plasmid and pBS-dsRed donor plasmid were coinjected to nos-Cas9 embryos (29). Microinjection was performed as described previously (30). 3xP3-dsRed was used for subsequent screening. For deletion of *eve* stripe 1 enhancer, two pCFD3 gRNA expression plasmids, pBS-GFP donor plasmid, and pBS-Hsp70-Cas9 plasmid (Addgene; 46294) were coinjected to *eve-MS2* embryos. 3xP3-GFP was used for subsequent screening.

Plasmids

*pCFD3-dU6-eve 3´UTR gRNA*

Two DNA oligos (5´-GTC GAG AGA GTG TGT GTG GAT CG-3´) and (5´-AAA CCG ATC CAC ACA CAC TCT CT-3´) were annealed and inserted into the pCFD3-dU6:3gRNA vector (addgene # 49410) using BbsI sites.
**pCFD3-dU6-eve stripe1 gRNA1**

Two DNA oligos (5’-GTC GGG CTT TTG ATA CGC TCC TGG-3’) and (5’-AAA CCC AGG AGC GTA TCA AAA GCC-3’) were annealed and inserted into the pCFD3-dU6:3gRNA vector (addgene # 49410) using BbsI sites.

**pCFD3-dU6-eve stripe1 gRNA2**

Two DNA oligos (5’-GTC GGA AGT TCC TCG GGG CCT CGC-3’) and (5’-AAA CGC GAG GCC CCG AGG AAC TTC -3´) were annealed and inserted into the pCFD3-dU6:3gRNA vector (addgene # 49410) using BbsI sites.

**pBS-eve5´-MS2-loxP-dsRed-loxP-eve3´**

A DNA fragment containing 3’ homology arm was amplified from genomic DNA using primers (5’-GGG GGT CTA GAT CCA CAC ACA CTC TCT CCC C-3’) and (5’-CCC CCG CGG CCG CCA TCC TTG TGC CTC GCT CGC ATC G-3´), and digested with Xbal and NotI. The resulting fragment was inserted between the NOTI and SpeI sites of pBS-MS2-loxP-dsRed-loxP (28). Subsequently, a DNA fragment containing 5’ homology arm was amplified from genomic DNA using primers (5’-TTC TTG TCG ACG ATT AGC ACC GTT CCG CTC AGG CTG-3´) and (5’-CCC CCG AAT TCT CGC GGG CTT ACG CCT CAG TC-3´), and digested with SalI and EcoRI. The resulting fragment was inserted between the SalI and EcoRI sites of the plasmid.

**pBS-eve5´-PP7-loxP-dsRed-loxP-eve3´**

A DNA fragment containing 3’ homology arm was amplified from genomic DNA using primers (5’-GGG GGT CTA GAT CCA CAC ACA CTC TCT CCC C-3’) and (5’-CCC CCG CGG CCG 3 CCA TCC TTG TGC CTC GCT CGC ATC G-3´), and digested with Xbal and NotI. The resulting fragment was inserted between the NOTI and SpeI sites of
pBS-PP7-loxP-dsRed-loxP (28). Subsequently, a DNA fragment containing 5´ homology arm was amplified from genomic DNA using primers (5´-TTC TTG TCG ACG ATT AGC ACC GTT CCG CTC AGG CTG-3´) and (5´-CCC CCG AAT TCT CGC GGG CTT ACG CCT CAG TC-3´), and digested with SalI and EcoRI. The resulting fragment was inserted between the SalI and EcoRI sites of the plasmid.

**pBS-loxP-GFP-loxP**

A DNA fragment containing loxP-3xP3-GFP was amplified from pHD-loxP-3xP3-eGFP SV40 (31) using primers (5´-TTT AAC TGC AGA TGC ATA AGG CGC GCC TAG G-3´) and (5´- GGA AAA GAT CTT AAG ATA CAT TGA TGT TGG AC-3´) and digested with PstI and BglII. The resulting fragment was inserted between the PstI and BglII sites of pBS-loxP (28).

**pBS-eve stripe1 5´-loxP-GFP-loxP- eve stripe1 3´**

A DNA fragment containing 3´ homology arm was amplified from genomic DNA using primers (5´-CGC GCG GCG GCC GCA GGC CCC GAG GAA CTT CCT G-3´) and (5´- CGC GCG GAG CTC GTA CAC CTC GCA TCT CCA ATC-3´), and digested with NOTI and SacI. The resulting fragment was inserted between the NOTI and SacI sites of pBS-loxP-GFP-loxP. Subsequently, a DNA fragment containing 5´ homology arm was amplified from genomic DNA using primers (5´-CGC GCG CTC GAG CTC CTG CTG TCC TTG GTG TG-3´) and (5´- CGC GCG GAT ATC GGA GCG TAT CAA AAG CCG CG -3´), and digested with XhoI and EcoRV. The resulting fragment was inserted between the XhoI and EcoRV sites of the plasmid.
Single-Color MS2 Live Imaging

MCP-GFP, His2Av-mRFP virgins were mated with homozygous males carrying MS2 alleles to image *ftz-MS2* or *eve-MS2* embryos (Figure 3-1 and 3-3 to 3-7). The resulting embryos were dechorinated, mounted between a semipermeable membrane (In Vitro Systems & Services) and a coverslip (18 mm × 18 mm), and then embedded in Halocarbon oil 27 (Sigma-Aldrich). Embryos were imaged with a Zeiss LSM 880 laser scanning confocal microscope at room temperature, using a Plan-Apochromat 40×/1.3 NA oil immersion objective. At each time point, a stack of 26 images separated by 0.5 μm were captured in 16 bits.

