Stony Brook University

Regulation of *wingless* transcription in the Drosophila blastoderm embryo

A Dissertation Presented

by

Kimberly Bell

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

December 2015

Stony Brook University

The Graduate School

Kimberly Bell

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

J. Peter Gergen – Dissertation Advisor
Professor, Department of Biochemistry & Cell Biology

Howard Sirotkin - Chairperson of Defense
Associate Professor, Department of Neurobiology & Behavior

Thomas Gingeras- Committee Member
Professor, Cold Spring Harbor Laboratory

Benjamin Martin- Committee Member
Assistant Professor, Department of Biochemistry & Cell Biology

Christine Rushlow- Committee Member
Professor, NYU Department of Biology

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Regulation of *wingless* transcription in the Drosophila blastoderm embryo

by

Kimberly Bell

Doctor of Philosophy

in

Genetics

Stony Brook University

2015

The Drosophila anterior/posterior segmentation pathway is a robust system for investigating transcription regulation in an *in vivo* setting. Runt, the founding member of the Runx family of transcription factors, is a master regulator of gene expression during development. Distinct combinations of Runt and other pair-rule factors act at two recently identified *cis*-regulatory modules (CRMs) DESE and PESE, to generate the initial 14-stripe *sloppy paired 1 (slp1)* pattern (Swantek & Gergen, 2004; Prazak et al, 2010). Each CRM alone drives a subset of the *slp1* pattern, but their combined action recapitulates early expression of *slp1* in a manner not expected from simple addition of their respective patterns. Runt is key in mediating these "non-additive" interactions, and its presence determines which enhancer regulates activity of the promoter in a particular cell context.

I investigated if Runt mediates non-additive interactions for the segment polarity gene, *wingless* (*wg)*. Experiments utilizing reporter constructs containing two putative pair-rule regulated CRMs for *wg* and double *in situ* hybridization confirm that the two CRMs are active at the blastoderm stage and moreover, respond to pair-rule regulation. Each CRM drives expression of all 14 *wg* stripes, however *wg2946* also drives ectopic expression posterior to the odd-numbered stripes, and the even-numbered stripes are weak. A composite *wg3911/2946* reporter recapitulates the endogenous *wg* pattern. Pair-rule loss and gain of function experiments reveal both similarities and differences to *slp1* regulation. This detailed functional analysis was complemented by efforts to investigate the physical basis of CRM/promoter interactions in different genetic backgrounds. 4C-seq and ChIP experiments can confirm if Runt mediates physical interactions and influences modENCODE identified enhancer signatures. Overall, these experiments provide a second example of the Runt-dependent regulation of enhancer promoter interaction and further suggest that the non-additive integration of inputs from different CRMs is a widespread aspect of regulating transcription in animal systems.

Table of Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank my wonderful husband Greg for being there for me through this journey. I am lucky to have him by my side.

Many heartfelt thanks to all of my friends and family who supported me along the way, especially my sister Ally for her humor and advice.

A warm thank you to my advisor, Peter Gergen, for always being patient, critical, and kind.

I greatly appreciate the help and support of my thesis committee: Howard Sirotkin, Benjamin Martin, Tom Gingeras and Chris Rushlow.

This all would not have been possible without the multiple levels of financial and academic support from Stony Brook University and the Genetics Program. Thank you to program directors Turhan Canli and Martha Furie and our amazing program coordinator, Kate Bell.

Thank you to all my past and present lab colleagues, particularly Kevin Chen, Lisa Prazak-Stockwell and Saiyu Hang whose experiments paved the way for mine, and to Michael Higgins and Yasuno Iwasaki who were always available for friendly discussion and to lend a helping hand.

Chapter 1

General Introduction

During development, the regulation of transcription or the process by which DNA is copied into mRNA for subsequent protein production, is essential to ensure that genes get turned on in the right cells at the right time. It is one of the most highly regulated components of gene expression (Tamkun, 2007). In fact, aberrant transcription underlies many disease states. In multicellular organisms, most cells contain the same sequence and amount of DNA and the development of complex structures and systems is ensured by properly and tightly regulated transcription. Successful transcription requires coordinating the activity of enhancers and promoters, segments of DNA responsible for gene activation (Berger, 2007). Promoters are typically located proximal to the start of the gene body (Perry et al., 2011), while enhancers can be located along the length of a gene locus (Frankel, 2012). A key component of enhancer activation, which allows their association with the gene promoter, is the binding or association of transcription factor proteins at enhancers (Paris et al., 2013), a type of *cis*-regulatory module (CRM). There are several families of transcription factors which recognize DNA with different binding domains. These include homeodomain, basic helix loop helix, leucine zipper, Runt domain, and zinc finger, among others (Klug, 1995; Kagoshima et al., 1993). Transcription factor binding site number and affinity affect the activity of CRMs (Tugrul et al., 2015).

Transcription in metazoans is comprised of three phases: initiation, where Pol II and general transcription factors are recruited to the promoter to form the initiation complex, elongation which extends the RNA transcript, continuing synthesis, and termination, where Pol II disengages from the promoter and the mRNA is released. Each phase is distinguished by epigenetic marks and recruitment of specific factors. Recent work indicates that this process is not only regulated at initiation, which was previously thought of as the critical step in transcription regulation. Pol II and other factors are often successfully recruited to the promoter

to initiate transcription, but the complex remains paused a short distance downstream of the transcription start site, such that elongation is blocked from proceeding (Fig. 1-1, Core and Lis, 2008; Koch et al., 2008; Wang et al., 2007). The experimental evidence for regulation of paused Pol II release includes effects on the phosphorylation status of the heptad repeats within the C-terminal domain (CTD) of the largest Pol II subunit. Phosphorylated serine 5 (Ser5P) marks initiation and Ser2P elongation; P-TEB phosphorylates serine 2 and is associated with elongation; negative elongation factor (NELF) is associated with pausing.

In order for transcription to be regulated in a context-dependent manner, different transcription factors bind *cis*-regulatory elements, genomic regions such as enhancers and insulators, and interact with gene promoters. The transcription factors bound to these elements influence how they interact with the promoters of their target genes to activate or repress transcription. A hallmark of enhancer elements is modular architecture: multiple binding sites for transcription factors, often multiple sites for the same factor, are clustered within a few hundred bases of DNA. Additionally, many enhancers exist for most developmentally relevant genes in order to fine-tune expression in a context dependent manner. Often, enhancers act over large distances and there is growing evidence for widespread instances of inter-chromosomal enhancer-promoter contacts (Dekker, 2008; Sanyal et al., 2013). Precisely how these characteristics of *cis*-regulatory elements relate to molecular details of enhancer-promoter interactions remains largely unknown.

There are currently two models of enhancer-promoter interaction, looping and tracking/linking (reviewed in Li et al., 2006). Looping of chromatin brings enhancers and any bound regulatory factors in physical contact with the promoter, potentially over long distances. Tracking involves propagation of signals from an enhancer to the promoter. Much evidence for

this model comes from experiments regarding regulatory sequences known as insulators, located in between enhancers and promoters that block communication from one to the other. Insulators contribute to the regulation of bithorax complex genes (Akbari et al., 2006). Studies at the *H19/IgF2* locus also provide evidence for a tracking model (Engel et al., 2008; Banerjee et al., 2001). However, there is also evidence that chromosome conformation changes underlie the influence of insulators on enhancer-promoter interactions at polycomb repressed domains (Sexton et al., 2012). A specialized form of tracking, called facilitated tracking, involves looping to achieve tracking (Fig. 1-2, reviewed in Li et al., 2006).

It is becoming recognized that long distance inter-chromosomal contacts, or loops, are common in transcription regulation (Dekker, 2008; Sanyal et al., 2013). Much work has been done regarding chromosome looping at the ß-globin locus, and other instances have been found in Drosophila (Splinter, 2006; Tolhuis et al., 2002) In fact, transvection in Drosophila was first observed by Lewis in 1954, in which an enhancer from one paired homologous chromosome contacts the promoter on the other chromosome (Duncan, 2002). A great deal of evidence for trans-action of enhancers, or inter-chromosomal contacts, has also come from studies on the bithorax complex (Drewell et al., 2002; Hopmann et al., 1995; Sexton et al., 2012). The development of chromosome conformation capture technology (Dekker et al., 2002) was key in demonstrating looping in the contexts mentioned above. More recently genome-wide studies such as ENCODE and modENCODE have shown that long-range looping contacts are quite common throughout both the human genome as well as Drosophila (Sanyal et al., 2013; Sexton et al., 2012). There is a gathering amount of evidence showing transcription factors mediate these long range and looping interactions (Deng et al., 2012; Nolis et al., 2009).

In addition to modular architecture, enhancers display predictable molecular signatures, similar to those found at promoters or regions of active chromatin, which are likely related to enhancer activity and may play a functional role in transcription regulation. For example, Pol II has been shown to be associated genome wide with enhancer elements (Koch et al., 2008). The methylation status of histones, such as H3K4Me3 and H3K4Me1 typically mark active enhancers, as does CBP binding and generation of enhancer RNA (eRNA), a novel class of RNA now associated with many enhancers (Natoli & Andrau, 2011). Much of this evidence was based on binding of initiation factors such as Pol II and activators such as CBP at enhancer regions. Sequencing data from the recent ENCODE and modENCODE projects show that RNA is directly generated from many enhancers genome wide (Sanyal et al., 2013). Most of this RNA is non-coding but PolyA+ transcripts are also derived from enhancers (www.modENCODE.org). Enhancer activity in general may be related to the output of eRNA, as has been observed for neuronal enhancers (Kim et al., 2010).

Transcription begins in Drosophila at 90 minutes post-fertilization, and coordinated activation of the embryonic genome is carried out by the transcription factor *zelda* (Nien et al., 2011). By 3 hours, the maternal to zygotic transition is complete, and the embryo is cellularized, all 6000 cells are provided positional information in the form of maternally expressed morphogen gradients. The Drosophila segmentation pathway lends itself to *in vivo* studies of early transcription regulation. We take advantage of this pathway and the fact that we can visualize gene expression in the three-hour blastoderm embryo with cellular resolution. Three classes of genes, gap, pair-rule and segment polarity, act in a hierarchical manner to decode the maternally provided information and establish the segmented body pattern in the blastoderm embryo. Broad expression of transcription factors encoded by gap genes set up the seven-striped

expression pattern of pair-rule genes (Fig. 1-3; Ingham, 1988). This system demonstrates the modular nature of enhancers. One classic example of this is the regulation of *even-skipped (eve)* stripe-two expression. Eve is expressed in a 7-stripe pattern of 4 cell-wide stripes. Gap gene products Bicoid and Hunchback activate *eve* and the anterior and posterior borders of stripe two are defined by the repressors Giant and Kruppel, respectively (Small et al., 1992). The enhancer described contains multiple binding sites for each transcription factor and is stripe specific. Stripe specific enhancers have also been described for pair-rule genes *hairy* and *runt* (Klingler et al., 1996; La Rosée et al., 1997).

In contrast to the gap to pair-rule transition, the pair-rule to segment polarity transition is not as clear-cut and most of the work done has been focused on later stages of segment polarity gene function in signaling pathways. Segment polarity genes in general do not have stripe specific elements, but the promoter interprets different combinatorial codes of enhancers and a set of overlapping transcription factors. For the 14 stripe patterns of segment polarity genes, there is one code for the even-numbered stripes and one for the odd-numbered stripes. The same set of pair-rule factors regulates each set, but the distinct combination determines which enhancer is active (Prazak et al., 2010). One particular hurdle to studying the pair-rule to segment polarity transition has been the size of the *cis*-regulatory regions for the segment polarity genes. For example, *engrailed (en)* has a *cis*-regulatory region of 70 kilobases (kb) of DNA and *wingless (wg)*, 30kb. Now, tools are available that make it easier to study these long distance interactions.

Much of the work in our lab has been built from and focused on the function of the pair-rule gene *runt,* the founding member of the Runx family of transcription factors. Runt functions as both an activator and repressor of transcription depending on the cellular context. Runt is not

only a key player in developmental transcription regulation in Drosophila, being involved in pattern formation and sex determination, but Runx family members are found in nearly all animals and play roles in blood, bone, neural and stomach development as well as cancer pathology (Gergen and Butler, 1988; Miyazono et al., 2004; Wang et al., 2010). The fact that that Eve, Hairy and Runt can influence the expression of other pair-rule genes makes it difficult to determine direct vs. indirect pair rule effects (Ingham, 1988). Investigating regulation of segment polarity genes, particularly *slp1*, by Runt has led us to some key findings regarding enhancer-promoter interactions. Indeed, other pair-rule factors are involved in regulating *slp1*, which highlights another challenge to studying segment polarity regulation: pair-rule interactions.

However, the *slp1* gene provides several advantages for studying the effects of Runt on enhancer-promoter interactions in the pair-rule to segment polarity transition. One is the relatively small size of its *cis*-regulatory region, which is about 10kb. Another is the relatively simple combination of pair-rule factors that work together to drive *slp1* expression. The initial 14-stripe pattern of *slp1* in the cellular blastoderm stage embryo is generated in response to combinatorial regulation by Runt and three other pair-rule transcription factors, the homeodomain proteins Eve and Ftz (the latter encoded by the pair-rule gene *fushi tarazu*) and the Zn-finger transcription factor Opa, encoded by *odd-paired* (Swantek and Gergen, 2004). The 14-stripe pattern consists of seven repetitive units. Each unit includes four types of cellular contexts for *slp1* transcription: type I cells correspond to the two cells located in the anterior half of odd numbered parasegments that do not express *slp1*; type II cells correspond to the two cells in the posterior half of odd numbered parasegments that comprise the odd-numbered slp1 stripes; type III cells are the two cells in the anterior half of even numbered parasegments that do not express

slp1; and type IV cells are the two cells in the posterior half of even numbered parasegments that comprise the even numbered slp1 stripes (Fig. 1-4A). The factors responsible for *slp1* regulation are different in each of these four different transcription contexts. Eve represses *slp1* transcription in type I cells, whereas repression in type III cells requires both Runt and Ftz. Activation of slp1 in type II cells requires Runt in combination with Opa. Activation in type IV cells also requires Opa, with a minor contribution from an as of yet unidentified Factor X (Swantek and Gergen, 2004).

The fact that all nuclei in the late blastoderm stage embryo are capable of regulating *slp1* in response to manipulations in the activity of these four pair-rule transcription factors opened the door for using biochemical approaches to investigate the mechanism of *slp1* regulation, leading to the discovery that repression by Runt and Ftz does not involve blocking recruitment of Pol II and transcription initiation at the *slp1* promoter, but instead is due to the regulation of transcription elongation (Wang et al., 2007). Recent work has identified two distinct *cis*-regulatory elements that are together responsible for mediating *slp1* regulation in response to Runt and the other pair-rule transcription factors (Prazak et al., 2010). The PESE (Proximal Early Stripe Element) enhancer, located between 3.1 and 2.5 kb upstream of the transcription start site drives expression in type IV cells, corresponding to the even-numbered *slp1* stripes. The DESE (Distal Early Stripe Element) enhancer, located between 8.1 and 7.2 kb upstream of the transcription start site drives expression in cells corresponding to both the odd- and even-numbered *slp1* stripes (type II and type IV cells) with a stronger than normal level of expression for the odd stripes (Fig. 1-4 A&B). Importantly, DESE also drives inappropriate expression in type I cells where expression is normally repressed by Eve. The inappropriate expression of the *DESE-lacZ* reporter gene construct in type I cells is due to the insensitivity of DESE to

repression by Eve (Prazak et al., 2010). A striking finding from this work was that a composite reporter gene construct containing both the DESE and PESE CRMs generates a pattern beyond what is expected from the combination of the two independent patterns; that is the inappropriate expression in type I cells that is observed when DESE is tested as an autonomous element is suppressed in the presence of PESE. Indeed the composite DESE+PESE reporter gene faithfully recapitulates the initial metameric expression of endogenous *slp1* in wild-type embryos as well as in response to numerous different manipulations in the activity of the Runt, Eve and Ftz transcription factors. This type of non-additive enhancer action has also been reported for the Drosophila *snail* gene (Dunipace et al., 2011).

We have proposed a model to account for this non-additive effect for *slp1* that involves Runt-dependent regulation of a competition between the two CRMs for interactions with the promoter (Prazak et al., 2010). In this model Runt plays a dual role in preventing functional interaction between PESE and the promoter while also contributing to DESE-dependent expression in cells that express both Runt and Opa and that do not express Ftz. The above findings have recently been elaborated upon by chromatin immuno-precipitation experiments (summarized in Fig. 1-4; Hang et al., submitted). In Runt and Ftz-expressing type III cells, DESE and PESE-dependent expression are repressed through different mechanisms. As found for endogenous *slp1*, Runt and Ftz repress DESE by blocking the release of the paused Pol II complex and preventing transcription elongation from progressing. In contrast, the Runt-dependent repression of PESE involves blocking the recruitment of Pol II and the initiation of transcription. Experiments exploring the mechanism of Eve-dependent repression in type I cells find that Eve blocks PESE-dependent expression by regulating the release of paused Pol II. These results have led to the proposal that when repressors interact with an enhancer to block

transcription elongation that this occurs in a manner that involves a sustained interaction between this enhancer and the promoter region that thereby prevents other enhancers from interacting with and potentially activating transcription from the same promoter.

Although many details of these enhancer-promoter interactions have been uncovered, the exact mechanism is still not known. In Fig 1-4 for example, the DNA surrounding PESE, DESE and the *slp1* promoter is drawn as looped, but these conformations have not been confirmed molecularly for the *slp1* locus or for other segment polarity genes. I describe preliminary experiments using chromosome conformation capture to investigate enhancer-promoter interactions for both the *slp1* and *wingless* loci.

