Stony Brook University

Role of Runt, Even-skipped, and Odd-paired in Regulating

Enhancer Activity in the *Drosophila* Embryo

A Dissertation Presented

by

Michael Luis Higgins

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

December 2016

Stony Brook University

The Graduate School

Michael Luis Higgins

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

J. Peter Gergen, Ph.D.
Professor
Department of Biochemistry and Cell Biology

Nancy Marie Hollingsworth, Ph.D.
Distinguished Teaching Professor
Department of Biochemistry and Cell Biology

Benjamin L. Martin, Ph.D.
Assistant Professor
Department of Biochemistry and Cell Biology

Ed Luk, Ph.D.
Assistant Professor
Department of Biochemistry and Cell Biology

Stephen J. Small, Ph.D.
Professor of Biology
Department of Biology
New York University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Role of Runt, Even-skipped, and Odd-paired in Regulating Enhancer Activity in the Drosophila Embryo

by

Michael Luis Higgins

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2016

The Drosophila *runt* gene is the founding member of the RUNX transcription factor family, a group of conserved genes that have vital roles in multiple developmental pathways throughout the animal kingdom. Runt plays important roles in segmentation, as well as in neurogenesis and sex determination in the Drosophila embryo. As a transcription factor, Runt is both an activator and a repressor of its target genes depending on the context. *Sloppy-paired (slp1)* is a target of *runt's* activity as a pair-rule transcription factor in the segmentation pathway that is both activated and repressed by Runt. Two cis-regulatory enhancers of *slp1*, a Distal Early Stripe Element (DESE) and a Proximal Early Stripe Element (PESE), that mediate regulation by Runt and other pair-rule transcription factors have been characterized. The transcription factor, encoded by the *odd-paired (Opa)* gene, Opa is an important activator for both DESE and PESE. This thesis describes progress towards identifying Opa binding sites in these enhancers. The PESE enhancer is repressed in cells that co-express Runt and the transcription factor encoded by

the *fushi-tarazu* (*Ftz*) gene. This thesis investigates the roles of Ftz and the Runt binding sites in PESE in this repression. Prior work on the regulation of the DESE and PESE enhancers revealed a non-additive interaction that could be explained if repression that involves preventing release of promoter-proximal paused RNA Polymerase II dominantly interferes with the ability of other enhancers to drive expression from this same promoter. This proposal is investigated by examining the effects of the *slp1* enhancers on expression driven by a heterologous enhancer from the *short-gastrulation* (*sog*) gene. The results of these experiments and their implications on our understanding of transcription regulation during animal development are discussed.

Dedication Page

To my mother, father, and sisters who have been there from the beginning. Thank you for everything!

Table of Contents

List of Figures/Tables/Illustrations

List of Abbreviations

Acknowledgments

This work was supported by a Research Grant from the National Institute of Health to Dr. J. Peter Gergen and by a supplemental NIH Diversity Fellowship grant R01GM094401 for Michael L. Higgins. Additional support was provided by a Turner Fellowship for Michael L. Higgins

I am deeply grateful for the advice and support of my advisor, Dr. John Peter Gergen, Ph.D. Additionally I am grateful to past and present members of the lab, Kimberly Bell, Yasuno Iwasaki, Saiyu Hang, and Lisa Prazak. I'm also grateful to the following individuals who helped, supported, and encouraged me: Emily Iocolano, Haiyue Zhang, Liujing Xing, Michael Klingener, Yelena Altshuller.

I am grateful for the advice and direction of my committee members, Nancy Hollingsworth, Ed Luk, Benjamin Martin, and Steve Small.

Chapter 1: Introduction

General background on the regulation of gene expression

I am interested in studying how gene expression is controlled by regulatory sequences called enhancers. Typically enhancers are short non-coding sequences of DNA from around 100-1000bp that can activate transcription at a promoter, from a relative distance and location, in both the forward or reverse orientation (Long, Prescott, & Wysocka, 2016). In mammals, transcription depends on multiple different *cis*-regulatory sequences such as enhancers (Stadhouders et al., 2012). It is not fully known exactly how the enhancers, which mediate transcription activation and repression by transcription factors, affect gene expression. Regulation of transcription is a changing process led by transcription factors that bind to cis-regulatory regions (Vermunt & Creyghton, 2016), such as enhancers and promoters, and spatiotemporally control gene expression. Promoters are sequences at the beginning of the gene, containing the transcription start site at the 5' end of the gene.

Tightly controlled gene expression is important because it sets up the body plan of the organism. The first step of gene expression, transcription, a process in which RNA Polymerase produces ribonucleic acid (RNA) transcribed from the deoxyribonucleic acid (DNA) plays a critical role in the expression of all genes. In the developing Drosophila embryo, a cell expressing the right set of genes to be a certain cell type depends on control of gene expression. Every embryo develops following a program of development, which progresses because of spatiotemporal domains of gene expression (Spitz & Furlong, 2012).

Though there are other things that affect transcription, in the early Drosophila embryo, transcription factors are the primary mode of control of gene expression. In eukaryotes, chromatin structure is important for transcription and other DNA processes. (G. Li & Zhu, 2015). Euchromatin or open chromatin are regions of the chromosome where genes are highly expressed, and heterochromatin, or densely packed DNA, is organized in a way that hinders gene expression. Modification of histone tails is one way that chromatin remodelers change chromatin and affect transcription. Also cellular signaling pathways, affect transcription. Signaling by receptor tyrosine kinase Torso also affects transcription at the poles of the embryo (Sopko & Perrimon, 2013). After the first few nuclear divisions in the development of the Drosophila embryo, the nuclei are totipotent, so there is no epigenetics, or cellular memory of chromatin structure. Transcription factors control the patterning and timing of gene expression and determine cell fate.

Later in development, transcription factors can control the transcription and ultimately the differentiation of cells either by repressing transcription initiation or preventing transcription elongation (Patel et al., 2013; Simmons & Bergmann, 2016). These mechanisms also affect the level of transcription in mammals in different ways (Day et al., 2016). There are several different mechanisms by which eukaryotic repressors may work. In competition, a repressor bound near an activator's binding site may crowd out the activator and prevent it from binding. For inhibition, a repressor may bind alongside an activator and interacts with it, preventing its activating region from working. Repressors bind to enhancers and affect transcription from a distance through interactions with the mediator and other transcriptional machinery. Repressors may

recruit histone modifiers such as histone deacetylase that inhibit transcription, indirectly repressing the gene.

Studying how enhancers work is a challenging, yet fascinating question, because there are different models of enhancer organization and function. The human β interferon gene has an enhancer bound by protein complexes, a structure known as the enhanceosome. Activators NF-κB, IRF, and Jun/ATF bind to the enhancer in a cooperative manner. They recruit CREB-binding protein (CBP), a coactivator, to these transcription factors that activates transcription. Another protein, HMGA1, binds the enhancer and straightens out the bent DNA. This enhancer is conserved in a variety of mammals, and even a minor change in distance between transcription factor binding sites causes loss of function of the enhanceosome (Jankowski et al., 2015).

In contrast in the "billboard" model the locations of binding sites relative to each other is more flexible and they act to display information (Arnosti & Kulkarni, 2005). The transcription factors, which bind to the enhancer, do not depend on each other for binding. Redundancy, or independently acting sub-elements, is another feature of enhancers (Arnosti & Kulkarni, 2005). Both activators and repressors bind to the enhancer, and different sites may be lost or replaced by another transcription factor through evolution.

Segmentation in Drosophila - a framework for investigating gene regulation

Segmentation, or the formation of repetitive parts or segments, is a process that is important for the proper embryonic development of many metazoans, or animals. It is important in vertebrate organisms such as humans for the organization of the spine into vertebrae, and each human normally has the same number of vertebrae. Many model

organisms exist for studying vertebrate segmentation, such as the embryos of chick, frog, zebrafish, and mice (Bénazéraf & Pourquié, 2013), but segmentation of the Drosophila embryo has been intensively studied in Drosophila more so than in other animals (Mike Levine, 2008).

The fruit fly *Drosophila melanogaster* provides a great model organism to work with, particularly for doing genetics to study enhancer function and segmentation, because of the genetic tools available and broad base of knowledge about Drosophila. One of these tools that is important, because it is used in flies throughout this work is the ΦC31 system. The ΦC31 integrase system is used to introduce a *lacZ* reporter construct, with the enhancers placed upstream the *lacZ* reporter gene, into flies (Fig.1-1). In contrast to the commonly used P-element transposon, ΦC31 integration is non-random, because the site of integration is determined by sequence. Specifically, "ΦC31 integrase mediates site-specific recombination between a bacterial attachment site (attB) and a phage attachment site (attP) to create stable recombinants." (Groth, Fish, Nusse, & Calos, 2003) It binds to these attachment sites, and recombines only with the combination of attP and attB (Thorpe, Wilson, & Smith, 2000). Some sequence homology between the two sites which allows for recombination is a 3bp central region where crossover occurs (Thorpe & Smith, 2000). Besides being site directed, the integration is unidirectional (Hillman & Calos, 2012). Finally the method is efficient, because integration occurs at a specified location and recombination is frequent.

The cloned lacZ reporter construct or plasmid I sent to a company to micro-inject embryos contains the *attB* site sequence and functional white gene. They inject it into selected ΦC31-containing *attP* docking site embryos, and they breed and identify

transformants. They send vials of larvae from the cross with these. Once the flies emerge, I select out of a mix of red and white-eyed flies the ones with the red eyes, or the transformants. As shown by Groth et. al, transgenic flies were created in *attP*-containing fly lines, at a frequency of 47% of fertile crosses, by co-injecting an *attB*-containing plasmid, along with integrase mRNA into Drosophila embryos (Groth et al., 2003). The recombination and integration involves cleavage of the *attP* and *attB* sites and joining them together to create hybrid product sites *attL* and *attR* (Hillman & Calos, 2012). The ΦC31 integrase system allows me to place enhancer elements at the same exact locus every time, eliminating position effects as an experimental variable.

The ease of use of Drosophila for genetics, and the homology of their genes and proteins with humans, also makes them a great model organism to use and learn from. Their life cycle is short, allowing for quick expansion of the population of flies with a genetic background of interest. While the flies are easy to work with, they share deoxyribonucleic acid (DNA) sequences similar to mouse and/or human DNA sequences for many genes. For example the signaling molecule encoded by the Drosophila *wingless* gene is homologous both for its profound role in development as well as its structure and molecular function as the mouse oncogene, *int-1* (Rijsewijk et al., 1987). In terms of pattern formation, the discovery of the conserved homeobox (Hox) family of transcription factors containing a highly homologous 61 amino acid DNA binding domain whose order of expression along the anterior-posterior (A-P) axis is conserved from Drosophila to man and is one of the most remarkable findings of modern molecular developmental biology (M Levine, Rubin, & Tjian, 1984; McGinnis, Levine, Hafen, Kuroiwa, & Gehring, 1984; Scott & Weiner, 1984).

The insect body plan is segmented with a conserved structure comprising eight abdominal segments, three thoracic segments and three gnathal segments anterior to the thorax and posterior to the unsegmented head region. Pioneering genetic studies of Drosophila embryogenesis carried out by Nusslein-Volhard, Wieschaus and Jurgens in the 1970s identified a large number of genes whose activities are required for the establishment and maintenance of the normal segmented body pattern during Drosophila embryogenesis (Wieschaus & Nüsslein-Volhard, 2016). Mutations that affect the activity of these so-called segmentation genes lead to defects in pattern formation that frequently also result in the death of the affected organism. Extensive genetic and molecular studies on these genes and the defects associated with perturbations in their activity reveal that segmentation of the Drosophila embryo is a hierarchical process involving genes that can be grouped into four categories: maternal, gap, pair-rule, and segment-polarity genes. The maternal-effect genes are expressed during oogenesis and provide positional information that is subsequently interpreted by genes that are expressed in the embryo. A classic example of this maternally provided information is the morphogenetic gradient of the maternally expressed transcription factor encoded by the *bicoid* gene (Driever & Nüsslein-Volhard, 1988). Bicoid activates the gap gene *hunchback* in a concentration dependent manner (Struhl, Struhl, & Macdonald, 1989) that results in expression in cells in the anterior third of the blastoderm embryo, i.e. all cells that have a Bcd concentration above the threshold level needed for *hb* activation.

The Bcd concentration gradient is one of the three different types of maternally provided positional information that are involved in creating the segmented body pattern along the A-P axis. The first response of the zygotic genome to the broad gradients of

maternally provided positional information is the localized expression of the gap segmentation genes such as *hb.* There are five gap genes that are each expressed in different over-lapping domains of the segmented region of the embryo in response to the maternal information, *hb*, *giant*, *Kruppel*, *knirps* and *tailless* (Jaeger, 2011). Loss of function mutations in each of these gap genes results in localized patterning defects that can be described as a gap in the segmentation pattern. There is a remarkable concordance between the localized defects associated with each gap gene mutant and the localized expression pattern of the gene, indicating that the generation of these expression patterns is central to creation of the segmentation pattern. It is notable that all of the gap-class segmentation genes encode DNA-binding transcription factors.

The second step in the segmentation gene hierarchy involves the regulated expression of the pair-rule genes by the gap gene transcription factors. The initial classification of the pair-rule genes was based on the defects observed in homozygous mutant embryos. These defects are spaced at two segment intervals. There are seven different pair-rule genes, and all but one of these are expressed in a series of seven stripes along the A-P axis, the exception being the *odd-paired* (*opa*) gene that is expressed throughout the pre-segmented region of the embryo. The pair-rule genes are also organized into somewhat of a mini-hierarchy based on their cross-regulatory interactions and the manner in which they receive inputs from the gap gene transcription factors. Three of the pair-rule genes, *even-skipped* (*eve*) *hairy* and *runt* have been termed primary pair-rule genes based on their roles in regulating secondary (and tertiary) pair-rule genes such as *fushi-tarazu* (*ftz*), *odd-skipped* (*odd*) and *paired* (*prd*) (P. Ingham & Gergen, 1988). These three primary pair-rule genes also cross-regulate each other. The striped

expression patterns of the three primary pair-rule genes are also generated in direct response to regulation by the gap genes with stripe-specific elements that drive expression in response to specific combinations of gap transcription factors some of which are responsible for activating expression and others of which block this activation thereby setting the boundaries of the stripes (M Fujioka, Emi-Sarker, Yusibova, Goto, & Jaynes, 1999; Howard & Struhl, 1990; Klingler, Soong, Butler, & Gergen, 1996). The classic example of this combinatorial regulation is provided by the element that specifically drives expression of *eve* stripe #2 (Small, Blair, & Levine, 1992; Stanojevic, Small, & Levine, 1991). This element is activated in cells that express both the Bcd and Hb transcription factors but this activation is blocked by either Giant or Kruppel, the gap gene transcription factors that set the anterior and posterior boundaries of eve stripe #2 expression, respectively. Similar combinatorial principles have been observed for other stripe-specific elements with additional observations that a gap gene transcription factor that serves as a repressor for one stripe element can function, either directly or indirectly as an activator for a different stripe element (Klingler et al., 1996).

The final step in the segmentation hierarchy involves generating the 14 stripe expression patterns of several of the segment-polarity genes. The generation of these patterns is in response to regulatory inputs of the pair-rule genes. The finding that all of the pair-rule genes also encode DNA-binding transcription factors indicate that, similar to the regulation of pair-rule gene expression by the gap genes, the initial regulation of segment-polarity gene expression is fundamentally a process of transcription regulation. Key to establishing the segmented body pattern is creating single cell-wide stripes of the segment-polarity genes *engrailed* (*en*), *gooseberry* (*gsb*), *hedgehog* (*hh*) and *wg.* The *en*

and *gsb* genes encode DNA-binding transcription factors that are required for the establishment and maintenance of the gene expression programs in the anterior- and posterior-most cells of each parasegment, respectively (P. W. Ingham, 2016). Expression of the *hh* and *wg* genes is also required in the anterior- and posterior-most cells of each parasegment, but in this case the genes encode secreted molecules that interact with receptors on neighboring cells to influence their gene expression programs. Indeed, the sustained expression of *gsb* and *wg* in cells in the posterior of each parasegment requires *hh*-signaling from their neighboring cells (P. W. Ingham, 2016) and sustained expression of *en* and *hh* in cells in the anterior of each parasegment requires *wg*-signaling from their neighbors (P. W. Ingham, 2016). Thus the juxtaposed stripe of cells along the posterior edge of each parasegment that co-express *gsb* and *wg* with a stripe of neighboring cells along the anterior edge of the adjacent parasegment that co-express *en* and *hh* comprises a unit that is responsible for maintaining these cell fate identities for the life of the organism. Although much work has gone into studying the regulation of these segment-polarity genes, a lot of this has involved studies on the *hh* and *wg* signaling pathways, processes that are conserved throughout the animal kingdom (P. W. Ingham, 2016). One complicating issue for dissecting the initial regulation of the segment-polarity genes are the complex cross-regulatory interactions of the pair-rule genes. A second complication comes from the large size of the flanking cis-regulatory regions that are needed for the proper regulation of these genes during development.

After the maternal-to-zygotic transition occurs, zygotic transcription starts and zygotic transcription factors encoded by genes such as the gap and pair-rule genes regulate the transcription of hundreds of genes at nuclear cycle 14 after cellularization

(Boija & Mannervik, 2015). Cell fate is determined in part by pair-rule transcription factors Runt and Even-skipped, which later set the planar polarity of the embryo (Zallen & Wieschaus, 2004).

The observations that segmentation involves the localized expression of the gap, pair-rule and segment-polarity genes and that the gap and pair-rule genes all encode sequence-specific DNA-binding proteins reveals the critical importance of transcription regulation in this developmental process.

The pair-rule transcription factor Runt is an important protein involved in segmentation, sex determination, and neurogenesis in Drosophila (Duffy & Gergen, 1994). Runt directly activates *Sex-lethal (Sxl,)* which determines the sex of the embryo (Kramer, Jinks, Schedl, & Gergen, 1999). Runt represses *orthodenticle* (*otd*) expression in the posterior of the embryo, and the conserved VWRPY motif that is necessary for interaction of Runt with Groucho, is not required for this repression. (C. C. Tsai, Kramer, & Gergen, 1998). Runt has gap gene properties and runt heat shock affects the expression of gap genes and pair-rule genes (C. Tsai & Gergen, 1994). *Runt* is expressed in a variety nerve progenitor cells and neurons and is required for the generation and cell fate specification of the Drosophila nervous system (Duffy, Kania, & Gergen, 1991).

Runt shares identity with the RUNX family, and the sequence of the DNA-binding domain of 128 amino acids is evolutionarily conserved (Akamatsu, 1997). The *RUNX1* or *AML1* gene is necessary for the production of hematopoietic stem cells and is often involved in chromosomal translocations associated with leukemia (Asou, 2003). The t(8;21) chromosomal translocation is a common mutation in acute myeloid leukemia, in which the Runx1 gene is fused to the *MTG8/ETO* gene, resulting in a fusion protein in

which DNA binding domain is fused to a different C-terminus, so it is not under normal control (Okumura, Peterson, Lo, & Zhang, 2007). Other RUNX family proteins in humans, RUNX2 and RUNX3, play important roles in development, including regulating bone growth and functioning in immunity and inflammation (Lotem et al., 2015; Wysokinski, Pawlowska, & Blasiak, 2015). RUNX family members are found in all animals, including sea anemones and mice.

Runt is a context-dependent transcription factor that can both activate and repress its target genes and that interacts with other proteins to affect gene expression (Vander Zwan, Wheeler, Li, Tracey, & Gergen, 2003). Runt needs its binding partner, the Brother protein, in order to bind to DNA, although Brother does not bind to DNA itself (L.-H. Li & Gergen, 1999). Although DNA binding is important for some gene activation or repression, in some contexts repression by Runt is DNA-binding independent, acting through other proteins to effect gene expression (Vander Zwan et al., 2003). The nanos-gal-tubulin (NGT) system is used to ectopically express a transgene with an upstream activation sequence (UAS). In this system males flies with the UAS transgene are crossed with with female flies with nanos-gal-tubulin (NGT), in which galactose four (Gal4) mRNA is maternally deposited in the egg and is translated throughout the embryo, activating the UAS transgene (Tracey, Ning, Klingler, Kramer, & Gergen, 2000). Experiments using the NGT system with the UAS RuntCK flies, in which Runt has two amino acids changed to prevent DNA binding, shows that DNA binding is necessary for activation of *slp1*, but not completely required for the initial repression of the *engrailed* odd stripes (Vander Zwan et al., 2003)). Another feature of Runx that is the C-terminal VWRPY motif which is necessary for interaction with its co-factor Groucho, and this

same motif and interaction are found in humans between the AML1 and TLE proteins (Aronson, Fisher, Blechman, Caudy, & Gergen, 1997; Javed et al., 2000). The VWRPY motif is involved in *slp1* activation but is not required for repression of *slp1*, and genetically Groucho does not regulate *slp1*, so the VWRPY motif may interact with other proteins to activate *slp1* (Walrad, Hang, Joseph, Salas, & Gergen, 2010). Another co-factor Hairless, but not Groucho, was found to contribute to *slp1* repression (Walrad, Hang, & Gergen, 2011).

Slp1 is a good model for studying Runt, because Runt can either activate or repress this gene depending on the context. Ectopic expression of Runt and the zinc finger transcription factors Opa using the NGT system activates *slp1* throughout the somatic blastoderm nuclei of the embryo in *ftz* mutants, while overexpressing Runt and Ftz together this way is enough to repress *slp1* throughout the blastoderm nuclei (Swantek & Gergen, 2004). Turning *slp1* on or off throughout the embryo allows for biochemical experiments like ChIP to be done on collections of whole embryos. ChIP experiments showed that Runt and Ftz repress *slp1* by inhibition of transcription elongation (Wang, Lee, Gilmour, & Gergen, 2007). ChIP experiments are needed to elucidate the cis-regulatory architecture of the Runt binding sites and other transcription factors in the enhancers of *slp1*.

Slp1 is also a good model for studying Runt because its cis-regulatory elements have been identified. A reporter gene combining two *slp1* enhancers, the distal early stripe element (DESE) and the proximal early stripe element (PESE), recapitulates the endogenous gene pattern. (Prazak, Fujioka, & Gergen, 2010). Other *slp1* stripe enhancers throughout the *slp1* locus have been identified and share overlapping activities, and

unlike the non-additive interactions between DESE and PESE which will be referred to in chapter two of this thesis, the authors of the study claim that most of the direct integration of the multiple enhancers is simply additive (Miki Fujioka & Jaynes, 2012)

Recent work has resulted in a model for regulation of *slp1* (Hang and Gergen, personal communication, see Fig.1-2). The *slp1* gene is expressed in a pattern of 14 stripes, each expressed in the posterior of each parasegment. The stripes are numbered by convention starting with stripe 0 near the head region and continue in a repetitive pattern of seven units until stripe 13, and the even-numbered stripes are more strongly expressed than the odd-numbered stripes. Regulation of *slp1* in four cell type contexts (Hang and Gergen, personal communication) make up each of the seven repetitive units across the embryo and are controlled by different transcription factors and enhancers. *Slp1* is repressed by blocking elongation by Eve mediated by PESE in type I cells, which are the two cells that do not express *slp1* anterior to the odd-numbered stripes of *slp1*. Type II cells comprise the odd stripes and *slp1* is activated by the combination of Runt and Opa mediated by DESE in this cell type context. Type III cells are located anterior to the *slp1* even-numbered stripes and are repressed through DESE by Runt and Ftz by preventing release of promoter proximal paused Pol II. Type IV cells are activated by Opa and not Runt, but also by an unidentified Factor X (Swantek & Gergen, 2004). Note that in cell type contexts I and III, *slp1* is repressed by preventing transcription elongation.

In the second chapter of this thesis, I will test the hypothesis of whether repression by an enhancer preventing release of paused Pol II is dominant and prevents another enhancer from activating the same promoter. In the third chapter I will discuss a mutagenesis of the Runt binding sites in the enhancer in a reporter gene, *PESE-lacZ*, and

a series of experiments with Ftz to assess the role of Runt binding sites and Ftz in PESE. In the fourth chapter I will show ChIP's of wild type flies with the *yw* mutation, using the customized Opa antibody to start to discover the DNA binding architecture of DESE and PESE for Opa. I will show a map of the putative binding sites identified by experiments to create a model of how DNA binding may be influencing regulation by pair-rule transcription factors of *slp1*.