Dual-Color MS2/PP7 Live Imaging

Dual-color MS2/PP7 live imaging was used to image embryos carrying *ftz-PP7/eve-MS2*, *eve-PP7/eve-Δstripe 1 enhancer-MS2*, and *eve-PP7/eve-MS2* alleles (Figure 3-8 and 3-10). MCP-GFP, mCherry-PCP, His2Av-eBFP2 virgins were mated with homozygous males carrying either the PP7 or MS2 allele (28). Resulting trans-heterozygote virgins were collected and mated with homozygous males carrying the MS2 or PP7 allele. The resulting embryos were prepared as described above for single-color live imaging.

In Situ Hybridization

Embryos were dechorionated and fixed in fixation buffer (0.5× PBS, 25 mM EGTA, 4% formaldehyde, and 50% heptane) for 20 min at room temperature. Antisense RNA probes labeled with digoxigenin (DIG RNA Labeling Mix 10× concentration; Roche) and biotin (Biotin RNA Labeling Mix 10× concentration; Roche) were used to detect *ftz* and *eve* RNAs, respectively. Template DNA for *ftz* probes was amplified from genomic DNA using primers (5′-CGT AAT ACG ACT CAC TAT AGG GTG GGG AAG AGA GTA ACT

80
GAG CAT CGC-3') and (5'-ATT CGC AAA CTC ACC AGC GT-3'). Template DNA for *eve* probes was amplified from genomic DNA using primers (5'-CGT AAT ACG ACT CAC TAT AGG GGT GTG TGG ATC GCG GGC TTA CGC C-3') and (5'-ACA CTC GAG CTG TGA CCG CCG C-3'). Hybridization was performed at 55 °C overnight in hybridization buffer (50% formamide, 5× SSC, 50 μg/mL heparin, 100 μg/mL salmon sperm DNA, and 0.1% Tween-20). Subsequently, embryos were washed with hybridization buffer at 55 °C and incubated with Western Blocking Buffer (Roche) at room temperature for 1 h. Then the embryos were incubated with sheep anti-digoxigenin (Roche) and mouse anti-biotin primary antibodies (Invitrogen) at 4 °C for overnight, followed by incubation with Alexa Fluor 488 donkey anti-sheep (Invitrogen) and Alexa Flour 555 goat anti-mouse (Invitrogen) fluorescent secondary antibodies at room temperature for 2 h. DNA was stained with Hoechst 33342 (Thermo Fisher Scientific), and embryos were mounted in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). Imaging was performed with a Zeiss LSM 880 confocal microscope, and maximum projections of Z-stacks are shown.

**Image Analysis**

All the processing and analysis were implemented in MATLAB R2017b (MathWorks).

**Nuclei Segmentation and Tracking.**

His2Av-mRFP and His2Av-eBFP2 were used to segment nuclei for single-color and dual-color imaging, respectively. Nuclei-labeled channel images were preprocessed with Gaussian filtering and adaptive histogram equalization to enhance the signal-to-noise contrast. Nuclei were then watershedded to further separate and distinguish neighboring nuclei. The number and positions of separate nuclei within a frame were obtained. Nuclei
tracking within nc 14 was done by finding the object with minimal movement across the frames of interest (from approximately 4 min into nc 14 to the onset of gastrulation).

**Recording of MS2 and PP7 Signals.**

Maximum projections of raw images were used to record fluorescence intensities. Fluorescence intensities within each segmented nucleus was extracted. After subtracting the background nuclear signal, the signals of MS2 and PP7 transcription foci were determined by taking an average of the top three pixels with the highest fluorescence intensity within each nucleus.

**Spatial Profile of ftz and eve Transcriptional Activity and mRNA Accumulation Along the AP Axis.**

To obtain spatial profiles of ftz and eve transcriptional activity along the AP axis, each embryo was divided into bins with equal width (10 μm per bin). In addition, to minimize the effect of ftz or eve curvature on the spatial profile, the middle 40-μm region along the dorsoventral axis was defined as the region of interest. Then, for each bin with a size of 10 μm × 40 μm, MS2 or PP7 intensity of all the nuclei within the bin was obtained. The average intensity value per bin was plotted for each frame, with an error bar representing the SEM of all the nuclei within the bin. mRNA accumulation over a specified period was obtained by taking the area of transcriptional trajectory per nucleus over the designated time. Then the spatial plot of mRNA accumulation was plotted in the same way, by taking the average mRNA accumulation value from all the nuclei within each bin.

**False Coloring**

Active nuclei were defined as those with an MS2 or PP7 signal above a threshold value. The threshold value was determined by taking 15% of the maximum fluorescence intensity
of each embryo throughout the entirety of nc 14. For each frame, all the nuclei with an MS2 or PP7 signal above the threshold value were identified. Using the segmentation mask, these active nuclei were colored either red or green, with a fixed pixel intensity, and layered over the raw His2Av-mRFP or His2Av-eBFP2 image in a given frame. Mature ftz and eve domains were defined as the nuclei showing active MS2 or PP7 signals above the threshold value after 40 min into nc 14 for longer than 2 min. Indices of these nuclei were obtained, and for each frame these nuclei were colored in a single color with a fixed intensity. False coloring of mRNA accumulation was done by coloring the nuclei with the pixel intensity value corresponding linearly to its mRNA accumulation.
REFERENCES


Figure 3-1. Dynamics of endogenous ftz expression pattern.

(A) Schematic of endogenous ftz gene locus and ftz enhancers. Here, 24 repeats of MS2 or PP7 stem loops are inserted into the 3′ UTR.