The segment polarity gene and homolog of *wnt-1, wingless (wg),* is expressed in 14 one-cell wide stripes in the blastoderm embryo, in the posterior most cell of each parasegment. Some details are known about the pair-rule regulation of this gene, however all developmentally relevant enhancers have not been identified, nor the enhancer promoter interactions underlying transcription regulation. Moreover, much of the work has focused on the role of *wg* after its initial pattern is established, as it plays a key role in cell signaling and parasegment identity, being expressed at the presumptive boundaries. Wingless expression overlaps with one cell of *slp1* expression and is expressed at the same time, so transcription of each may be regulated by similar enhancer-promoter interactions.

Regulation of *wingless* is quite dynamic within the first few hours of development, with its initial pattern established upon cellularization. Prior studies on *wg* have shown that Ftz, Eve and Prd all participate in generating the initial endogenous expression pattern. Ftz and Eve have been shown to repress endogenous *wg* and Paired is presumed to be an activator (Ingham & Hidalgo, 1993; Ingham, 1988). Runt has also been shown to play a dual role in regulating *wg*,

acting as either and activator or repressor depending on context (Swantek and Gergen, 2004). A 4.5 kb region upstream of the transcription start site of *wg* was found to drive expression in the embryo but again was studied at a later time point and not assessed for its involvement in pair-rule regulation, but instead for its role in maintaining parasegment identity (Lessing & Nusse, 1998). In addition, the 4.5kb segment is a fairly large region and may encompass multiple enhancers.

Later in development, ending at 6 hours, *slp1* is required for maintenance of *wg* expression (Fig. 1-5). Additionally, *hedgehog (hh)* is required to maintain *wg,* which in turn is required for *engrailed (en)* maintenance (Cadigan et al., 1994). *wingless* also plays a role in mesoderm patterning, wing and eye development (Lee &Frasch, 2000; Pereira et al., 2006). Multiple enhancers have been predicted and several identified for these later stages (MacArthur et al, 2009; Pereira et al., 2006).

Unlike what has been observed for *slp1, prd* and *odd* are implicted in *wg* regulation (Mullen and DiNardo, 1995). The expression of *prd* and *odd* is also dynamic, as their patterns change from 7 to 14 stripes between 3 and 4 hours (Fig. 1-5). It is unclear what combinatorial codes are acting to regulate *wg* and what CRMs are active at the blastoderm stage. Therefore, in this study I sought to verify predicted CRMs for *wg* active at the blastoderm stage, and investigate if non-additive enhancer promoter interactions also underlie regulation of *wg* transcription. This type of regulation potentially governs transcription of the entire class of segment polarity genes. The experiment described provide further insight into the developmental relevance of having of multiple CRMs, and how they can individually interact with the promoter.

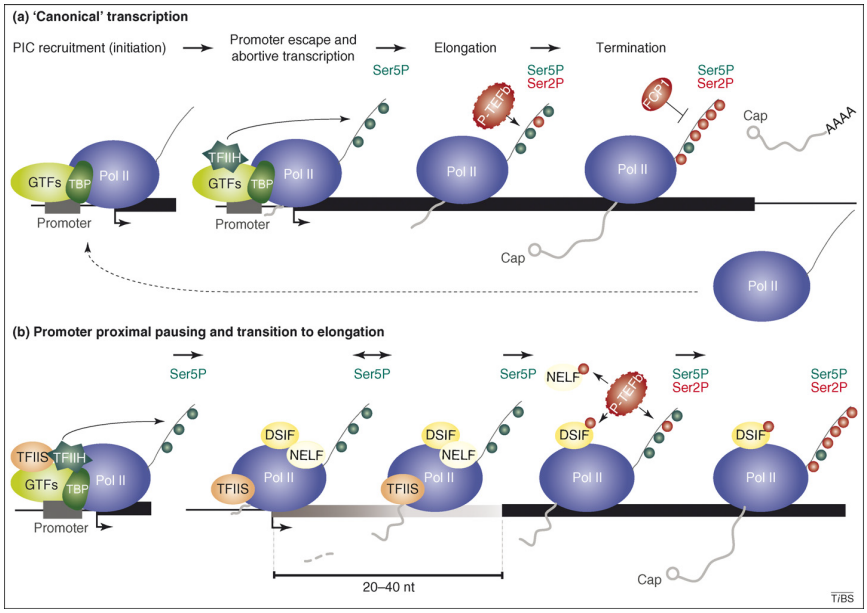

Figure 1-1. Promoter-proximal pausing

In (a) above, canonical transcription is depicted as a constant transition from initiation to elongation to termination. In (b) promoter proximal pausing is shown, where the initiation complex is blocked from elongation by associated factors NELF and DSIF, remaining paused 20-40nt downstream of the promoter (Koch et al., 2008).

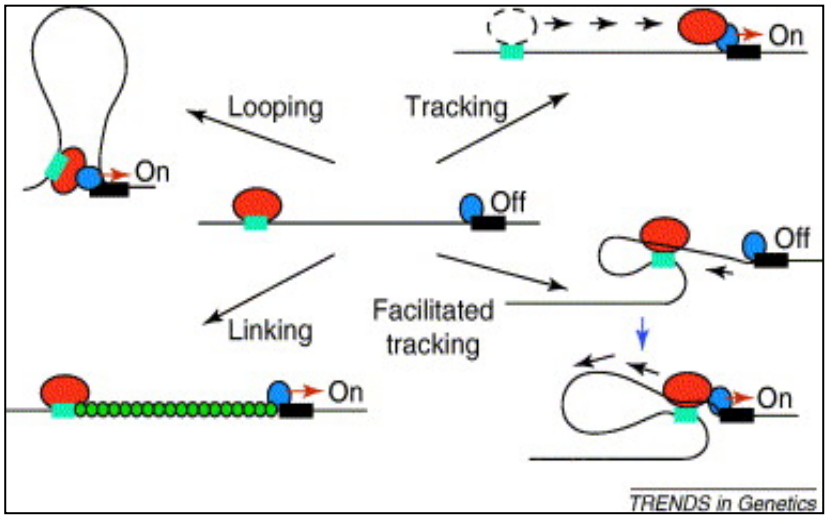

Figure 1-2. Tracking vs. Looping

Transcription regulation by looping involves a change in chromosome conformation to bring regulatory elements in close physical proximity to the promoter. Tracking primarily involves the propagation of a signal from a regulatory element to the promoter and variations include changes in chromosome conformation. (Li et al., 2006)

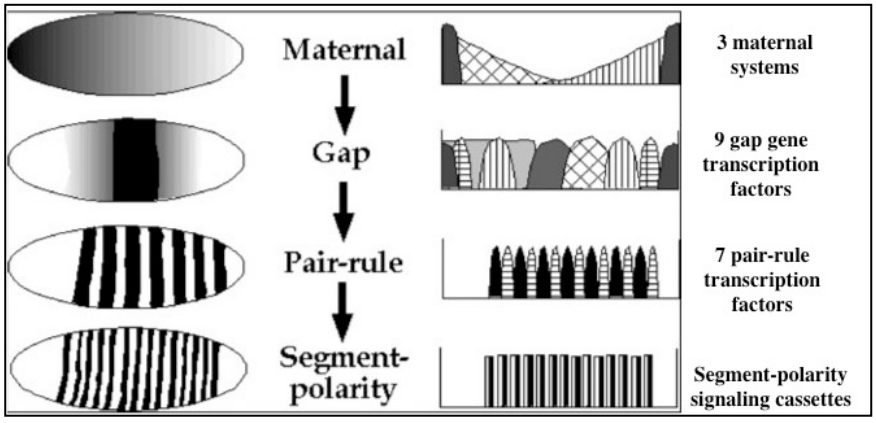

Figure 1-3. Drosophila segmentation

Segmentation along the anterior-posterior axis of the blastoderm embryo is carried out through a hierarchical expression of genes. Maternal factors set up domains of expression for gap genes, which encode gap transcription factors that then act to establish the pair-rule pattern. The pair-rule transcription factors then act in a combinatorial manner to set up the pattern of expression for the segment polarity genes, many details of which remain unknown.

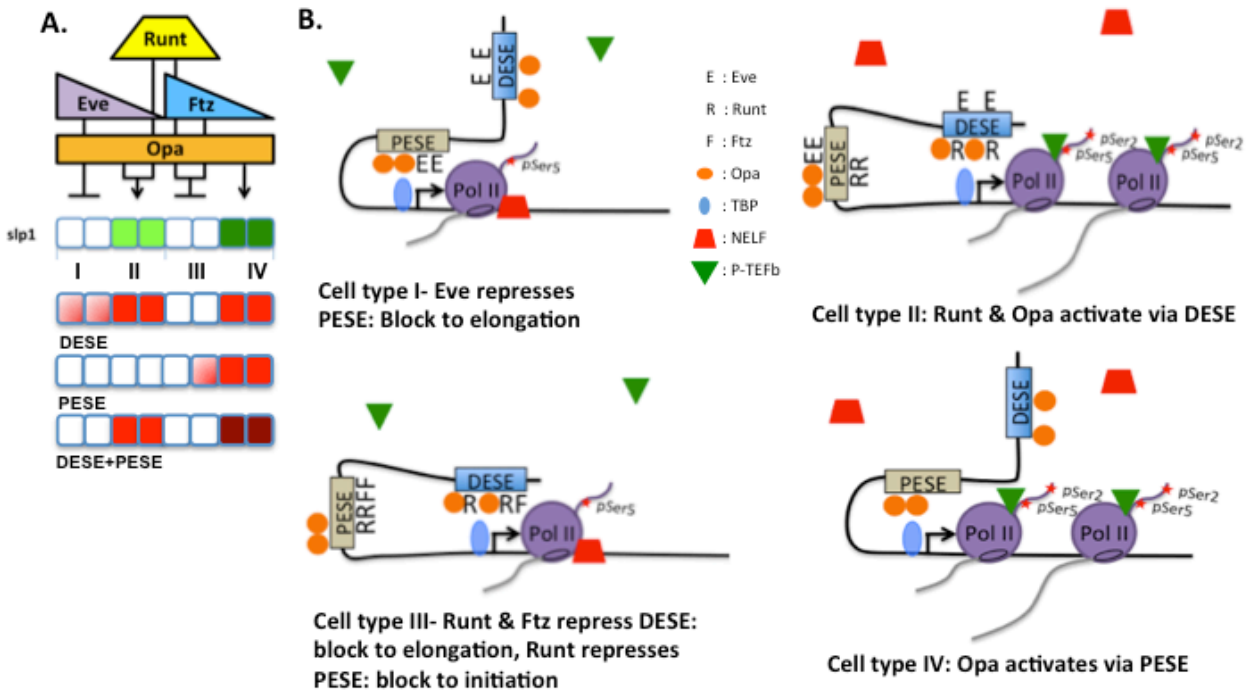

Figure 1-4. *slp1* enhancer activity summary

A. Schematic diagram representing one of the seven repetitive units of the fourteen number stripe pattern of slp1, indicating the inputs from Opa, Runt, Eve and Ftz responsible for regulating *slp1* in each of the different four transcription contexts. Expression of slp1 is indicated by green shading of cells. The patterns produced by different *lacZ* (red) reporter gene constructs are indicated below. As indicated, expression of a reporter containing both the DESE and PESE enhancers is not explained by a simple addition of the patterns produced by the two independent enhancers. **B.** The four diagrams depict the proposed chromatin conformations and transcription status in each of the four different cellular contexts. Pol II (purple) with the pSer5 modification is paused and associated with NELF downstream of the promoter in cells where DESE is mediating repression by Runt and Ftz, as well as in cells where PESE is mediating repression by Eve.

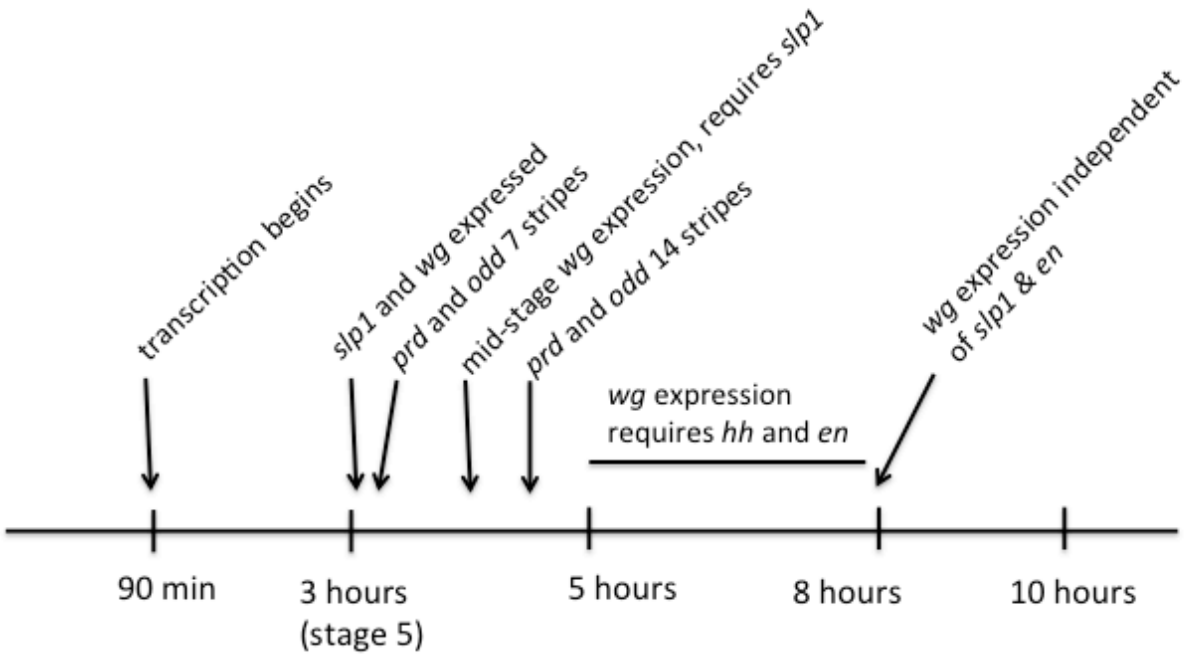

Figure 1-5. Timing of *wg* transcription in Drosophila

Timeline of embryonic development from 90 minutes post-fertilization to 10 hours is represented. Each arrow indicates a transcription event, time points indicated are 90 minutes, 3 hours, between 3-5 hours, and 5-8 hours.

Chapter 2

Regulation of *wingless* transcription in the Drosophila blastoderm embryo

Abstract

Distinct combinations of Runt and other pair-rule transcription factors act at two recently identified *cis*-regulatory modules (CRMs), DESE and PESE, to generate the initial 14-stripe pattern of *sloppy paired 1* (*slp1)* (Swantek & Gergen, 2004; Prazak et al, 2010). DESE and PESE each drive a subset of the *slp1* pattern. Interestingly, their combined action recapitulates early expression of *slp1* in a manner not expected from simple addition of their respective patterns. Runt is key in mediating these non-additive interactions, and its presence or absence determines which CRM controls activity at the promoter in a particular cell context (Prazak et al, 2010). We sought to investigate if Runt also mediates non-additive interactions for the Runt-sensitive segment polarity gene, *wingless (wg)*. We confirmed two putative CRMs active at the blastoderm stage for *wg*: *wg3911*, located from -3.9 to -1.1kb upstream of the *wg* transcription start site (tss) and *wg2946,* located from 2.9 to 4.6kb downstream of the tss, in intron two of the *wg* gene. The upstream CRM drives all 14 *wg* stripes*,* as does *wg2946*, however this CRM drives ectopic expression posterior to the odd-numbered stripes and generates weak even-numbered stripes. A composite reporter recapitulates the endogenous *wg* pattern, providing evidence that non-additive interactions underlie *wg* regulation. Pair-rule loss and gain of function experiments show that Runt and an additional transcription factor, paired (Prd), are key regulators of *wg* expression. Overall, these experiments provide a second example of the Runt-dependent regulation of enhancer promoter interaction and further suggest that the non-additive integration of inputs from different CRMs is a widespread aspect of regulating transcription in animal systems.

Introduction

The segment polarity gene *wingless (wg),* like *slp1,* is also sensitive to Runt regulation (Swantek and Gergen, 2004). It is expressed in 14 one-cell wide stripes in the blastoderm embryo, in the posterior most cell of each parasegment. All developmentally relevant enhancers have not been identified for *wg*, as well as the full combination of pair-rule factors regulating its expression. Moreover, much of the work has focused on the role of *wg* after its initial pattern is established, as it plays a key role in cell signaling and parasegment identity, being expressed at the presumptive boundaries.

The pair-rule factors Ftz, Eve and Prd all participate in generating the endogenous *wg* expression pattern. Ftz and Eve have been shown to repress endogenous *wg* and Paired is presumed to be an activator (Ingham & Hidalgo, 1993; Ingham, 1988). A 4.5 kb region upstream of the tss of *wg* was found to drive expression in the embryo but was studied at a later time point and not assessed for pair-rule regulation, but for their roles in maintaining parasegment identity (Lessing & Nusse, 1998). In addition, the 4.5kb segment is a fairly large region and may encompass multiple enhancers. Runt has been shown to play a dual role in regulating *wg*, acting as either and activator or repressor depending on context (Swantek and Gergen, 2004). Perhaps Runt is also mediating non-additive promoter-enhancer interactions for *wg*. This type of regulation potentially governs transcription of the entire class of segment polarity genes.

Two putative CRMs for *wg* were identified using the whole-genome ChIP-on-chip data set from the Berkeley Drosophila Transcription Network project (MacArthur et al., 2009), based on overlapping peaks of pair-rule transcription factor association. The impetus for this approach comes from the observation that the *slp1* PESE and DESE CRMs, which were initially identified by scanning the *slp1* locus for early stripe elements using a series of reporter gene constructs

(Prazak et al, 2010), both show clear overlapping peaks of pair-rule transcription factor association in this data set. These putative pair-rule responsive CRMs were tested for their activity in *lacZ* reporter constructs, each made into stable lines after being cloned and integrated into the same location on chromosome III using ΦC mediated recombination. One putative CRM, *wg3911*, is located from -3.9kb to -1.1kb upstream the *wg* tss and the second, *wg2946*, is located in intron 2 of *wg,* from 2.9 to 4.6kb downstream of the tss. In addition, a composite *wg3911/2946-lacZ* was generated. Double fluorescent *in situ* hybridization experiments show that *wg3911-lacZ* responds to both gain and loss of function of pair rule genes and moreover, behaves differently from endogenous *wg* in response to manipulations of Prd and Odd activity. Similar experiments *wg2946-lacZ* show that this CRM produces an expression pattern that is different from endogenous *wg* in wild-type embryos. Additional genetic manipulation experiments with these two reporters as well as the composite *wg3911/2946-lacZ* reporter show that the composite reporter containing both CRMs more faithfully recapitulates the response of endogenous *wg* than is observed for either single element when they are tested autonomously.