Figures

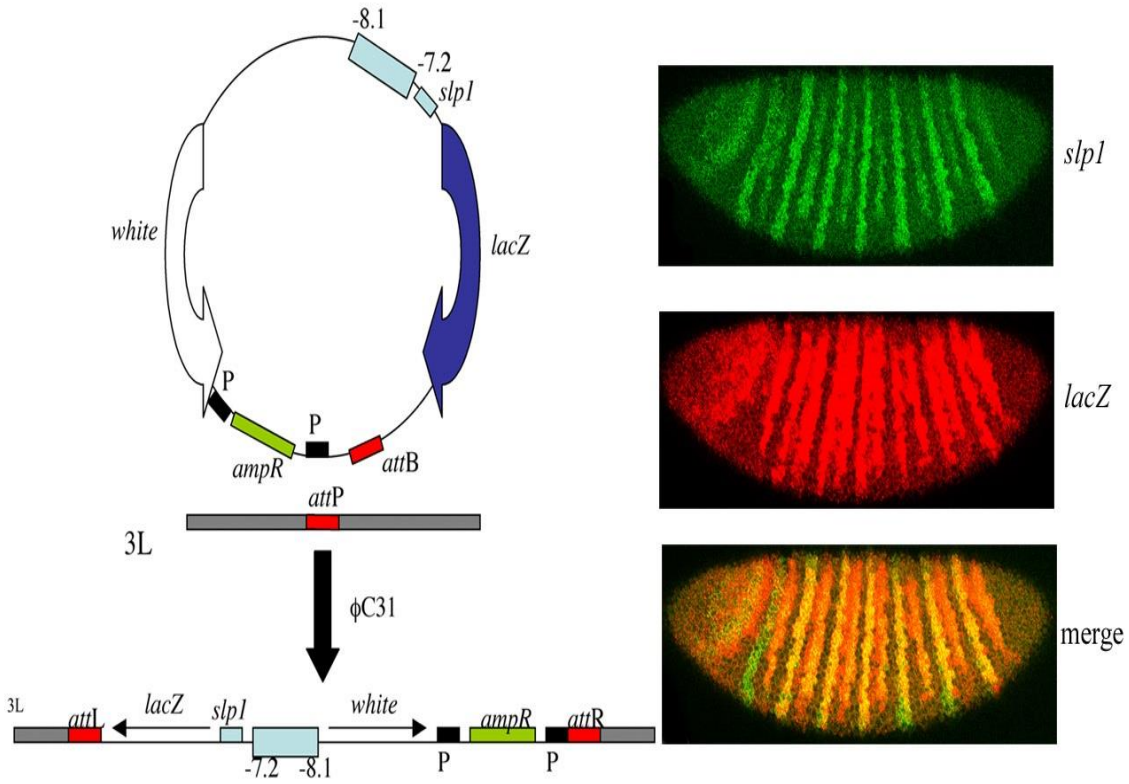

Figure 1-1: ΦC31 integration of lacZ reporter gene. (figure from Lisa Prazak)

A plasmid with an attB site and the lacZ reporter gene is injected into fly embryos with an attP site on chromosome III and the ϕC31 site-specific integrase that integrates the reporter gene in into the attP site to generate a transgenic fly with the reporter gene.

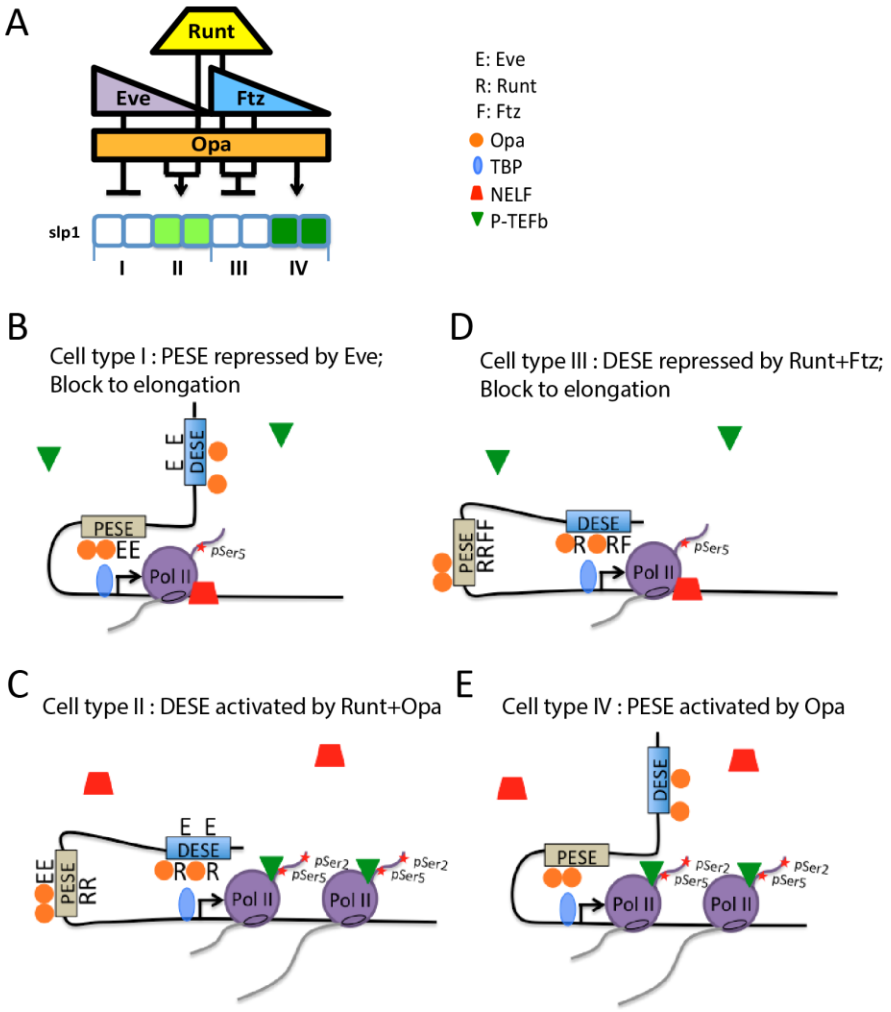

Figure 1-2. Model of slp1 cell type contexts (Hang and Gergen, personal communication)

A. Schematic of pair-rule genes Runt, Eve, Ftz, and Opa regulating slp1 in four different cell type contexts. Arrows indicate activation while perpendicular lines indicate repression. Legend to right shows letters and shapes representing transcription factors, and TATA binding protein (TBP), Negative elongation factor (NELF), and positive transcription elongation factor (P-TEFb). B. In cell type I, Pol II is paused with NELF downstream the transcription start site (TSS). TBP is bound at the promoter bound to the TATA box sequence indicating formation of the transcription pre-initiation complex (PIC) and PolII CTD phosphoserine-5 modifications indicate transcription initiation. PESE enables the start of transcription initiation, but in this cell type context PESE mediates repression by Eve, which prevents PolII from releasing into elongation. C. In type II cells, DESE mediates activation by Runt and Opa and Runt prevents PESE interaction with the promoter. PolII CTD phosphoserine2 modification by P-TEFb allows for transcription elongation. D. In type III cells, DESE mediates repression by Runt and Ftz, inhibiting P-TEFb recruitment and preventing promoter proximal paused Pol II

release. E. In type IV cells, PESE mediates activation by Opa and the absence of Eve and Runt allows for transcription elongation.

Chapter 2: Enhancer-mediated repression associated with promoter-proximal paused RNA polymerase is dominant

Abstract

The genetic hierarchy responsible for generating the segmented body pattern of the Drosophila embryo provides a powerful framework for investigating the in vivo regulation of gene transcription during development. The series of two cell-wide stripes in the posterior half of each of the fourteen parasegments in *sloppy-paired-1* (*slp1*) gene, is a well-studied model in the segmentation pathway that is expressed in response to regulation by the pair-rule transcription factors Even-skipped (Eve), Fushi Tarazu (Ftz), Oddpaired (Opa), and Runt. Two distinct enhancers upstream of the *slp1* gene, the proximal (PESE) and distal (DESE) early stripe elements recapitulate the endogenous *slp1* expression pattern in a *lacZ* reporter construct in a manner not expected from the simple addition of their respective patterns. (Prazak et al., 2010) Experiments investigating the mechanism of PESE and DESE regulation by these transcription factors have led to a proposal that repression of an enhancer that results in a block to release of promoter-proximal paused RNA Polymerase II (Pol II) prevents other enhancers from activating transcription at the same promoter (Hang and Gergen, personal communication). This hypothesis was tested by generating composite reporter gene constructs that combine the *slp1* PESE and DESE enhancers with an enhancer from the *short gastrulation (sog)* gene that is normally not sensitive to regulation by the pair-rule transcription factors. In wild-type embryos the composite *PESE:sog-lacZ,* reporter is repressed in cells where PESE normally mediates repression by Eve and blocks release of promoter-proximal paused Pol II. Similarly, the *DESE:sog-lacZ* reporter is repressed in

cells where DESE mediates repression by Runt and Ftz that results in a block to paused Pol II release. I further show that the PESE and DESE enhancers prevent all *sog*-driven expression as a specific response to the ectopic expression of Eve or the combination of Runt and Ftz, respectively. I conclude that repression mediated by the DESE and PESE enhancers that results in a block to release of promoter-proximal paused Pol II dominantly interferes with the ability of other enhancers to activate transcription at this same promoter.

Introduction

Enhancers are DNA elements that can be located upstream or downstream, distal or proximal to the transcription start site (TSS) of a gene that integrate the inputs from sequence-specific DNA-binding transcription factors to regulate the activity of Pol II at the gene promoter. Studies on enhancer structure and function reveal they act to integrate both positive and negative inputs from combinations of different sequence-specific DNA-binding transcription factors, frequently with the presence of multiple binding sites for the key regulatory factors (Long et al., 2016). However, there is no clear understanding on how the integrated information from these regulatory inputs is transmitted to control the activity of Pol II at the gene promoter. This problem is compounded for genes in metazoan organisms such as humans and Drosophila that frequently have multiple distinct enhancers, that can each contribute to the spatial and temporal regulation of transcription in different cellular contexts. Amongst the types of mechanisms that have been put forward to explain the regulation of enhancer-promoter interactions are the control of looping, perhaps with a promoter sampling the potential inputs from different enhancers via a billboard model (Arnosti & Kulkarni, 2005). Another class of mechanisms involves cis-regulatory DNA elements termed insulators that prevent a more distal enhancer element from driving expression at a promoter (Bushey, Dorman, & Corces, 2008; Vogelmann et al., 2014).

An example of the potential complications associated with the presence of multiple enhancers is the non-additive interaction between two enhancers from the Drosophila *slp1* gene that integrate regulatory inputs from a small set of transcription factors during the establishment of the segmented body pattern in the early embryo

(Prazak et al., 2010). The *slp1* pattern normally comprises 14 two cell-wide stripes in the posterior half of each parasegment in the late blastoderm stage embryo. A survey of the *slp1* locus for DNA elements that contribute to this expression identified two distinct elements, PESE and DESE that drive periodic expression of *lacZ* reporter genes during this developmental stage (Miki Fujioka & Jaynes, 2012; Prazak et al., 2010). The PESE enhancer produces seven two cell-wide stripes of expression that overlap with the strong stripes of *slp1* expression observed in the posterior half of the even-numbered parasegments at this stage (Figure 2-1). In contrast, DESE drives expression of 14 stripes that overlap expression of the *slp1* gene, but with increased expression of the reporter in odd parasegments, resulting in odd- and even-numbered stripes of equivalent intensity. The *DESE-lacZ* reporter also shows inappropriate expression anterior to these stronger than normal odd-numbered stripes (Prazak et al., 2010). Interestingly, when DESE and PESE are combined in a composite reporter gene construct the inappropriate expression driven by DESE anterior to the odd-numbered stripes is eliminated and the intensity of reporter gene expression within cells comprising the odd-numbered stripes is reduced to levels more comparable to the endogenous *slp1* gene (Fig. 2-1). These observations suggest the PESE enhancer is able to prevent the abnormal activity of the DESE enhancer in cells comprising the odd-numbered parasegments. Prazak et. al (2010) have termed this as a non-additive interaction because the final pattern is not the sum of the individual expression patterns produced by the DESE and PESE enhancers.

Comprehensive studies on *slp1* regulation by the different pair-rule transcription factors provide a robust framework for further investigating the non-additive interaction between the PESE and DESE enhancers (Swantek & Gergen, 2004). This identifies four

different cellular contexts for *slp1* regulation within the segmented region of the embryo. Expression of the odd-numbered stripes requires Runt in combination with the zinc-finger transcription factor Opa (Figure 2-1). The even-numbered *slp1* stripes are also reduced in *opa* mutant embryos with some residual expression in these cells driven by an as of yet unknown Factor X. More important to understanding the non-additive interaction between PESE and DESE are the rules for repressing *slp1* expression in the anterior half of each parasegment. Repression of *slp1* in the anterior half of the odd parasegments requires Eve whereas repression in the anterior half of the even parasegments requires both Runt and Ftz (Swantek & Gergen, 2004). Both PESE and DESE are repressed in cells that co-express Runt and Ftz. In contrast, only PESE but not DESE is sensitive to repression by Eve (Fig. 2-1). Thus key to understanding the non-additive interactions between these two enhancers is determining how Eve-dependent repression of PESE interferes with the ability of DESE to drive expression in cells comprising the odd parasegments.

Recent studies on the molecular mechanisms of repression of the PESE and DESE enhancers by Runt and Eve provide a potential explanation for their non-additive interaction. Runt and Ftz repress expression of a *DESE-lacZ* reporter gene by blocking release of promoter-proximal paused Pol II (Fig 2-2 A), just as was found for the endogenous *slp1* gene (Wang et al., 2007). In contrast, Runt-dependent repression of PESE involves a block to recruitment of Pol II and the initiation of transcription (Fig 2-2 B). More interesting was the finding that Eve represses PESE via the same mechanism associated with Runt-dependent repression of DESE, that is, by blocking release of promoter-proximal paused Pol II (Fig 2-2 C, Hang and Gergen, personal

communication). Based on this observation Hang and Gergen proposed that Eve-dependent repression mediated by PESE that blocks release of promoter-proximal paused Pol II prevents DESE from activating expression at the *slp1* promoter.

Here I test the hypothesis that repression mediated by one enhancer that results in a block to release of promoter-proximal paused Pol II prevents other enhancers from activating transcription at the same promoter. To do this I investigated the effects of the PESE and DESE enhancers on the activity of an enhancer from the *sog* gene that is normally regulated by factors involved in dorsal-ventral patterning and not sensitive to regulation by the pair-rule transcription factors. The expression of these composite reporters in wild-type embryos indicates expression driven by the *sog* enhancer is specifically prevented in cells where PESE and DESE normally mediate repression by Eve, or the combination of Runt and Ftz, respectively. The response of these composite reporters to ectopic expression of Eve or the combination of Runt and Ftz further indicates that the dominant interference with expression driven by the *sog* enhancer is specific both to the *slp1* enhancer element and the mechanism by which it is repressed. The potential widespread importance of enhancer dominant repression due to the regulation of promoter-proximal Pol II pausing is discussed.

Results and Discussion

The molecular studies on the different modes of repression mediated by the *slp1* PESE and DESE enhancers in response to Eve and Runt were carried out using a basal promoter segment spanning from 261 basepairs (bp) upstream to 121 bp downstream of the *slp1* transcription start site (TSS). As a first step in investigating the effect of these two enhancers on a heterologous enhancer I first examined expression of a reporter gene containing a previously characterized enhancer from the *sog* gene (Crocker, Tamori, & Erives, 2008; Hong, Hendrix, & Levine, 2008) placed immediately upstream of this same *slp1* basal promoter region (Figure 2-3). As previously found for this DNA element, the *sog-lacZ* reporter containing the *slp1* basal promoter was expressed throughout the ventral lateral ectoderm in a broad stripe that weakens dorsally with some variable modulation in expression levels along the anterior-posterior axis (Figure 2-4 A). As previously described for other reporter gene constructs containing this element, the *sog-lacZ* reporter is repressed ventrally with a sharp border along the edge of the ventral lateral ectoderm. This result indicates the *sog* enhancer is capable of regulating expression at the *slp1* promoter. With respect to regulation on the anterior-posterior axis it is notable that the *sog-lacZ* reporter was expressed in cells both anterior and posterior to the strong, even-numbered *slp1* stripes (Fig. 2-4 A).

A reporter gene containing the PESE enhancer upstream of the *slp1* promoter (Fig 2-3) expressed the *lacZ* mRNA in cells corresponding to the more strongly expressed, even-numbered *slp1* stripes (Fig. 2-4B). I tested the effect of PESE on the activity of the *sog* enhancer in a composite reporter gene construct that contains the PESE enhancer inserted upstream of the *sog* enhancer (Fig 2-3). This composite *PESE:sog-lacZ* reporter

was most strongly expressed in cells corresponding to the even-numbered *slp1* stripes (Fig. 2-4C). Although there was some expression throughout the lateral ectoderm, there was clear evidence of modulation along the anterior-posterior axis with the most obvious breaks in expression occurring in cells posterior to the even-numbered *slp1* stripes, i.e. in cells where Eve normally represses. Importantly, the *PESE:sog-lacZ* reporter showed no evidence of repression in cells anterior to the even-numbered *slp1* stripes (Fig. 2-4C), i.e. in cells where Runt and Ftz repress PESE by a mechanism that does not involve regulation of promoter-proximal Pol II pausing (Hang and Gergen, personal communication). These observations provide evidence that PESE can dominantly interfere with the ability of the *sog* enhancer to drive expression at the promoter of the composite reporter gene in a manner that also depends on the mode of PESE repression.

I generated an analogous set of constructs to examine the effect of DESE on the activity of the *sog* enhancer (Fig. 2-3). As described previously (Prazak et al., 2010), a *DESE-lacZ* reporter gave expression in cells corresponding to both the odd- and even-numbered *slp1* stripes with evidence of inappropriate expression anterior to the weakly expressed odd-numbered *slp1* stripes (Fig. 2-4D). The composite *DESE:sog-lacZ* reporter also showed evidence of regulation along the anterior-posterior axis, but in this case the strongest repression occurred in cells anterior to the strong even-numbered *slp1* stripes (Fig. 2-4E), i.e. in cells where the combination of Runt and Ftz represses *slp1* via a DESE-mediated block to release of promoter-proximal paused Pol II (Hang and Gergen). The contrasting loss of *sog*-driven expression in cells anterior and posterior to the even-numbered *slp1* stripes observed for the *DESE:sog-lacZ* and *PESE:sog-lacZ* reporters,

respectively, demonstrates each *slp1* enhancer interferes with *sog* enhancer activity in a cell context-dependent manner.

The above results are based on the expression of the composite reporter gene constructs in wild-type embryos where the regulatory rules for *slp1* are well understood and where there is good evidence that PESE and DESE are involved in mediating repression in different specific cellular contexts by blocking promoter proximal paused Pol II release (Hang and Gergen, personal communication). The response of these same reporter gene constructs to the ectopic expression of the different pair-rule transcription factors was examined to further establish the specificity of these dominant interfering effects. Ectopic Eve expression had no discernible effect on *sog-lacZ* reporter gene expression in embryos where the even-numbered *slp1* stripes were repressed (Fig. 2-5A). As described previously (Prazak et al., 2010), the *PESE-lacZ* reporter was very sensitive to repression by ectopic Eve (Fig. 2-5B). Importantly, the *PESE:sog-lacZ* reporter was also sensitive to ectopic Eve with little to no expression visible throughout the pre-segmental region of the embryo (Fig. 2-5C). The effect of ectopic Eve was also tested on the different *DESE*-containing reporters. Expression from *DESE-lacZ* reporter was not affected in embryos where the even-numbered *slp1* stripes were repressed by ectopic Eve (Fig. 2-5D). Consistent with this, the *DESE:sog-lacZ* reporter also showed no response to ectopic Eve (Fig. 2-5E). These results provide compelling evidence that PESE is required for the repression of the composite *PESE:sog-lacZ* reporter observed in response to ectopic Eve. Based on these results I conclude that Eve-dependent repression of PESE dominantly interferes with the ability of the *sog* enhancer to drive expression at the composite reporter gene promoter.

I examined the response of this same set of reporter genes to ectopic co-expression of Runt and Ftz. There was no evidence for regulation of the *sog-lacZ* reporter in embryos where *slp1* expression was nearly eliminated by ectopic co-expression of Runt and Ftz (Fig. 2-6A). Although the *PESE-lacZ* reporter was repressed in response to Runt and Ftz (Fig. 2-6B), expression of the composite *PESE:sog-lacZ* reporter was not affected (Fig. 2-6C). In contrast, expression from both the *DESE-lacZ* and *DESE:sog-lacZ* reporters was efficiently repressed by ectopic co-expression of Runt and Ftz (Fig. 2-6 D and E). These results are analogous to those found for Eve and PESE but in this case indicate that the repression of DESE in response to Runt and Ftz dominantly interferes with the ability of the *sog* enhancer to drive expression at the composite reporter gene promoter. An important additional conclusion from this last set of experiments is that this enhancer dominant repression depends on the mode of repression. Although both PESE and DESE are repressed in cells that co-express Runt and Ftz, only the repression mediated by DESE interfered with the activity of the *sog* enhancer (Fig. 2-6 C and E). Runt and Ftz repress PESE by preventing Pol II recruitment, whereas DESE-mediated repression in response to these two factors involves regulation of promoter-proximal Pol II pausing (Hang and Gergen, personal communication). Based on these observations, the conclusion is that enhancer-mediated repression that results in a block to release of promoter-proximal paused Pol II dominantly interferes with the ability of other enhancers to activate transcription at this same promoter.

The results described here provide compelling evidence that enhancer-mediated regulation of promoter-proximal Pol II pausing can dominantly interfere with the ability of other enhancers to drive expression from the same promoter. Regulation of promoter-

proximal Pol II pausing was initially thought to confer advantages in facilitating rapid changes in gene expression (Lis, 1998), although other roles for this mode of regulation in contributing to the fidelity and/or dynamics of enhancer-promoter interactions have been demonstrated (Boettiger & Levine, 2009; Chopra, Cande, Hong, & Levine, 2009; Lagha et al., 2013). The dominant repressive properties described for the PESE and DESE enhancers are in some ways analogous to those ascribed to so-called insulator elements but with the important exception that they interfere with the activity of the *sog* enhancer when placed further upstream of the promoter. This mechanism may explain other non-additive interactions, such as the ability of a distal 'shadow-enhancer' to block the action of a proximal enhancer on the *snail* promoter (Dunipace, Ozdemir, & Stathopoulos, 2011). Individual enhancers frequently drive expression in cells that do not express the endogenous gene (Pfeiffer et al., 2008), and in silico attempts to model integration of cis-regulatory information at promoters strongly suggest there are non-autonomous mechanisms for silencing unwanted activation (Kim et al., 2013; Samee & Sinha, 2014). Dominant repression of transcription from a promoter via an enhancer-dependent block to paused Pol II release provides a straightforward route for silencing such spurious transcription that may have widespread significance.