(B) Snapshots from an embryo expressing ftz-MS2. Nuclei that will form mature ftz stripes are false-colored in red. Mature ftz stripes are defined as the nuclei that exhibit active transcription after 40 min into nc 14. Nuclei that show active transcription in a given frame are false-colored in green. Yellow nuclei represent those that show active transcription within the mature ftz domain. Nuclei were visualized with His2Av-mRFP and are shown in blue. Note that stripes are initially formed at one to three cells posterior to their final position. Time represents minutes after the onset of nc 14.

(C) Average spatial profile of ftz transcriptional activity along the AP axis during nc 14. Each plot corresponds to the snapshot shown in B. The AP axis was divided into 35 bins, and an average transcriptional activity from all the nuclei within each bin was plotted. The error bar represents ±SEM of all the nuclei within each bin. Blue columns represent the location of mature ftz stripes 1–7.
Figure 3-2. *In situ* images of *yw*, *ftz*-MS2, and *eve*-MS2 embryos.
Embryos are stained with *eve* (red), *ftz* (green), and DAPI (blue). Insertion of MS2 stem loops does not disrupt the endogenous *eve* or *ftz* patterns.
Figure 3-3. Snapshots from an embryo expressing *ftz-MS2*.
Nuclei are visualized with His2Av-mRFP and are shown in blue. *ftz-MS2* allele is visualized with MCP-GFP and the transcripts are shown in green. Nuclei within the mature *ftz* stripe domain are false-colored in red. Mature *ftz* stripes are defined as the nuclei that show active transcription after 40 min into nc 14. Time represents minutes after the onset of nc 14. The same snapshots are shown in Figure 3-1, with *ftz-MS2* active nuclei false-colored in green.
Figure 3-4. Dynamics of endogenous *ftz* expression.

(A) A *ftz-MS2* embryo that shows *ftz* mRNA accumulation during early and late nc 14. Each nucleus is false-colored with respect to the level of corresponding mRNA accumulation. Red represents the mRNA output between 10-25 min into nc 14, and green represents the mRNA output between 30-45 min into nc 14.

(B) Average spatial profile of *ftz* mRNA accumulation along the AP axis, between 10-25 min (red) and 30-45 min (green) into nc 14. Error bars represent standard error of the mean of all the nuclei that are within each bin (35 bins along the AP axis). The data came from the embryo shown in (A). Snapshots of the same embryo are shown in Figure 3-1.

(C) Another replicate of *ftz-MS2* embryo where nuclei are false-colored with respect to the mRNA accumulation during early (red) and late (green) nc 14.

(D) Average spatial profile of *ftz* mRNA accumulation along the AP axis, from an embryo shown in (C).

(E-F) *ftz-PP7* embryos visualizing stripe 1-4 (E) and 5-7 (F), where mRNA accumulation during early and late nc 14 are false-colored in red and green, respectively.

(G-H) Average spatial profile of *ftz* mRNA accumulation along the AP axis, from an embryo shown in (E) and (F).
Figure 3-5. Dynamics of endogenous eve expression pattern.
(A) Schematic of endogenous eve gene locus and eve enhancers. Here, 24 repeats of MS2 or PP7 stem loops are inserted into the 3’ UTR.
(B) Snapshots from an embryo expressing eve-MS2, showing stripes 1–4. Nuclei that will form mature eve stripes are false-colored in red, and nuclei with active transcription in a given frame are false-colored in green. Yellow represents actively transcribing nuclei within the mature eve domain. Nuclei were visualized with His2Av-mRFP and are shown in blue. Mature eve stripes are defined as the nuclei that exhibit active transcription after 40 min into nc 14. Each stripe shifts from posterior to anterior, and anterior stripes shift earlier than posterior stripes. Time represents minutes after the onset of nc 14.
(C) Average spatial profile of eve transcriptional activity along the AP axis during nc 14. The plot corresponds to the snapshot shown in B. Error bars represent ±SEM of all the nuclei within each of the 25 bins along the AP axis. Blue columns represent the location of mature eve stripes 1–4.
Figure 3-6. Anterior shift of *eve* expression patterns.

(A) Schematic of endogenous *eve* gene locus and *eve* enhancers. 24 repeats of MS2 or PP7 stem loops are inserted into the 3’ UTR.

(B) Snapshots from an embryo expressing *eve-MS2*, visualizing stripes 5-7. Nuclei that will form mature *eve* stripes are false-colored in red, and nuclei with active transcription in a given frame are false-colored in green. Yellow represents actively transcribing nuclei within the mature *eve* domain. Nuclei are visualized with His2Av-mRFP and are shown in blue.

(C) Average spatial profile of *eve* transcriptional activity along the AP axis during nc 14. Error bars represent standard error of the mean of the nuclei that are within each bin (25 bins along the AP axis). Blue columns represent the location of mature *eve* stripes 5-7.
Figure 3-7. Dynamics of endogenous eve expression.

(A-B) A eve-MS2 embryo that shows eve mRNA accumulation during early and late nc 14 in red and green, respectively. Each embryo visualizes eve stripe 1-4 (A) and stripe 5-7 (B). Snapshots of the same embryo are shown in Figure 3-5 and Figure 3-6.

(C-D) Average spatial profile of eve mRNA accumulation along the AP axis during nc 14 for the embryos shown in (A) and (B). Error bars represent the standard error of the mean of all the nuclei that are within each bin (25 bins along the AP axis).

(E-F) Biological replicates of eve-MS2 embryos that show eve mRNA accumulation

(G-H) Average eve mRNA accumulation of embryos shown in (E) and (F).
Figure 3-8. Relative dynamics of *ftz* and *eve*.