Methods

Construction of *wg* reporters

The putative *wg3911* CRM was identified from ChIP-on-chip data from the Berkeley Drosophila Transcription Network Project . There are two peaks of overlapping pair-rule transcription factor binding (MacArthur et al., 2009). One putative CRM is located 3.9 to 1.1kb upstream of the *wg* tss and the other is in intron 2 of the gene. A *lacZ r*eporter construct made from the upstream CRM candidate region (Fig. 2-2). The putative CRM was cloned into a pCaSpeR backbone upstream of the *wg* basal promoter and the lacZ reporter gene, integrated into the $y^{1}w^{67c23}$;P{CaryP}attP2 at location 68A4 (3L) by ΦC31 mediated recombination, in forward and reverse orientations (3911 and 1139). We focus here on 3911. Transformants were recovered and made into homozygous stocks, then combined with the nanos-GAL4-tubulin driver (NGT40) for ectopic expression experiments, the final stocks being *yw;NGT40;wg3911* and *yw;NGT40;wg1139*. Reporter constructs *wg2946-lacZ* and *wg3911/2946-lacZ* were produced by in fusion cloning to integrate into pCaSPeR and Best Gene carried out transformation. Male recombinants recovered and bred to make homozygous stocks and also combined with NGT40 for gain of function experiments.

Drosophila mutants and genetics

Gain of function lines:

Ectopic expression of pair-rule transcription factors was achieved using the NGT (nanos-GAL4-tubulin) expression system. The second chromosome linked P{GAL4 nos.NGT}40(NGT40) driver and the P{UAS-runt.T}15, transgenes have been described

previously (Li and Gergen, 1999; Swantek and Gergen, 2004; Tracey et al., 2000). The P*{UAS-eve}12* stock were given to us by John Reinitz (Zallen and Wieschaus, 2004). Embryos were collected from crosses between females homozygous for *NGT40* and for the different third chromosome-linked reporter genes and males homozygous for different UAS transgenes. *yw;NGT40;wg3911*, *yw;NGT40;wg2946* and *yw;NGT40wg3911/2946* virgins were crossed to UAS-male lines for ectopic expression. UAS lines used were as follows: *UAS-prd1*, *UAS-odd1A2/TM3*, *UAS-runt15*, *UAS-eve12*, and *UAS-ftz261/CyO,*

Loss of function lines:

The pair-rule gene mutations used were *eve1* (also known as *eve[ID19]*), *run29* (*run[YP17]*), *prdx3*, and *ftz11 (ftz[W20]* , and *slp1Δ34b*. Reporter gene expression in runt mutants was examined in embryos from a cross between *cv v run29/y w* females and males homozygous for the reporter gene of interest. To generate embryos homozygous for *eve1* and containing at least one copy of a reporter, females homozygous for different third chromosome-linked reporters were crossed to *eve1*/CyO males, creating flies doubly heterozygous for *eve1* and the reporter gene. Female and male progeny were then backcrossed generating embryos where 3/16 are homozygous for *eve1* and contain at least one copy of the reporter. Reporter gene expression in *ftz* mutants was determined by generating recombinant chromosomes containing the *ftz11* mutation and different third chromosome linked reporter genes balanced over *TM3*.To generate *odd* mutants, 2 overlapping deletion lines (resulting deletion spans the entire *odd* gene) were used: *Df(2L)BSC292* and *Df(2L)exel7018*. Recombinant stocks were made from reporters and *ftz11pp/TM3* for all 3 reporters to conduct *ftz* loss of function experiments.

Whole-mount in situ hybridization and imaging

Embryos were collected as described (Tsai and Gergen, 1994). Embryos from experiments with temperature sensitive alleles were collected for 1.5 hours at 25°C, allowed to develop at 18°C for four hours then shifted to a non-permissive temperature of 30°C for 20 minutes immediately prior to fixation. In situ hybridization was carried out as described in (Swantek and Gergen, 2004) using the digoxigenin-labeled riboprobe for *lacZ* as described in (Tsai and Gergen, 1994). Images are taken using a camera-mounted DIC microscope at 20x magnification.

Fluorescent in situ hybridization was carried out as described (Janssens et al., 2005). The fluorescein labeled *lacZ* riboprobe was synthesized with fluorescein-12-UTP (Roche). After hybridization, *lacZ* mRNA was visualized by sequential incubation with Rabbit Anti- fluorescein (1ug/ml final) and Alexa Fluor 647 Donkey Anti-rabbit (1ug/ml) antibodies (Molecular Probes). Digoxigenin labeled wg probe was detected using Mouse Anti- Digoxigenin antibody (Roche, 1.25ug/ml final) followed by Alexa Fluor 555 Goat Anti- mouse (1ug/ml) and Alexa Fluor 555 Donkey Anti-goat (1ug/ml) antibodies (Molecular Probes). Blocking was done in 2x Western Blocking Reagent (Roche) diluted in PBT. All antibodies were pre-absorbed at a 10x concentration in PBT with 1/10 volume of 0 to 12 hour embryos. Embryos were washed in PBS:glycerol (1:1) prior to mounting in 2.5% Dabco (Sigma), 50mM Tris (pH 8.0) and 90% glycerol. Images were obtained on a Leica TCS SP2 Spectral Confocal Microscope system and were scanned 8x to obtain averaged images with reduced noise

Results

Two putative *wingless* enhancers are bona fide pair-rule CRMs

The observation that non-additive interactions between *slp1* pair-rule response elements (CRMs) PESE and DESE establish the complete *slp1* pattern, led me to investigate if this was a general mechanism for segment polarity gene regulation by pair-rule transcription factors. The Berkeley Drosophila Transcription Network Project (MacArthur et al, 2009) pair-rule factor binding data matched well with the identified CRM regions for *slp1*. Using this data, two putative CRMs for *wg* in the vicinity of the promoter were identified (Fig. 2-2A). One is located from -3.9 to -1.1kb upstream of the tss, and the second is in intron 2 of the *wg* gene, from 2.9 to 4.6kb downstream of the tss. Reporter constructs were made from each putative CRM alone, *wg3911-lacZ* for the upstream region and *wg2946-lacZ* for the downstream region, and combined in a composite reporter in the endogenous order, but with both CRMs upstream of the tss, *wg3911/2946-lacZ*. (Fig. 2-2B).

A 4.5kb fragment upstream of the transcription start site of *wg* was previously cloned and tested as an *in vivo* CRM. Its activity was studied after the start of gastrulation, and was not assessed for pair-rule regulation, but for its role in maintaining parasegment identity. We investigated a more focused region with the *wg3911-lacZ* reporter. In the blastoderm embryo, the *wg3911-lacZ* reporter pattern is comprised of fourteen 1-cell wide stripes, and largely recapitulates the initial metameric endogenous *wg* expression (Fig. 2-3B).

The overlapping pair-rule binding presented in the ChIP-on-chip data (Fig. 2-2A), indicates a second candidate region for a *wg* CRM, located in intron 2 of the gene. We investigated if a reporter made from this region also can also drive expression, and if it responds

to pair-rule manipulation. The *wg2946-lacZ* reporter drives ectopic expression posterior to *wg* odd numbered stripes into the first cell that express both Runt and Ftz (Fig. 2-3E). The expression of even-numbered stripes is weaker relative to odd-numbered stripes.

Endogenous *wg* and *wg* reporter activity with and without *slp1* loss of function

Knowing that *wg* is expressed in the posterior cell of each 2-cell wide stripe of *slp1* expression, and that *slp1* is directly regulated by pair-rule transcription factors, I sought to determine whether *slp1* plays a role in regulating *wg* and the reporter gene expression at the blastoderm stage. Therefore I conducted *slp1* loss of function experiments and examined the expression of *wg, wg3911* and *wg2946*. Double fluorescent *in situ* hybridization in a *slp1* loss of function background shows that *wg* expression remains unaffected (Fig. 2-3A). All 14 stripes are present and are 1-cell wide as expected. The *wg3911-lacZ* reporter pattern recapitulates endogenous *wg* expression (Fig. 2-3C) and is likewise unaffected by *slp1* lof (Fig. 2-3D). The *wg2946-lacZ* reporter pattern is variable by stage, with odd stripes coming on by early stage 5 and even stripes in late stage 5 (data not shown). In the early stage 6 embryo (Fig. 2-3F) the even stripes of *wg2946-lacZ* are weakly expressed, but the initial basic metameric pattern is still apparent.

Double fluorescent *in situ* hybridization experiments show that each putative CRM can in fact drive at least a portion of the endogenous *wg* pattern, thus suggesting both as *in vivo* CRMs. The wild-type expression patterns for each reporter alone are different from each other, and for *wg2946-lacZ,* different from the endogenous *wg* pattern; this CRM "misbehaves". A composite *wg3911/2946-lacZ* reporter recapitulates endogenous *wg* expression, thus further suggesting that non-additive interactions underlie the regulation of *wg* as well as *slp1*.

The *wg3911-lacZ* reporter responds to pair-rule loss of function

Because *wg3911*-lacZ can drive full expression of the *wg* pattern, I wanted to investigate if this reporter also responds to pair-rule gene manipulation. I conducted pair-rule loss of function experiments on blastoderm stage embryos to determine how the different *cis*-elements respond to mutations in the pair-rule transcription factors. With *ftz* loss of function, the endogenous *wg* pattern and *wg3911-lacZ* pattern match (Fig. 2-4B). The odd stripes of *wg* and *wg3911-lacZ* are unaffected. There is de-repression in cells that normally express *ftz*, resulting in a 5-cell wide expression pattern. Loss of function of *eve* results in matching patterns of endogenous *wg* and the *wg3911-lacZ* reporter; the even and odd stripes of *wg* are unaffected, but de-repression is seen in other cells of the odd parasegments, where *eve* is normally expressed (Fig. 2-4C). Loss of function of *runt* results in matching patterns of 4-cell wide stripes expanding posterior to the odd numbered stripes, and overlapping the endogenous even numbered *wg* stripes for both endogenous *wg* and *wg3911-lacZ (*Fig. 2-4D). Eve can still repress *wg* and *wg3911-lacZ* in the odd parasegment cells that do not express *runt*. De-repression is seen throughout the odd parasegment, as *ftz* can no longer repress *wg* and *wg3911-lacZ* without *runt*. Eve and Ftz have important roles in repressing *wg* and *wg3911* in anterior portions of the odd and even parasegments, respectively. This supports previous findings that Runt acts as a repressor or activator depending on cell context. The even numbered stripes in the posterior most cell are unaffected as *prd* is sufficient to activate *wg* in this cell type.

Loss of function of *prd* results in a misbehaving pattern of *wg3911-lacZ*; the pattern does not match that of endogenous *wg* (Fig. 2-4E). The *wg3911-lacZ* reporter shows a strong reduction in odd stripe expression, and the even stripes are present, with some stripe specific reduction in expression. However, the odd stripes of endogenous *wg* are present and cross the

mesoderm, and the even stripes are repressed. The latter result agrees with prior observations that Prd activates *wg*. With *odd* loss of function, all stripes of endogenous *wg* are slightly expanded. (Fig. 2-4F). The *wg3911-lacZ* reporter shows repression of the odd stripes, but they are still present. The even stripes are strongly expressed and are 2-cells wide.

The *wg2946-lacZ* reporter responds to pair-rule loss of function

One of the most striking observations when looking at the overall *wg2946-lacZ* expression patterns is expression in the mesoderm. There is consistent mesodermal expression in wild-type embryos as well as in responses to different pair-rule gene manipulation. Changes in expression are also seen in the ectoderm. The *wg2946-lacZ* reporter misbehaves with *ftz* loss of function. This is most evident in the merged image (Fig. 2-5B). When *wg* and *lacZ* align, cells with expression are yellow-orange. Differences observed for the *wg2946* reporter are evident as cells that express either *wg* or *lacZ* but not both. The even numbered stripes of the reporter are repressed and the odd stripe expands posteriorly. (Fig. 2-5B).

Manipulation of *eve* results in a misbehaving pattern for *wg2946-lacZ* with loss of function (Fig. 2-5C). The *wg2946-lacZ* reporter has loss of expression with *eve* loss of function throughout the segmented region, but expression is strong in the presumptive mesoderm. With *runt* loss of function, the expression pattern of *wg2946-lacZ* does not match endogenous *wg* (Fig. 2-5D). The even stripes of *wg* and *wg2946-lacZ* are unaffected, as *prd* is sufficient to activate these stripes. The *wg2946-lacZ* reporter generates only a subset of stripes with *prd* loss of function- the odd numbered stripes. A ventral view shows that the reporter is expressed throughout the presumptive mesoderm as well. The odd numbered *wg2946-lacZ* reporter stripes

are repressed The *wg2946-lacZ* reporter drives expression in 2-cell wide odd-numbered stripes and even stripes are weak (Fig. 2-5F).

The composite *wg3911/2946-lacZ* reporter largely recapitulates *wg* expression

A composite *wg3911/2946-lacZ* reporter was constructed to investigate the possibility that non-additive interactions underlie transcription regulation of *wg,* as they do for *slp1*. The two CRMs each drive at least a subset of the *wg* pattern and respond to pair-rule loss of function. We show that indeed, there are non-additive interactions between the CRMs. With *ftz* loss of function, the composite *wg3911/2946-lacZ* reporter matches the endogenous *wg* pattern. There are 5-cell wide stripes throughout the segmented region of the embryo (Fig. 2-6B). The expression patterns match between the composite *wg3911/2946-lacZ* reporter and endogenous *wg* with *eve* loss of function*;* de-repression is seen in the anterior 2 cells of the odd parasegments, while the odd and even stripes are unaffected (Fig. 2-6C). The composite reporter, *wg3911/2946-lacZ* is affected similarly to the *wg3911-lacZ reporter* with *runt* loss of function, and the resulting expression pattern matches that of endogenous *wg* (Fig. 2-6D). With *prd* loss of function, a composite *wg3911/2946-lacZ* reporter odd numbered stripes are present, and the even stripes are reduced. This matches the expression pattern of endogenous *wg* (Fig. 2-6E). The *wg3911/2946-lacZ* reporter recapitulates endogenous *wg* expression with *odd* loss of function (Fig. 2-6F). All pair-rule loss of function experiments show that when the two CRMs are together, the resulting expression patterns match, suggesting that there are non-additive interactions underlying regulation *wg* expression.

The *wg3911-lacZ* reporter responds to pair-rule gain of function

We also investigated how pair-rule gain of function affects each reporter; the loss of function experiments provide information about requirements for pair-rule regulation at each CRM, while gain of function experiments assess if the factors are sufficient to produce changes in expression. With *ftz* gain of function, the even stripe of *wg3911-lacZ* is unaffected. Alone, *ftz* is not sufficient to repress the even stripes. However, the odd stripes of both *wg* and the *wg3911-lacZ* reporter are repressed. Again, the reporter pattern matches endogenous *wg* (Fig. 2-7B). With *eve* gain of function, the odd numbered stripes of both the *wg3911-lacZ* reporter and endogenous *wg* are repressed, but the even stripes are largely unaffected as *prd* is sufficient to activate this stripe (Fig. 2-7C). Gain of function for *runt* results in 4-cell wide stripes expanding anterior to and overlapping the odd numbered stripes for both *wg* and the *wg3911-lacZ* reporter. There is also a stripe specific effect, with little expression seen in the 6th wide stripe for both *wg* and *wg3911-lacZ*. (Fig. 2-7D). Endogenous *wg* is not affected by *prd* gain of function, as all 14 stripes are presents, and are the correct width and strength. For the *wg3911-lacZ* reporter, the odd stripes are repressed, whereas the even stripes are stronger than usual, being 2-cells wide in some locations (Fig. 2-7E). With *odd* gain of function, all endogenous *wg* stripes are present, but are weaker than normal. The *wg3911-lacZ* reporter shows that all stripes are present but weak, with odd stripes weaker than the even stripes (Fig. 2-7F).

The *wg2946-lacZ* reporter responds to pair-rule gain of function

The *wg2946-lacZ* reporter misbehaves with *ftz* gain of function; all 14 stripes are present, but are weak (Fig. 2-8B), the notable area of mismatching expression is in the mesoderm, where the odd stripe is more strongly driven by the *wg3911-lacZ* reporter than in *wg.* Matching patterns

are seen for the *wg2946-lacZ* reporter and endogenous *wg* with *eve* gain of function. The odd numbered stripes of both the *wg2946-lacZ* reporter and endogenous *wg* are repressed, and the even stripes are largely unaffected. The reporter expression of the even stripes is stronger than the endogenous gene (Fig. 2-8C). Gain of function of *runt* affects *wg2946-lacZ* in the same was as it does *wg3911-lacZ*, and the resulting expression pattern matches endogenous *wg* (Fig 2-8D). With *prd* gain of function the composite *wg2946-lacZ* reporter generates only even stripes as well, but they are strong. The stripes are 2-cells wide, and expression expands into the cell anterior to the even stripe. Odd numbered *wg2946-lacZ* reporter stripes are repressed (Fig. 2-8E). With *odd* gain of function the *wg2946-lacZ* reporter shows overall repression, which does not match the pattern of endogenous *wg* (Fig. 2-8F).