Figures

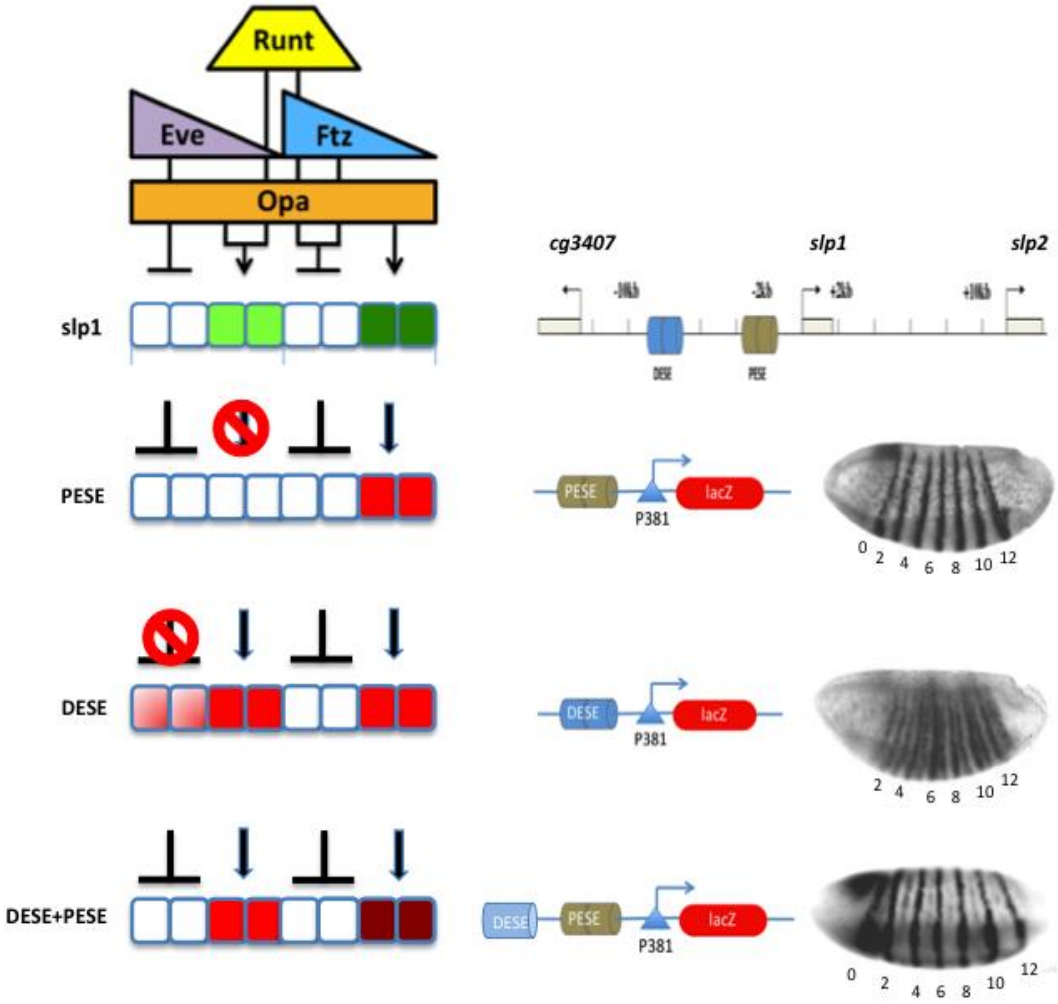

Figure 2-1. Non-additive interactions between the slp1 enhancers DESE and PESE recapitulate the endogenous slp1 pattern.

Repeating unit of cells exhibiting the endogenous *slp1* expression is shown underneath the diagram of combinatorial regulation by pair-rule transcription factors Runt, Eve, Ftz, and Opa. Arrows indicate activation, while vertical lines with perpendicular horizontal lines represent repression. Map showing two enhancers, DESE and PESE is shown in the *slp1* locus. On the left cells in which each *lacZ* gene is repressed or activated are shown, with red shading to indicate expression of *lacZ* reporter genes. Repression or activation marks are crossed out for cells that differ from the endogenous *slp1* pattern. In the middle a diagram with the enhancer element, *slp1* promoter, and *lacZ* gene represented. On the right next to each diagram is an embryo showing the pattern for each reporter gene.

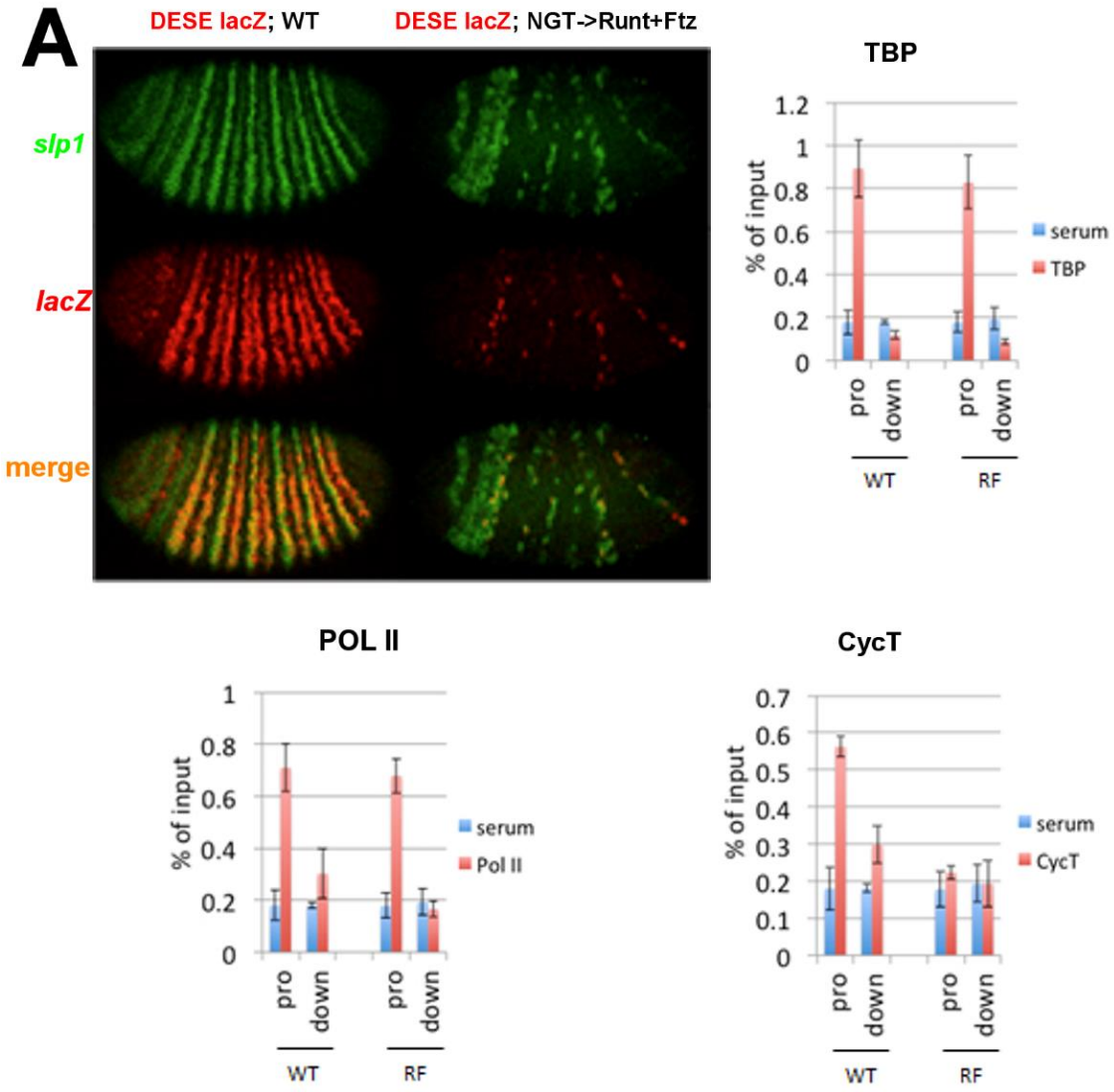

Figure 2-2A. (from Hang and Gergen, personal communication) ChIP assays show modes of repression of DESE and PESE

A. Fluorescent in situ hybridization (FISH) of embryos containing *DESE-lacZ* with wildtype and ectopic expression of Runt and Ftz, is shown on left. ChIP of TBP, Pol II and CyclinT are shown in red with percentage input on the Y-axis, and regions tested in the X-axis shown for two primer pairs for two regions, the *slp1* promoter, and the downstream structural *lacZ* gene, for both genotypes, wildtype (WT) and Runt and Ftz (RF). Rabbit serum (blue) was used as a negative control for all ChIPs.

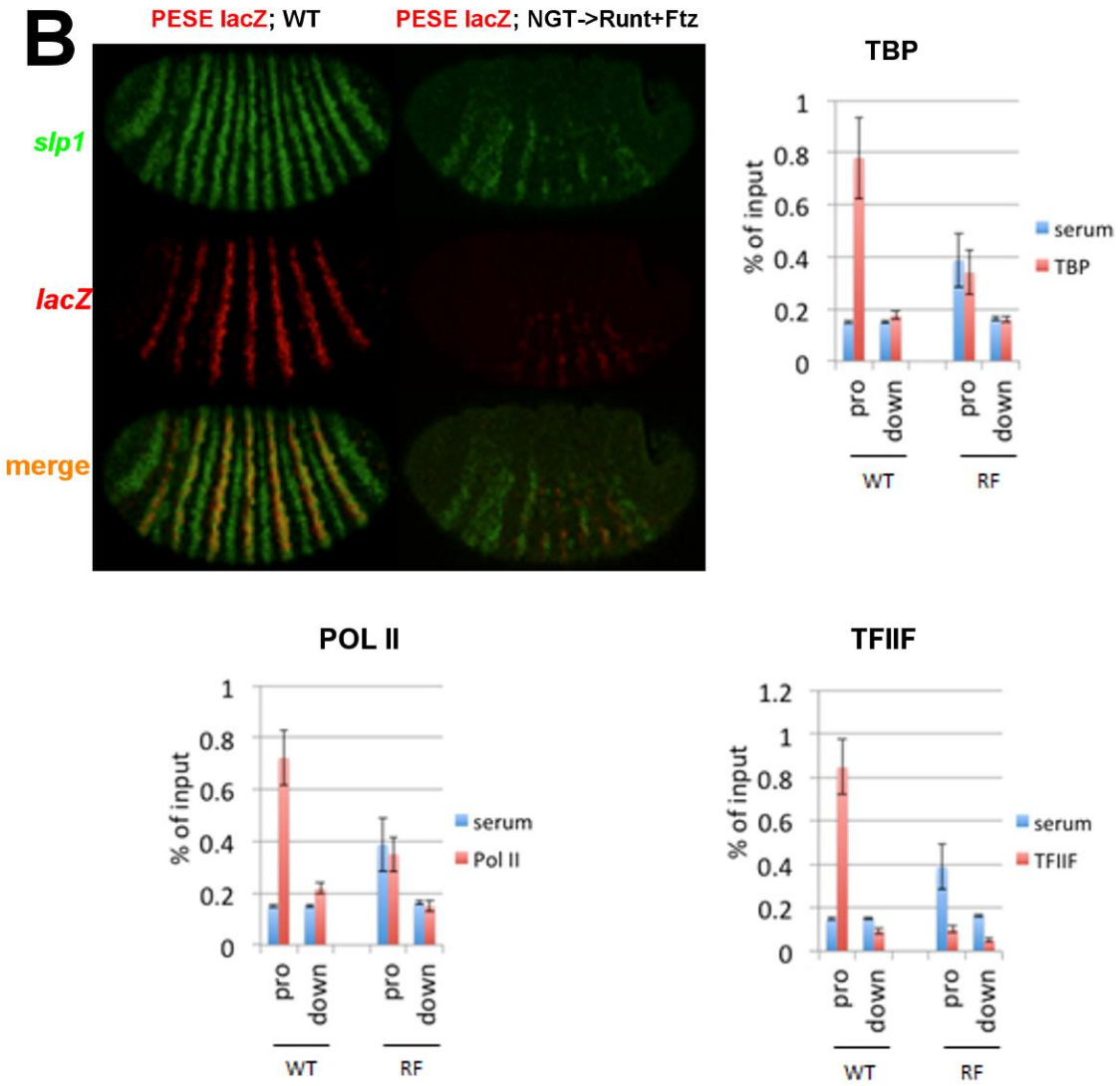

Figure 2-2B. (from Hang and Gergen, personal communication) ChIP assays show modes of repression of DESE and PESE

B. FISH of embryos containing the *PESE-lacZ* reporter with or without ectopic expression of Runt and Ftz. The *slp1* (green) and *lacZ* (red). The percentage of input chromatin of ChIP with TBP, Pol II, and transcription factor II F (TFIIF) in red are shown for two primer pairs for two regions, the *slp1* promoter, and the downstream structural *lacZ* gene, for both genotypes, wildtype (WT) and Runt and Ftz (RF). Rabbit serum (blue) was used as a negative control for all ChIPs.

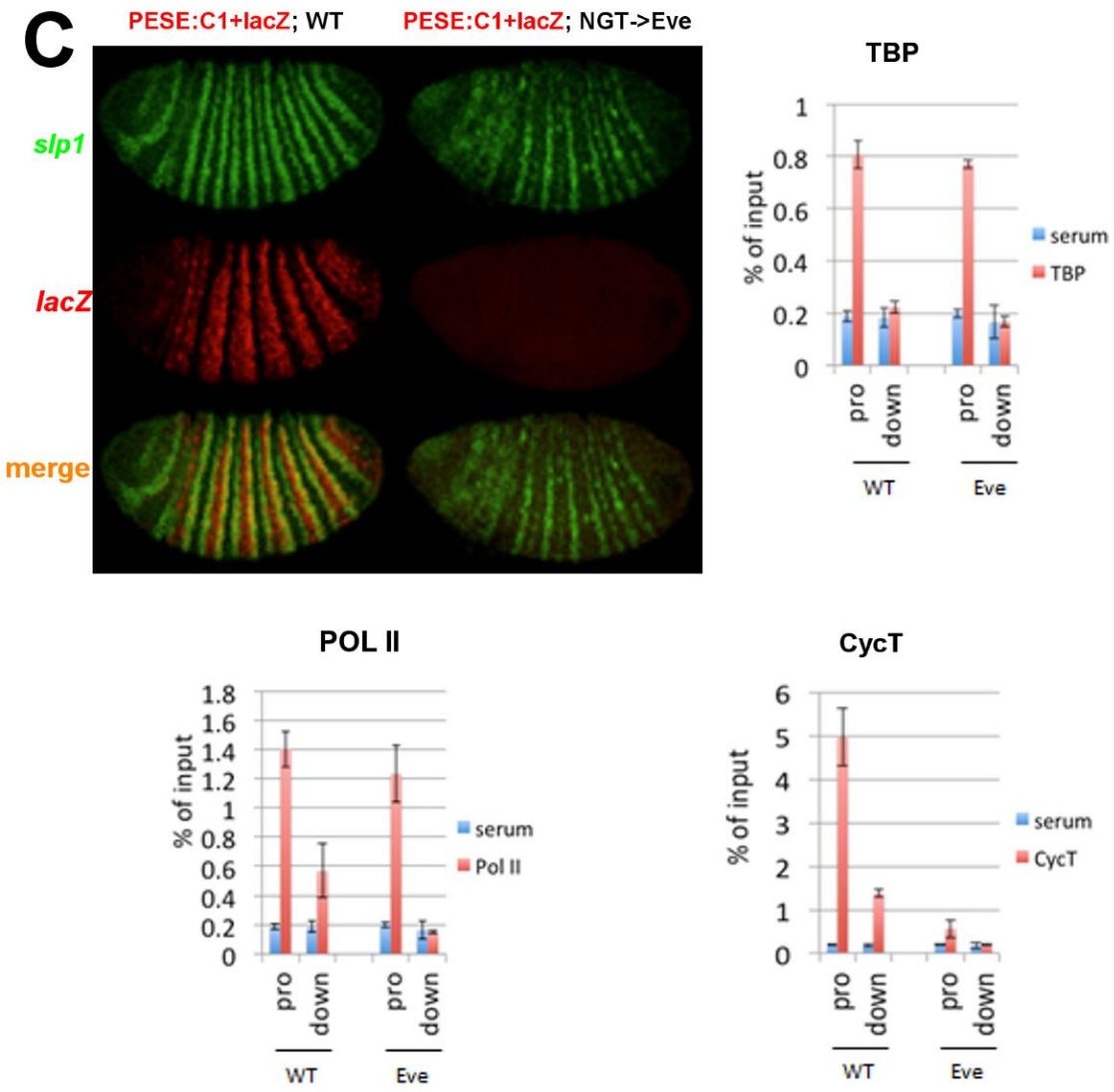

Figure 2-2C. (from Hang and Gergen, personal communication) ChIP assays show modes of repression of DESE and PESE

FISH of *PESE-C1+-lacZ*, wildtype and with ectopic expression of Eve is shown in image. ChIP with TBP, Pol II, and Cyclin T is shown in red at both regions (lacZ gene promoter and structural gene) for both genotypes (WT and Eve). Rabbit serum is shown in blue.

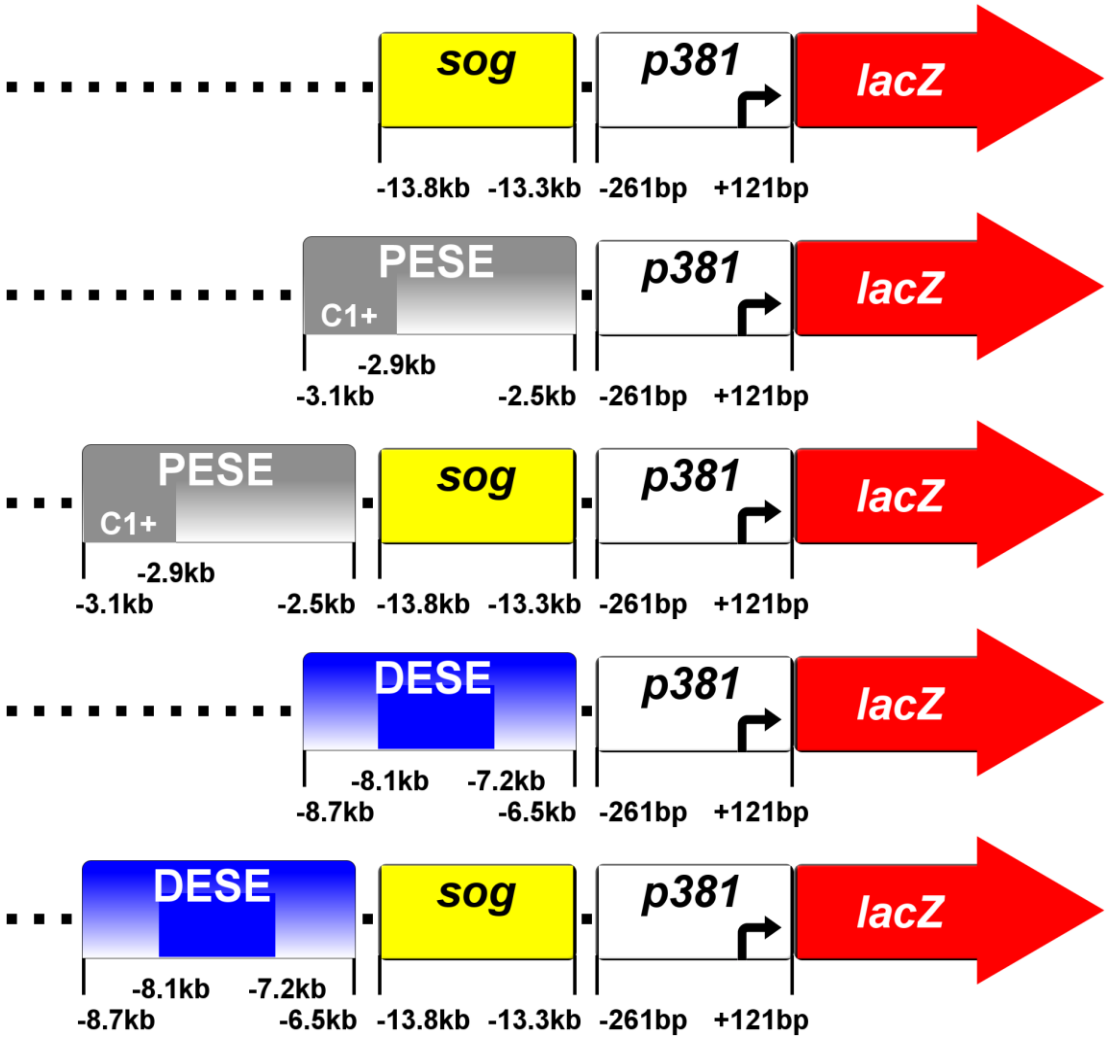

Figure 2-3. LacZ reporter constructs

PESE-lacZ, PESE-sog-lacZ, DESE-lacZ, and *DESE-sog-lacZ lacZ* reporter genes are PhiC31 integrated into attP2 site on third chromosome, while *sog-lacZ* is integrated into the attP1 site on second chromosome.

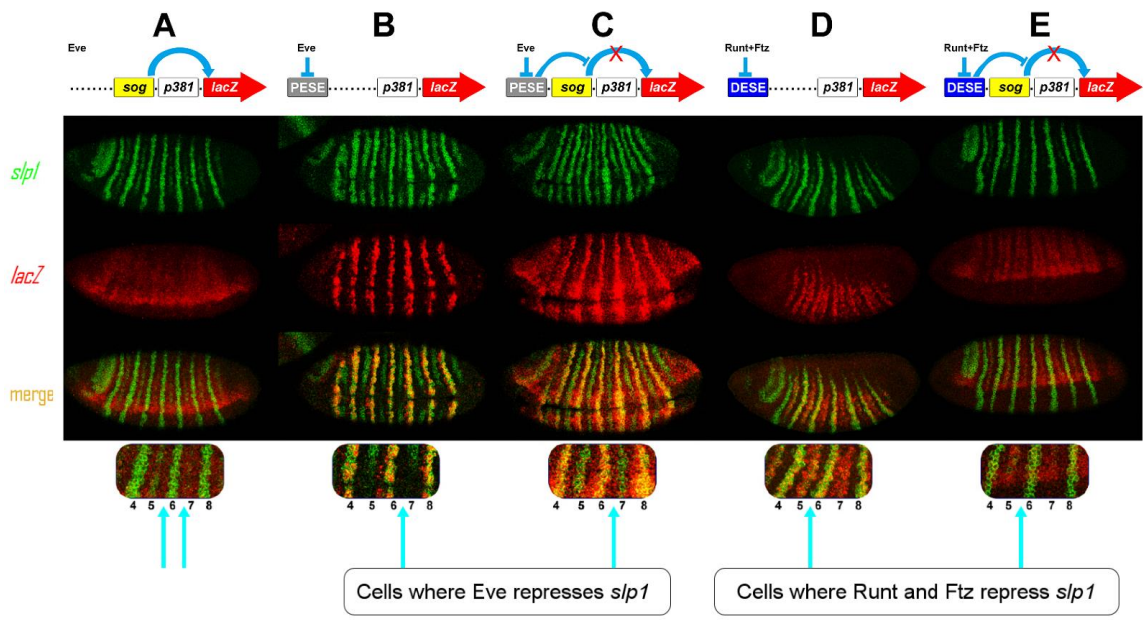

Figure 2-4. Sog enhancer is repressed in slp1 cell type contexts when combined with DESE or PESE.

The mRNA expression pattern of *slp1* (green) and *lacZ* (red) is shown for five *lacZ* reporter constructs. From left to right an embryo is shown for each lacZ reporter construct: A. *sogp381-lacZ*, B. *PESE p381-lacZ*, C. *PESE-sog p381-lacZ*, D. *DESE-p381-lacZ*, and E. *DESE-sog p381-lacZ*. Above each image is a construct diagram representing the reporter gene, and below each is an enlarged section of embryo from stripes numbered 4 to 8.

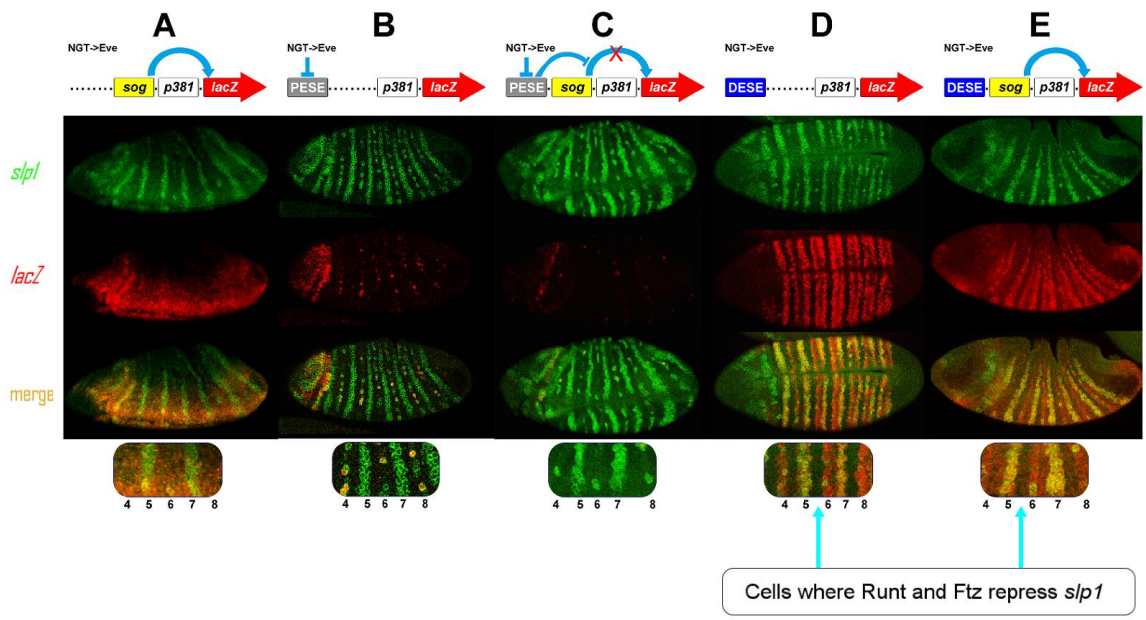

Figure 2-5. PESE-dependent repression by Eve interferes with activation driven by an enhancer from the short-gastrulation (sog) gene.