(A) Snapshots from an embryo expressing *ftz*-PP7 and *eve*-MS2, showing stripes 1–4. Nuclei that exhibit active *ftz* and active *eve* transcripts in a given frame are false-colored in red and green, respectively. Nuclei are shown in yellow when both *ftz* and *eve* are expressed in a given nucleus. Nuclei were visualized with His2Av-eBFP2 and are shown in blue. Time represents the minutes after the onset of nc 14.

(B) Average spatial profile of *ftz* (red) and *eve* (green) transcriptional activity along the AP axis during nc 14. Error bars represent ±SEM of all the nuclei within each of the 25 bins along the AP axis. Red and green columns represent the location of mature *ftz* and *eve* stripes 1–4, respectively.
Figure 3-9. Relative dynamics of endogenous ftz and eve.
(A) Fraction of actively transcribing nuclei that express ftz-PP7 (red) or eve-MS2 (green) within the mature eve stripe 2-6 domain.
(B) Fraction of actively transcribing nuclei that express ftz-PP7 (red) or eve-MS2 (green) within the mature ftz stripe 2-6 domain.
Figure 3-10. Dynamics of eve in the absence of the stripe 1 enhancer.

(A, Top) Schematic of eve locus and enhancers, with eve stripe 1 enhancer deleted via CRISPR/Cas9. Here, 24 copies of MS2 stem loops are inserted at the 3’ UTR.

(A, Bottom) Schematic of wild-type eve locus and enhancers with 24 copies of PP7 stem loops inserted at the 3’ UTR.

(B) Snapshots from an embryo expressing eve-Δstripe 1-MS2 (green) and eve-PP7 (red). Overlap of MS2 and PP7 alleles makes false-colored nuclei appear yellow. Nuclei were visualized with His2Av-eBFP2 and are shown in blue. Time represents minutes after the onset of nc 14. Note the precocious MS2 expression anterior to the eve stripe 2.

(C) Snapshots from an embryo expressing eve-MS2 (green) and eve-PP7 (red). Overlap of MS2 and PP7 alleles makes false-colored nuclei appear yellow. Nuclei were visualized with His2Av-eBFP2 and are shown in blue. Time represents minutes after the onset of nc 14.
Chapter 4

Large distances separate co-regulated genes in living *Drosophila* embryos.

Author Statement:

The work in Chapter 4 has been submitted for publication to *Proceedings of the National Academy of Sciences of the United States of America*. My contributions for this work included designing the study, performing experiments, analyzing data, and writing the final manuscript. I previously presented this work at the 13\(^{th}\) EMBL Conference on Transcription and Chromatin (2018)

Reference:

Heist T, Fukaya T, and Levine M (2019). Large distances separate co-regulated genes in living *Drosophila* embryos. In review at PNAS.
**ABSTRACT**

Transcriptional enhancers are short segments of DNA that switch genes on and off in response to a variety of cellular signals. Many enhancers map quite far from their target genes, on the order of tens or even hundreds of kilobases. There is extensive evidence that remote enhancers are brought into proximity with their target promoters via long-range looping interactions. However, the exact physical distances of these enhancer-promoter interactions remain uncertain. Here, we employ high-resolution imaging of living *Drosophila* embryos to visualize the distances separating linked genes that are co-regulated by a shared enhancer. Co-transvection assays (linked genes on separate homologs) suggest a surprisingly large distance during transcriptional activity: at least 100-200nm. Similar distances were observed when a shared enhancer was placed into close proximity with linked reporter genes *in cis*. These observations are consistent with the occurrence of “transcription hubs”, whereby clusters (or condensates) of multiple RNA Polymerase II complexes and associated cofactors are periodically recruited to active promoters. The dynamics of this process might be responsible for rapid fluctuations in the distances separating the transcription of co-regulated reporter genes during transvection. We propose that enhancer-promoter communication depends on a combination of classical looping and linking models.

**SIGNIFICANCE**

The prevailing model of gene activation is the looping of a distal enhancer to its target promoter, although it remains uncertain whether enhancers must come into direct contact. Recent studies suggest that clusters of RNA Polymerase II are recruited to active genes, raising the possibility that enhancers cannot get too close to their target promoters due to
molecular crowding. Here, we use live imaging methods to visualize genes that are simultaneously regulated by a single enhancer both in cis and in trans. We present evidence that these co-regulated genes are separated by a substantial distance – at least 100-200nm – during transcription. These observations challenge the textbook view of enhancer function and instead suggest an action-at-a-distance model for gene expression.

**INTRODUCTION**

There is emerging evidence for transcription hubs, whereby gene activation is mediated by the recruitment of clusters or condensates of multiple RNA Polymerase II (Pol II) complexes (reviewed in 1). Transient clusters of Pol II were detected in living mammalian cells using high-resolution live imaging methods (2). At least some of these clusters appear to be associated with foci of active transcription, visualized by the insertion of MS2 RNA stem loops into defined genes (3, 4). *In vitro* and *in vivo* assays suggest that the low complexity domains contained in many components of the pre-initiation complex (PIC) mediate liquid-liquid condensation, including Pol II, TAF, and different subunits of the Mediator complex (5-7). Coordinate activation of linked reporter genes by a shared enhancer lends further support for the occurrence of transcription hubs (8, 9).

This hub model contrasts with the classical view of gene expression, which suggests the formation of a single PIC and recruitment of individual Pol II complexes at active promoters (10). An important implication of the hub hypothesis is that distal enhancers need not, and indeed cannot, come into close contact with their target promoters due to molecular crowding caused by the collective size of multiple Pol II complexes and other components of the PIC (11, 12).
Hi-C (chromosome conformation capture) contact maps have been interpreted to suggest close contact of distal enhancers with their target promoters (13). However, recent high-resolution imaging methods raise the possibility that distal enhancers can map quite far during transcriptional activation. For example, activation of a synthetic lacZ reporter gene by endogenous even-skipped (eve) stripe enhancers is detected in living embryos over apparent distances of 300-400 nm (14). Similar distances (~300-350nm) were observed between the Sox2 control region (SCR) enhancer and Sox2 promoter in mouse embryonic stem cells (15). High-resolution fixed tissue methods are consistent with such distances in the Bithorax complex in Drosophila embryos (16).