The composite *wg3911/2946-lacZ* reporter responds to pair-rule gain of function

With *ftz* gain of function, the composite *wg3911/2946-lacZ* reporter once again matches the endogenous *wg* pattern. There are 5 cell wide stripes with Ftz lof, and repressed odd stripes with Ftz gof (Fig. 2-9B). With *eve* gof, the odd stripes of *wg* and *wg3911/2946-lacZ* are repressed and the even stripes remain expressed, as Prd is sufficient to activate this stripe set (Fig. 2-9C). The composite reporter, *wg3911/2946-lacZ* is also affected similarly to *wg3911-lacZ* with Runt gain of function, and the resulting expression pattern matches endogenous *wg* (Fig. 9D). With Prd gof, however, the composite reporter pattern misbehaves. All 14 stripes are present, but the odd numbered ones are weak and the even numbered stripes show strong expression. The even numbered stripes are 2-cells wide and expand into cell 7 (Fig. 2-9E). The *wg3911/2946-lacZ* reporter recapitulates endogenous *wg* expression with *odd* gain of function (Fig. 2-9F).

In general, we confirm that *eve* and *ftz* repress *wg*. As seen for *slp1*, *runt* acts as both an activator and a repressor of *wg.* Different from *slp1, prd*, & *odd*, are important for *wg* regulation. *Prd* is an activator of the even number stripes and *odd* represses *wg* in the even parasegment. There are some key differences in *wg3911* and *wg2946* driven expression as compared to endogenous *wg* expression. For most pair-rule manipulations, *wg3911* recapitulates *wg* expression, except in response to *prd* and *odd.* For *wg2946,* ectopic expression is observed posterior to the odd-numbered stripe, and the reporter misbehaves with each pair-rule loss of function.

Discussion

In the experiments presented, I have confirmed that two predicted CRMs for the segment polarity gene *wingless* can drive striped expression patterns *in vivo*. Moreover, results suggest that these two CRMs act in a non-additive manner to drive the full expression pattern of endogenous *wg.* A composite reporter more faithfully recapitulates endogenous *wg* expression than the individual CRMs. Each reporter working on its own cannot faithfully recapitulate *wg* expression when pair-rule gene expression is manipulated. The *wg3911* reporter drives full 14 stripe *wg* expression, and *wg2946* drives ectopic expression in the cell posterior to the odd stripe. Another notable feature of *wg2946* is that is drives expression in the mesoderm as, and this is not regulated by pair-rule genes. When subjected to pair-rule loss or gain of function, the resulting expression patterns driven by the composite reporter are more faithful than the patterns driven by either single CRM. This observation, taken together with our previous findings that similar interactions underlie the pair-rule regulation of *slp1,* suggests that this may be a common method of enhancer interaction for the segment polarity genes.

Six cell contexts for *wg* expression are suggested by the observed results. Each cell from anterior posterior in the 8 cell schematic representation of one odd and even parasegment will be referred to as 1-8 (Fig. 2-10). The odd parasegments are comprised of the first two cell contexts. The first cell context is made up of cells 1, 2 and 3 and is repression by *eve* (Fig. 2-10A). This repression is thought to be mediated by the *wg3911* CRM, with *wg3911* interacting with the promoter, preventing access of *wg2946*. This is a similar mechanism to what is observed for *eve* repression of PESE. Loss of expression of *wg2946-lacZ* is observed with *eve* loss of function (Fig. 2-5C), suggesting that *eve* is required for *wg2946* expression. Further experiments are

suggested to fully elucidate the role of *eve* in *wg2946* mediated expression. The second context is cell 4, or the even-numbered stripe of *wg* expression. This is thought to be mediated by *wg3911,* which interacts with the promoter to drive expression when Runt is present (Fig. 2-10B). In addition to Runt, Opa likely plays a role in activating *wg* here as it is expressed uniformly throughout the pre-segmental region of the embryo. Additionally, *opa* is required for activation of *slp1* (Swantek & Gergen, 2004) and *wg* is delayed in *opa* mutants (Benedyk et al., 1994). The observation that expression of *wg3911* is lost in cell 4 in the absence of *runt,* also suggests it is an activator for the odd-numbered stripe of *wg* expression.

The even parasegment is comprised of 4 additional cell contexts. In cell 5, Runt+Ftz is suggested to mediate repression by *wg3911* (Fig. 2-10C), analogous to Runt+Ftz repression of *slp1.* This explains why the ectopic expression driven by *wg2946* in cell 5 is not seen in endogenous *wg.* There may be a block to the release of paused Pol II, due to repression mediated by *wg3911,* and since this interaction between CRM and promoter is sustained, *wg2946* cannot access the promoter. Additionally, *ftz* cannot repress *wg2946* (Fig. 2-5B and 2-8B), suggesting that repression by *ftz* in cell 5 is mediated by *wg3911.* In cell 6, *odd* and *runt* are expressed, and repression is mediated by *wg3911.* Repression is also seen for *wg2946* with *odd* manipulation, however repression mediated by *wg3911* is thought to be dominant, blocks the release of paused Pol II and therefore prevents *wg2946* from accessing the promoter (Fig. 2-10D). In cell 7, repression is suggested to be mediated by *wg2946*, which is repressed by *odd (*Fig. 2-5F). Ectopic expression of *wg3911* is seen with *odd* loss of function (Fig. 2-4F). This is not seen, however for endogenous *wg.* A possible explanation for this observation is that a sustained interaction between *wg2946* and the promoter prevents *wg3911* from interacting with the promoter (Fig. 2-10E). Finally, cell 8 is the even-numbered stripe of *wg* expression (Fig. 2-10F).

As seen in *prd* gain of function, the even stripe is expanded for both *wg3911* and *wg2946,* indicting that *prd* is sufficient for even stripe expression (Fig. 2-7E and 2-8E). Even stripe expression may be mediated by both CRMs, following the billboard model of CRM activity (Kulkarni and Arnosti, 2003). This is suggested by the observation that with *prd* loss of function, *wg3911* is still capable of maintaining even numbered stripe expression so may also have access to the promoter.

For *runt, eve,* and *ftz* loss of function experiments, the resulting expression patterns of the *wg3911-lacZ* reporter and *wg* match. However, this CRM does not recapitulate *wg* expression for 3 tested manipulations: *prd* and *odd* mutants and *prd* gain of function, in other words the pattern of the reporter misbehaves in these cases. These results suggest that *wg3911* is not the only element capable and perhaps necessary for generating the full *wg* expression pattern. The behavior of this CRM can at least partially be explained by the activity of the *wg2946* CRM. The *wg2946* CRM drives a misbehaving expression pattern without pair-rule manipulation: ectopic expression 1-cell posterior to the odd stripe. It is likely that another CRM is working in conjunction with *wg3911* and *wg2946* to drive full *wg* expression. There is evidence of at least 1 such region in the ChIP-on-chip data from (MacArthur et al., 2009). Just upstream of the promoter, there is another peak of binding, overlapping for *hairy*, *runt*, and *prd*.

Even though the activity of each enhancer can at least partially explain the misbehaving expression seen when the CRMs are acting alone to regulate transcription of *wg,* there are other possibilities for the observations seen. Positioning of CRMs may contribute to ectopic expression for example the composite reporter is constructed with one the *2946* CRM downstream of *3911,* the context of the intron the CRM usually resides in may be important. However, order of PESE

and DESE is a reporter construct was shown irrelevant, they behaved the same way regardless of the position they were cloned into the reporter gene.

Another explanation for the different patterns produced by each enhancer is that *wg2946* is acting more similarly to a shadow enhancer. Levine's group coined this term in 2008 to describe a particular type of enhancer. They found from ChIP-chip experiment data that there were multiple enhancers predicted for several *dorsal* target genes (Hong et al., 2008). When tested as CRMs in *vivo,* these enhancers act as secondary enhancers and are able to drive expression that overlaps with that of the primary enhancer. Primary and shadow enhancers are proposed to act together to refine the expression of developmentally regulated genes. Shadow enhancers are found to be large distances away from promoters, some being over 5kb away, such as one enhancer for brinker *(brk)* located 15kb updtream of the tss. Although, the 2 identified *wg* CRMs are located within 5kb of the *wg tss*, this evidence supports the idea *wg3911* and *wg2946* are acting as primary and secondary enhancer. Since *wg3911-lacZ* more faithfully recapitulates endogenous *wg* expression, this suggests that *wg3911* is the primary enhancer for *wg* at the blastoderm stage.

It is difficult to distinguish what is acting as a secondary enhancer, shadow enhancers, redundant enhancer, etc. however, there is no shortage of examples of genes regulated by CRMs that when tested alone, can produce redundant expression patterns. In addition to our findings regarding *slp1,* the *snail* gene, which is a regulator of ventral patterning, has 2 enhancers that were shown not to be redundant, and moreover, act to fine tune each other in a non-additive manner. A dual enhancer module is described to regulate transcription in Drosophila neural precursor genes (Miller et al., 2014). For several neural precursor genes, pairs of CRMs, notably found on relatively close to the promoter, and on either side of it, act together to produce a fine-

tuned pattern of *neur, phyl, sens,* and *nvy* expression. This organization is similar to that of the *wg* locus, where *wg3911* is upstream of the tss and *wg2946* is downstream.

In order to verify the models suggested for both *slp1* and *wg* transcription regulation, physical confirmation of enhancer promoter interactions is required. One recent method developed to investigate physical contacts between enhancers and promoters is chromosome conformation capture (Dekker et al., 2000). In addition, probing the binding site architecture of the confirmed *wg* CRMs would provide more insight into DNA binding dependent vs. independent effects mediated by pair-rule transcription factors.

Overall, these experiments provide two additional examples of CRMs that mediate both activation and repression by Runt and further suggest that the non-additive integration of inputs from different CRMs is a widespread aspect of regulating transcription in animal systems.

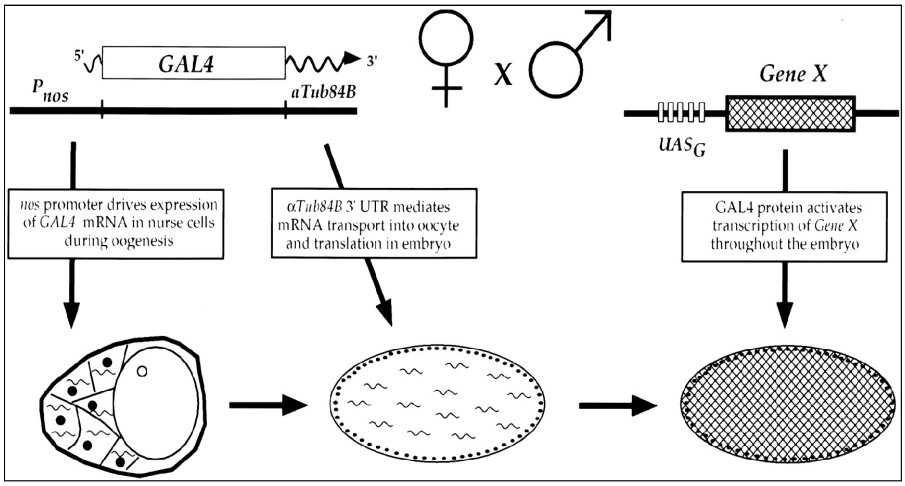

Figure 2-1. Maternal GAL4 system. GAL4 is expressed under control of the *nanos* promoter in nurse cells and transported into oocyte through the tubulin 3' UTR. Crossing NGT females to UAS males results in GAL4 driven ectopic expression of gene of interest in all cells. (Tracey et al., 2000)

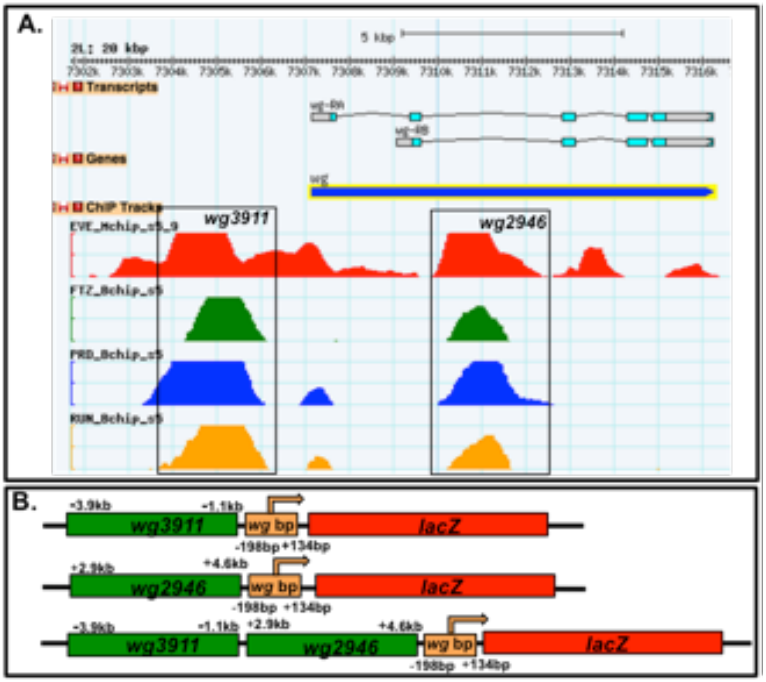

Figure 2-2. Two regions of overlapping pair-rule binding are candidates for CRMs.

Panel A shows ChIP-on-chip data from the Berkeley Drosophila Transcription Network Project (McArthur et al, 2009). Boxes are drawn around regions of overlapping pair-rule binding. Panel B shows schematics of report constructs made from the regions of overlapping binding, The upstream region is named *wg3911* (-3.9 to -1.1kb upstream of transcription start site or tss) and the downstream region is named *wg2946* (2.9 to 4.6kb downstream of tss).

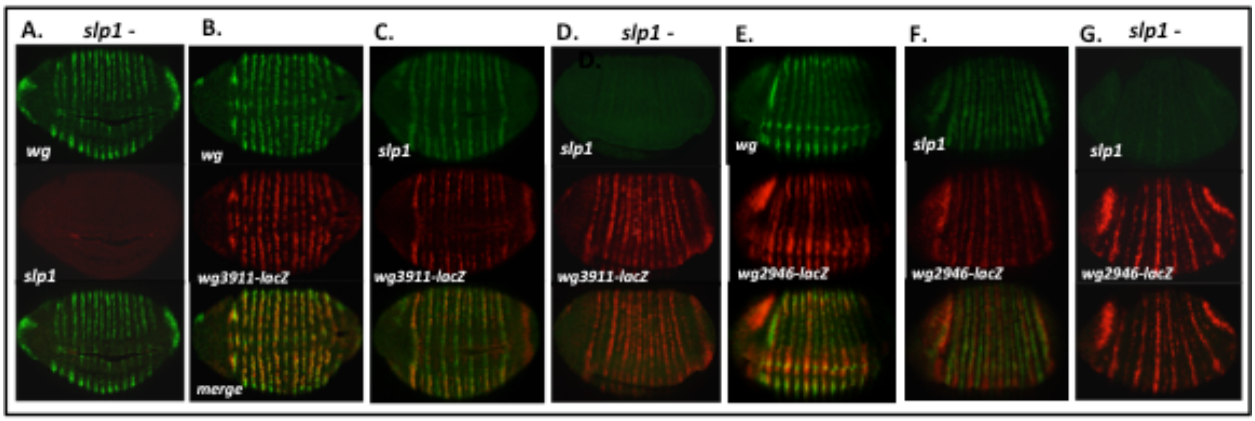

Figure 2-3. Effects of *slp1* loss of function on endogenous *wg* and *wg* reporter expression

A. Double fluorescent *in situ* hybridization shows endogenous *wg* expression is unaffected by *slp1* lof. Embryos are approximately 3.5 hours old, or cycle 14 and are oriented anterior to the left. Endogenous *wg* is depicted in green, and *slp1* in red, which shows null expression. Merge image in the bottom image for each panel. B. Wild type expression driven by *wg3911-lacZ* . C. Expression of *slp1* (top) compared to *wg3911-lacZ.* D. Expression of *wg3911-lacZ* in a *slp1* loss of function background. E. Expression driven by *wg2946-lacZ.* F. Endogenous *slp1* expression top) compared to *wg2946* driven expression. G. Expression driven by *wg2946* in a *slp1* lof background.

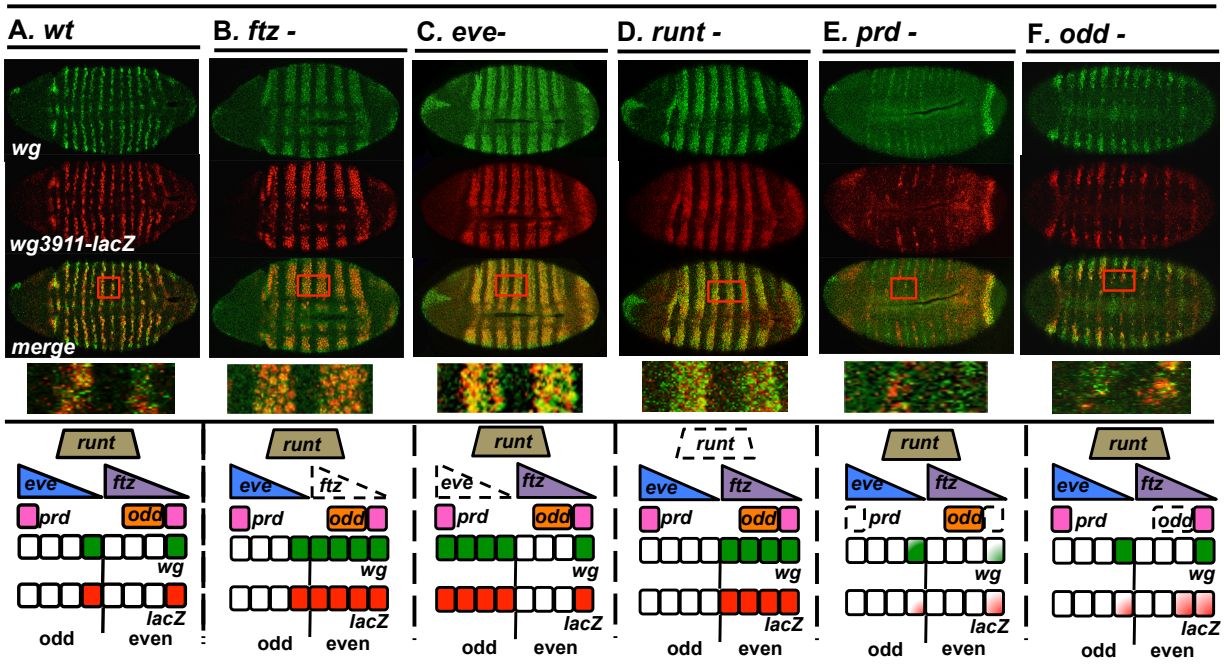

Figure 2-4. *wg3911* is a pair-rule response element

Double fluorescent *in situ* hybridization shows the response of the *wg3911-lacZ* reporter to pair-rule loss of function. All embryos are in the blastoderm stage and oriented anterior to the left. Endogenous *wingless* expression is shown in green at the top of each panel, reporter expression is shown in red in the middle, and the merged image on the bottom. Insets below show zoomed in views of one repeating even and odd stripe set. Schematics below the image include the phasing of all 5 pair rule gene expression patterns and also represent one repeating unit of *wg* and reporter expression. The endogenous and reporter patterns largely match, the response of each to pair rule loss of function is the same, except with *prd* in panel E and *odd* in panel F loss of function, in which the *wingless* and *wg3911-lacZ* do not respond the same.