Expression of *slp1* (green), *lacZ* (red) and merged images for the reporter genes A. *sog-lacZ*, B. *PESE-lacZ,* C. *PESE/sog-lacZ,* D. *DESE-lacZ, and* E. *DESE-sog-lacZ* with ectopic expression of Eve. Ectopic expression was accomplished using the NGT maternal GAL4 system. Above the images are construct diagrams of each reporter, and below each image is an enlarged image of stripes 4-8.

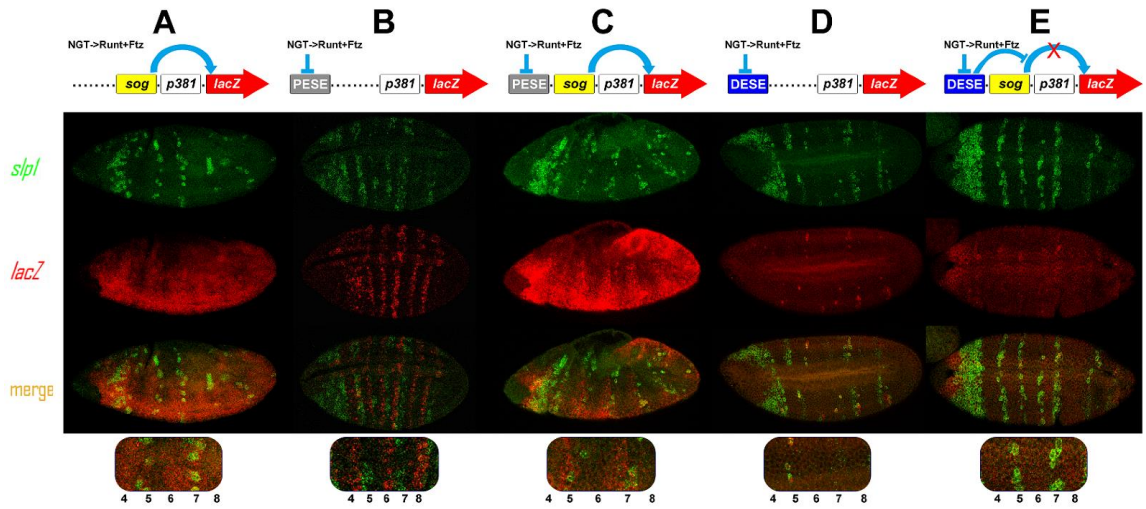

Figure 2-6. DESE-dependent repression by Runt and Ftz interferes with activation driven by an enhancer from the short gastrulation (sog) gene.

In situ hybridization shows expression of slp1 (green) and lacZ (red) in embryos with ectopic expression of Runt and Ftz. Ectopic expression was accomplished using the NGT maternal GAL4 system. The embryos are shown containing a A. sog-lacZ, B. PESE-lacZ, C. PESE-sog lacZ, D. DESE-lacZ, and E. DESE-sog-lacZ reporter gene. The insets below these images shown on the bottom row show magnified merged images for a region of lateral ectoderm spanning the 4^{th}- 8^{th} stripes of slp1 expression.

Chapter 3: Role of putative Runt binding sites and Ftz in PESE.

Abstract

Ectopic expression of Runt and Ftz represses expression of the PESE-containing *PESE[3918]/p126-lacZ-lacZ* reporter gene. In this chapter I investigate the role of six putative Runt binding sites within PESE in mediating Runt-dependent repression. Comparison of the expression of the wild-type PESE [3125]/p381-lacZ with that of the mutant PESE[3125m1-6]/p381-lacZ revealed a modest lost of repression at early stages that is restored as the embryos develop. The implications of these findings on understanding the Runt-dependent regulation of PESE are discussed. I also explored the requirement for Ftz in the repression of PESE. Embryos mutant *for ftz* did not show de-repression of the *PESE[3125]/p381-lacZ* reporter gene similar to that observed for endogenous *slp1* demonstrating *ftz* is not necessary for repression of PESE. Other experiments suggest Runt may be sufficient to repress PESE without Ftz, because ectopic expression of Runt and Opa in a *ftz* mutant background results in complete repression of the *PESE[3918]/p126-lacZ* reporter in embryos shortly following the completion of cellularization and the onset of germband extension.

Introduction

Runt and Ftz repress the endogenous *slp1* gene by preventing release of promoter proximal paused RNA Polymerase II (Wang et al., 2007). Ectopic expression of Runt and Ftz results in repression of the endogenous *slp1* gene as well as in the repression of reporter genes containing either the *slp1* DESE or PESE enhancers (Prazak et al., 2010). As described in Chapter 2, the Runt-dependent repression of DESE involves blocking the release of a paused Pol II complex from the promoter, as was found for the Runt-dependent repression of endogenous *slp1*, whereas Runt-dependent repression of PESE occurs upstream of Pol II recruitment and the initiation of transcription. Although both Runt and Ftz are normally expressed in the type III cells anterior to the even-numbered *slp1* stripes where PESE is repressed it remains an open question as to whether both Runt and Ftz are necessary for this repression. The immediate loss of *slp1* repression in response to transient elimination of Runt in experiments that used a temperature sensitive allele provides strong evidence that *runt* plays a critical role for Runt in repression of PESE (Prazak et al., 2010). Is Runt sufficient to repress PESE? Previously it was found that there was some de-repression of PESE in a *ftz* mutant background (Prazak et al., 2010). What contribution does Ftz have if any? Can we simultaneously observe repression of PESE by Runt in the absence of Ftz? These are some of the questions I addressed regarding the repression of PESE by Runt and Ftz in this chapter.

The crystal structure of the RUNX1-CBFβ DNA-binding domain complex bound to DNA was determined (Bravo et. al 2001). This domain is homologous to the Runt DNA binding domain, and the structure of the complex gives a good example of a well-defined RUNX structure binding to DNA. As shown in the study, some of the six

nucleotides which match the consensus motif have direct nitrogenous base contacts and phosphate backbone contacts with the amino acid side chains, in particular the pair of double CC's and GGs. Later in this chapter I will describe the mutations that were made to the core GG/CC bases for the Runt binding site mutations in PESE. The Big Brother:Runt (Bgb:Run) consensus binding motif TAACCGCAA was identified using bacterial 1-hybrid system. (Meng et. al 2005, JASPAR core database). Putative Runt binding sites in DESE and PESE were derived and scored from a position weight matrix. A position weight matrix (PWM) is a representation of sequence patterns in biological sequences, and it is a matrix of scores that are weighted to match any sequence of fixed length.

Prior work investigating the expression of DESE-lacZ reporters with and without Runt binding site mutations led to the finding that five Runt binding sites in DESE contributed to repression, because when these sites were mutated, the stripes of expression expanded into the type III cells that normally co-express Runt and Ftz (Prazak et. al, personal communication). Furthermore, it was found when comparing the composite *DESE/PESE-lacZ* with a mutated *DESE [m1-5]/ PESE -lacZ* that the Runt sites contributed to activation, because the odd stripes were weaker for mutated binding site reporter gene (Prazak et al., personal communication). In this chapter I describe reporter constructs with six putative Runt binding sites in PESE mutated to test the role of Runt DNA binding in repression of PESE and in the interaction with DESE in a composite reporter gene construct. I also show results of genetic experiments exploring the role of Ftz in PESE regulation.

Results

Stage-specific effects of mutations to Runt binding sites in PESE.

Mutations to six binding sites in a *PESE-lacZ* reporter were made in a PESE sequence inserted upstream of a *lacZ* reporter gene with the large p381 *slp1* basal promoter to assess the role of DNA binding by Runt in the regulation of PESE. Two core cytosines of each of the six binding sites were changed to adenines to eliminate most of the important direct nucleotide base contacts of the Runt DNA binding domain with the binding sites (Fig.3-1). One of the binding sites is a double overlapping binding site, so for this binding site an extra base was changed. (Fig.3-1) These mutations should eliminate binding because they disrupt specific nucleotide base contacts of the protein with the DNA (Bravo, Li, Speck, & Warren, 2001).

The experiment with *PESE[3125]-p381-lacZ* and PESE[*3125 m1-6]- p381-lacZ* indicates there is a stage-specific effect affect of the Runt binding site mutations on PESE regulation. In early blastoderm stage embryos, identified by the weak expression of the odd-numbered *slp1* stripes, there is evidence for reporter gene expression anterior to the more strongly expressed even-numbered *slp1* stripes (Figure 3-2). In contrast, in later stage embryos that are identified by stronger expression of the odd-numbered *slp1* stripes and the presence of the head-fold there is no significant difference in the expression of the two reporters. (Figure 3-3) These results suggest that Runt binding to PESE can paly a role in the repression of this element. However, the finding that the loss of repression is not as severe as that produced by elimination of *runt* activity indicates that there are other factors that contribute to the Runt-dependent regulation of PESE. The implications of these results are discussed later.

The putative Runt sites in PESE do not affect it's interaction with DESE.

Composite reporter genes containing both DESE and PESE revealed a non-additive interaction between these two *slp1* enhancers (Prazak et al., 2010) and studies investigating the mechanisms by which Runt and Eve repress these enhancers have led to the model that enhancer-mediated repression associated with a block to release of promoter-proximal paused Pol II dominantly interferes with the ability of other enhancers to activate expression at this same promoter (Hang and Gergen, personal communication). The motivation for this model was to explain the ability of PESE to impose Eve-dependent repression in Eve-expressing type I cells where DESE is capable of driving expression except in the presence of PESE. The Runt-dependent repression of PESE does not involve regulation of Pol II pausing, but instead acts at a more upstream point and affects Pol II recruitment to the promoter and transcription initiation. As shown above, mutations in six putative Runt binding sites in PESE result in a modest loss of repression in cells anterior to the *slp1* even-numbered stripes, i.e. in the type III cells where Runt and Ftz are required for *slp1* repression. The model Hang and Gergen have put forward to explain *slp1* regulation involves two distinct modes of repression in these type III cells. The first is prevention of PESE-dependent recruitment of Pol II and transcription initiation, potentially due to disruption of functional interactions between PESE and the *slp1* promoter. The second proposed mode of regulation is the DESE-mediated block to release of promoter-proximal paused Pol II. Chapter 2 presented results demonstrating that repression of DESE by Runt and Ftz dominantly interferes with the ability of the *sog* enhancer to drive expression in a composite *DESE/sog-lacZ* reporter. I generated a composite reporter containing DESE and the mutant PESEm1-6

enhancer to determine whether the putative Runt sites in PESE had a role in the non-additive interaction between these two *slp1* enhancers. In the images from the three composites of DESE and PESE, there is little difference between the three constructs (Fig. 3-4). They all recapitulate the endogenous *slp1* gene pattern with no obvious de-repression anterior to the even stripes (Fig.3-4). These results suggest that any de-repression caused by mutagenesis of the Runt binding sites is dominantly suppressed by the presence of DESE, a result consistent with the findings in Chapter 2.

Ftz is not necessary for Runt-dependent repression of PESE.

While expression of *slp1* differs greatly between wildtype and *ftz* mutant embryos, the various PESE reporter genes do not differ (Fig.3-5). The *slp1* RNA in the *ftz* mutants is expanded to 6 cell wide stripes, but the different PESE-containing *lacZ* reporter genes are not de-repressed. For the three PESE-containing reporters that retain sensitivity to repression by Runt (3918, 3125 and 3118) each stripe stays 1-2 cells wide and did not expand or shift in a *ftz* mutant background. Only the truncated PESE[C1+] reporter showed expanded expression in cells anterior to the normal even-numbered *slp1* stripes, and a similar expansion was observed in the presence or absence of Ftz. These experiments provide strong evidence that Ftz is not necessary for repression of *PESE-lacZ.*

I also examined the role of Ftz in the Runt-dependent repression of PESE by comparing the response of the PESE[3125]-p126-lacZ reporter to ectopic Runt versus ectopic co-expression of Runt and Ftz. This reporter showed nearly complete repression in both genotypes (Figure 3-6). However, the embryos overexpressing Runt also have

some endogenous levels of Ftz. To truly test if Runt is sufficient to repress PESE without Ftz, it was necessary to overexpress Runt in a *ftz* mutant background.

Ectopic expression of Runt and Opa in a *ftz* mutant background activates the *slp1* gene throughout the embryo (Swantek & Gergen, 2004), but does it also activate PESE throughout the embryo? To test if Runt alone is sufficient to repress PESE without *ftz*, this same ectopic expression was performed in a cross with the full length PESE reporter. All embryos of the cross overexpressed Runt and Opa. Approximately ¾ of the embryos showed the typical response of *slp1* to ectopic co-expression of Runt and Opa with seven 4 cell wide stripes in the segmented reion of the embryo and activation throughout head (Figure 3-7). As expected, a subset of these embryos corresponding to the presumed balancer chromosome homozygotes (Figure 3-8A) showed no evidence of the reporter gene. Approximately one quarter of the embryos showed essentially uniform expression of the endogenous *slp1* gene (Figure 3-8). I presume that these embryos correspond to the *ftz* mutants. Amongst the embryos with uniform *ftz* expression were one class that also showed relatively uniform expression of *lacZ* (Figure 3-8A), and a second class that showed complete repression of *lacZ* (Figure 3-8B). Both of these homozygous *ftz* mutant embryos also should be homozygous for the PESE-containing lacZ reporter gene as it was on the same chromosome as the *ftz* mutation in these crosses. Therefore I assume the dramatic difference in expression is due to a difference in developmental stage. There were several embryos that were homozygous *ftz* mutants with *lacZ* expressed everywhere as in Figure 3-8A and only a few embryos in which endogenous *slp1* was expressed throughout the embryo and *lacZ* had no expression as in 3-8B. Consistent with the idea that there is a shorter window of time in which Runt can effectively can repress PESE in

a Ftz-independent manner, *PESE-lacZ* was repressed in a younger embryo in which the mesoderm was not invaginated, whereas an older embryos had *lacZ* activated throughout the embryo. This observation suggests that later on in development, Ftz maybe required to prevent Opa-dependent activation of PESE in cells that also express Runt. The complicated response of the PESE-containing reporter to Runt and Opa in *ftz* heterozygous embryos (Figure. 3-7B) provides additional evidence that numerous factors contribute to the response of the PESE enhancer in the early Drosophila embryo.

Discussion

The result of mutations to Runt binding sites in PESE in a *lacZ* reporter gene, in which there was no expansion of the reporter with the mutated binding sites more than the reporter without mutations, in the dorsal region of the embryo (Figure 3-3 A and B), can be interpreted several ways that lead to different hypotheses. One possibility is that these Runt binding sites are not the ones that affect the mRNA transcription pattern. There may be other binding sites within the same enhancer to which Runt binds and affects the function of PESE in regulating the expression of the gene. The other possibility is DNA binding is not required for Runt to repress *PESE-lacZ*. However, structurally, these binding sites may still be required for Runt binding and interacting with other proteins, but their mutation does not affect expression of *PESE-lacZ*. Note that there may be an effect in the ventral region of the embryo, where the stripes of PESE m1-6 p381 lacZ are wider. (Figure 3-2B)

The composite *lacZ* reporters of DESE and PESE with and without mutations of the Runt binding sites in PESE are expressed in a 14-stripe pattern recapitulating the *slp1* pattern of expression (Figure 3-4). There is no inappropriate expression anterior to the

odd stripes in the embryos at the stage imaged. The embryos of composite reporters of DESE and PESEm1-6 at younger stages before the even-numbered *slp1* stripes are formed or are just forming do not have expression of *lacZ* stripes in type 3 cells in most the embryos. At the onset of the odd-numbered *slp1* stripes, there are some cells expressing *lacZ* and not *slp1* anterior to the even-numbered stripes, but these stripes are no more expanded or misexpressed than the composite with the normal *PESE 3125 p381 lacZ* sequence. The stripes are either one or two cells wide with occasionally a few extra cells. The mutated binding sites do not affect composite reporter expression either because they have no affect, or the dominant repression by Runt and Ftz mediated by DESE in type III cells prevents the stripe expansion in the ventral part of the embryo.

Four different reporter *ftz* mutant stocks were examined to see if Ftz was necessary for repression of PESE and if so which regions were necessary. All *lacZ* PESE reporters regardless of *ftz* mutation had seven 1-2 cell wide stripes, while the endogenous slp1 gene had 6 cell wide stripes due to de-repression (Figure 3-5). While *3918 p126 lacZ* showed slight de-repression due to loss of Ftz, *3125 p126 lacZ* in a *ftz* mutant background showed none. Neither did *3118-p126-lacZ* or *C1+-p126-lacZ* when comparing the control to *ftz* mutant. The loss of function experiments for the full length and truncated PESE reporters suggest that Ftz is not necessary for repression of PESE.

One result shown (Figure 3-5) suggests that ectopic expression of Runt alone completely represses minimal PESE with the small basal promoter, *3125 p126 lacZ*. I overexpressed Runt alone, and Runt and Ftz, but adding Ftz to the overexpression did not change the effect on the *lacZ* reporter pattern. Since there was no further repression with the addition of Ftz, this suggests Ftz does not contribute to repression of PESE by Runt,

and that Runt may be sufficient to repress. However, the experiment with ectopic expression is not conclusive because those flies, even the Runt overexpressed alone, have natural levels of Ftz, which may be necessary for repression by Runt.

Finally, in an experiment that was repeated with many embryos, I tested if Runt required *ftz* activity to repress PESE. In this experiment I overexpressed Runt and Opa with PESE-*lacZ* in a cross where ¼ of the progeny were also homozygous mutant for *ftz*. The internal control was the *slp1* pattern. If the embryo was homozygous for the *ftz* mutation, it would be fully activated by Runt and Opa, and the green *slp1* signal would be observed everywhere, but it was unknown what would happen to the *lacZ*. My hypothesis was that overexpression of Runt would be sufficient to repress *PESE-lacZ* and that there would be no red signal in the embryo for *lacZ*. Indeed there are a few embryos of this phenotype in both experiments. However, at later time points, though it is hard to see with the saturated signal, the *lacZ* gene is completely activated in several embryos (Fig.3-8). This result suggests that Opa even in the presence of Runt activates PESE and that sometimes, depending on stage and level of Opa, this activation is dominant over repression by Runt. Although I have interpreted these differences as due to the developmental stage further work should be done to confirm this. Careful staging requires using morphology such as the headfold and ventral furrow to time age of embryo. Another way to track changes over time is live imaging of embryos in real time.

Further evidence for such complications comes from the observations on the heterozygous *ftz* mutants obtained in this cross. As expected, *slp1* had 7 wide stripes, whereas the PESE-containing reporter showed evidence of both incomplete repression as

well as some ectopic activation (Fig.3-7B). Could Runt be cooperating with Opa to activate PESE as it does for DESE, and is Ftz necessary for repression?

One explanation for the activation of *lacZ* throughout the embryo at later stages is autoregulation by endogenous *slp1* that has been uniformly activated in these embryos. The more *slp1* is produced, the more the transcription factor is available to activate it and other genes like the *lacZ* reporter genes with the *slp1* promoter. The overexpression of Opa activates PESE and the overexpression of Runt represses PESE. The experiment to test if Runt was sufficient to repress PESE without Ftz could be repeated by just overexpressing Runt and not overexpressing Opa, but another control other than the *slp1* pattern would be necessary to identify embryos that are homozygous *ftz* mutants with the reporter.

Figures

	Wildtype	Mutant
0001	GGACTAGTgtcgacaggaca	GGACTAGTgtcgacaggaca
0021	atgactgaaaggcaaagtgc	atgactgaaaggcaaagtgc
0041	gaggatgaggatgagggcga	gaggatgaggatgagggcga
0061	aaagacgactcgctgtgcgc	aaagacgactcgctgtgcgc
0081	gacaaatgccgaagatatcc	gacaaatgccgaagatatcc
0101	tgctgatgtcgcagccatgt	tgctgatgtcgcagccatgt
0121	gacagtctccgaggatcgga	gacagtctccgaggatcgga
0141	tcgttgaggataggactcgc	tcgttgaggataggactcgc
0161	agcgagtcctgttgacagcg	agcgagtcctgttgacagcg
0181	aggttcctcaaaaggaatat	aggttcctcaaaaggaatat
0201	cgcacattacgatcaattat	cgcacattacgatcaattat
0221	gagactgaatgctcacccac	gagactgaatgctcacccac
0241	agtggaacgaaaacgggcga	agtggaacgaaaacgggcga
0261	aagtcctgaaggagggaaat	aagtcctgaaggagggaaat
0281	cttgt**gg**cgttaattgaaaa	cttgt**TT**cgttaattgaaaa
0301	ttacatgcacattgtctgt**g**	ttacatgcacattgtctgt**T**
0321	**gc**cacggtgtctacttggaa	**TA**cacggtgtctacttggaa
0341	ttcgcttatgcatctccata	ttcgcttatgcatctccata
0361	gttttccatagttttccata	gttttccatagttttccata
0381	gttttctatagtttccctac	gttttctatagtttccctac
0401	gtttttgtgagctaatgagg	gtttttgtgagctaatgagg
0421	caggcgtagcgttgcctcga	caggcgtagcgttgcctcga
0441	agcgtgaagctcgccaagga	agcgtgaagctcgccaagga
0461	aacgcctaagtgt**gg**ttttg	aacgcctaagtgt**TT**ttttg
0481	gcaatgaaaccataatgtca	gcaatgaaaccataatgtca
0501	aacagcgtggtcctgatcgc	aacagcgtggtcctgatcgc
0521	gtcctgatcacccaatcagc	gtcctgatcacccaatcagc
0541	**gg**ccacatgctctcggtctg	**TT**ccacatgctctcggtctg
0561	atccacgctcgattatgctc	atccacgctcgattatgctc
0581	aaggtgtgccctcaaggatc	aaggtgtgccctcaaggatc
0601	tcgaatcgctaatcaattcg	tcgaatcgctaatcaattcg
0621	ttttaccacgacgtcACTAG	ttttaAAacgacgtcACTAG
0641	TGG	TGG

Figure 3-1. Synthesized sequences of insert for lacZ reporter constructs for PESE and PESE m1-6 constructs, and composites with DESE

The 0.6kb sequence, spanning from -3.1kb to -2.5kb upstream the *slp1* TSS and flanked by SpeI sites and two guanosines, is exhibited for PESE normal (wild type) and PESEm1-6. The critical basepairs of each binding site that were mutated are highlighted in blue, with the normal sequence left, and the mutations on the right. G was changed to T, and C changed to A.