Here, we use a recently reported transvection assay and high-resolution imaging of living Drosophila embryos to measure the distances separating co-expressed reporter genes that are regulated by a single enhancer across homologous chromosomes (9). These studies reveal large distances of nascent transcripts associated with active alleles, on the order of 250-300 nm. These distances include a relatively large standard deviation (~120nm). We present evidence that this variance is not strictly due to technical or biological noise in our measurements, but instead might arise from rapid fluctuations in the distances separating coordinately expressed alleles. We discuss the implications of these measurements with the hub hypothesis and the dynamics of Pol II aggregation/condensation at active promoters.

RESULTS AND DISCUSSION

As a first step towards estimating the distance between a distal enhancer and its target promoter during gene activation, we performed high-resolution measurements of
coordinately active alleles during transvection (summarized in Figure 4-1A). Briefly, this assay employs a PP7 reporter gene with a closely-linked upstream snail (sna) shadow enhancer, which mediates expression in the presumptive mesoderm (17, 18). The other allele contains an MS2 reporter gene that lacks its own enhancer. Consequently, MS2 expression depends on pairing of the alleles to allow the sna shadow enhancer on the PP7 allele to act in trans (9). To improve the frequency of pairing and transvection, gypsy insulator DNAs were placed upstream of the two reporter genes.

To address this question of distances separating enhancers and target genes, we measured the center-to-center distances of active MS2 and PP7 foci in XY dimensions (representative nuclei are shown in Figure 4-1B). We focused on the XY axes due to technical difficulties in achieving similar resolution in the axial (Z) dimension. To compensate for this emphasis on XY resolution, we observed a large number of transcription foci to ensure reliable capture of the distances separating dynamically expressed alleles. As a control, we analyzed a reporter gene that contains interleaved MS2 and PP7 stem loops (19). These measurements suggest a high degree of localization precision, incorporating both technical noise and noise arising from the MS2/PP7 system, with 32 ± 22 nm in the X axis and 26 ± 20 nm in Y (Figure 4-2).

During transvection, the two foci of transcription do not coincide, but instead display considerable spatial separation. Rapid fluctuations cause some variations in the distances separating the PP7 and MS2 nascent transcripts within individual nuclei during the acquisition window (e.g., Figure 4-1B). We discuss these fluctuations in more detail below.
Surprisingly large distances were found to separate the two alleles during transvection: 282 ± 124 nm (Figure 4-1C), although they are considerably smaller than those seen for MS2 and PP7 alleles that are regulated by separate enhancers (Figure 4-3). This underscores the fact that coordinate transcription of PP7 and MS2 by a shared enhancer depends on pairing of the two alleles during transvection and the resultant increase in approximate physical proximity. Of course, these measurements represent only an approximation of the distance separating the sna shadow enhancer and MS2 target promoter since we are visualizing nascent transcripts. However, previous studies suggest that RNAs are tightly associated with the chromatin templates from which they are derived (20) and therefore provide a reasonable proxy for visualizing enhancer-promoter interactions across homologous chromosomes.

The distances separating transvecting PP7 and MS2 alleles are consistent with the hub hypothesis, since they are in the range of the combined Stokes radii of ~10 Pol II complexes (11). Further support for this view stems from measurements of coordinately expressed PP7 and MS2 alleles linked in cis in close proximity on the same chromosome (Figure 4-4). We measured the distances separating coordinately expressed MS2 and PP7 reporter genes organized in distal (Figure 4-4A) or proximal (Figure 4-4B) arrangements. The promoter regions of the reporter genes map ~16.5kb away from one another in the distal arrangement. In this configuration, the distance between the transcription foci is an average of 202 ± 111 nm (Figure 4-4C, E), somewhat closer than that seen in the transvection assay. Surprisingly, the distance separating the transcription foci is larger for the proximal arrangement, whereby the PP7 and MS2 promoters are separated by just ~3.5kb. The
average distance that was measured is 268 ±119nm, more similar to the distance seen in the transvection assay (Figure 4-4D, E).

Why would the proximal arrangement exhibit a larger distance than the distal arrangement? These distances correlate with the level of transcriptional activity (Figure 4-4F). In the proximal orientation, the sna shadow enhancer is located immediately upstream of both the MS2 and PP7 reporter genes and the total fluorescence intensity (burst size) is ~33% higher than that observed when the enhancer is positioned in the distal orientation (8). We suggest that this difference in burst size is due to the different numbers of Pol II complexes that are recruited. More Pol II is recruited when the enhancer is proximal, leading to a larger separation of the two reporter genes. In contrast, fewer Pol II complexes might be recruited when the shared enhancer is located in the distal orientation. It is possible that this same phenomenon might occur between transgenic constructs inserted at different genomic locations (Figure 4-5). An alternative possibility is that the greater distances separating the proximal vs. distal orientations of the PP7 and MS2 reporters are due to divergent elongation of Pol II along reporter genes positioned in opposing directions (see below).