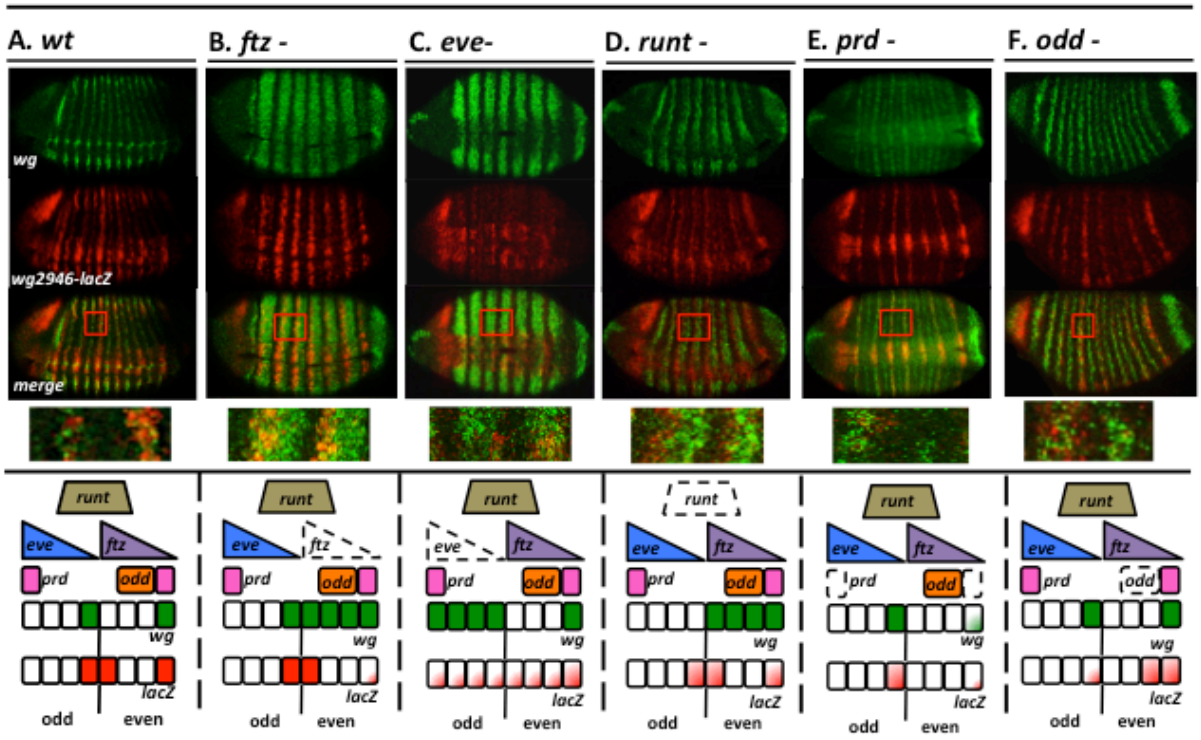

Figure 2-5. *wg2946* responds to pair-rule manipulation

Double fluorescent *in situ* hybridization shows the response of the *wg2946-lacZ* reporter to pair-rule loss of function. The endogenous and reporter do not match for any of the pair-rule manipulations. There are specific differences in mesoderm expression between *wg* and *wg2946-lacZ* in panels C, D and E.

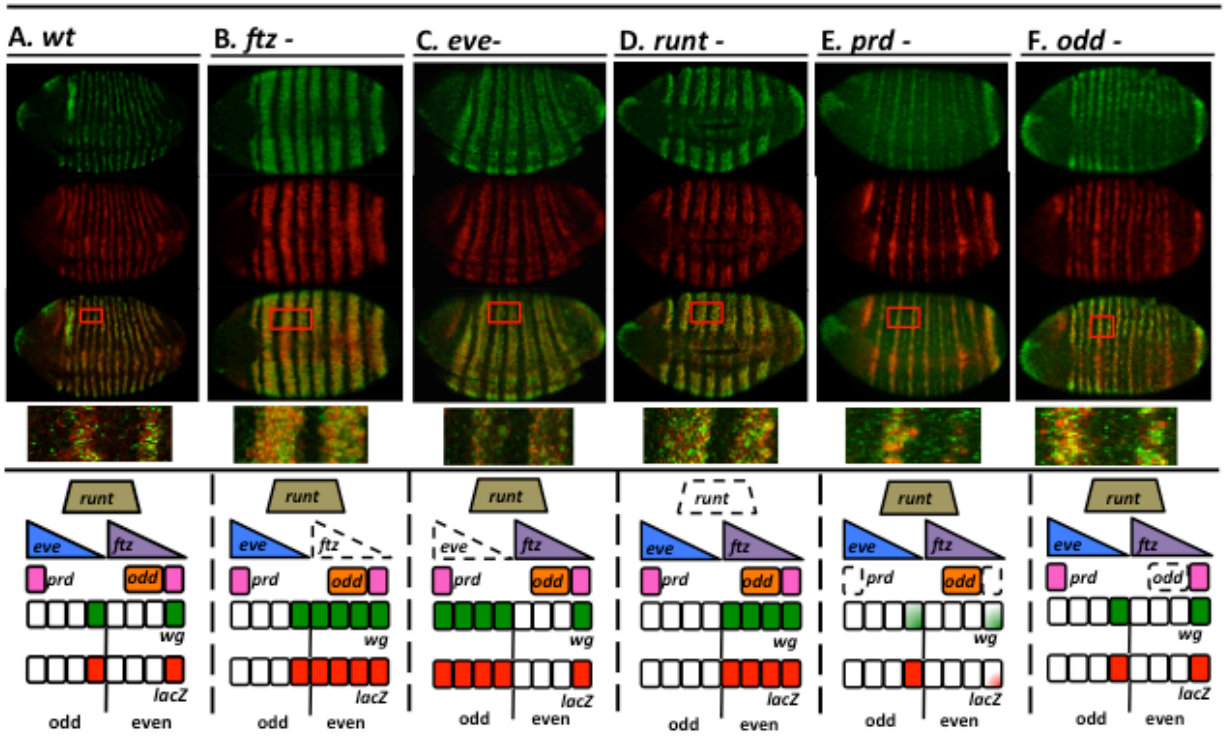

Figure 2-6. *wg3911/2946* recapitulates endogenous *wingless* expression

Double fluorescent *in situ* hybridization shows the response of the *wg3911-lacZ* reporter to pair-rule loss of function. The endogenous and reporter patterns largely match, the response of each to pair rule loss of function is the same, with some differences in head expression.

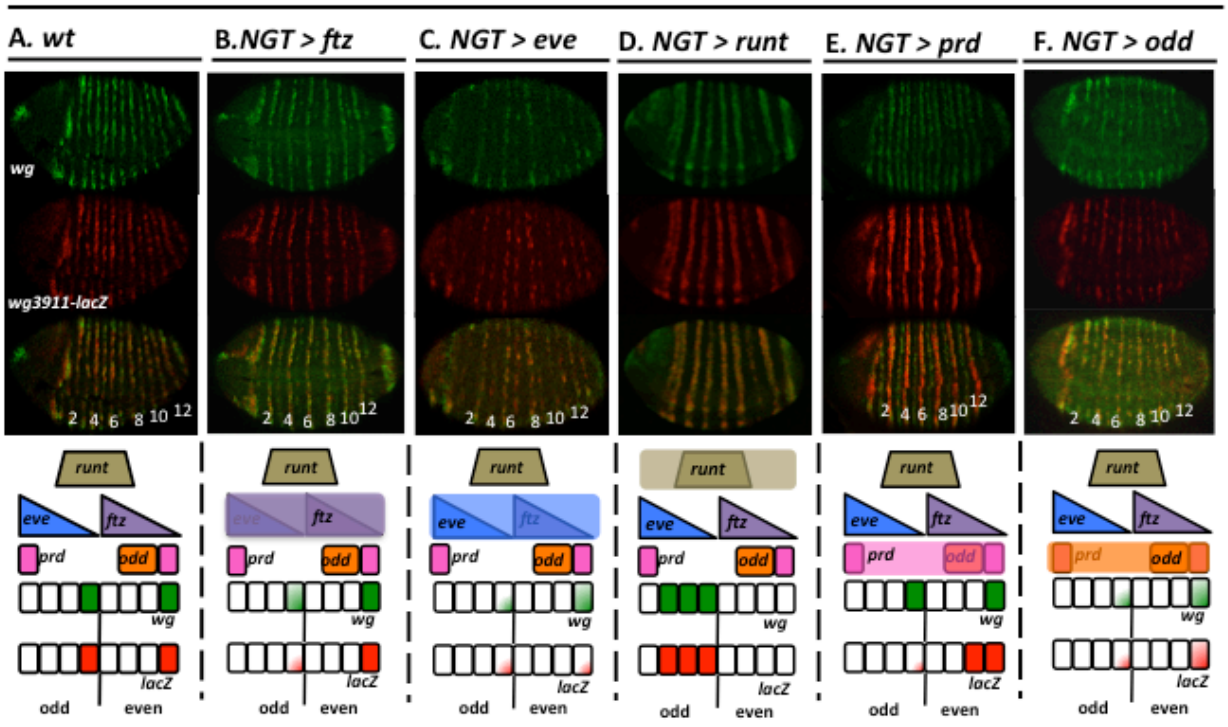

Figure 2-7. Pair-rule gain of function and *wg3911* expression

Double fluorescent *in situ* hybridization shows the response of the *wg3911-lacZ* reporter to pair-rule gain of function. Numbers in merge image represent the even stripes of each expression pattern. The endogenous and reporter gene expression patterns match, except with *prd* gain of function. In panel E, *wg* expression is normal, however, the reporter drives only the even stripes.

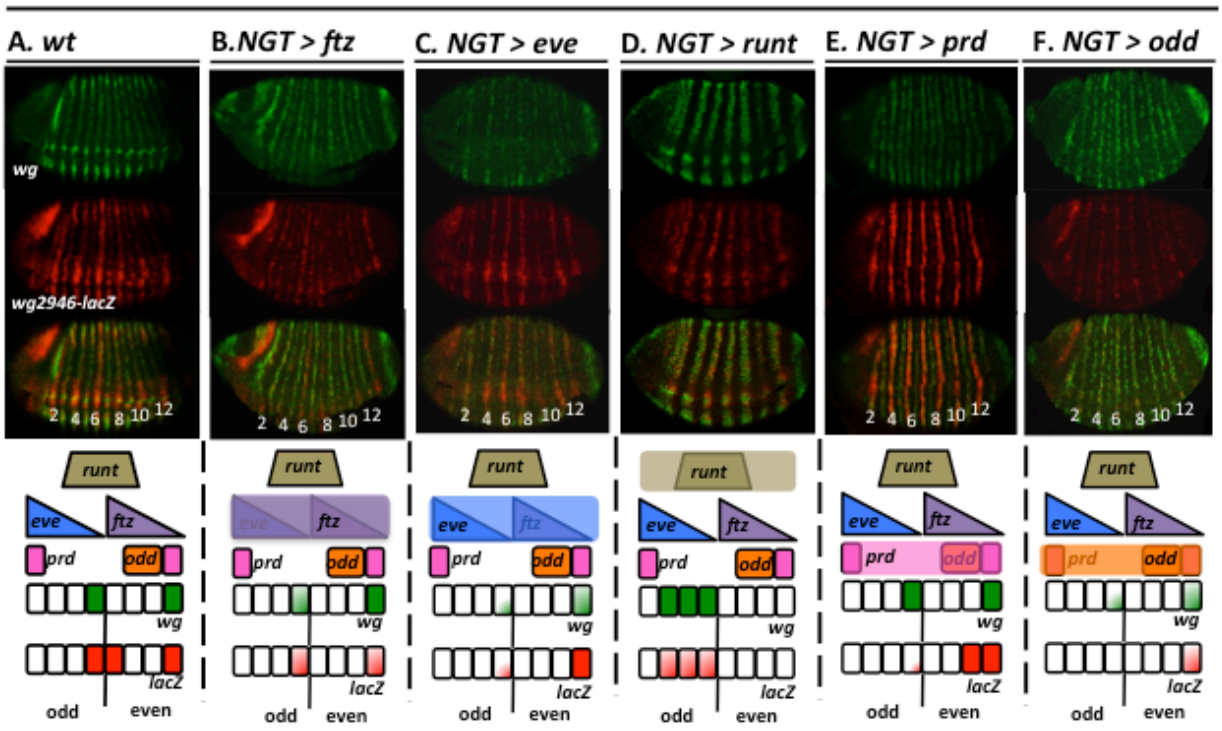

Figure 2-8. *wg2946* responds to pair-rule gain of function

Double fluorescent *in situ* hybridization shows the response of the *wg2946-lacZ* reporter to pair-rule gain of function. The endogenous and reporter expression patterns do not match in the wild-type embryo in panel A, as well as in all 5 pair-rule gain of function experiments.

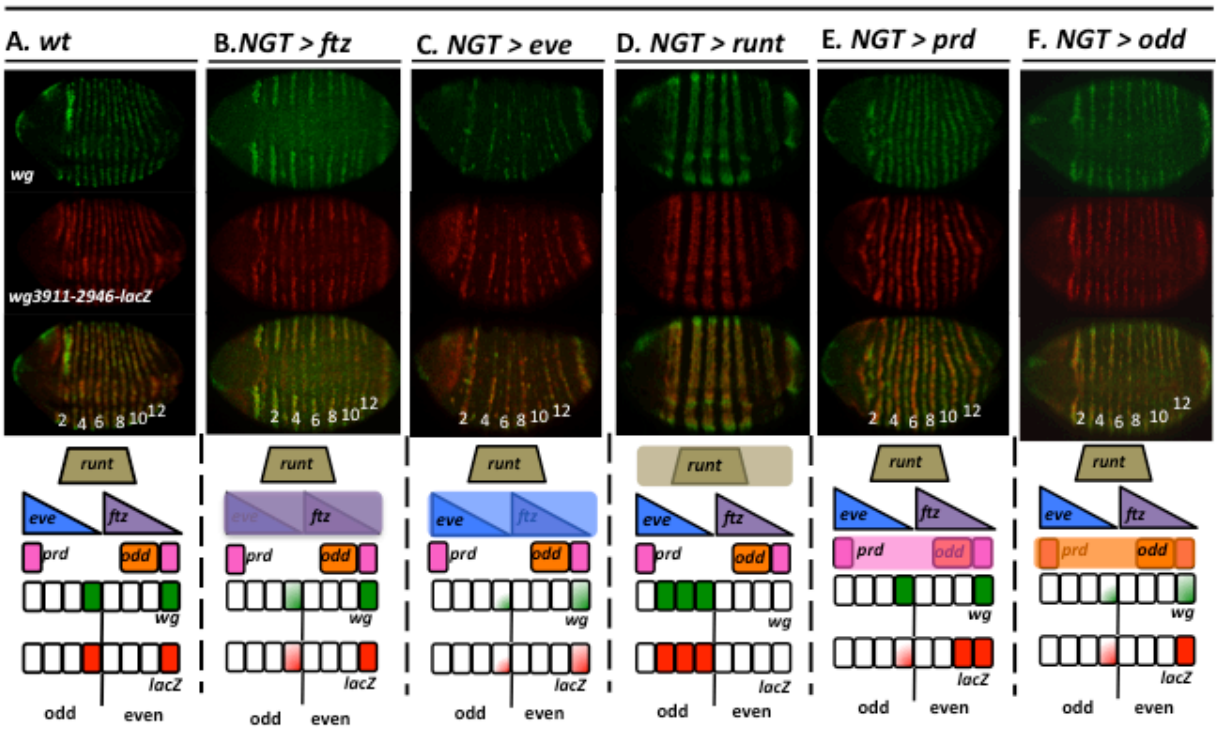

Figure 2-9. *wg3911/2946* recapitulates endogenous *wingless* expression.

Double fluorescent *in situ* hybridization shows the response of the *wg3911/wg946-lacZ* reporter to pair-rule loss of function. The endogenous and reporter patterns largely match, the response of each to pair rule loss of function is the same, with some differences in head expression. With *odd* gain of function, there are subtle differences in mesoderm expression.