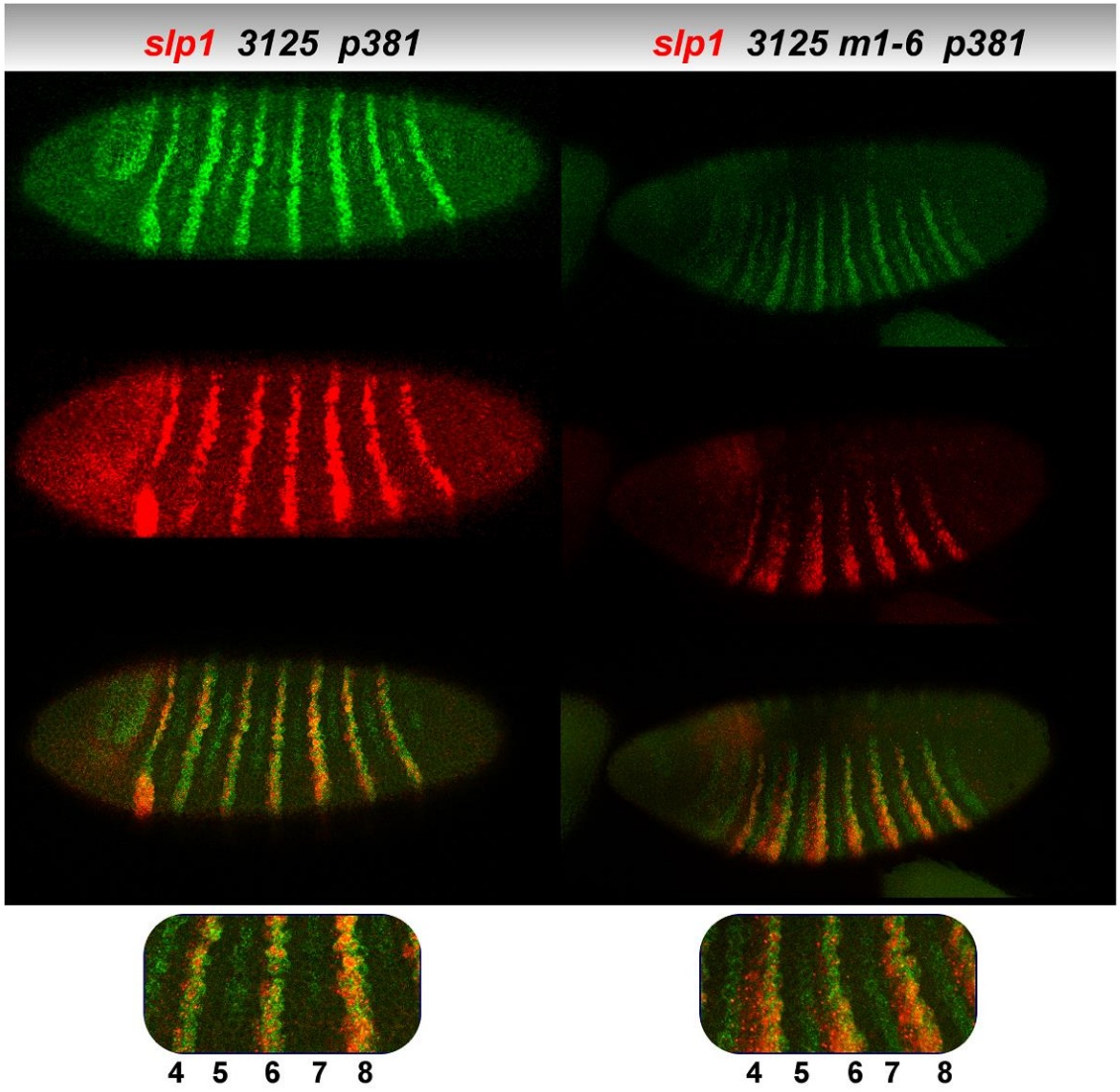

Figure 3-2. Runt binding site mutations in PESE result in early expansion of PESE stripes.

The *slp1* (green) and *lacZ* (red) mRNA expression patterns are shown for two *lacZ* reporter gene construct, *3125-p381-lacZ* and *3125m1-6-p381-lacZ.* The PESE enhancer sequence of *3125-m1-6-p381-lacZ* contains mutations in six putative Runt binding sites. In the image of stripes labeled 4, 6, and 8 below shows a section of the merged image enlarged to show derepression into type III cells anterior to the even stripes.

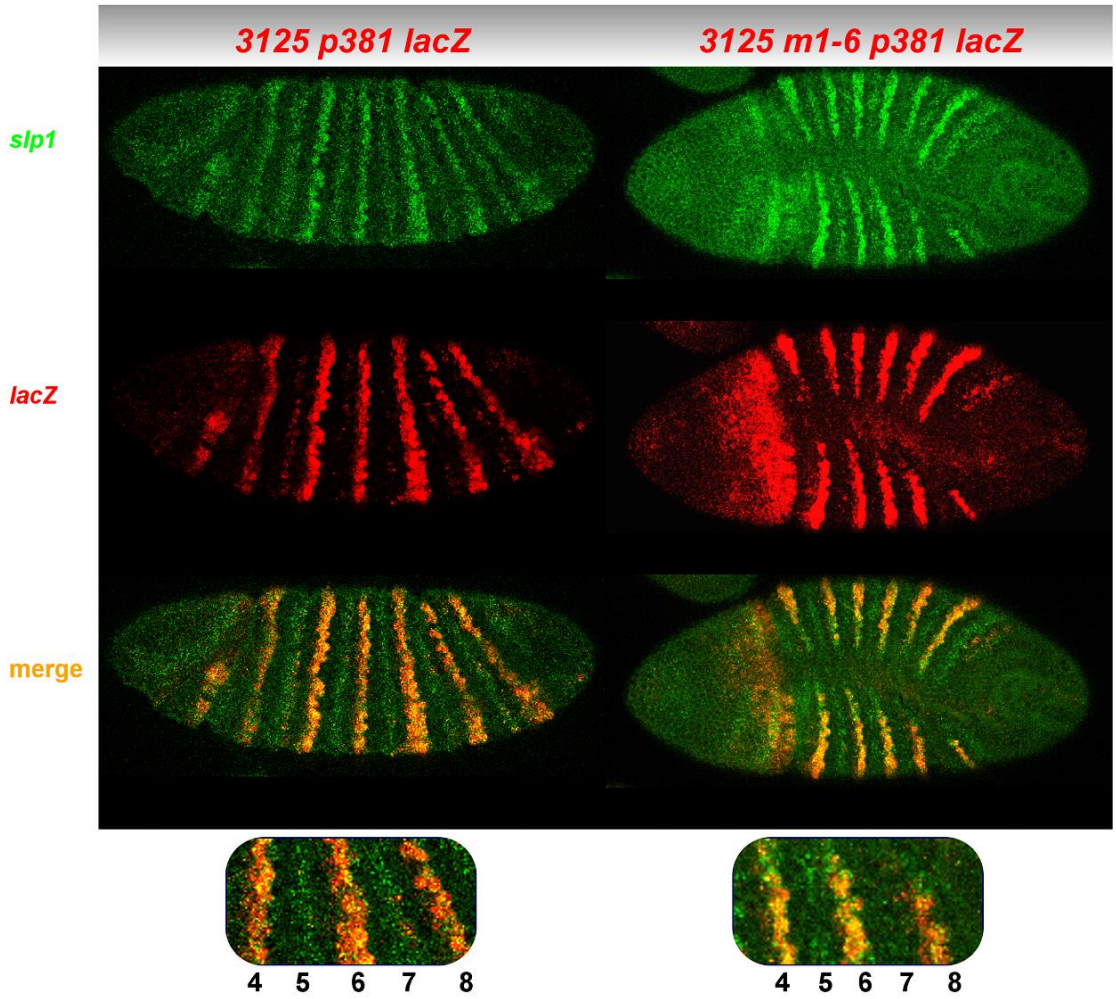

Figure 3-3. Six putative Runt binding sites do not affect later expression of PESE.

The mRNA expression pattern of *slp1* (green) and *lacZ* (red) fluorescently labeled embryos A. *3125 p381-lacZ* without Runt binding site mutations B. *3125 m1-6 p381* with Runt binding site mutations in the gastrulated stage. In both embryos, *slp1* is expressed in 14 stripes and *lacZ* is expressed in 7 stripes corresponding to the even-numbered stripes of *slp1*. The magnified image below shows stripes 4, 6, and 8 enlarged to show slight expansion of *lacZ* anterior to the even-numbered stripes for both reporter genes.

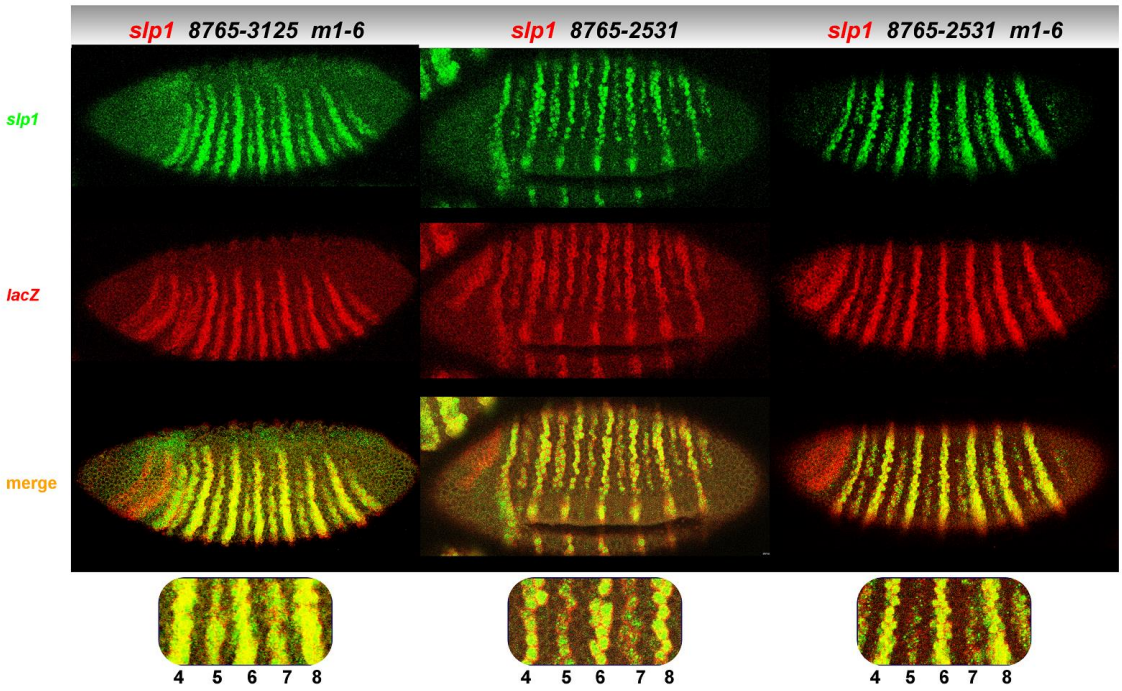

Figure 3-4. Runt binding site mutations in PESE do not affect non-additive interactions with DESE.

The mRNA expression patterns of *lacZ* (green) and *slp1* (red) are shown above. Three *lacZ* reporters, *8765-3125m1-6-p381-lacZ*, *8765-2531-p381-lacZ*, and *8765-2531m1-6-p381-lacZ* recapitulate the endogenous gene pattern, except for two *slp1* stripes in the head region.

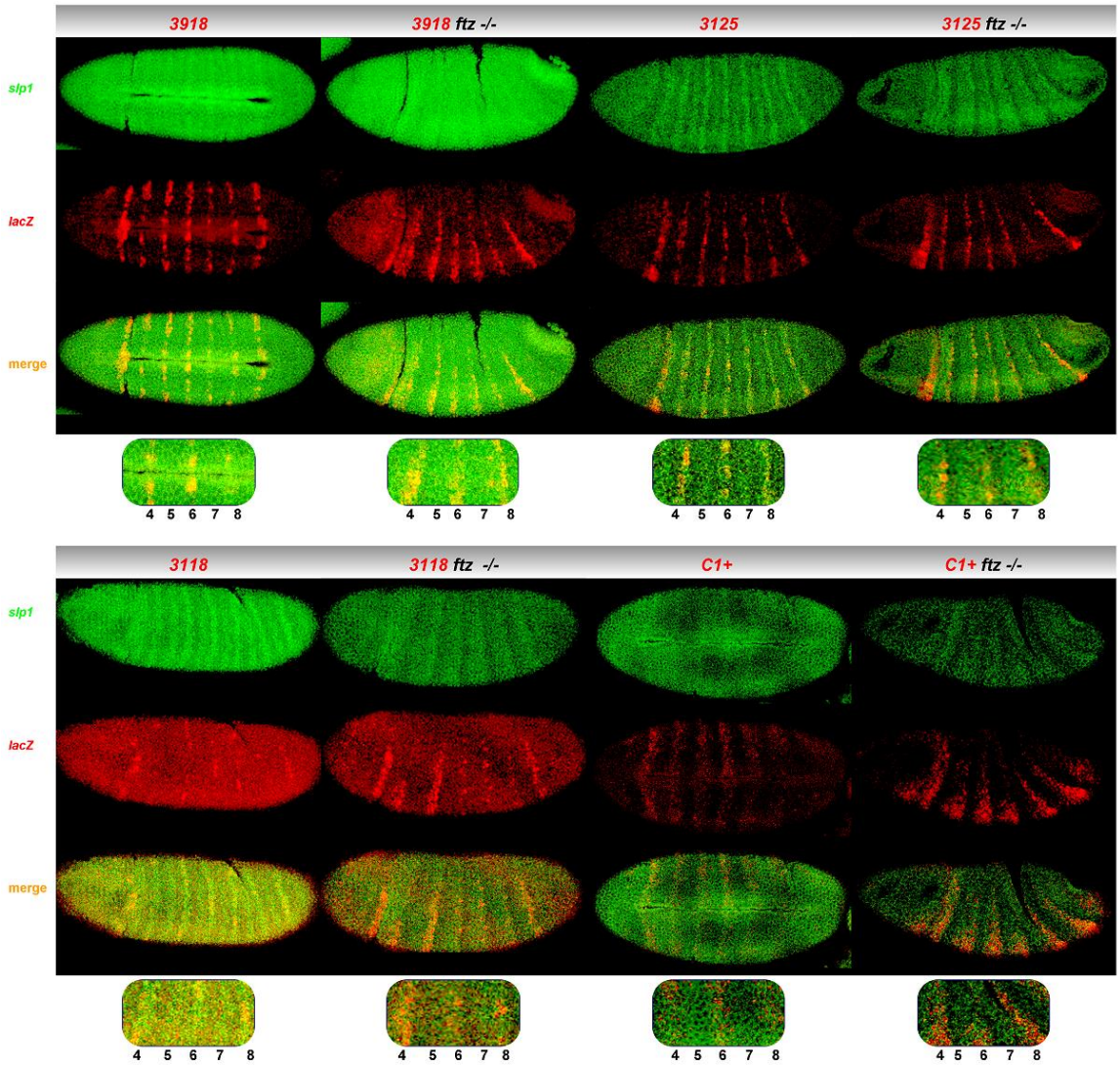

Figure 3-5. PESE p126 is not significantly derepressed in ftz mutant embryos.

Embryos of four reporters of *PESE: 3918, 3125, 3118* and *C1+* are in wild-type and *ftz* mutant backgrounds with *slp1* (green) and *lacZ* (red) fluorescently labeled. Though *3918-p126-lacZ* slightly expands into type 3 cells anterior to the even stripes in a *ftz* mutant background, the *3125-p126-lacZ* reporter does not have expanded stripes. Both *3118-p126-lacZ* and *C1+-p126-lacZ* show slight derepression in the *ftz* mutant

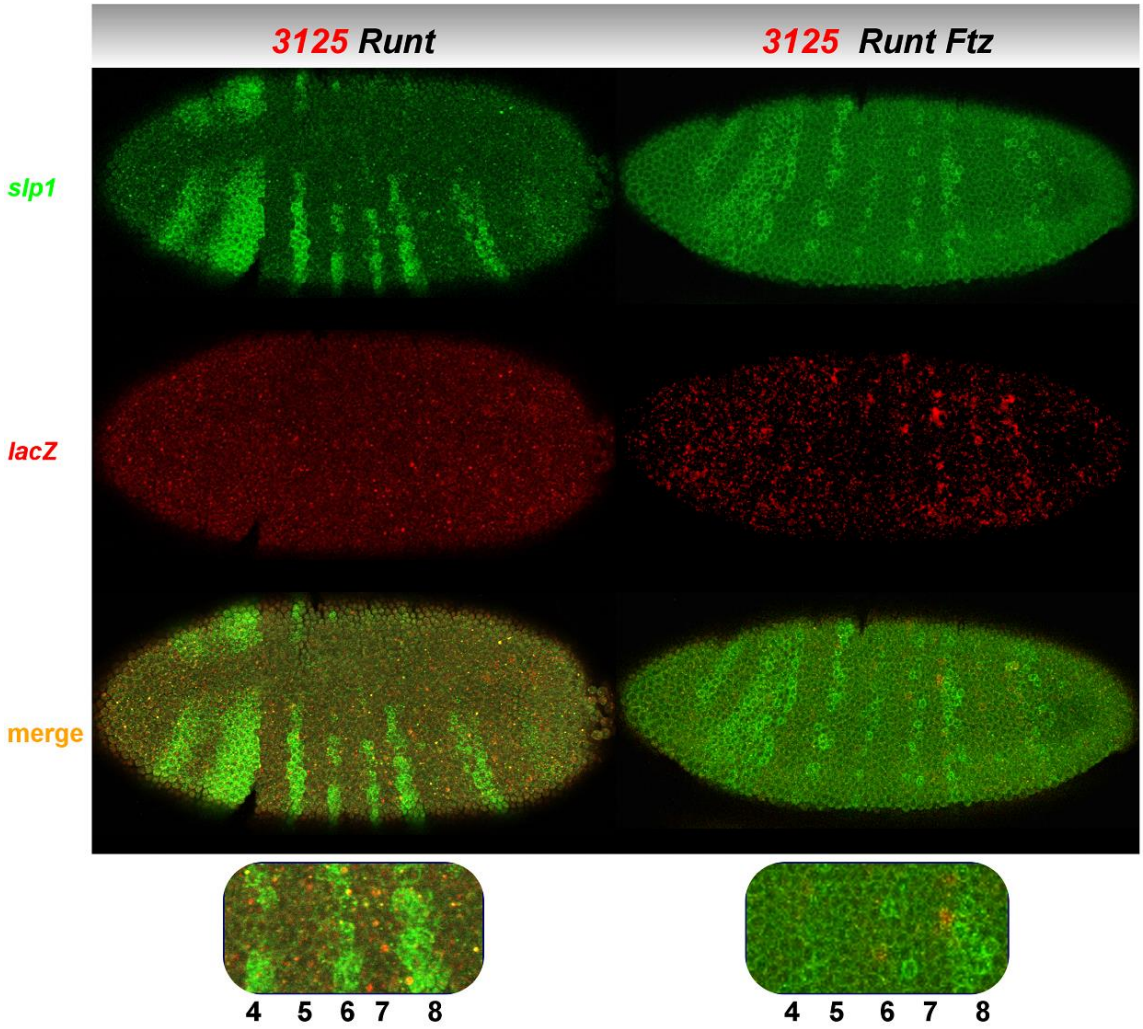

Figure 3-6. PESE p126 is effectively repressed by overexpression of Runt alone, as well as Runt and Ftz.

NGT40 3125 p126 lacZ mRNA expression is repressed completely by ectopic expression Runt as shown on the left. Ectopic expression of Runt and Ftz also fully represses the reporter gene.

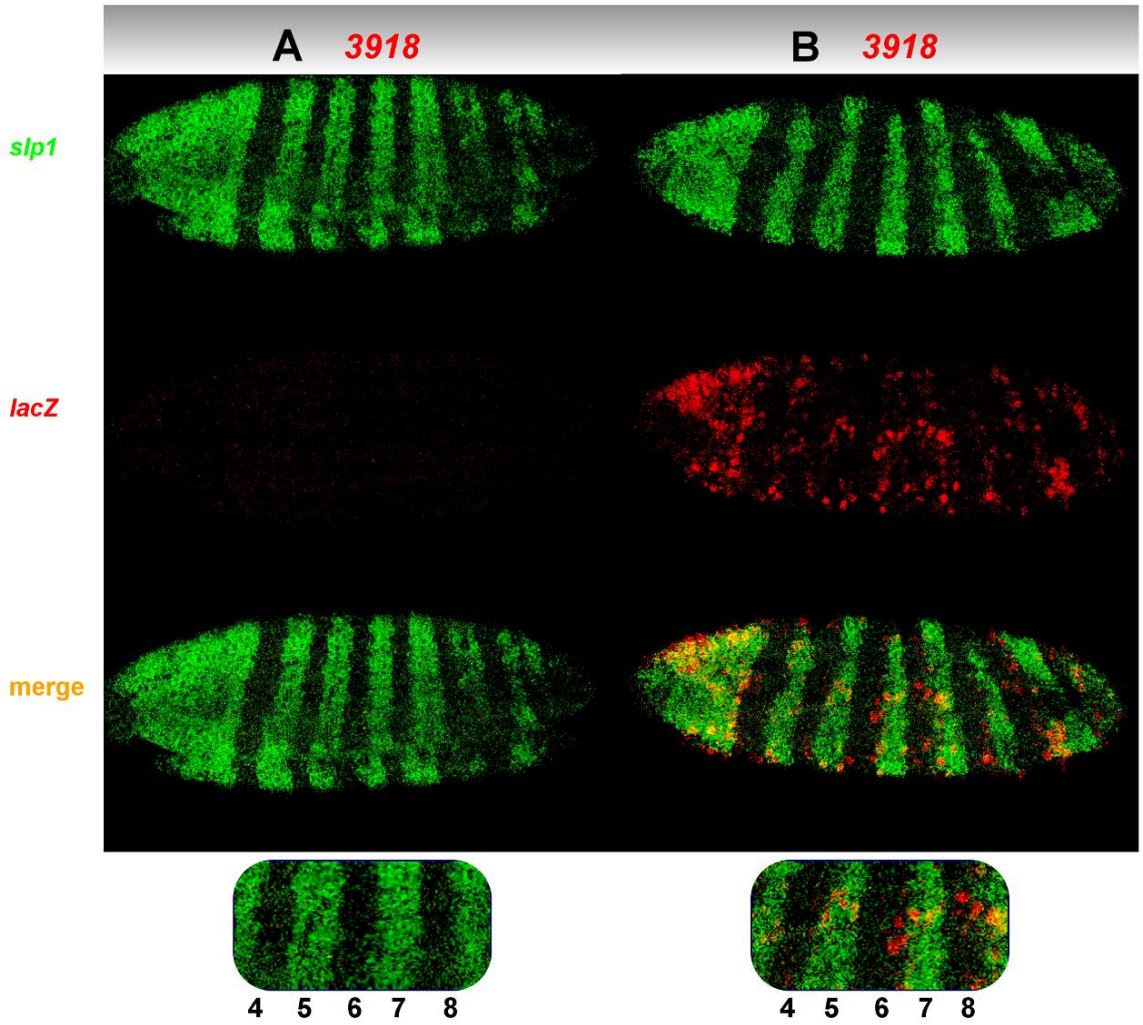

Figure 3-7. PESE expression is altered by overexpression of Runt and Opa in a heterozygous ftz mutant background.

The embryo on the left lacks the reporter gene, but the *slp1* pattern is altered by ectopic expression of Runt and Opa. The embryo on the right has the reporter, which is activated in the head region and has spotty expression throughout the pre-segmented region. The presence of red cells in the merged image reveals cells in which the endogenous *slp1* gene, but not the PESE-containing reporter, is effectively repressed. In contrast, cells in this same embryo that co-express *slp1* and *lacZ* reveal ectopic activation of the PESE-containing reporter in cells where the PESE-reporter is usually not expressed.