As discussed earlier, there is a relatively large standard deviation in our measurements, ~110-120nm. This does not appear to be strictly due to technical and biological noise (Figure 4-2), but rather, arises (at least in part) from rapid fluctuations in the physical separation of the PP7 and MS2 reporter genes during their coordinate expression (Figure 4-6). We examined these fluctuations in >100 living nuclei and several examples are presented in Figure 4-6 (A-D), e.g., the nucleus shown in panel D exhibits a full fluctuation
cycle over an interval of ~40 seconds. The timing and distances of these fluctuations are variable, with distances of 50 to 200nm and time intervals of ~10 to 40 seconds (Figure 4-7).

It is possible that these fluctuations are caused by the dynamics of Pol II recruitment. Perhaps Pol II clusters of variable sizes are recruited to the hub. Large clusters (e.g., 10-20 Pol II complexes) lead to a larger separation (e.g., ~200nm) of the active reporter genes as compared with small clusters (5-6 Pol II complexes; ~100nm). As Pol II is released and undergoes elongation, the reporter genes may come into closer proximity with one another. An alternate possibility is that these fluctuations might be a consequence of dynamic ‘kiss and run’ interactions, whereby an enhancer comes into sporadic and transient contact with its target promoter in order to trigger transcription (4). However, we do not favor this view as we observed several instances of co-regulated transcription foci that are separated by substantial distances (>200nm) during the entirety of a fluctuation cycle (Figure 4-7).

We believe that the transvection assay (Figure 4-1) provides the best opportunity for estimating the distance separating an enhancer from its target promoter due to the tight linkage of the sna shadow enhancer with the PP7 reporter gene and the uncoupling of the MS2 target gene on the other homologous chromosome. As a result, the PP7 signals provide a reasonable estimate for the position of the enhancer. As discussed above, the mean distance separating PP7 and MS2 signals in the transvection assay is ~250-300nm. However, this might overestimate the actual distance between the PP7-linked enhancer from the MS2 target promoter, as we are measuring nascent transcripts that are located
downstream of the transcription start state. If we conservatively apply our error in localization (Figure 4-2) to our measurements, the minimal distance separating the two alleles is then on the order of ~200nm. Further, with the fluctuations as described above, it is possible that minimum distances of ~100nm separate the enhancer from its target promoter at transient periods during transcription, although distances of 200-300nm (or more) are equally plausible. We argue that even a distance of ~100nm is larger than what is expected by traditional looping models.

In summary, we have presented evidence that large distances separate coordinately-expressed reporter genes regulated by a shared enhancer. These distances are consistent with the transcription hub model, whereby clusters or condensates—not individual complexes—of Pol II are recruited to active foci. Further support for this view is the correlation between the distances of separation and the size of transcriptional bursts (summarized in Figure 4-8). In particular, proximally-linked reporter genes (Figure 4-8C) display a greater separation than distal linkage (Figure 4-8B). We propose that this difference is due in part to the number of Pol II complexes recruited to active foci. The magnitude of the distances documented in this study are consistent with several recent studies that have employed various measurement methods, in both fixed and living tissues (14-16). The picture that is emerging from these studies is that enhancer-promoter communication is mediated by a combination of the classical looping and linking models. Once enhancers come into proximity, they influence transcription via many protein-protein interactions within a hub across the distance of the enhancer to its promoter.
MATERIALS AND METHODS

Fly Strains

The fly strains gypsy-sna shadow enhancer-evePr-PP7-lacZ (9), gypsy-sna shadow enhancer-evePr-MS2-lacZ (9), gypsy-evePr-MS2-lacZ (9), evePr-MS2-yellow-sna shadow enhancer-yellow-PP7-evePr (8), snaPr-MS2-yellow-sna primary enhancer-yellow-PP7-snaPr (8), hb proximal enhancer-hb P2 promoter-12x(PP7-MS2)-lacZ (19) used in this study were previously published and have all been integrated into the third chromosome (VK00033). The pbphi-snaPr-MS2-yellow-sna primary enhancer-yellow-PP7-snaPr plasmid (8) was also injected and integrated into the second chromosome (VK00002) by BestGene.

Cloning and Transgenesis

The pbphi-yellow-MS2-snaPr-linker-sna shadow-linker-snaPr-PP7-yellow plasmid was constructed as previously described (8), except the insertion of the DNA fragment containing linker-snaPr-MS2-yellow in the opposite orientation was selected after sequencing. Injection was done as previously described (21) into the third chromosome (VK00033).

Live Imaging

For the assays containing both MS2 and PP7 transgenes on the same allele, MCP-GFP, mCherry-PCP, His2Av-eBFP2 homozygous virgins (9) were mated with homozygous males containing the transgenic construct. For the assays containing MS2 and PP7 transgenes on separate alleles, MCP-GFP, mCherry-PCP, His2Av-eBFP2 homozygous virgins were first mated with homozygous males carrying the PP7-containing transgenic construct. Resultant trans-heterozygote virgins were obtained and mated with homozygous
males carrying the MS2-containing transgenic construct. For all crosses, resulting embryos were treated in preparation for imaging as previously described (9). All imaging was done at room temperature using a Zeiss LSM 880 with the Fast Airyscan module and using a Plan-Apochromat 63x/1.4 NA oil immersion objective. The 405nm, 488nm, and 561nm lasers were used across all imaging. For the transvection assay, 30 z-slices of 0.17um were acquired per time point. For all other experiments, either 45 or 60 z-slices of 0.17um were acquired per time point. For the experiments examining the two cis-linked reporter genes at different integration sites (Figure 4-5), 0.7X Nyquist XY sampling was used. For all other experiments, 1.8X Nyquist XY sampling was used. For assays in which we compared fluorescence intensities across constructs (Figure 4-4), laser power was kept constant. Images were captured in 8 bit.