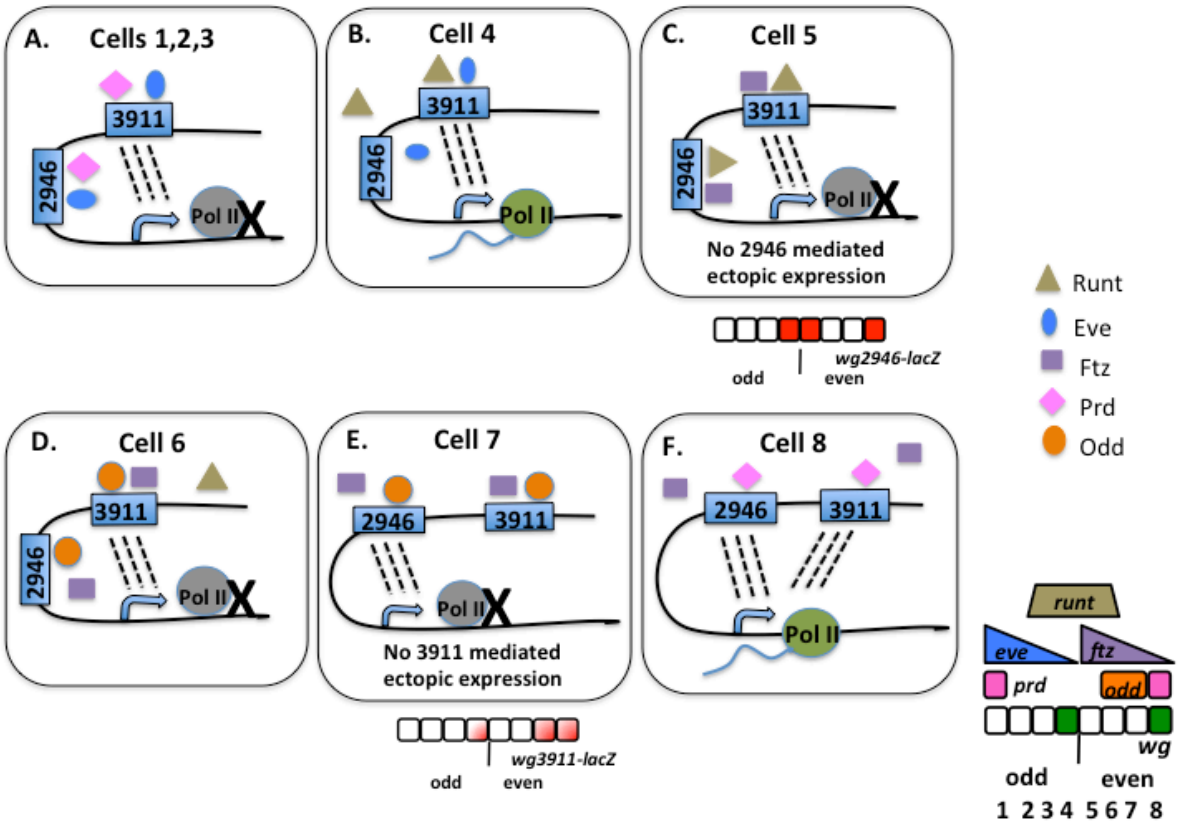

Figure 2-10. Cell context model of *wg* regulation by pair-rule factors

Schematics A-F represent a summary of pair-rule regulation of *wg* by cell context. Dotted lines indicate interaction with the promoter (arrow). Green Pol II is active, gray is paused. Blue boxes are *wg* CRMs *wg3911* and *wg2946.* Colored shapes, labeled on the right are pair-rule transcription factors. Schematics of reporter activity are below C and E. Overall schematic of endogenous *wg* expression and pair-rule inputs is on the bottom. Cells in the odd and even parasegments are numbered 1-8, consistent with the labels in A-F.

Acknowledgements: Jinelle Wint provided technical assistance in generating the yw;wg2946/wg2946 homozygous stock, and helped with wild type *wg2946* in situ hybridization experiments. Yelena Altshuller carried out the molecular cloning for the *wg2946* and *wg3911/2946* constructs.

Chapter 3

Molecular basis of enhancer-promoter interactions in the *slp1* and *wingless* loci

Abstract

There is growing evidence for long range DNA looping as a molecular mechanism for transcription regulation. Extensive work done on the β-globin locus paved the way for genome wide studies of enhancer-promoter interactions (Noordermeer et al., 2008; Kim & Dean, 2010). These loops can occur between enhancers and promoters, and are often associated with paused polymerase (Ghavi-Helm et al., 2014). Evidence from our lab regarding the *slp1* gene indicates that there are cell context dependent enhancer-promoter configurations, mediated by the transcription factor Runt. To investigate the molecular mechanisms underlying Runt-mediated expression, I conducted 4C-seq experiments. I first generated wild-type (*yw)* 4C-seq libraries from 3-4 hour old embryos and sequenced with short reads as a preliminary experiment. Results looked promising and I carried out a larger scale experiment using a slightly modified protocol, starting with less raw material. Results among viewpoints were inconsistent, however, and to address some of the issues, a higher resolution version of the preliminary experiment protocol was developed.

Introduction

The original 3C method was developed in 2002 (Dekker et al., 2002). In the years since, several variations on the original method have evolved which take advantage of the ability to conduct genome-wide surveys of chromosome conformation. 4C-seq, allows the detection of all contacts between a specific genomic locus, or viewpoint, and the rest of the genome, by next generation sequencing. It is possible that this method will uncover new enhancers for *slp1* or other genes investigated. Indeed, enhancer prediction data sets, such as those from RedFly (Gallo et al., 2006) suggest there are many enhancers for *slp1*, which have not been confirmed *in vivo*. Whether these are all relevant to the blastoderm embryo remains to be seen.

It is suggested from *in situ* and ChIP data that PESE and DESE are being repressed in different manners depending on cell context. As shown in Fig 1-4 (Hang et al., submitted), which summarizes these experiments, is it suggested that Runt regulates PESE and DESE interactions with the *slp1* promoter. In cells that have *runt+ftz*, or cell type III DESE repressed by being blocked at transcription elongation, while DESE is interacting with the promoter, but Pol II is paused, PESE cannot interact at the same time, and so in the same cell context PESE is repressed by a block to transcription initiation. It is preferred in 4C-seq experiments to use cells that are all of the same context, and these experiments are often done in cultured cells. However, taking advantage of the NGT40 system provides us with a way of shifting cell contexts and therefore a way to compare in different *runt* backgrounds, interactions between PESE, DESE and the *slp1* promoter. These embryos will provide an indication of the potentially wide range of interactions available to the *slp1* promoter, including those with enhancers that control expression in the head at this stage (Fujioka & Jaynes, 2012).

Methods

To investigate chromosome contacts at the *slp1* locus, I chose to use a modified version of the original chromosome conformation capture (3C) method, circular chromosome conformation capture coupled to next generation sequencing or 4C-seq (Gheldof, et al., 2012; Splinter et al., 2012; see Fig. 3-1 for overview) to generate a control 4C-seq library from wild-type *yw* embryos. 4c-seq gives us an unbiased way to investigate enhancer interactions. More traditional 3C methods require some previous knowledge of both the bait and the presumed interacting fragment. In 4C-seq, 1 viewpoint can capture up to 1 million interactions. Briefly, chromatin from ~80 milligrams of 2.5-3.5 hour old *yw* embryos, representing $5x10^7$ cells, was cross-linked with 3% formaldehyde for 30 minutes. Chromatin was first digested with EcoRI (NEB) at 37°C overnight in a shaking incubator, ligated for 2 hours at 16°C , and crosslinks were reversed. The resulting DNA was then digested with MseI (NEB) as before, and ligated under dilute conditions for 4 hours at 16°C. Inverse PCR (iPCR) was carried out using the Roche Expand Long Template PCR System with primers with single-end Illumina adapters attached to the 5' ends. To obtain sufficient amplification, 16 iPCR reactions were done per viewpoint, starting with 200ng of 4C library per reaction. This represents 3.2µg of ligated DNA, and the 16 reactions were pooled per viewpoint for sequencing. PCR clean up was done with the Qiaquick PCR cleanup kit either once or twice, and also followed by another clean up with the Agencourt AMPure XP PCR clean up kit, as residual primer removal was not sufficient with the Qiagen kit. Moving forward, 8 reactions will be pooled and cleaned up with the Qiagen column based kit, followed by Agencourt beads. Several presumed ligation products were verified by traditional PCR amplification and DNA sequencing. The library was visualized on a 1.5% agarose gel (Fig. 6A) and analyzed with an Agilent 2100 Bioanalyzer, using the high-sensitivity DNA kit (Fig.

6B). Paired-end sequencing is done and then results can be analyzed in several ways, but it is suggested that a custom fragment end library is generated for mapping. Relative frequency of interactions is measured by read number. Viewpoints included the *slp1 promoter,* PESE, DESE, the *slp2* promoter and the *cg3407* promoter, the gene just distal to the *slp1* locus.

For a larger scale experiment using 3 genotypes: wild-type *yw*, ectopic expression of Runt+Opa (RO) and ectopic expression of Runt+ Ftz (RF), another protocol was found which used less staring material, thus allowing for the inclusion of more viewpoints in the experiment. 4C-seq libraries were made according to (Harmen et al., 2012). This protocol suggests starting with $1x10^7$ nuclei, 5 times less than the previous protocol. The primary 6-cutter restriction enzymes were EcoRI or BamHI. The secondary restriction enzymes used were MseI or CviQI. Viewpoints captured by EcoRI and MseI were the *slp1 promoter*, PESE, *slp2* promoter, *wg2946*, and a putative *prd* CRM. EcoRI and CviQI viewpoints were PESE 2, DESE, the *engrailed (en)* promoter/overlap with CRM, and the *gooseberry (gsb)* promoter. Finally, BamHI/MseI viewpoints were the *twist (twi)* promoter, and the *wg3911* CRM. Statistical analysis is done to assess the significance in the differences in interaction frequencies. Final data can be represented graphically (Gheldof et al., 2012; Splinter et al., 2012). 16 base-pair reads should offer enough uniqueness to map to the Drosophila genome without the use of a fragment end library, so traditional mapping was used.

Results from this experiments prompted me to design a third experiment to overcome some shortcomings of the previous method. First, embryos were fixed in 2% paraformaldehyde for 20 minutes. Viewpoints were narrowed to the *wg* and *slp1* loci, two genotypes were used: yw and RO genotypes, and resolution was increased by using two 4-cutter restriction endonucleases instead of one 6 cutter and one 4 cutter. Working primer concentration is suggested to be lower

than the previous concentration of 1.12nmol, followed by clean up with magnetic beads in addition to column based purification (as done for the preliminary experiment). The protocol otherwise follows (Gheldof et al., 2012). The primary restriction enzyme used was DpnII for all libraries, two samples of 5x10^7 nuclei from each genotype, *yw* and RO. The secondary restriction enzyme was either CvQI or NlaIII. Viewpoint fragments captured by DpnII-Csp6I were the *slp1* promoter, DESE, *slp2* promoter, *wg* promoter, and *wg2946.* Viewpoint fragments captured by DpnII-NlaIII were PESE, DESE, and *wg3911.* We can use single-end, sequencing, with 100bp reads, with to investigate physical interaction on an Illumina Hi-seq Genome Analyzer. Refer to appendix for list of primers used. An online pipeline for data analysis can be found at http://htstation.vital-it.ch/.

Results & Discussion

Library construction and sequencing was successful for the preliminary 4C-seq experiment on *yw* embryos. Figure 3-2 shows a gel image that indicates efficient overall digestion and ligation of the *yw* library. Primers were largely depleted. Paired-end Illumina sequencing was done on the MiSeq in the Leatherwood/Futcher lab at Stony Brook University. Sequence quality was good and overall read number was substantial. There were 4,983,783 reads, but the read length was too short to map after trimming of the primer sequences. Despite these setbacks, this method looked promising to investigate *runt* dependent changes in enhancer promoter interaction.

To expand this experiment, 3 different genotypes were used to investigate Runt-dependent changes in physical interactions enhancer-promoter interactions for segment polarity genes. Genotypes were verified by immuno-histochemical *in situ* hybridization with *slp1-dig* probe (Fig. 3-3), NGT40 driven ectopic expression of Runt+Opa restuls in expansion of *slp1* expression, and Runt+Ftz represses expression. Viewpoints were chosen from several different segment polarity gene loci in addition to *slp1* to potentially uncover novel enhancers or interactions for other Runt dependent genes. Libraries were well digested and ligated, and final products are shown in Figure 3-4. Faint smears representing a mix of interacting fragments are seen for the 3 genotypes. Paired-end sequencing returned 15.73 million reads with good quality scores.

No significant differences were discernible with analysis of the sequencing results for this experiment, however. The paired-end sequencing analysis indicated low mapping efficiency, and results were inconsistent between viewpoints. Figures 3-5 through 3-7 summarize the results

from the 4C-seq experiment. Figure 3-5 is a summary of the *slp1* locus viewpoints. A major viewpoint of interest, the *slp1* promoter, only had 256 reads, 3 of which mapped to the genome. This library did have lower relative concentration, but was diluted to 2nM. One successful viewpoint was DESE in the *yw* genotype. There was 83% coverage and nearly 150,000 reads. However, not all viewpoints that returned large read numbers mapped efficiently. For example, Figure 3-6, which summaries sequencing data for the *wg* locus, shows that for the *wg2946* viewpoint in *yw* embryos, there were almost 1 million reads, but only 1% coverage.

One contributing factor which may have caused low read number or poor mapping efficiency for particular viewpoints are poor digestion with primary restriction endonuclease. Further analysis of restriction digestion analysis, assessed by qPCR with primers that span a restriction site, show that not all viewpoints were equally digested. Another possibility is that the recognition sequence for MseI is too common (TTAA) so the enzyme, due to its long incubation time, had non specific effects. It is recommended that the cut site be comprised of all 4 nucleotides (Gheldof et al., 2012). Also, low library concentration could contribute to improper dilutions or high primer concentrations could overwhelm the libraries with primer dimers. Paired-end sequencing is not the most commonly used type with 4C-seq, but long read single-end reads are suggested. Finally, mapping without the use of a custom generated fragment end library may have resulted in low numbers of mapped reads. Analysis of restriction digestion analysis shows that viewpoints were not equally represented.

3C libraries were constructed using the higher resolution version of (Gheldof et al, 2012) 3C-ibraries largely show efficient digestion and ligation (Fig. 3-7). One of the *yw* libraries has a low concentration, as seen in the gel image, and by Qubit quantification, and is suggested to

remake it before moving on with sequencing. Concentration for *yw2* was less than 56ng/ul and over 100ng/ul for the remaining 3 libraries

Another genotype to be considered in future experiments is Eve over-expression. Eve represses PESE in type I cells and results in a block to elongation of transcription. However, in type III cells, Runt represses PESE resulting in a block to transcription initiation. Ectopic Eve expression will convert Type IV cells to type I, a subset of the total embryo. There may be detectable differences though in contacts between PESE and the *slp1* promoter in Eve overexpressing embryos compared to wild-type.

The model predicts that a shift in context of *slp1* expression will result in an increase in interactions that are associated with Runt-dependent regulation of DESE (cell types II and III). It is of great interest to determine if there are any differences in conformation between the activated and repressed *slp1* expression states. In a simple interpretation of the current model, there may be no difference in chromatin conformation between these two expression states. However, the model predicts differences between these two states and the collection of expression states that exist in wild-type embryos. Cell types II and III account for approximately 35% of the cells in wild-type embryos, whereas cells in RO cell type II number increases and cells in RF embryos are nearly all cell type III. Thus, there is the opportunity for an approximate 3-fold increase in interactions associated with DESE mediated *slp1* regulation. Conversely, and perhaps more importantly, there may be significant reductions in interactions associated with regulation mediated by PESE or by enhancers that normally control *slp1* expression in the non-segmented regions of the embryo.

Looking at embryos that ectopically express Runt+Opa will allow us to determine if these factors alone are capable of producing detectable changes in the conformation of the *slp1* locus.

One advantage that has been instrumental to our studies on *slp1* is the relatively simple combinatorial rules for its regulation. These simple rules are unique to *slp1*, as other factors such as Paired and Odd-skipped play important roles in generating the single cell wide stripes of *engrailed*, *wingless* and other segment-polarity genes (*wg* results summarized in Chapter 2, Manoukian & Krause, 1993; Miskiewicz et al., 1996; Morrissey et al., 1991; Mullen and DiNardo, 1995).

Opa was in the above experiment based on the observation that the levels of promoter associated Pol II increase in response to co-expression of Runt, Opa and Hairy as compared to wild-type (Hang et al., submitted). Opa is uniformly expressed throughout the pre-segmental region of the blastoderm stage embryo (Benedyk et al., 1994), and although Opa does not seem to provide positional information that discriminates between the different cell states within this region it may play an important role in defining this region by 'enabling' the activity of enhancers that mediate pair-rule regulation

Taking advantage of the details of *slp1* pair-rule regulation allowed me to generate embryos that shift cell context to ectopically express or show repression of *slp1*. The maternal GAL4 expression system was used to co-express either Runt & Opa (RO) to activate *slp1* expression and increase the proportion type II cells or Runt & Ftz (RF) to repress *slp1* throughout the embryo to turn nearly all cells into cell type III**.** I attempted to carry out 4C-seq experiments to I investigate chromatin conformation at the *slp1* locus in the Drosophila embryo. The frequency of interactions between the *slp1* regulatory elements and other genomic loci is still to be compared in late blastoderm stage Drosophila embryos of yw, RO, RF embryos 4C-seq can be used to assess contacts between the *slp1* promoter and other noted locations in the *slp1* locus as well as in the rest of the genome.