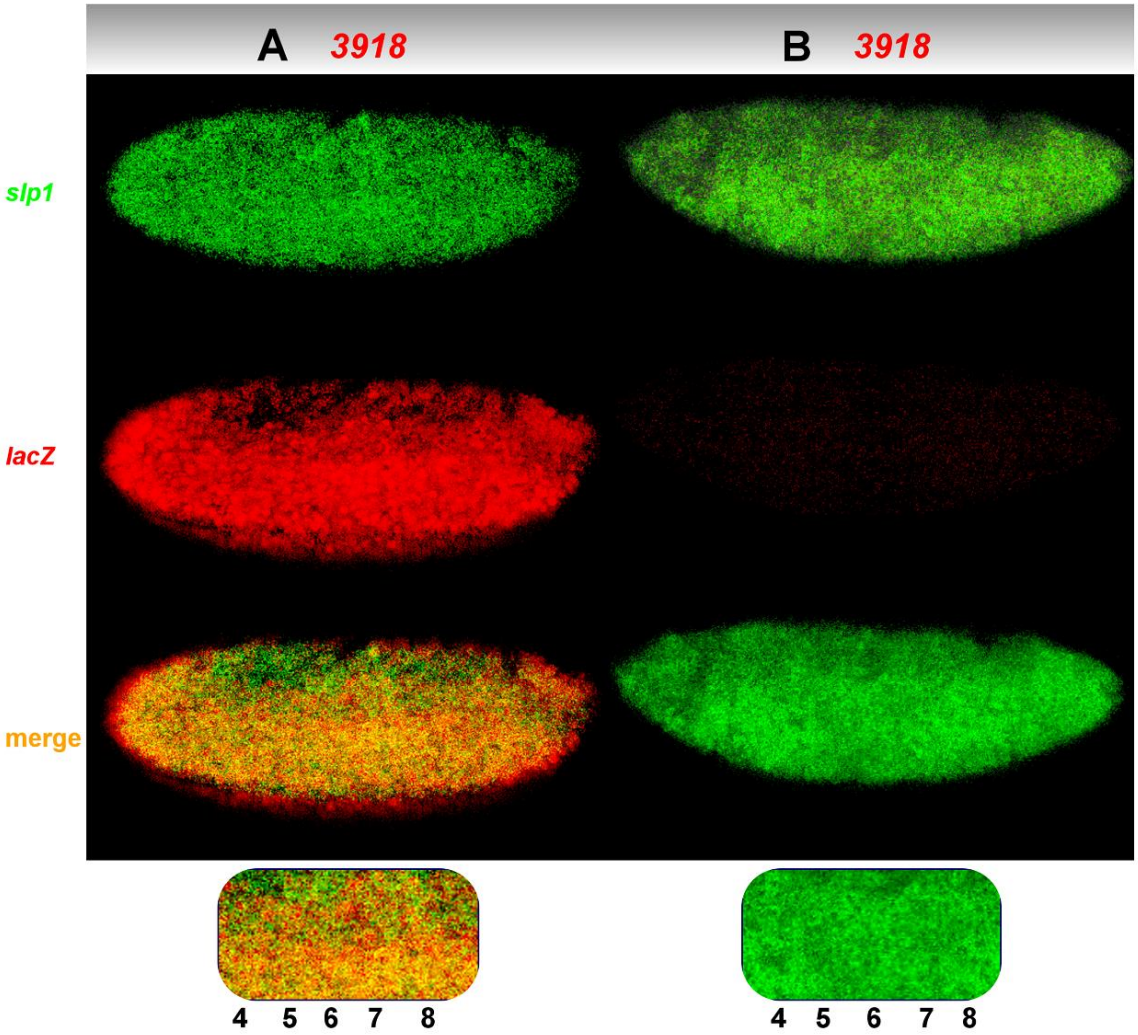

Figure 3-8. Runt and Opa both activate and repress PESE-lacZ in homozygous ftz mutant background.

The mRNA expression of *slp1* and *lacZ* are shown. In both embryos *slp1* is activated by Runt and Opa in the absence of *ftz* throughout the embryo. Embryos homozygous for the *ftz* mutation and containing *PESE (3918)-lacZ* are of two phenotypes; *lacZ* activated throughout the embryo as shown on the left (red), and repressed throughout the embryo, as shown on the right.

Chapter 4: Progress towards identification of binding sites for Opa

Abstract

Opa is a transcription factor that shares homology to the zinc finger in the cerebellum (ZIC) family of proteins that are important in human development. Opa is an important activator of both the *slp1* DESE and PESE enhancers. Runt potentiates DESE-mediated Opa-dependent activation, but has the opposite effect of repressing PESE-mediated activation by Opa. Understanding the architecture of the Runt and Opa binding sites within these two enhancers is likely to be key to understanding Runt's differential effect on the Opa-dependent activation of these two enhancers. The function and location of bona fide Opa binding sites has yet to be determined for any genes in Drosophila. In this work Opa antibodies were developed and tested through Western blot, immunostain, and ChIP experiments. Although immunostaining experiments were inconclusive a Western Blot demonstrates the specificity of Opa antisera raised against a peptide. Preliminary ChIP experiments with a different antibody raised against a truncated recombinant Opa protein suggest that Opa binds to a functionally important region of DESE.

Introduction

Opa is a Drosophila ortholog of the zinc finger in the cerebellum (ZIC) family of transcription factors, which are important in human development (Houtmeyers, Souopgui, Tejpar, & Arkell, 2013). In humans, mutations to ZIC genes cause malformations of the brain and head such as holoprosencephaly, Dandy Walker malformation, and coronal craniosynostosis. (Sen, Stultz, Lee, & Hursh, 2010; Twigg et al., 2015). ZIC2 regulates transcription by binding to enhancers, interacts with a chromatin remodeling complex, and is required to control embryonic stem cell specification (Luo et al., 2015). Opa has zinc fingers homologous to the zinc finger transcription factor of the human glioblastoma oncogene, GLI, for which a crystal structure of a complex with the five zinc fingers bound to DNA was determined (Benedyk, Mullen, & DiNardo, 1994; Pavletich & Pabo, 1993). This structural information potentially provides a basis for design of experiments that further explore the role of DNA-binding by Opa.

In Drosophila, Opa is an important activator of *slp1* and other segment-polarity genes during segmentation and also plays a role in adult head development (Sen et al., 2010). Opa contributes to the activation of both the *slp1* DESE and PESE enhancers. The odd-numbered *slp1* stripes, as well as the odd-stripes produced by the DESE-containing reporters are eliminated in *opa* mutant embryos. The even-numbered *slp1* stripes are also reduced in *opa* mutants, and this residual expression is presumably due to DESE as the expression of PESE-containing reporters is also eliminated in the absence of *opa* function. This residual DESE-dependent expression of the even-numbered stripes is

eliminated in *opa* mutant embryos that are also mutant for *unpaired*, a ligand for the JAK-STAT signaling pathway (Liujing Xing and JP.Gergen, personal communication).

Opa has been shown to bind to DNA in electromobility shift assays, and a DNA binding motif has been determined by systematic evolution of ligands by exponential enrichment (SELEX) to be CGGGGGGTC.(Sen et al., 2010) A synthetic reporter containing part of DESE from 7670 to 7153 bp upstream of the *slp1* promoter with a natural Opa binding site closely matching the SELEX motif, did not lose activity caused by overexpressing Opa when the binding sites were mutated in late stage Drsophila embryos (Sen et al., 2010) Though the authors show that direct DNA-binding was not required for activation of a synthetic reporter construct with this one Opa binding site, they conclude that another DNA binding partner may enable Opa to activate the gene, because there is a conserved sequence directly upstream of the Opa binding site (Sen et al., 2010).

Though Opa was shown to bind this binding site in *in vitro* assays, ChIP assays are necessary to demonstrate the binding of Opa *in vivo*. I used two companies to generate polyclonal anti-Opa antibodies from rabbits: Genscript and Abclonal. The antibody from Genscript was generated from immunization with a peptide, and the antibody from Abclonal used a recombinant truncated Opa protein as the antigen for immunization. These antibodies were tested in three different assays: Western blots, immunostaining of embryos, and ChIP experiments. In this chapter, I will show an experiment that validates the specificity of the Genscript anti-peptide antibody with two recombinant Opa proteins from bacteria. I will also show a pilot experiment with the Abclonal anti-Opa antibodies that detect specific binding in the DESE enhancer of *slp1*.

Results

Generation of Opa antibodies

We generated several Opa antibodies using two companies. The Genscript antibody was generated from a peptide and affinity purified. For Abclonal, we supplied the company a plasmid with the full length *opa* gene. They cloned this gene and two truncated forms into other vectors for protein expression. The construct with the best expression was a truncated *opa* gene, spanning from amino acids 125-507 of the Opa protein. This was used as the antigen for immunizing the rabbits. Blood was taken from three different rabbits and the antibodies were affinity purified using the recombinant protein antigen. Three different assays were done to see if the custom Oddpaired antibody bound to the Opa protein. In this chapter I will discuss Western blots, immunostainings, and ChIP's.

Immunostain (data not shown)

All of the antibodies were tested in fluorescent immunostains of Drosophila embryos at different dilutions. Regardless of dilution, the antibody stain showed either strong staining across the whole embryo, or background signal. There was no unique staining pattern that matched the in situ of the Opa mRNA at cellular blastoderm stage, in which Opa is expressed in the presegmented body region and not the head or tail. Two antibodies, the Abclonal E992 and Genscript anti-peptide antisera, were also tested in an immunohistochemical immunostain. Neither of these antibodies showed a distinct pattern in this assay as well.

Western blot

Bacterial extracts run on an SDS-PAGE gel were from pMal-p2x, the negative control with maltose binding protein (MBP), and p-Mal-p2x-FL Opa, the plasmid with the full length Opa protein tagged with maltose binding protein (Opa-MBP). The recombinant Opa protein truncated and spanning from 125-507 amino acids from the antigen was also run on the gel. In a Western blot with Opa antibody generated from a peptide (by Genscript), the MBP protein extract negative control had no bands (Fig.4-1 left). The MBP-Opa protein in the second lane ran above 100kDa, and the antigen from Abclonal ran at 70kDa (Fig.4-1 lane 3). On another membrane with the same protein samples, but with the Genscript Opa antibody blocked with peptide, there were no bands. This experiment provides evidence that the anti-peptide antibody specifically binds to peptide-containing region of the MBP-Opa and recombinant truncated Opa proteins.

Chromatin Immunoprecipitation

After first testing TBP and H3K4me3 antibodies as positive control antibodies and obtaining 0.1% input at the promoter region and actin5c gene respectively, I then obtained in a separate experiment 0.1% input for Opa antibody E990 at the Opa binding site in DESE, using TBP as a positive control antibody. Levels of Opa matching the background serum were found for the *slp1* promoter, the control region of *odorant receptor 42B*, and *actin5c* gene. These served as negative controls. No locus was used as a positive control other than the experimental regions of slp1: DESE, PESE-C1+, and slp1 promoter.

Discussion

The negative results of the immunostaining experiments suggest two possibilities. If the fluorescent immunostains were done correctly, the strong signal across the embryo suggests that Opa is expressed in high levels in every cell of the embryo. More likely however, this is background signal from an experiment that needs optimization. Diluting the antibody did not improve the signal. However, the possibility remains that the epitopes of the polyclonal antibodies are hidden by protein folding in the native in vivo conformation.

The Western blot with Genscript anti-Opa antibody raised against a peptide indicated it was specific for Opa protein and peptide. While this is an in vitro experiment, it detects both full length and truncated protein. Importantly, the peptide sequence use to generate the antibody is within the truncated protein.

The Opa ChIP described above (Fig.4-2) was the first experiment to show DNA occupancy or binding at the specific binding site for Opa in the DESE enhancer i*n vivo*. The binding site, AGGGGGGTA, is a close match with the SELEX consensus sequence, CGGGGGGTC, and Opa bound to it in EMSA (Sen et al., 2010). However it was not known whether Opa bound this region *in vivo* in the cellular blastoderm stage. While the Sen et. al paper assayed activation of a reporter with the binding site and a few hundred basepairs of the DESE *slp1* DNA, the ChIP tests occupancy at the locus of the endogenous *slp1* for three different regions. While the exact concentration and percentage input may have a margin of error, the binding of Opa at the DESE Opa binding site is above background above the margin of error.

Figures

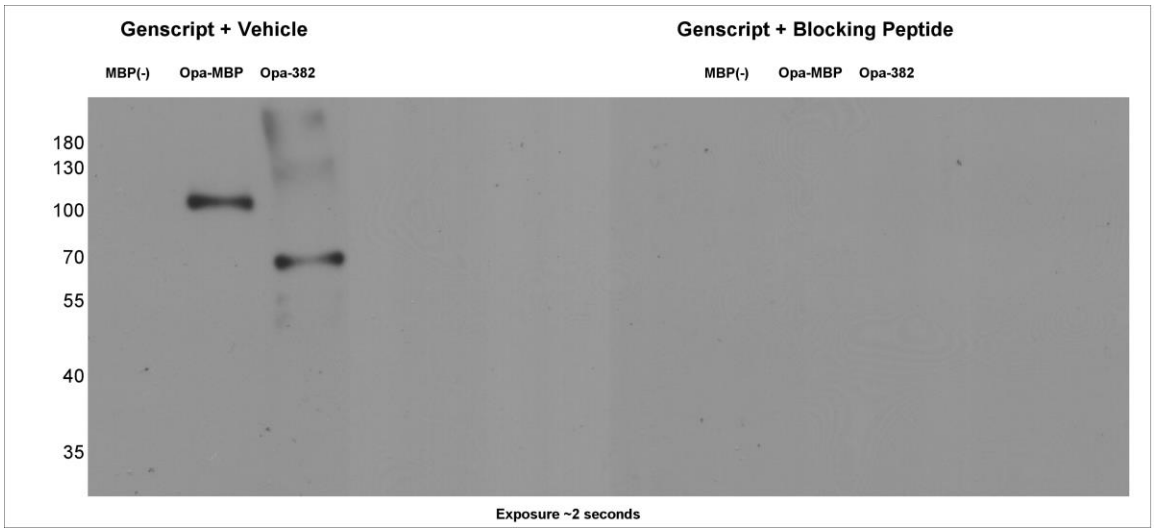

Figure 4-1. The Genscript Oddpaired antibody is specific for the two Opa proteins recombinantly expressed from bacteria and for the peptide.

A Western blots of two membranes is shown; the membrane on the left is incubated with Opa antibody and vehicle (water and proteinase inhibitor) and the membrane on the right is incubated with Opa antibody and blocking peptide (in water and proteinase inhibitor). Each membrane was from a gel running two bacterial extracts, a maltose binding protein (MBP) and a full length Opa protein tagged with maltose binding protein (Opa-MBP), and truncated recombinant Opa protein (Opa-382).

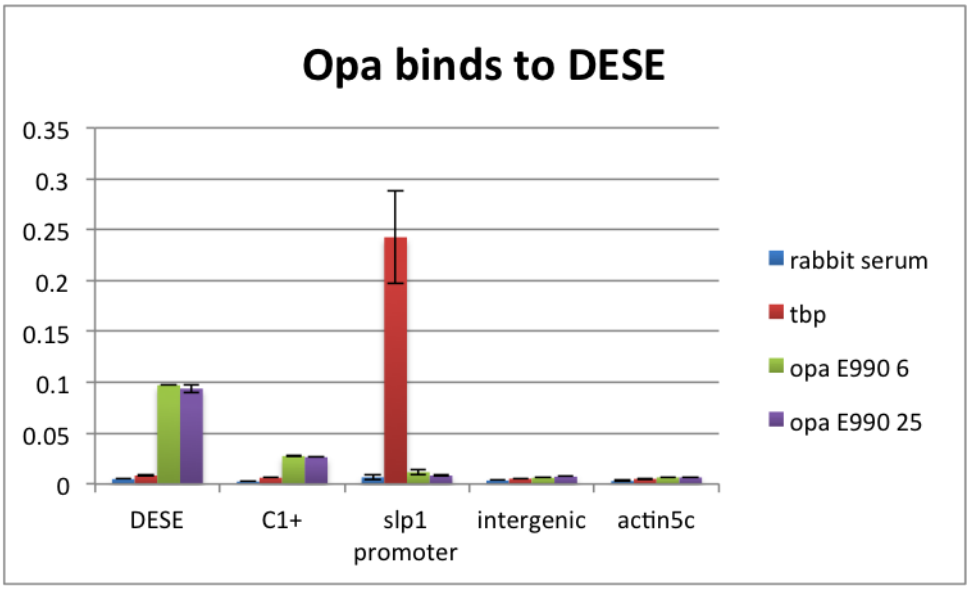

Figure 4-2. Opa binds to a region including an Opa binding site in DESE.

A bar graph showing the percent input of 4 ChIPs is shown: rabbit serum (blue), TBP (red), and Opa E990 with two different amounts is shown 6µl (green) and 25µl(purple) for five regions; *DESE*, *C1+*, the *slp1* promoter, and two control regions (*odorant receptor 42 B* and *actin5c*). The Y-axis is the percentage of the input chromatin, and the X-axis is the region amplified by specific primers. Rabbit serum was used as a negative control and TBP as a positive control.

Chapter 5: Characterization of the PESE C1 region and its interaction with DESE

Abstract

Prior studies on the PESE enhancer identified a 155 bp region referred to as C1 (most upstream deletion within the Central region of PESE) that was required for expression in a reporter gene containing the small p126 basal promoter region (Prazak et al., 2010). A truncated 272 bp sub-element of PESE referred to as PESE[C1+] that spans the C1 and that contains distal and proximal extensions of 44 and 73 bp, respectively is able to drive expression that is sensitive to repression by Eve, but not to repression by Runt and Ftz. (Prazak et al., 2010). Conversely, the DESE enhancer of the *DESE-p381-lacZ* construct was shown to be insensitive to repression by Eve, and is repressed by Runt and Ftz (Prazak et al., 2010). In this thesis, the interactions of DESE and the PESE[C1+] sub-element are investigated in reporter gene constructs with both the small (p126) and large (p381) *slp1* basal promoters. A PESE-containing construct that lacks the C1 region, but in combination with the large p381 basal promoter, *PESE[3918ΔC1]-p381-lacZ*, has delayed and weakened expression of lacZ, but, unlike *3918ΔC1-p126-lacZ* does not lose complete activation of the stripes, suggesting that C1 contributes to activation by PESE in a manner that depends on the size of the basal promoter region. Interestingly, both the *DESE-PESE[C1+]-p126-lacZ* and *DESE-PESE[C1+]p381-lacZ* reporters recapitulate the endogenous slp1 pattern, suggesting that Eve-dependent repression mediated by C1+ blocks DESE-driven inappropriate expression in type I cells, and further that repression of DESE by Runt and Ftz blocks expression driven by C1+ in type III cells.

Introduction

Prior work by Lisa Prazak showed that the C1 region of PESE, a 272 bp subelement including DNA from 3.1 to 2.9 kb upstream of *slp1* TSS, is necessary for activation of *PESE 3918 p126* and that the *C1+-p126-lacZ* construct has expression in 4-cell wide stripes that is sensitive to repression by Eve, but not to Runt and Ftz (Prazak et al., 2010). The DESE enhancer is sensitive to repression by Runt and Ftz, but not Eve. She also found that the full length DESE enhancer did not work with the small basal promoter, p126, from -72 to +57 bp from the *slp1* TSS, and drove transcription with a larger basal promoter, p381, which spans from -260 to +121 from *slp1* TSS.

In this chapter I describe generation of flies combining DESE and C1+ in composite *lacZ* reporters, one with p126 and the other with p381. I show images of each reporter, with images of the individual enhancers previously described as controls. I also show full length PESE with the C1 region deleted and with the large *slp1* basal promoter. I compared the expression of constructs with the large basal promoter to constructs with the small basal promoter. Finally, I discuss future experiments to test enhancer dominance by repression by preventing the release of paused Pol II.

Results

3918 Δ C1-p381-lacZ has 7-stripes of expression in the absence of the C1 region and head activation, suggesting the region contributes to, but is not necessary for PESE stripe expression.

The 155 bp C1 section of PESE is necessary for expression when PESE is tested with the smaller p126 *slp1* basal promoter (Prazak et al., 2010). *The PESE[3918 Δ C1]-p381-lacZ* reporter also showed greatly reduced expression in early stages, but with clear evidence of stripes in head-fold and later stage embryos (Figure 5-1). This result reinforces the importance of the C1 region for early PESE-driven expression within the segmented region of the embryo. It is interesting to note that this construct has activation throughout the head region.

DESE-C1+ lacZ recapitulates the endogenous *slp1* pattern

In this work, two composite *lacZ* reporters, *DESE-PESE[C1+]-p126-lacZ,* and *DESE-PESE[C1+]-p381-lacZ,* recapitulate the endogenous *slp1* pattern (Figure 5-2, right). The controls from prior work, *DESE-p381-lacZ,* and *PESE[C1+-]p126-lacZ,* are shown on the left, and these two reporters have different patterns. *DESE-p381-lacZ* has 14 stripes with inappropriate expression in type I cells anterior to the odd stripes. *PESE[C1+-]p126-lacZ* has expanded stripes, because it is not repressed by Runt and Ftz in type III cells. When these enhancers are combined in *DESE-PESE[C1+]-p126-lacZ* and *DESE-PESE[C1+]-p381-lacZ* they lose the inappropriate expression of each reporter; there is no expression in type I cells and type III cells. Prior work shows that each enhancer is sensitive to different transcription factors, *DESE-p381-lacZ* is repressed

by Runt and Ftz, and *PESE[C1+]-p126-lacZ* is repressed by Eve (Prazak et al., 2010). This experiment shows that both of these transcription factors and enhancers are at work in the composite reporter, with either a large or small *slp1* basal promoter.

Discussion

3918 Δ C1 p381 has weak expression of the even stripes, suggesting C1 contributes to, but is not necessary, for activation of the even stripes PESE.

Unlike the experiment, *PESEΔC1 p126 lacZ* construct, the *PESEΔC1 p381 lacZ* described here still has expression of the even stripes. Other regions of PESE outside of the C1 region may contribute to the activation of the even-numbered stripes of *lacZ*, and the larger slp1 basal promoter enables activation of *PESE-lacZ* without the C1 region.

The C1+ region of PESE is sufficient to repress the inappropriate expression of DESE in cell type I in the reporter *DESE-C1+ lacZ*.

Prior work showed that C1+ is completely repressed by ectopic expression of Eve, but it is not repressed by Runt and Ftz (Prazak et al., 2010). Two experiments to further test the hypothesis of the dominance of repression would be to ectopically express Eve, and Runt and Ftz in flies containing DESE-C1+. Would repression by Eve mediated by C1+ prevent activation by DESE, which is not normally sensitive to repression by Eve? The experiment in Fig.5-2 shows that the addition of C1+ to the DESE reporter results in a loss of expression in type I cells, so perhaps Eve is enabled to repress these cells through C1+. It appears DESE only needs a little space from the promoter and the repression by Eve conferred by C1+ to generate the *slp1* pattern.

An intriguing hypothesis for future experiments is that Eve would repress expression of *DESE-C1+lacZ*, because C1+ is completely repressed by Eve and this repression by preventing release of promoter proximal paused Pol II would be dominant to DESE. Increasing the amount of Eve throughout the embryo might repress both the odd and even stripes as it does when Eve is ectopically expressed with C1+. Repressing *DESE-C1+ lacZ* by overexpressing Runt and Ftz, would test the same rule of dominance of repression for the other enhancer. Based on the experiment with *DESE-sog-p381-lacZ*, Runt and Ftz could repress *DESE-C1+ lacZ* at the later stage of the cellular blastoderm when the ventral furrow is formed, by preventing release of paused Pol II through DESE.

It is interesting to note that DESE normally does not function efficiently with the smaller p126 basal promoter. In this light it is unexpected that the *DESE-PESE[C1+]-lacZ* reporter produces what looks like a nearly wild-type *slp1* pattern. One interpretation of this finding is that the C1+ region provides an activity that enables DESE to interact more productively with the truncated p126 basal promoter region, perhaps by aiding in the recruitment and association of other proteins to the promoter region that enhance the affinity for DESE.

Figures

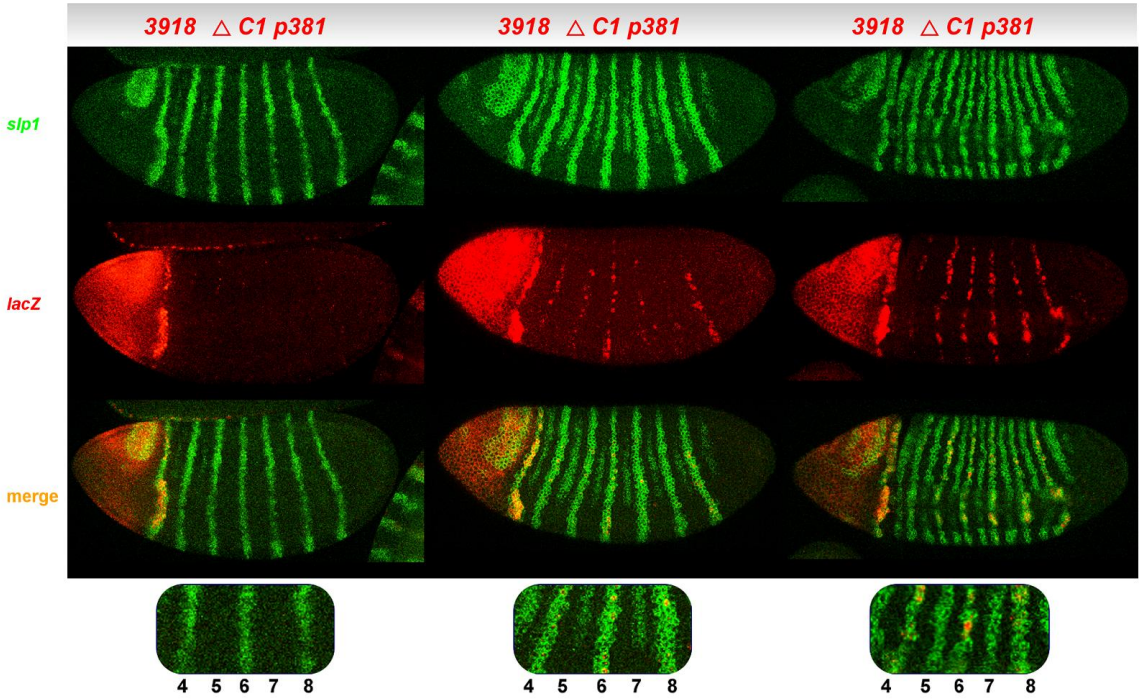

Figure 5-1. The PESE C1 region is important for early reporter gene activation in cells expressing the even-numbered slp1 stripes.

mRNA expression pattern from fluorescent in situ hybridization with probes for *slp1* (green) and *lacZ* (red) is shown for *3918ΔC1-p381-lacZ* reporter construct. Different embryos are arranged in order by stage from the onset of the odd-numbered slp1 stripes to the full length 14 stripes during the cellular blastoderm stage. The C1 region is a 155 bp sub-element of PESE identified by Prazak and co-workers that was absolutely required for expression with the smaller p126 *slp1* basal promoter fragment (Prazak et al., 2010).