**Image Analysis**

All images were first deconvolved and processed in Zen Black, version 2.3 (Zeiss). Processed images were then moved to Imaris, version 9.2.1 (Bitplane). Objects were fit to MS2 and PP7 foci using the Surfaces tool and tracked across time in the processed movie. For each time point in each nucleus, the centers of both foci were recorded. Time points in which a single MS2 or PP7 focus was split into two distinct foci, likely due to sister chromatid separation, were excluded from further analysis by manual curation. Four-color 200nm diameter Tetraspeck beads (Thermo Fisher) were imaged as described above and the centers of each bead signal upon 488nm and 561nm excitation were calculated. The optical offsets in these channels were used to fit a second order polynomial transform in MATLAB R2016b (Mathworks) using the cp2tform function (22). This mapping was used to correct the offset between the two color channels in the experimental images. For Figure
4-4F, fluorescence intensity was determined as the highest intensity pixel found in each 
MS2 surface per time point and nucleus (this is due to the higher photostability of GFP vs 
mCherry).

**Statistical Analysis**

All plots and statistical tests were done in R, version 3.5.1. A two-sided Mann-Whitney-U 
test was used to calculate statistical significance of the difference between two distributions 
of fluorescence intensity (Figure 4-4F) and between two distributions of distance (Figure 
4-4E and 4-5D).
REFERENCES


**Figure 4-1. Large distances separate co-regulated genes \textit{in trans}.

(A) A schematic of the transvection assay. On the top allele, there is a \textit{gypsy} insulator, the \textit{sna} shadow enhancer, and a \textit{PP7-lacZ} reporter gene. On the bottom allele, there is a \textit{gypsy} insulator and an \textit{MS2-lacZ} reporter gene.

(B) Stills of two representative nuclei taken from maximum projected Airyscan-processed images. \textit{MS2} foci were visualized by MCP-GFP (green), \textit{PP7} foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 500nm.

(C) Distribution of XY distances separating \textit{MS2} and \textit{PP7} foci during transvection (n = 2441). Bin widths plotted were 10nm.
Figure 4-2. Distances separating MS2/PP7 foci in an interleaved reporter gene.

(A) A schematic of the interleaved MS2/PP7 reporter gene, wherein the \textit{hb} primary enhancer and \textit{hb} P2 promoter is driving a \textit{12x(PP7-MS2)}-\textit{lacZ} reporter gene.

(B) Stills of two representative nuclei taken from maximum projected Airyscan-processed images. MS2 foci were visualized by MCP-GFP (green), PP7 foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 500nm.

(C-E) Distribution of distances across XY (C), X (D), or Y (E) axes separating MS2 and PP7 foci (n = 2084). Bin widths plotted were 2nm.
Figure 4-3. Distances separating independently regulated MS2 or PP7 reporter genes on separate alleles.

(A) A schematic of the assay. On the top allele, there is a *gypsy* insulator, the *sna* shadow enhancer, and a *PP7-lacZ* reporter gene. On the bottom allele, there is a *gypsy* insulator, the *sna* shadow enhancer, and an *MS2-lacZ* reporter gene.

(B) Stills of two representative nuclei taken from maximum projected Airyscan-processed images. *MS2* foci were visualized by MCP-GFP (green), *PP7* foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 1µm.

(C) Distribution of XY distances separating *MS2* and *PP7* foci (n = 2140). Bin widths plotted were 50nm.
Figure 4-4. Distal and proximal configurations of two cis-linked reporter genes yield unique distance distributions.

(A-B) Schematics for two configurations with two cis-linked reporter genes. The sna shadow enhancer is regulating MS2-yellow and PP7-yellow reporter genes with either distal eve promoters (A) or proximal sna promoters (B). The promoters are separated by either approximately 16.5kb (A) or approximately 3.5kb (B).

(C-D) Stills of representative nuclei with either the distal (C) or proximal (D) configuration taken from maximum projected Airyscan-processed images. MS2 foci were visualized by MCP-GFP (green), PP7 foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 500nm.

(E) Distribution of XY distances separating MS2 and PP7 foci with the distal (blue; n = 2098) or proximal configuration (red; n = 2016). Bin widths plotted were 10nm. These two distance distributions were determined to be significantly different (p = 2.0 x 10^-85, Mann-Whitney-U test).

(F) Boxplot showing distribution of fluorescence intensity of MS2 foci between proximal (red; n = 832) or distal configuration (blue; n = 1066). The box represents the lower (25%) and upper (75%) quantile and the solid line represents the median. Whiskers cover up to the 10th and 90th percentile of each distribution. P-value was calculated using a Mann-Whitney-U test.
Figure 4-5. Distance distributions of two cis-linked reporter genes affected by genomic location.

(A) A schematic of two cis-linked reporter genes. The sna primary enhancer is regulating MS2-yellow and PP7-yellow reporter genes with distal sna promoters.

(B, C) Stills of representative nuclei from the construct integrated at VK00033 (B) and VK00002 (C) taken from maximum projected Airyscan-processed images. MS2 foci were visualized by MCP-GFP (green), PP7 foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 500nm.

(D) Distribution of XY distances separating MS2 and PP7 foci at the VK00002 (red; n = 2415) or VK00033 integration site (blue; n = 1996). Bin widths plotted were 10nm. These two distance distributions were determined to be significantly different ($p = 1.4 \times 10^{-8}$, Mann-Whitney-U test).
Figure 4-6. Distance dynamics between co-regulated genes during transvection.