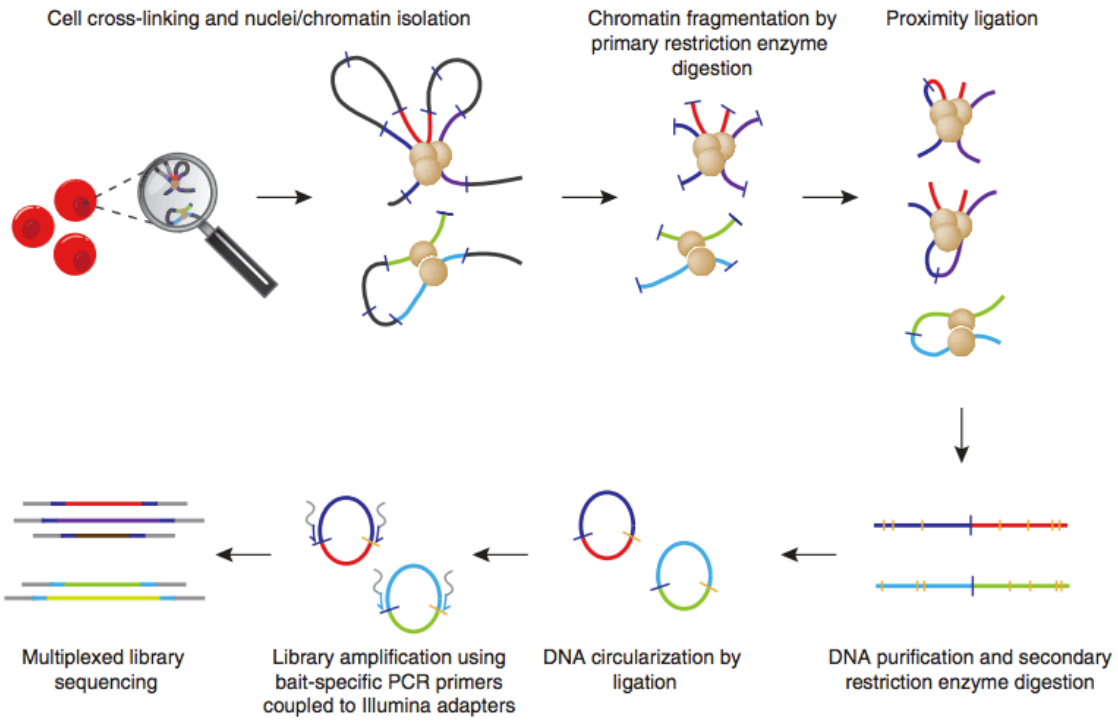

Fig 3-1. Overview of 4C-seq method

The general strategy for 4C-seq is outline above (Gheldof et al., 2012). Both methods used in the described experiments follow the same general principles; any important differences are explained in the text. To uncover interacting fragments of DNA, chromatin is isolated and subject to digestion with a primary restriction endonuclease. The first round of proximity ligation is followed by digestion with a secondary restriction enzyme. DNA is the circularized, viewpoint specific primers are synthesized with the addition of Illumina sequencing adapters, and inverse PCR is done to amplify the fragments The interactions are analyzed by calculating the relative frequencies of the reads.

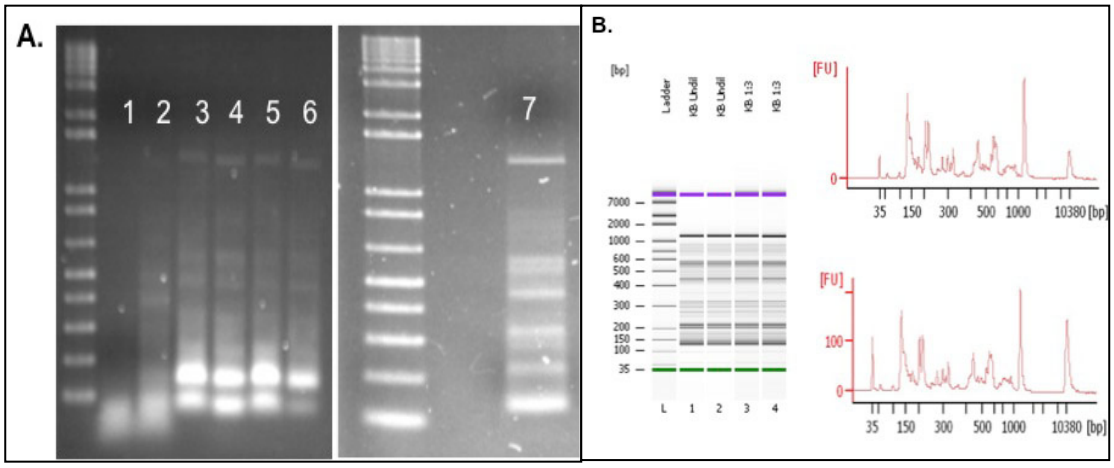

Figure 3-2. Preliminary 4C library

A. Agarose gel images of iPCR products. Lane 1: negative control; Lane 2: iPCR, primers without Illumina adapters, 10µl from 1 reaction loaded onto the gel; Lane 3: iPCR with primers that include the adapters; Lane 4: PCR product pooled from 4 reactions, purified once with Qiaquick PCR purification kit, 5µl loaded onto the gel; Lane 5: same as lane 3; Lane 6: same as lane 4 except purification was done twice; Lane 7: pooled product from 8 reactions cleaned up twice with the Qiagen kit, followed by the Agencourt AMPure XP PCR cleanup kit, note full depletion of residual primers, 10µl loaded onto gel **B.** Bioanalyzer results from the high sensitivity DNA kit (samples run in duplicate); the gel image looks similar to the agarose gel; graph to the right represent concentration peaks for each distinct band; graph areas without distinct peaks represent smear of PCR product. Top graph is undiluted product from Lane 7 in panel A, and bottom graph is diluted 1:3

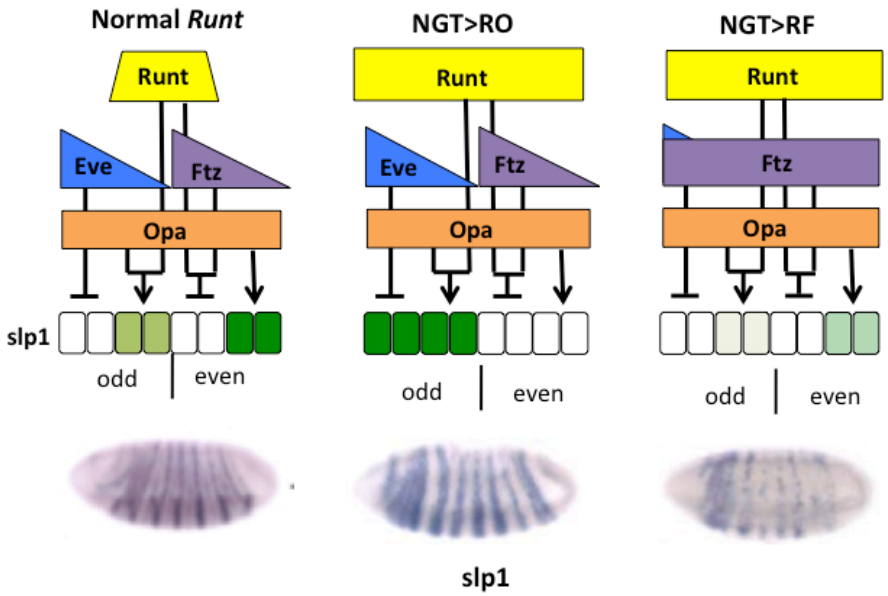

Fig 3-3. Verification of genotypes for 4C-seq experiments

Schematics represent pair-rule inputs that regulate *slp1* expression over an 8-cell division of the embryo comprised of the posterior portion of the odd parasegment and the anterior portion of the even parasegment. Immunohistochemical *in situ* hybridization on Stage 5 embryos, oriented anterior to the left, verifies ectopic expression of *slp1* with NGT>Runt and loss of expression of *slp1* with NGT>Runt+Ftz.

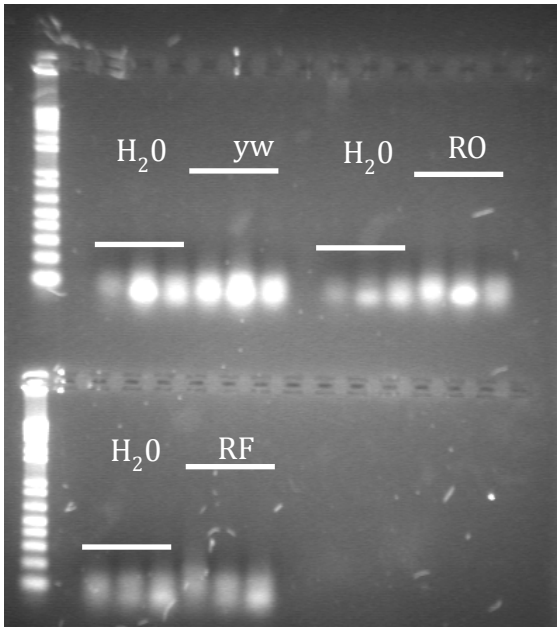

Fig. 3-4. 4C-libraries for Experiment #2 (3 genotypes)

4C-seq libraries analyzed on a 1% agarose gel, after 2 rounds of enzyme digestion and dilute ligation. A test inverse PCR was done with primers for the *slp1* promoter viewpoint. Each library had a water control.Top left set of lanes if for the *yw* genotype , top right is RO, and bottom left is RF. For each library, sample was run compared to

Viewpoint	slp1 pro.			PESE			PESE2			DESE			slp2 pro.		
Genotype	yw	RO	RF	yw	RO	RF	yw	RO	RF	yw	RO	RF	yw	RO	RF
Total read number	253	7715	104824	801906	1295179	2947942	787837	354302	438854	124917	40097	18435	5893492	658758	239894
Mapped read number	3	3879	5056	311252	2206	309	7359	5328	650	103567	7556	15222	3433204	162552	97849
% coverage	1.19	50.28	4.82	38.81	0.17	0.01	0.93	1.5	0.15	82.91	18.84	82.57	58.25	24.68	40.79
% undigested 6 cutter	92	43	49	36	28	21				28	42	12	18	11	41
Concentration of library	.139 ng/ul	.110 ng/ul	.104 ng/ul	.420 ng/ul	1.55 ng/ul	6.82 ng/ul	1.39 ng/ul	.272 ng/ul	.802 ng/ul	2.99 ng/ul	.132 ng/ul	.252 ng/ul	.111 ng/ul	.516 ng/ul	.44 ng/ul
6 cutter left or right	right EcoRI	right EcoRI	right EcoRI	left EcoRI	left EcoRI	left EcoRI	right EcoRI	right EcoRI	right EcoRI	left EcoRI	left EcoRI	left EcoRI	right EcoRI	right EcoRI	right EcoRI
Size of viewpoint frag	278bp			318bp			347bp			458bp			406bp		

Figure 3-5. Summary of 4C-seq results for *slp1* locus

Results for paired-end 4C-seq are summarized for the *slp1* locus. The five viewpoints included the *slp1* promoters, the *slp1* enhancers PESE and DESE, and the *slp2* promoter. Two of the viewpoints contained PESE. Total sequence read number, mapped read number, % coverage, % undigested, concentration measured with a Qubit fluorometer, position of 6 cutter site, and size of viewpoint fragment is listed for each viewpoint.

Viewpoint	wg 3911			wg 2946		
Genotype	yw	RO	RF	yw	RO	RF
Total read number	1227813	166550	149614	1492843	602738	408972
Mapped read number	649201	44894	50634	16021	1632	29
% coverage	52.87	26.96	33.84	1.07	0.27	0.01
% undigested 6 cutter	35	78	5	34	49	29
Concentration of library	.968 ng/ul	.128 ng/ul	.128 ng/ul	3.87 ng/ul	.106 ng/ul	1.24 ng/ul
6 cutter left or right	left BamHI	left BamHI	left BamHI	right EcoRI	right EcoRI	right EcoRI
Size of viewpoint frag	1092bp			722bp		

Figure 3-6. Summary of 4C-seq results for *wg* locus

Results for paired-end 4C-seq are summarized for the *wg* locus. The 2 viewpoints included the *wg3911* and *wg2946* .Total sequence read number, mapped read number, % coverage, % undigested, concentration measured with a Qubit fluorometer, position of 6 cutter site, and size of viewpoint fragment is listed for each viewpoint.

Viewpoint	prd CRM			en pro.			gsb CRM			twi pro.		
Genotype	yw	RO	RF	yw	RO	RF	yw	RO	RF	yw	RO	RF
Total read number	1004339	602738	408972	322409	5352833	1358106	873881	789709	86713	549404	106561	229221
Mapped read number	523342	1632	29	10564	2064	819	631621	384171	50808	4363	430	848
% coverage	52.11	17	14.91	3.28	0.04	0.06	72.28	48.65	58.59	0.79	0.4	0.37
% undigested 6 cutter	27	23	34	3	4		19	2	28		61	34
Concentration of library	.989 ng/ul	.204 ng/ul	.318 ng/ul	2.38 ng/ul	4.94 ng/ul	.984 ng/ul	3.54 ng/ul	1.36 ng/ul	.264 ng/ul	1.29 ng/ul	.136 ng/ul	.148 ng/ul
6 cutter left or right	left EcoRI	left EcoRI	left EcoRI	left EcoRI	left EcoRI	left EcoRI	right EcoRI	right EcoRI	right EcoRI	right BamHI	right BamHI	right BamHI
Size of viewpoint frag	482bp			3891bp			238bp			624bp		

Figure 3-7. Summary of 4C-seq results for *prd, en, gsb, and twi* loci

Results for paired-end 4C-seq are summarized for the *wg* locus. The 4 viewpoints included the *a* predicted *prd* CRM identified by overlapping regions of pair-rule transcription factor binding (Macarthur et al., 2009), the *engrailed (en)* promoter, a predicted CRM of *gooseberry (gsb)*, and the *twist* (*twi)* promoter. Total sequence read number, mapped read number, % coverage, % undigested, concentration measured with a Qubit fluorometer, position of 6 cutter site, and size of viewpoint fragment is listed for each viewpoint.

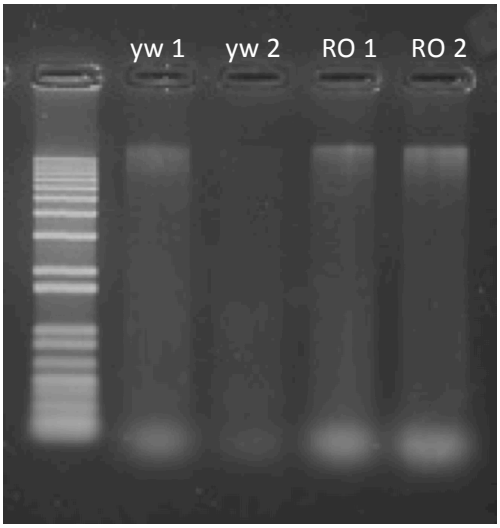

Fig 3-8. Digested and ligated 3C libraries

Chromatin from 3-4 hour old *yw* or NGT>RO embryos was digested with DpnII as the primary restriction enzyme. All viewpoints have DpnII as the primary cutter. The concentration of the yw2 library is low relative to the others. A Qubit fluorometer measured the library at 56ng/ul. Distinct band of viewpoint ligation with the proximal fragment is visible, with any undigested DNA remaining at the top.

Chapter 4

Miscellaneous Work

Molecular signatures of *slp1* enhancer activity

Abstract

In addition to transcription factor binding, enhancers are also predicted by certain molecular signatures, such as methylation at lysine 4 of histone 3 (H3K4me1) and enhancer RNA (eRNA) output, short messenger RNAs transcribed from enhancer DNA (www.modENCODE.org; Fig 4-1). Active transcription or initiated and paused transcription can also be predicted, and one way is by assessing cohesin complex binding among other methods. Rad21 is one member of the cohesin complex. To determine whether these signatures can be detected, I conducted ChIP and experiments to detect H3K4Me1 and Rad21, a cohesin complex component, and RT-qPCR to detect eRNA. In comparing yw to RO embryos, only low levels of H3K4Me1 and Rad 21 were detected with no significant differences, and a higher resolution method, such as RNA-seq, is suggested to detect eRNA output from PESE and DESE.

Introduction

Amongst the attributes of enhancers identified by the ENCODE and modENCODE projects are mono-methylation of lysine 4 on histone H3 (H3K4me1), association of cohesin with Pol II at initiated and paused sites of transcription (Schaaf et al,. 2013) and perhaps most intriguingly, the synthesis of a novel class of enhancer RNAs (eRNAs) (Fig. 4-1; Nègre et al.,2012). Recruitment of Creb-binding protein (CBP; Nejire in Drosophila), a common transcriptional co-activator, and eRNA synthesis have both been implicated with enhancer activity (Kim et al., 2010). Our knowledge of *slp1* regulation allows us to manipulate the activity of the DESE and PESE enhancers throughout the blastoderm embryo, thereby opening the door to experiments that investigate the relationship between the epigenetic status of these two enhancers and their activity. Data from the modENCODE project reveal peaks of H3K4me1 on both the DESE and PESE enhancers in chromatin samples from 0–4 hour Drosophila embryos (Fig. 4-1). It is notable that the patterns of association are somewhat different for these two enhancer regions. In PESE the signals are coincident and peak precisely within the C1 interval essential for the activity of this enhancer (Prazak et al., 2010). In contrast, within DESE a central peak of H3K4me1 association is flanked by two peaks of Nejire interacting regions, both contained within the currently defined minimal full length 900 bp DESE enhancer (from 8.1 to 7.2 kb upstream of the transcription start site).

Our knowledge of *slp1* provides us several approaches to investigating the relationship between the epigenetic signatures of these two regions and the activity of the enhancers. One approach is using the maternal GAL4 expression system to convert a subset of blastoderm cells into either *slp1* cell type II (RO embryos) or type III (RF embryos). In both of these embryo

genotypes the PESE enhancer is inactive, and potentially physically prevented from interacting with the *slp1* promoter.

Data from the modENCODE project also indicates polyA+ RNAs are expressed from regions corresponding to both the DESE and PESE enhancers in 2–4 hour Drosophila embryos. In both cases, these polyA+ RNAs lie just downstream of the currently defined minimal regions required for the activity of these enhancers (Fig. 4-1, www.modENCODE.org). Not all enhancers produce polyA+ RNAs. Indeed the more prevalent form of eRNAs are non-polyA+ RNAs that are divergently transcribed from a site corresponding to the center of CBP association, and interestingly the level of these short eRNAs correlates with enhancer activity (Kim et al., 2010). We will use this information to design primers to detect eRNAs from the *slp1* DESE and PESE enhancers in wild-type embryos as well as in embryos where the activity of the enhancers is manipulated either by genetics, or by mutagenesis of reporter gene constructs. It will be of particular interest to determine if eRNA synthesis is affected by perturbations that alter functional interactions between the two *slp1* enhancers and the promoter. The results of these experiments should provide new information on the relationship between eRNAs and enhancer-promoter interactions. The functional role of eRNAs is beginning to be elucidated, largely related to promoting accessibility of chromatin. Two examples are that eRNAs have been shown to associate with the mediator complex at the promoter, and to bind specifically with Rad21 and cohesin to mediate DNA looping (reviewed in Mousavi et al., 2014). Further investigation into eRNA output from *slp1* and *wg* enhancers may provide further insight into the molecular mechanisms underlying regulation of segment polarity genes.