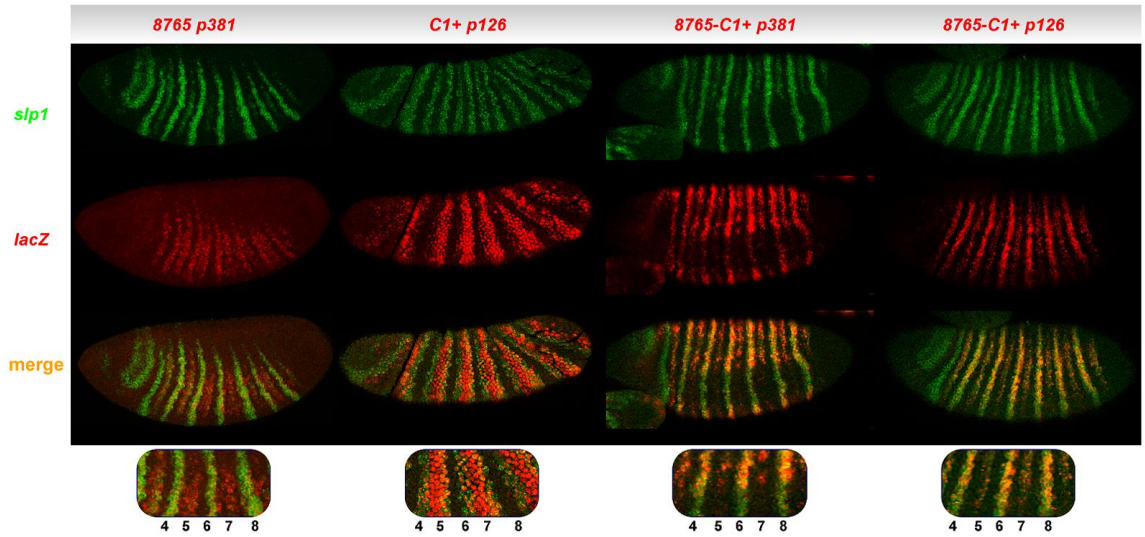

Figure 5-2. Non-additive interactions between DESE and PESE[C1+]

The wild type expression patterns of four *lacZ* reporter constructs are shown with *slp1* in green and *lacZ* in red. From left to right, A 8765-p381, B. C1+-p126, C. 8765-C1+ p381, and D. 8765-C1+ p126.

Chapter 6: Discussion and Future Work

Evidence for hypothesis of dominant repression

The location of the reporter gene in the genome, location of the enhancers and promoters relative to each other, and the promoter for the *lacZ* constructs are all factors that may affect the results of the in situ experiments shown in this thesis. To control for the position effects of location, four of the five constructs of chapter 2, *DESE-lacZ, PESE-lacZ, DESE-sog-lacZ,* and *PESE-sog-lacZ* were integrated on the third chromosome in the AttP2 site. Only *sog-lacZ* was integrated on the second chromosome in the AttP1 site, but the reporter gene is stably integrated into that site for all the reporter genes. Although the ϕC31 integrations are stable, it is possible the different location of *sog-p381 lacz* could change its expression. To rule out that variable, the sog-p381 reporter could be integrated on the third chromosome, or if this did not work as happened before, the other constructs could be integrated on the second chromosome.

Though *sog* is regulated on the dorsal-ventral (DV) axis, *slp1* is regulated by the anterior-posterior (A-P) axis hierarchy of genes, and the *slp1* promoter is important for creating the pattern of the segment-polarity gene. While some enhancers prefer promoters with a TATA box, other enhancers like the *rho* NEE, similar in pattern to the *sog* NEE, do not preferentially work with promoters with or without this sequence element, and upstream activators like Ftz prefer the TATA-containing promoters, and other activators like Dorsal do not prefer either (Ohtsuki, Levine, & Cai, 1998). Perhaps the ability of the *sog* enhancer with dorsal binding sites to work with either type of promoter allows for the expression of *sog* with the *slp1* promoter, and while the *slp1* promoter may introduce some A-P modulation, it matches the expression of other *sog-lacZ* constructs, which have

broad lateral stripes through the ventral ectoderm (Foo et al., 2015; Markstein, Markstein, Markstein, & Levine, 2002).

ChIP, in situ hybridization, and chromosome conformation capture assays are all future experiments that would reveal more about the molecular basis of the mechanism of repression mediated by each enhancer of the reporter genes. ChIP on the composite reporters and *sog-p381-lacZ* with antibodies against Pol II, NELF-E, TBP, and Cyclin-T, just as was done by Hang et. al to measure occupancy of these proteins at the promoter of each reporter gene and downstream in the structural body of the gene, would investigate and verify whether repression by preventing release of promoter-proximal paused Pol II was indeed responsible for the repression observed in the fluorescent in situs. As mentioned in Chapter 2, the short gastrulation gene is repressed by Snail in the mesoderm by preventing release of promoter proximal paused Pol II (Bothma, Magliocco, & Levine, 2011). UAS Snail lines generated by Emily Iocolano from a fly stock from the Mannervik lab have a distorted cuticle phenotype when crossed with NGT40, but do not alter the *sog-lacZ* pattern in several stages. Testing this line on the endogenous *sog* gene or other targets of snail could validate the lines and reveal whether the *sog* NEE is insensitive to Snail or whether this UAS Snail line does not work when crossed with NGT40. Chromatin conformation capture such as 3C-seq or 4C-seq may show a physical molecular basis for dominant repression. If fragments of DNA from the enhancer and promoter or other regions are in close proximity to one another, these interactions can be captured and identified by sequence.

The *slp1* basal promoter may affect expression of *lacZ* constructs and sensitivity to repression by Runt

Experiments investigating the role of Runt binding sites and the role of Ftz in regulating the PESE enhancer are shown in chapter 3. The runt binding site mutations had an effect on some embryos but not others, indicating the contribution of the Runt binding sites to repression through PESE has a variable phenotype, that sometimes the *PESEm1-6* results in a wider expansion of the stripes into type III cells where Runt and Ftz normally represses *slp1*. The composite reporters of *DESE-PESEm1-6* do not have expression in these cells, suggesting that DESE mediates repression by Runt and Ftz in the cell type III context in this reporter. This same effect is seen in the expression of the *DESE-C1+* reporter shown in Chapter 5; in addition to the non-additive interaction in cell type I, there is a non-additive interaction in cell type III for these composites with misbehaving PESE. As suggested by Lisa Prazak's work with C1+ (Prazak et al., 2010), this subelement may lack Runt binding sites that are necessary for repression. These observations suggest that the Runt binding sites do play a role in the regulation of PESE.

These experiments warrant further investigation on the role of the basal promoter in enhancer-promoter interactions. While the PESE containing Runt binding site constructs were made with the large basal promoter, the DNA from -260 to +121 bp of the TSS, the experiment with *ftz* mutants and overexpressing PESE with Runt, and Runt and Ftz, were done on *lacZ* reporter constructs with the small basal promoter, from -72 to +57bp from *slp1* TSS. Creating constructs of each type of promoter and comparing them

is a way to investigate how the size and sequence of the slp1 basal promoter affects the lacZ reporter gene's sensitivity to repression mediated by PESE or DESE.

Search for Opa binding sites

The experiments of Chapter 4 investigated the role of Opa and the binding sites for Opa. Western blots replicated the result obtained by Abclonal, that the antigen band ran at around 70 kDa. This was the size of the antigen expressing full-length Opa protein, not the size of the protein from expression of the truncated Opa protein, from amino acids 125 to 507, which was around 57 kDa. The antibody definitely detects Opa protein in a Western blot, at least region from 125 aa-507 aa of the protein, and the Opa protein expressed by the company (57kDa protein detected at 70kDa in western blot) was smaller than the protein fusion of maltose-binding protein with full length Opa that was around 100kDa. Furthermore the specificity of the antibody generated from a peptide was shown in a Western blot, in which antibody incubated with peptide was blocked from binding. (Chapter 4, Fig. 4-1)

Future experiments to investigate the role of Opa and the Opa binding sites are ChIP-seq which would identify in vivo binding sites genome wide using the antibody validated by Western blot. The Western blot experiments with bacterial extracts suggest the antibody is specific, and it would be possible to do more western blots on Drosophila extracts to see if the antibody was selective for the endogenous Oddpaired protein out of all the proteins of the fly and to optimize the immunostain. This would complete the validation of the antibody with Western and immunostain, and after ChIP qPCR with positive controls for the Opa binding sites, the DNA could be processed for sequencing. The ChIP qPCR with Opa antibody E990 suggests the Opa binding site in DESE

previously identified (Sen et al., 2010), may be bound by Opa in 3-4 hour embryos and could serve as a positive control or experiment to repeat. These experiments could be done on wildtype flies or in different genetic backgrounds, such as Opa mutant or overexpressing Opa. The expression of reporters with and without Opa binding sites could be compared after generation of transgenic flies.

ChIP experiments can investigate the binding at Opa binding sites, but genetic experiments and mutations to binding sites in a reporter construct are necessary to assess if these binding sites identified *in vivo* are functional, and whether occupancy changes when the level of Opa in the embryo is altered. The *cis*-regulatory architecture of Runt sites relative to Opa sites in the enhancers DESE and PESE can be predicted and tested by ChIP, and changes in occupancy in different genetic backgrounds could assay the functionality and occupancy at binding sites. My primers for the PESE reporter genes in the Runt binding site constructs can be used to investigate the binding of Runt in PESE and ChIP with Runt could also be sequenced.

Map of Runt, Opa, Eve, and Ftz binding sites for DESE and PESE and analysis

The predicted binding sites of Runt, Opa, Eve, and Ftz for DESE and PESE using the JASPAR database and a PWM from Katsua Shigesada for Runt are shown in (Fig 6-1, 6-2). JASPAR is a relevant database for the binding sites of these four *Drosophila melanogaster* transcription factors, because it is based on position weight matrices derived from bacterial-1 hybrid system (Meng, Brodsky, & Wolfe, 2005).

The JAPSAR website warns after conducting the scan of the enhancer sequence for transcription factor binding sites that the analysis of the sequence has high sensitivity, but the selectivity is abysmal. This means that while there are few if any false negatives

or true binding sites that are missed, there may be many false positives, or binding sites that are are not true Opa binding sites. Authors from the JASPAR update claimed that DNA-binding proteins often recognize a wide variety of related sequences, only a fraction of which are bound (Badis et al., 2009). My search for Opa in DESE identified many binding sites, including a binding site in DESE where I detected Opa binding or association with the enhancer by ChIP qPCR. While not all sites identified by this tool are necessarily in vivo functional sites, some of them may be. Runt binding sites identified by Katsua Shigesada's position weight matrix are shown for DESE and PESE, and they match sites predicted by JASPAR at 65% selectivity. There are many Eve and Ftz sites that overlap, because these are both homeodomain transcription factors that often recognize the same motif.

The C1 region was shown to be important for activation of PESE and drove expression of the *C1+-lacz* reporter (Prazak et al., 2010). This region is not necessary for expression of stripes of *PESE-lacZ* with the larger basal promoter, however there is another reason it may be important for activation. In the JASPAR sequence map of PESE (Fig.6-2), many Opa binding cites are concentrated in the C1 region. These binding sites are separate from the high scoring Runt sites found downstream in PESE. Only one Runt and Opa binding site overlap in the map of PESE. In DESE the distribution of binding sites is more mixed and less modular, underscoring the importance of ChIP seq and mutagenesis of binding site experiments for identifying these binding sites.

Figures

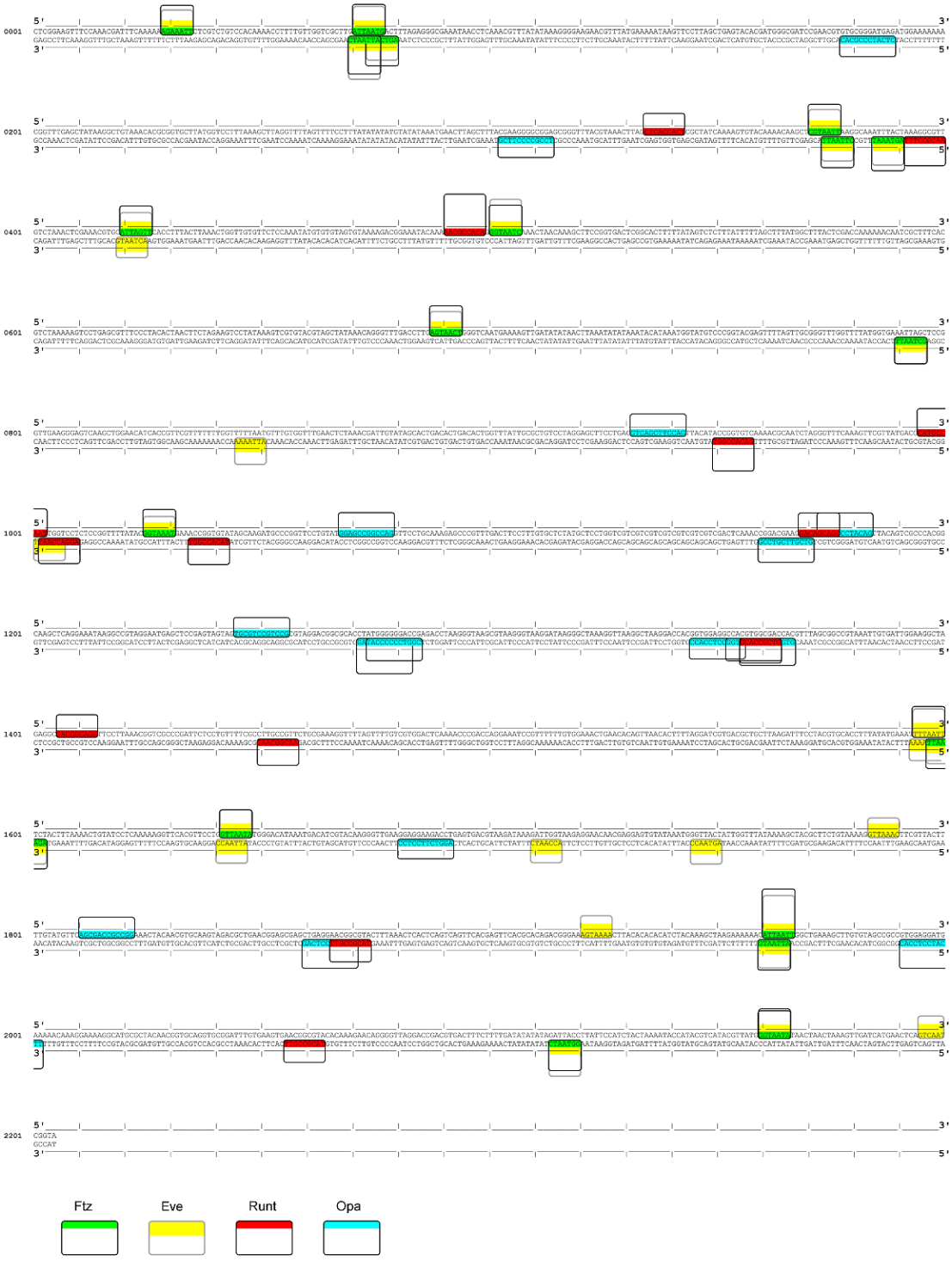

Figure 6-1. Putative transcription factor binding sites of Runt, Opa, Eve, and Ftz in the distal early stripe element (DESE) of slp1.

Binding sites shown are Ftz (green), Eve, (yellow), Runt (red), and Opa (blue) were identified with the most current version of JASPAR. A relative profile score

threshold of 80% for Eve and Ftz, and 70% for Opa was used. For Runt only the sites identified by Katsua Shigesada's PWM are shown and match binding sites predicted by JASPAR at a lower selectivity threshold. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2015). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016 44: D110-D115.

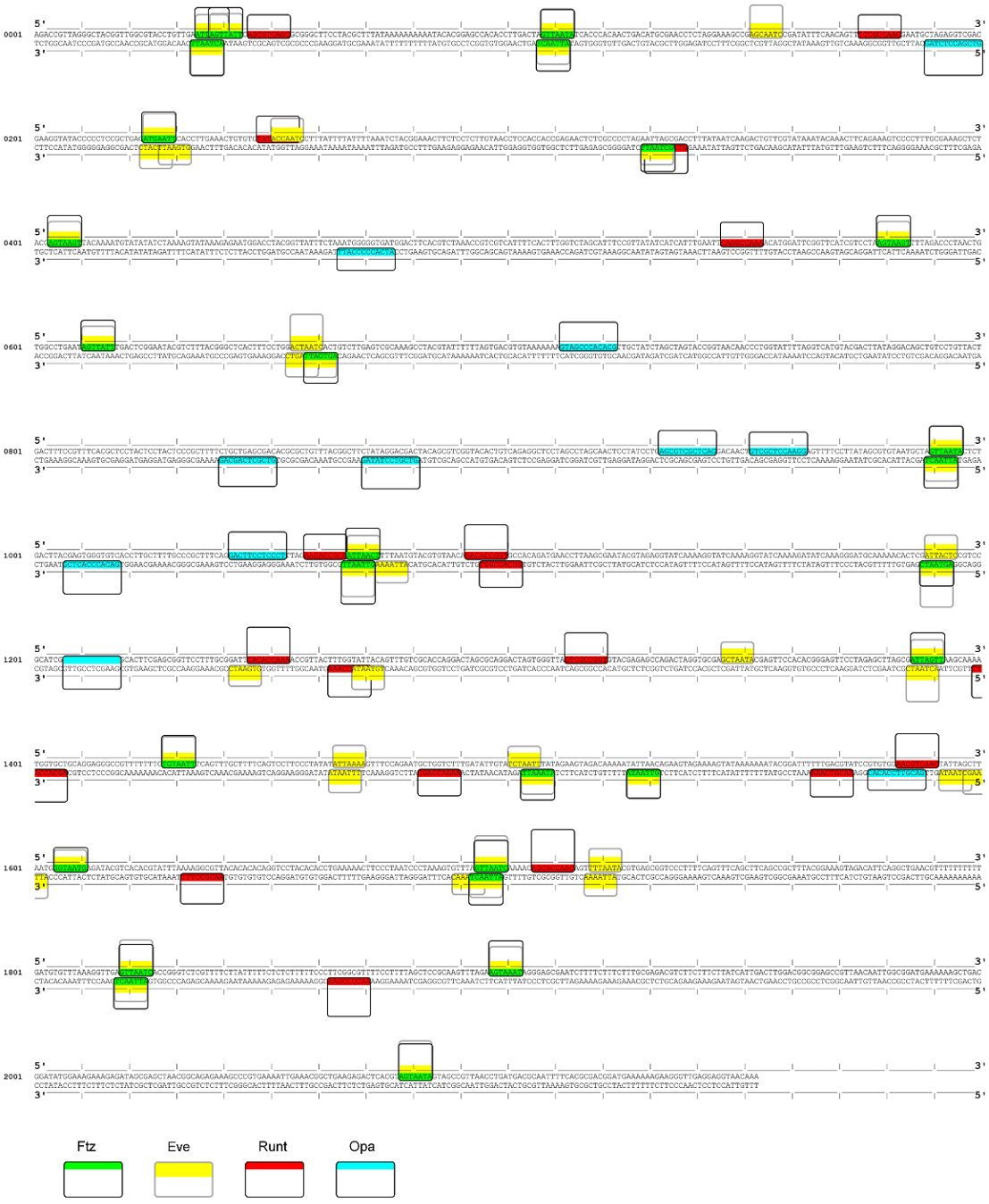

Figure 6-2. Runt, Opa, Eve, and Ftz putative binding sites in PESE.

Binding sites were identified with the most current version of JASPAR at relative profile score threshold of 80% for Eve and Ftz, and 70% for Opa. Setting the threshold at 80% results in no binding sites for Opa. Runt binding sites predicted from Katsua's PWM are shown and matched Runt binding sites predicted by JASPAR at a lower selectivity threshold. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi,

W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2015). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016 44: D110-D115.

Chapter 7: Methods and Materials

Cloning and transformants

LacZ reporter constructs *DESE-lacZ, DESE-sog-lacZ, PESE-lacZ,* and *PESE-sog-lacZ* reporter genes are *PhiC31* integrated into *attP2* site on third chromosome, while *sog-lacZ* is integrated into the attP1 site on second chromosome. The *sog* enhancer, containing the neurogenic ectoderm enhancer (NEE) from the sog locus, 13756 to 13287 bp upstream of the TSS (Crocker et. al 2008) was obtained by PCR amplification with addition of flanking SpeI sites and ligated into SpeI digested, pC:slp1[p381]lacZatt and pC:slp1[8765/p381]lacZatt and pC:slp1[3125/p381]-lacZatt constructs (Prazak et. al 2010, Hang and Gergen, personal communication).

Fluorescent in situ hybridization (FISH)

Two fluorescently labeled probes, *slp1* coupled to digoxigenin (DIG) and *lacZ* coupled with fluorescein isothiocyanate (FITC) and fluorescent antibodies were used to label the embryos green for *slp1* and red for *lacZ.* Embryos are first treated with Xylene, acetone, and hybridization buffers before incubating with probe overnight. Pre-absorbed primary, secondary and tertiary antibodies are added in sequence and incubated and washed over the next two days, and finally the slides are mounted with 1,4-diazabicyclo[2.2.2] octane (DABCO). (Janssens et al., 2006)

NGT40 stock construction

The first cross is the reporter gene flies with flies with the curly (Cyo), second chromosome balancer. This is so that the flies are balanced for NGT40, which is on the second chromosome. The second cross is with a TM3 stock that is homozygous for NGT40. This is to introduce NGT40 and to balance the reporter with TM3 (also known as stubble), a third chromosome balancer. Then virgins and males are collected containing both the balancers (curly and stubble) flies and crossed. Finally stubble plus curly plus virgin females and males are collected and crossed to make a homozygous NGT40 stock. (Tracey et al., 2000)

Ectopic expression of Eve, and Runt and Ftz

Crosses of male flies with a UAS transgene and female flies homozygous for the *lacZ* reporter and the NGT40 construct. The female flies were *NGT40 DESE p381*, *NGT40 DESE-sog p381*, *NGT40 sog p381*, *NGT40 PESE-sog p381*, *NGT40 PESE p381*, are each crossed with UAS Eve^{12}/cyo and UAS $Runt^{15}$ UAS Ftz^{263}.

Antibody generation

The template gene for the Opa antibody from Abclonal was derived from a vector (pMal-p2x-OpaFL, a gift from Deborah Hursh) and cloned into another vector, pGEX4T-1, to express the antigen. Truncated versions of the Opa gene for the antigen were expressed in Pet32a vectors. The Opa gene in a truncated form (125-507 codons) was successfully expressed in large quantities for the immunization. This antigen contains the 5 zinc fingers of the transcription factor DNA binding domain. The antigen was immunized in three rabbits 5 times. During this time the rabbits were bled several times,

serum was affinity purified with the recombinant protein antigen, and the company sent us the antibodies. One of the rabbits died early and that antibody E992 was from a third bleed from that rabbit. The three antibodies sent were from rabbits E990, E991, and E992.

The antibody from Genscript was generated from a peptide matching sequence near the fifth Opa zinc finger, amino acids 378-391. The sequence of the peptide is CVDEKSPSHGYDSEG, the cysteine is added, and the peptide is conjugated to keyhole limpet hemocyanin (KLH) a carrier protein to generate an immune response to make antibodies. Two rabbits were bled for the antibody.