(A-D) (left) Stills of representative nuclei during transvection taken from maximum projected Airyscan-processed images. *MS2* foci were visualized by MCP-GFP (green), *PP7* foci were visualized by mCherry-PCP (red), and nuclei were visualized using His2Av-eBFP2 (blue). Scale bar indicates 500nm. (right) XY distance trajectories between *MS2* and *PP7* foci over time. Time points that are indicated by blue correspond to a still image shown on the left.
Figure 4-7. Distance trajectories for transvecting nuclei.
Trajectories of XY distance over time is plotted for each transvecting nucleus in which at least five frames of continuous measurements could be analyzed. Time is relative to when acquisition started and any gaps in a nucleus’ trajectory is due to an inability to resolve MS2 or PP7 foci at those time points.
Figure 4-8. A model for the distances separating co-regulated genes in different configurations.

(A-C) In each assay, we propose that clustering and aggregation of RNA Polymerase II (Pol II) and activators leads to the separation seen between co-regulated reporter genes. With the cis-linked assays, recruitment of fewer Pol II molecules might explain the smaller distance observed between reporter genes in the distal configuration (B) as compared to the proximal configuration (C).
Chapter 5

Discussion
CONCLUDING REMARKS

The molecular mechanisms by which cis-regulatory elements interact at large genomic distances to elicit precise and specific transcriptional initiation and how this occurs at an endogenous locus with many overlapping enhancer-promoter interactions remains poorly understood. My dissertation focused on this crucial step of transcriptional regulation using the early *Drosophila* embryo as a model system. The topics that I pursued were: how enhancer-promoter communication can be achieved at large distances, specifically between homologous chromosomes (Chapter 2); how transcriptional regulation takes place within an endogenous locus and the potential role of enhancer-enhancer interactions in driving specific gene expression (Chapter 3); and finally, what distance separates genes co-regulated by the same enhancer during transcription (Chapter 4). One thread that has connected all of these studies is the proposal of a transcription hub model, wherein an enhancer’s “area of effect” on a promoter is quite large due to assembly of clusters containing Pol II and other activators. This view of a transcriptional hub nicely encompasses characteristics of both the classical looping and linking models of transcription (reviewed in Furlong and Levine 2018). While I believe that the transcription hub is the best interpretation to both the results presented here and other recent work that will be discussed below, it is important to note another competing view recently employed to explain these results. This is the so-called “kiss-and-run” model, in which an enhancer directly contacts a promoter only infrequently during transcription to recruit or release Pol II for initiation (Cho et al., 2018). This model differs from the transcription hub view of transcription in that it invokes direct one-to-one contact between transcription machinery bound at the promoter and activators bound at the enhancer. The “kiss-and-run” model
might interpret the finding that a single enhancer can regulate multiple genes at the same time (e.g., Fukaya, Lim, and Levine 2016) as rapid movement of the enhancer between each promoter to initiate Pol II elongation. Since the completion of the research described in the earlier chapters, there have been recent studies that nicely complement my work that I would like to discuss in the context of the transcriptional hub model.

As described earlier, the *Shh* locus holds a remarkable example of long-range enhancer-promoter communication, where the ZRS enhancer, which directs a specific expression domain of *Shh* in limb buds, is located nearly 1Mb away from the promoter. The underlying TAD structure of the locus has been previously described, where boundary elements bound by CTCF delineate an approximately 900kb domain that encompasses *Shh* and all associated enhancers (Anderson et al., 2014). Recent work has further examined the role of this TAD structure particularly on the interaction between the *Shh* promoter and the ZRS enhancer (Williamson et al., 2019). The physical distance between these two elements was found to be significantly increased by DNA FISH upon endogenous deletion of CTCF boundaries in the locus likely in response to a restructured genomic organization. Surprisingly, however, this large increase in physical distance did not affect *Shh* expression in the ZRS domain, resulting in normal limb development. This suggests that productive transcription can occur at physical distances even larger than what was observed earlier in Chapter 4. In another recent study, researchers utilized the mouse alpha globin locus, whereby all the enhancers seem to be interacting simultaneously with the alpha globin promoters in a hub-like complex (Oudelaar et al., 2018), to address the question of promoter competition. Upon deletion of flanking CTCF sites at a TAD boundary
encompassing the locus, interactions between the alpha globin enhancers and promoters are not negatively affected, suggesting that that TAD structure is not crucial for establishing or maintaining theses contacts (Oudelaar et al., 2019). Indeed, it appears that promoters from neighboring genes that were originally outside the TAD are recruited to the alpha globin locus hub by interacting with its promoters and enhancers; thus, competition between an enhancer and available promoters does not seem to be driven by genomic organization. Overall, these studies further suggest “action-at-a-distance” and “many-to-many” models for enhancer function during transcriptional initiation at promoters, which are both consistent with our transcriptional hub model. However, there do exist some questions that will still need to be addressed in order to get closer to a complete understanding of transcriptional regulation.

One crucial finding across recent studies has been the occurrence of a potential further compaction of the enhancer-promoter interaction upon initiation of transcription. This phenomenon seems to have been observed in both fixed and live imaging (Chen et al., 2018; Mateo et al., 2019), where, upon transcriptional activation, the average physical distance between the enhancer and promoter decreases. It is unclear what could be driving this observation. Could it possibly be that release of Pol II at the promoter shrinks the distance periodically? Further work will need to be done to address this ambiguity. If we are to believe that transcription takes place through activation by cis-regulatory elements that are all brought proximal in recruitment to a hub-like structure, how do we achieve promoter specificity and avoid potential enhancer crosstalk? Again, this will need to be a critical area for additional investigation in the future. In conclusion, I believe that the work
presented in this dissertation has provided key insights into enhancer-promoter communication and its role in transcriptional regulation during development. Further, I think that these observations and resulting models for transcriptional regulation provide a potential roadmap for further investigation into the underlying molecular mechanisms for enhancer function.
REFERENCES