Fixing embryos

Embryos were dechorionated in 50% bleach, fixed in 10% paraformaldehyde (PFA) and heptane, and washed devitylinized in methanol. (C. Tsai & Gergen, 1994)

Immunohistochemical in situ hybridization (IHC)

The lacZ-DIG probe was used to label the RNA of the embryos, and the embryos were permeabilized with protease K so the probe could enter the nucleus. They were post-fixed with 10% paraformaldehyde (PFA) and washed with phosphate-buffered saline and Tween 20 (PBT). Blocking was done for an hour with heat inactivated normal goat serum, incubated for 1.5 hours using pre-absorbed mouse anti-DIG antibody, and stained for the color reaction using a commercial solution of NBT_BCIP. The embryos were dehydrated and mounted on glass slides with Histomount. (C. Tsai & Gergen, 1994)

Western Blot

The final dilutions of each protein sample for the peptide specificity experiment shown were 1:10 for each sample. The dilutions of antibody were 1:8000 for the primary rabbit Opa antibody and 1:5000 for the secondary antibody, a goat-anti-rabbit antibody conjugated to horseradish peroxidase (HRP). The proteins were run on an 8% acrylamide gel, transferred to a polyvinylidene difluoride (PVDF) membrane, blocked with 2% bovine serum albumin (BSA), and incubated with primary antibody overnight. The next day they were incubated with secondary antibody for one hour, washed, and then developed using a chemiluminescent horseradish peroxidase (HRP) substrate. The concentration of the peptide used for the experiment was .3ug/ul.

Chromatin immunoprecipitation (ChIP)

The chromatin for the experiment is obtained by fixing a collection of 3-4 hour old embryos. I prepared chromatin by homogenizing embryos with a rotastator and pestle, adding Tris-ethylenediaminetetraacetic acid that is Tris-EDTA (TE) and proteinase inhibitor first and later 1% SDS after homogenization. Next I vortexed, put the lysates on ice, followed by sonication. Sonication was on a medium power setting (2.5) for four minutes pulse time (Hang and Gergen, personal communication). Glass beads are used either during the sonication with probe tip. The chromatin was pelleted, and the supernatant was transferred. A measurement of around 2000 ng/µl is taken at the start of each ChIP to set 200µg for each ChIP and 80µg of input. The input was set aside in the -80°C freezer, but the other supernatant samples are diluted with dilution buffer, 10% protease inhibitor cocktail tablet is added, and the mixture is incubated with antibody overnight at 4°C. The immunoprecipitation with the Protein A agarose beads is done for 2 hours followed by washes and elution. The protein-DNA crosslinks were reversed

overnight in a water bath at 65°C, and then protein digestion for two hours at 45°C, phenol chloroform extraction, and ethanol precipitation were carried out. Finally, qPCR was done using a standard curve of the input to determine concentration as percent input.

Mutations to Runt binding sites in a *PESE-lacZ* reporter

Two sequences of PESE from 3.1 to 2.5kb upstream of the *slp1* transcription start site were synthesized with one matching the wild type enhancer and one with mutations to six putative Runt binding sites. These sites were a good match with high ranking on a position weight matrix from Katsua Shigesada. Two basepairs were changed in each site; the two core guanosines or cytosines were changed to thymines or adenines respectively. The second and third binding sites overlap, so in this double site we mutated ggc to tta. (Fig. 3-1)

Generation of flies containing *lacZ* reporter gene with Runt binding site mutations in PESE

Two versions of this enhancer sequence were synthesized: one wildtype and one with Runt binding site mutations. (see Fig. 3-1) The synthesized PESE (3125) enhancer sequences were cut out of pUC57 plasmids from Genscript using SpeI, inserted into an SpeI site in the Casper vector directly upstream the *slp1* large basal promoter, p381, (-260 to +121bp from TSS) and is adjacent to the *lacZ* gene in the construct. The plasmids used for backbone vectors were pC:slp1[p381]lacZatt and pC:slp1[8765/p381]lacZatt. The composite reporters were generated by inserting each sequence into a Casper vector with the full length DESE reporter in the SpeI site in between DESE and p381. The single enhancer constructs are minimal PESE with the large basal promoter, which we will refer

to as *3125-p381-lacZ,* and the same sequence with Runt binding site mutations denoted as *3125-m1-6-p381lacZ.* Furthermore three composites with DESE were generated which are *8765-3125m1-6 p381 lacZ, 8765-2531 p381, and 8765-2531 m1-6.* These were all sent to Best Gene to be injected into fly embryos and PhiC31 integrated into the AttP2 site on the third chromosome.

Loss of function of *ftz* in *PESE-lacZ* reporters

PESE-lacZ reporters were crossed into recombinant *ftz* mutants over a TM3 balancer to make a balanced stock. Each reporter had a *kinked* (*ki*) mutation, a *ftz* mutation, and a *pink peach* mutation on the third chromosome with the reporter, and a TM3 balancer for the other third chromosome. Though the flies are heterozygous for the *ftz* mutation, backcrossing these flies to themselves resulted in a quarter of the progeny being homozygous *ftz* mutants. The balanced recombinant *ftz* mutants were of the genotypes slp1 [3918] *ki ftz pp*/TM3, *slp1* [3118] *ki ftz pp*/TM3, *slp1* [3125] *ki ftz pp*/TM3, *slp1* [PESE:C1+] *ki ftz pp/*TM3. All *lacZ* stocks had a small *slp1* basal promoter, p126, spanning from -72 to +57 bp.

Ectopic expression of Runt and Runt and Ftz

Virgin females of PESE-lacZ reporters homozygous for NGT40 were crossed with UAS Runt^{15} and UAS Runt^{15} UAS Ftz^{263} males. The stocks were NGT40; *slp1* [3918] *lacZ,* NGT40; *slp1* [3925] *lacZ,* NGT40; *slp1*[3125] *lacZ.* All *lacZ* stocks had a small basal promoter, p126, spanning from -72 to +57bp.

Ectopic expression of Runt and Opa in a *ftz* mutant background with *PESE 3918 lacZ* reporter

Virgin females of the compound stock NGT40;*slp1 [3918] ki ftz pp*/TM3 were crossed with *UAS Runt*^{15} *UAS Opa*^{14}*: ftz* ^{11} *e/TM3* males. The resulting phenotypes of the embryos were *slp1* [3918] *ki ftz pp*/ftz^{11}*e*, *slp1[3918] ki ftz pp*/*TM3*, *TM3/ ftz*^{11}*e*, and *TM3/TM3*. The females are homozygous for NGT40 and the males are homozygous for *UAS Runt*^{15} *UAS Opa*^{14} *ftz* ^{11} *e*. This reporter had the small basal promoter, p126, spanning from -72 to +57bp of *slp1* TSS.

Primers for Opa ChIP

All sequences given 5'– 3'

SKB_140 intergenic (negative control in *odorant receptor 42 B* gene)
TCAAGCCGAACCCTCTAAAAT

SKB_141 intergenic (neg. control *odorant receptor 42 B* gene)
AACGCCAACAAACAGAAAATG

SKB_144 actin5C
CGAAGAAGTTGCTGCTCTGGTTGTCG

SKB_145 actin5C
GGACGTCCCACAATCGATGGGAAG

DESEopaF
TGCCGTTCGAGTCCTTTATT

DESEopaR
CGGAGATCGGAAGGTTAGTG

C1+ F
TATGAGACTGAATGCTCACCCACA

C1+ R
CTGCCTCATTAGCTCACAAAAACG

slp1 promoter F
GGGCTCTCTTCGTGTAGACTTCGT

slp1 promoter R
GGAGAAGTTGCTCTTGAATTCCATT

Bibliography

Akamatsu, Y. (1997). Redox Regulation of the DNA Binding Activity in Transcription Factor PEBP2. The roles of two conserved cysteine residues. *Journal of Biological Chemistry*, *272*(23), 14497–14500. http://doi.org/10.1074/jbc.272.23.14497

Arnosti, D. N., & Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, *94*(5), 890–898. http://doi.org/10.1002/jcb.20352

Aronson, B. D., Fisher, A. L., Blechman, K., Caudy, M., & Gergen, J. P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. *Molecular and Cellular Biology*, *17*(9), 5581–7. http://doi.org/10.1128/MCB.17.9.5581

Asou, N. (2003). The role of a Runt domain transcription factor AML1/RUNX1 in leukemogenesis and its clinical implications. *Critical Reviews in Oncology/Hematology*, *45*(2), 129–150. http://doi.org/10.1016/S1040-8428(02)00003-3

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., … Bulyk, M. L. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science (New York, N.Y.)*, *324*(5935), 1720–3. http://doi.org/10.1126/science.1162327

Bénazéraf, B., & Pourquié, O. (2013). Formation and segmentation of the vertebrate body axis. *Annual Review of Cell and Developmental Biology*, *29*, 1–26. http://doi.org/10.1146/annurev-cellbio-101011-155703

Benedyk, M. J., Mullen, J. R., & DiNardo, S. (1994). *odd-paired*: a zinc finger pair-rule protein required for the timely activation of *engrailed* and *wingless* in Drosophila embryos. *Genes and Development*, *8*, 105–117. http://doi.org/10.1101/gad.8.1.105

Boettiger, A. N., & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science*, (22), 23–25.

Boija, A., & Mannervik, M. (2015). A time of change: Dynamics of chromatin and transcriptional regulation during nuclear programming in early *Drosophila* development. *Molecular Reproduction and Development*, *746*, n/a-n/a. http://doi.org/10.1002/mrd.22517

Bothma, J. P., Magliocco, J., & Levine, M. (2011). The snail repressor inhibits release, not elongation, of paused Pol II in the Drosophila embryo. *Current Biology*, *21*(18), 1571–1577. http://doi.org/10.1016/j.cub.2011.08.019

Bravo, J., Li, Z., Speck, N. A, & Warren, A J. (2001). The leukemia-associated AML1 (Runx1)--CBF beta complex functions as a DNA-induced molecular clamp. *Nature Structural Biology*, *8*(4), 371–378. http://doi.org/10.1038/86264

Bushey, A. M., Dorman, E. R., & Corces, V. G. (2008). Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Molecular Cell*, *32*(1), 1–9. http://doi.org/10.1016/j.molcel.2008.08.017

Chopra, V. S., Cande, J., Hong, J. W., & Levine, M. (2009). Stalled Hox promoters as chromosomal boundaries. *Genes and Development*, *23*(13), 1505–1509.

http://doi.org/10.1101/gad.1807309
Crocker, J., Tamori, Y., & Erives, A. (2008). Evolution acts on enhancer organization to fine-tune gradient threshold readouts. *PLoS Biology*, *6*(11), 2576–2587. http://doi.org/10.1371/journal.pbio.0060263
Day, D. S., Zhang, B., Stevens, S. M., Ferrari, F., Larschan, E. N., Park, P. J., & Pu, W. T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. *Genome Biology*, *17*(1), 120. http://doi.org/10.1186/s13059-016-0984-2
Driever, W., & Nüsslein-Volhard, C. (1988). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, *54*(1), 95–104. http://doi.org/10.1016/0092-8674(88)90183-3
Duffy, J. B., & Gergen, J. P. (1994). Sex, segments, and the central nervous system: common genetic mechanisms of cell fate determination. *Advances in Genetics*, *31*, 1–28.
Duffy, J. B., Kania, M. a, & Gergen, J. P. (1991). Expression and function of the Drosophila gene *runt* in early stages of neural development. *Development*, *113*(4), 1223.
Dunipace, L., Ozdemir, A., & Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila *snail* expression. *Development*, *138*(18), 4075–4084. http://doi.org/10.1242/dev.074377
Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., Brien, K. O., Kirov, N., … Rushlow, C. A. (2015). Zelda potentiates morphogen activity by increasing chromatin accessibility. *24*(12), 1341–1346. http://doi.org/10.1016/j.cub.2014.04.032.Zelda
Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T., & Jaynes, J. B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development (Cambridge, England)*, *126*(11), 2527–38.
Fujioka, M., & Jaynes, J. B. (2012). Regulation of a duplicated locus: Drosophila *sloppy-paired* is replete with functionally overlapping enhancers. *Developmental Biology*, *362*(2), 309–319. http://doi.org/10.1016/j.ydbio.2011.12.001
Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2003). Construction of transgenic Drosophila by using site-specific integrase from phage ΦC31.
Hillman, R. T., & Calos, M. P. (2012). Site-specific integration with bacteriophage ϕC31 integrase. *Cold Spring Harbor Protocols*, *7*(5), 609–614. http://doi.org/10.1101/pdb.prot069211
Hong, J.-W., Hendrix, D. A., & Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science*, *321*(5894), 1314. http://doi.org/10.1126/science.1160631
Houtmeyers, R., Souopgui, J., Tejpar, S., & Arkell, R. (2013). The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences*, *70*(20), 3791–3811. http://doi.org/10.1007/s00018-013-1285-5
Howard, K. R., & Struhl, G. (1990). Decoding positional information: regulation of the

pair-rule gene hairy. *Development*, *110*(4), 1223–31.
Ingham, P., & Gergen, P. (1988). Interactions between the pair-rule genes runt, hairy, even-skipped and fushi tarazu and the establishment of periodic pattern in the Drosophila embryo. *Development 104 Supplement*, *60*, 51–60.
Ingham, P. W. (2016). Drosophila segment polarity mutants and the rediscovery of the hedgehog pathway genes. *Current Topics in Developmental Biology*, *116*, 477–488. http://doi.org/10.1016/bs.ctdb.2016.01.007
Jaeger, J. (2011). The gap gene network. *Cellular and Molecular Life Sciences*, *68*(2), 243–274. http://doi.org/10.1007/s00018-010-0536-y
Jankowski, A., Obara, P., Mathur, U., Tiuryn, J., Panne, D., Ford, E., … Dalke, A. (2015). Enhanceosome transcription factors preferentially dimerize with high mobility group proteins. *BMC Systems Biology*, *10*(1), 14. http://doi.org/10.1186/s12918-016-0258-3
Janssens, H., Hou, S., Jaeger, J., Kim, A. R., Myasnikova, E., Sharp, D., & Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even-skipped* gene. *Nature Genetics*, *38*(10), 1159–1165. http://doi.org/10.1038/ng1886
Javed, a, Guo, B., Hiebert, S., Choi, J. Y., Green, J., Zhao, S. C., … Stein, G. S. (2000). Groucho/TLE/R-esp proteins associate with the nuclear matrix and repress RUNX (CBF(alpha)/AML/PEBP2(alpha)) dependent activation of tissue-specific gene transcription. *Journal of Cell Science*, *113 (Pt 1*, 2221–2231.
Kim, A. R., Martinez, C., Ionides, J., Ramos, A. F., Ludwig, M. Z., Ogawa, N., … Reinitz, J. (2013). Rearrangements of 2.5 kilobases of noncoding DNA from the Drosophila *even-skipped* locus define predictive rules of genomic cis-regulatory logic. *PLoS Genetics*, *9*(2). http://doi.org/10.1371/journal.pgen.1003243
Klingler, M., Soong, J., Butler, B., & Gergen, J. P. (1996). Disperse versus compact elements for the regulation of runt stripes in Drosophila. *Developmental Biology*, *177*(1), 73–84. http://doi.org/10.1006/dbio.1996.0146
Kramer, S. G., Jinks, T. M., Schedl, P., & Gergen, J. P. (1999). Direct activation of Sex-lethal transcription by the Drosophila runt protein. *Development (Cambridge, England)*, *126*(1), 191–200.
Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., … Levine, M. S. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell*, *153*(5), 976–987. http://doi.org/10.1016/j.cell.2013.04.045
Levine, M. (2008). A systems view of Drosophila segmentation. *Genome Biology*, *9*(2), 207. http://doi.org/10.1186/gb-2008-9-2-207
Levine, M., Rubin, G. M., & Tjian, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila. *Cell*, *38*(3), 667–73.
Li, G., & Zhu, P. (2015). Structure and organization of chromatin fiber in the nucleus. *FEBS Letters*, *589*(20), 2893–2904. http://doi.org/10.1016/j.febslet.2015.04.023
Li, L.-H., & Gergen, J. P. (1999). Differential interactions between Brother proteins and Runt domain proteins in the Drosophila embryo and eye. *Development*, *126*(15), 3313–3322.
Lis, J. (1998). Promoter-associated pausing in promoter architecture and

postinitiation transcriptional regulation. *Cold Spring Harbor Symposia on Quantitative Biology*, *63*, 347–356.
Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, *167*(5), 1170–1187. http://doi.org/10.1016/j.cell.2016.09.018
Lotem, J., Levanon, D., Negreanu, V., Bauer, O., Hantisteanu, S., Dicken, J., & Groner, Y. (2015). Runx3 at the interface of immunity, inflammation and cancer. *Biochimica et Biophysica Acta*, *1855*(2), 131–143. http://doi.org/10.1016/j.bbcan.2015.01.004
Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., … Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Molecular Cell*, *57*(4), 685–694. http://doi.org/10.1016/j.molcel.2015.01.007
Markstein, M., Markstein, P., Markstein, V., & Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc Natl Acad Sci U S A*, *99*(2), 763–768. http://doi.org/10.1073/pnas.012591199\r012591199 [pii]
McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., & Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature*, *308*(5958), 428–433.
Meng, X., Brodsky, M. H., & Wolfe, S. A. (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature Biotechnology*.
Ohtsuki, S., Levine, M., & Cai, H. N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. *Genes and Development*, *12*(4), 547–556. http://doi.org/10.1101/gad.12.4.547
Okumura, A. J., Peterson, L. F., Lo, M. C., & Zhang, D. E. (2007). Expression of AML/Runx and ETO/MTG family members during hematopoietic differentiation of embryonic stem cells. *Experimental Hematology*, *35*(6), 978–988. http://doi.org/10.1016/j.exphem.2007.03.002
Patel, M. C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., … Ozato, K. (2013). BRD4 Coordinates Recruitment of Pause Release Factor P-TEFb and the Pausing Complex NELF/DSIF To Regulate Transcription Elongation of Interferon-Stimulated Genes. *Molecular and Cellular Biology*, *33*(12), 2497–2507. http://doi.org/10.1128/MCB.01180-12
Pavletich, N. P., & Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, *261*(5129), 1701 LP-1707. JOUR.
Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., … Rubin, G. M. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(28), 9715–9720. http://doi.org/10.1073/pnas.0803697105
Prazak, L., Fujioka, M., & Gergen, J. P. (2010). Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors. *Developmental Biology*, *344*(2), 1048–1059. http://doi.org/10.1016/j.ydbio.2010.04.026

Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D., & Nusse, R. (1987). The Drosophila homology of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell*, *50*(4), 649–657. http://doi.org/10.1016/0092-8674(87)90038-9

Samee, M. A. H., & Sinha, S. (2014). Quantitative modeling of a gene's expression from its intergenic sequence. *PLoS Computational Biology*, *10*(3). http://doi.org/10.1371/journal.pcbi.1003467

Scott, M. P., & Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(13), 4115–9. http://doi.org/10.1073/pnas.81.13.4115

Sen, A., Stultz, B. G., Lee, H., & Hursh, D. A. (2010). Odd paired transcriptional activation of *decapentaplegic* in the Drosophila eye/antennal disc is cell autonomous but indirect. *Developmental Biology*, *343*(1–2), 167–177. http://doi.org/10.1016/j.ydbio.2010.04.003

Simmons, A. R., & Bergmann, D. C. (2016). Transcriptional control of cell fate in the stomatal lineage. *Current Opinion in Plant Biology*, *29*, 1–8. http://doi.org/10.1016/j.pbi.2015.09.008

Small, S., Blair, A., & Levine, M. (1992). Regulation of *even-skipped* stripe 2 in the Drosophila embryo. *The EMBO Journal*, *11*(11), 4047–57.

Sopko, R., & Perrimon, N. (2013). Receptor tyrosine kinases in Drosophila development. *Cold Spring Harbor Perspectives in Biology*, *5*(6), 1–31. http://doi.org/10.1101/cshperspect.a009050

Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, *13*(9), 613–626. http://doi.org/10.1038/nrg3207

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., & Soler, E. (2012). Transcription regulation by distal enhancers: who's in the loop? *Transcription*, *3*(4), 181–186. http://doi.org/10.4161/trns.20720

Stanojevic, D., Small, S., & Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science (New York, N.Y.)*, *254*(5036), 1385–1387.

Struhl, G., Struhl, K., & Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, *57*(7), 1259–1273. http://doi.org/10.1016/0092-8674(89)90062-7

Swantek, D., & Gergen, J. P. (2004). Ftz modulates Runt-dependent activation and repression of segment-polarity gene transcription. *Development*, *131*(10), 2281–2290. http://doi.org/10.1242/dev.01109

Thorpe, H. M., & Smith, M. C. M. (2000). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family.

Thorpe, H. M., Wilson, S. E., & Smith, M. C. M. (2000). Control of directionality of the site-specific recombination system of the Streptomyces phage ΦC31.

Tracey, W. D., Ning, X., Klingler, M., Kramer, S. G., & Gergen, J. P. (2000). Quantitative analysis of gene function in the Drosophila embryo. *Genetics*, *154*(1), 273–84.

Tsai, C. C., Kramer, S. G., & Gergen, J. P. (1998). Pair-rule gene *runt* restricts orthodenticle expression to the presumptive head of the Drosophila embryo. *Developmental Genetics*, *23*(1), 35–44. http://doi.org/10.1002/(SICI)1520-6408(1998)23:1<35::AID-DVG4>3.0.CO;2-7

Tsai, C., & Gergen, J. P. (1994). Gap gene properties of the pair-rule gene *runt* during Drosophila segmentation. *Development (Cambridge, England)*, *120*(6), 1671–1683.

Twigg, S. R. F., Forecki, J., Goos, J. A. C., Richardson, I. C. A., Hoogeboom, A. J. M., Van Den Ouweland, A. M. W., … Wilkie, A. O. M. (2015). Gain-of-function mutations in ZIC1 are associated with coronal craniosynostosis and learning disability. *American Journal of Human Genetics*, *97*(3), 378–388. http://doi.org/10.1016/j.ajhg.2015.07.007

Vander Zwan, C. J., Wheeler, J. C., Li, L. H., Tracey, W. D., & Gergen, J. P. (2003). A DNA-binding-independent pathway of repression by the Drosophila Runt protein. *Blood Cells, Molecules, and Diseases*, *30*(2), 207–222. http://doi.org/10.1016/S1079-9796(03)00026-3

Vermunt, M. W., & Creyghton, M. P. (2016). Transcriptional dynamics at brain enhancers: from functional specialization to neurodegeneration. *Current Neurology and Neuroscience Reports*, *16*(10), 94. http://doi.org/10.1007/s11910-016-0689-7

Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., … Nöllmann, M. (2014). Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. *PLoS Genetics*, *10*(8). http://doi.org/10.1371/journal.pgen.1004544

Walrad, P. B., Hang, S., & Gergen, J. P. (2011). Hairless is a cofactor for Runt-dependent transcriptional regulation. *Molecular Biology of the Cell*, *22*, 1364–1374. http://doi.org/10.1091/mbc.E10-06-0483

Walrad, P. B., Hang, S., Joseph, G. S., Salas, J., & Gergen, J. P. (2010). Distinct contributions of conserved modules to Runt transcription factor activity. *Molecular Biology of the Cell*, *21*(13), 2315–2326. http://doi.org/10.1091/mbc.E09-11-0953

Wang, X., Lee, C., Gilmour, D. S., & Gergen, J. P. (2007). Transcription elongation controls cell fate specification in the Drosophila embryo. *Genes & Development*, (631), 1031–1036. http://doi.org/10.1101/gad.1521207.GENES

Wieschaus, E., & Nüsslein-Volhard, C. (2016). The Heidelberg screen for pattern mutants of *Drosophila* : a personal account. *Annual Review of Cell and Developmental Biology*, *32*(1), annurev-cellbio-113015-023138. http://doi.org/10.1146/annurev-cellbio-113015-023138

Wysokinski, D., Pawlowska, E., & Blasiak, J. (2015). RUNX2: A master bone growth regulator that may be involved in the DNA damage response. *DNA Cell Biol*, *0*(0), 1–11. http://doi.org/10.1089/dna.2014.2688

Zallen, J. A., & Wieschaus, E. (2004). Patterned gene expression directs bipolar planar polarity in Drosophila. *Developmental Cell*, *6*(3), 343–355. http://doi.org/10.1016/S1534-5807(04)00060-7