Preliminary experiments are described here, which investigated enhancer signatures in different genetic backgrounds. No significant differences were observed for H3Kme1 or Rad21

association at PESE and DESE in different genetic backgrounds based on ChIP experiments, and eRNA output from PESE and DESE could not be detected by RT-qPCR. In light of the seemingly wide-spread role of eRNAs contributing to transcription regulation, further studies are warranted.

Methods

Chromatin immuno-precipitation and quantitative reverse transcription polymerase chain reaction (qRT-PCR) were used to investigate the relationship between enhancer activity and select molecular signatures of enhancers as identified by the modENCODE project (www.modENCODE.org). The *slp1* DESE and PESE enhancers are marked by H3K4me1 and express a novel class of enhancer RNA (eRNA) (Fig. 4-1). In addition, the cohesin Rad21 has been shown to regulate transcription and associate with Pol II (Hallson et al., 2008; Mckee, 2008). Rad21 has also been found to regulate *runx1* and *runx3* transcription in zebrafish (Horsfield et al., 2007)

Antisera specific for H3K4me1 and Rad21 that are useful for these experiments are commercially available, purchased from AbCam. ChIP was be carried out as in (Wang et al., 2007) with slight modifications: staged embryos are dechorionated, fixed with 2% formaldehyde for 20 min at room temperature, washed, and snap-frozen at −80°C for storage (Orlando et al., 1998). Approximately 100 mg of embryos are homogenized for 1 min in 10 mM EDTA and 50 mM Tris (pH 8.1). After addition of SDS to a final concentration of 1% and incubation on ice for 10 min, glass beads are added and the homogenate is sonicated to give sheared chromatin preparations with an average DNA size of 300–400 bp. Chromatin preparations from the equivalent of ~25 mg of embryos will be used for each immuno-precipitation using coated agarose beads (Millipore) with Rad21 (AbCam) or H3K4Mme1 (AbCam) antisera or rabbit serum. The relative amount of immuno-precipitated DNA is quantified using real-time PCR (Roche Light-Cycler) using primer pairs with a Tm in the range of 59°C–66°C that generate products between 150 and 211 bp. The percent precipitation values that are reported were

calculated using a dilution series of input chromatin to determine the relative efficiency for each primer pair. qPCR will be done in triplicate on samples from at least two independent immuno-precipitation assays.

In order to assess eRNA output from PESE and DESE, I extracted RNA using commercially available Qiagen kits and use the SYBR Green one-step qRT-PCR (reverse transcription and PCR) kit for quantitation. Primers targeted cDNA specific sequences corresponding to PESE and DESE, see appendix for primer list.

Preliminary Results & Discussion

I assessed if manipulations that affect the activity of these enhancers result in changes in H3K4me1 and Rad21 association as well as eRNA output. Higher resolution methods than the ones described here are required to detect any differences among these signatures, such as ChIP-seq and RNA-seq. It remains to be confirmed if Runt is sufficient for altering eRNA profiles on the *slp1* enhancers.

Rad21 ChIP on yw, RO, RF gave relatively low signals. In wild-type *yw* embryos, signal was higher at DESE and PESE than the promoter. For RO and RF this trend is reversed and the signal at the promoter is higher, however the signal overall is relatively low and doesn't differ from background in most cases. Each data point for *yw* represents 3 replicates (except for PESE 9537) and RO and RF represent 2 replicates. H3KMe1 ChIP on yw and Runt overexpressing embryos showed minor differences at the *slp1* promoter vs. PESE and DESE. Signal was higher at PESE and DESE than at the *slp1* promoter. Replicates are needed to confirm this.

Using publicly available data from modENCODE, cDNA sequences for eRNA generated from PESE and DESE were used for primer design. RNA was extracted with the Roche High Pure kit from *yw* embryos and cDNA synthesized. RT-qPCR did not detect these sequences. These experiments were preliminary, and so a standard curve was not used for quantification (data not shown).

The RT-qPCR signals were too low to detect any significant changes at PESE and DESE, but if manipulations in Runt activity are found to alter the eRNA profile of the target *slp1* then RNA-sequencing (RNA-seq) would be a powerful approach for identifying enhancers whose activity is regulated by Runt at a genome-wide level. Using this higher resolution method to

further assess Runt induced changes in molecular signatures of enhancer activity is suggested.

This area warrants more investigation, as there is an increasing amount of evidence for the role of eRNA in gene expression regulation. Transcription of long non-coding RNA precedes transcription activation of genes in the β-globin locus (Kim et al., 2015). Long non-coding RNAs can also regulate NELF release. Knockdown of ln-RNA expressed at neuronal enhancers for *arc,* impairs transient release of NELF from the specific target promoters during transcriptional activation. In this case, the enhancer-promoter interaction was unaffected by eRNA knockdown. (Schaukowitch et al. 2014).

The data from modENCODE also indicates peaks of Nejire/CBP binding at PESE and DESE. In addition to H3K4Me1 and eRNA output future experiments would include using ChIP to assay CBP binding. Another approach for investigating the relationship between epigenetic status and enhancer activity takes advantage of our functional dissection of these two enhancers. Our studies in this area have used ΦC31 transgenic approaches (Groth, et al., 2004) to determine the effects of specific mutations in the DESE and PESE enhancers on the expression of a *lacZ* reporter gene in the Drosophila embryo, i.e. within a normal physiological context. One example of an informative reporter gene is the *slp1[PESE:ΔC1]-lacZ* construct described in Prazak et al. This construct is deleted for a 155 bp region near to the middle of a 2.1 kb segment of DNA containing the PESE enhancer and fails to express l*acZ* mRNA in late blastoderm stage embryos (Prazak et al., 2010). ChIP experiments can be carried out to compare H3K4me1 and Nejire association in embryos with the wild-type *PESE-lacZ* and *slp1[PESE:ΔC1]-lacZ* reporter genes using primer pairs specific for these constructs. It will also be interesting to determine whether these (and other) reporter genes respond to genetic manipulations in the activity of Runt and other pair-rule transcription factors. For example, do *PESE-lacZ* reporters emulate the epigenetic

response of the endogenous *slp1* PESE enhancer to Runt, and if so how do different mutations in the PESE enhancer affect this response? These experiments have the potential to provide important new insights into the regulation of enhancer activity in a well-defined and physiologically relevant developmental context.

It will be very interesting to determine if there are any differences in the epigenetic status of the DESE enhancer in RO embryos where this enhancer is driving expression, versus in RF embryos where DESE-dependent expression is blocked at an elongation step and P-TEFb association with the *slp1* promoter is reduced. Furthermore, Eve expression in cell type I represses activation of *slp1* by PESE at the elongation step, the same mechanism as DESE mediated repression in cell type III. Ectopic *eve* expression, while not able to convert all cells to Type I as DESE is not repressed by Eve, will convert Type IV cells (PESE active) to Type I (PESE repressed) there may be differences in signatures associated with PESE mediated repression of elongation compared to wild type.

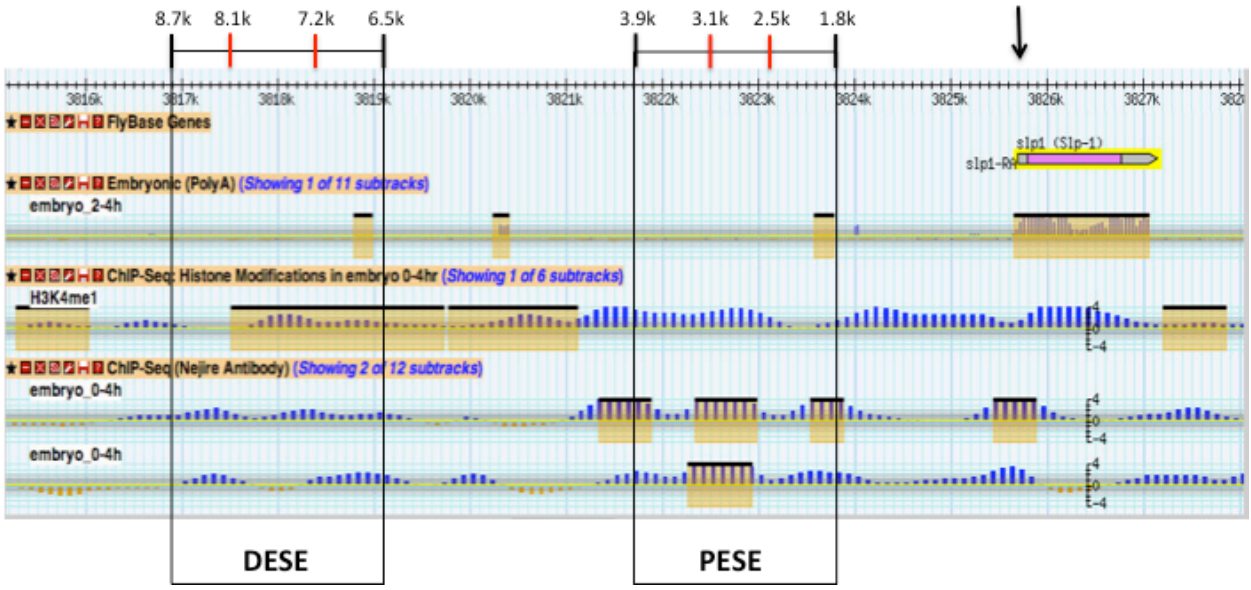

Fig 4-1. H3K4me1 and eRNA output at PESE and DESE

A screenshot of the *slp1* locus from the UCSC genome browser. Chosen tracks show peaks at PESE and DESE of H3K3Me1 (0-4h) and the output of eRNAs (2-4h). Peaks are represented by blue bars. For Nejire association, yellow boxes topped with black bars indicate maximum detectable association. A black box indicates the boundaries of DESE and PESE, the red lines on the top border of the boundaries represent the minimal regions required for PESE and DESE activity.

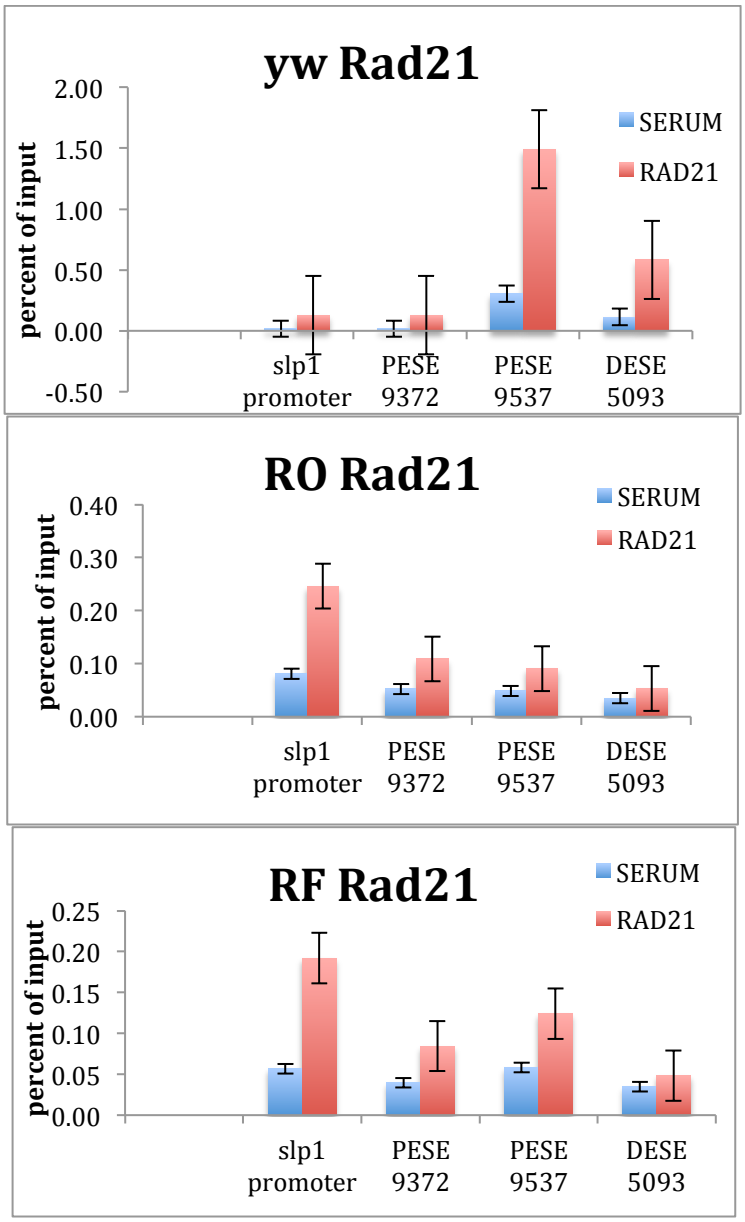

Figure 4-2. Rad21 association at the *slp1* promoter, PESE and DESE

Chromatin immunoprecipitation with antibody for Rad21 (AbCam) on chromatin isolated from 3.5-4.5 hour old embryos from 3 genotypes: yw, NGT>RO, and NGT>RF. Each data point for *yw* represents 3 replicates (except for PESE 9537) and RO and RF represent 2 replicates Primer pairs detected one region from the *slp1* promoter, 2 regions within PESE and one within DESE.

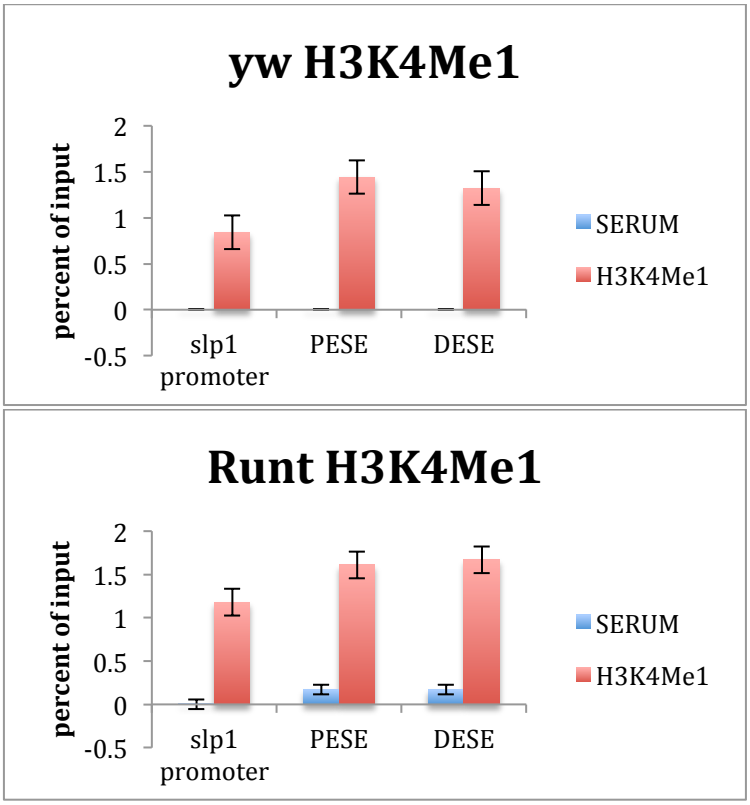

Fig. 4-3. H3KMe1 association at the *slp1* promoter, PESE and DESE

Chromatin immunoprecipitation with antibody for H3Kme1 (AbCam) on chromatin isolated from 3.5-4.5 hour old embryos from 2 genotypes: yw, NGT>Runt. Each data point represents the average of 3 technical replicates. Primers are the same as in figure 4-2 (PESE 9537).

Acknowledgement: Brian Kinney and Anthony Rampello provided technical assistance for ChIP experiments. Karthik Rao provided technical assistance for RT-qPCR experiments.

Chapter 5

General Discussion and Future Perspectives

Drosophila melanogaster shares 70% of disease related genes with humans, making fruit flies an excellent model for studying the underlying mechanisms of relevant disease states. For example, the human transcription factor RUNX1 or AML1 (acute myeloid leukemia 1) is closely related to Drosophila Runt. AML1 is responsible for regulation the differentiation of hematopoietic stem cells into mature blood cells (Okuda et al., 2001). The t(8;21) translocation was implicated in AML (Miyoshi et al., 1991) and ultimately shown that this mutation causes aberrant transcription regulation, preventing maturation of white blood cells. RUNX 2 and 3 are also implicated in craniofacial abnormalities and gastric cancer, respectively (Zhang et al., 2000; Li et al., 2002). Successful transcription regulation is essential for proper development. It is carried out by a combination of trans-acting transcription factors which can activate or repress target genes and the *cis*-regulatory DNA regions which these factors interact with to control when and where genes are expressed, enhancers are one example. Enhancers are classically defined as *cis*-acting DNA regulatory elements that stimulate transcription, independent of their position and orientation with respect to the transcriptional initiation site [Banerji et al., 1981]. Importantly, transcription factors can also mediate repression by acting at cis-acting DNA regulatory elements. Based on these dual properties and their importance in developmental regulation of transcription, the term emerging to describe well defined regulatory DNA-elements is cis-regulatory module, or CRM. CRMs often contain clusters of different transcription factor binding sites and need to integrate the regulatory inputs to regulate activation or repression (Ben-Tabou de-Leon and Davidson, 2007). Moreover, the promoter for a particular gene typically must interact with multiple CRMs and interpret different combinatorial codes to properly initiate transcription.

In Drosophila, Runt plays a role as a master developmental regulator, and can mediate both activation and repression of transcription. It remains unclear by what molecular mechanism Runt is influencing transcription early in development. Transcriptional regulation of *slp1* gene illustrates this basic principle of how *cis*-regulatory elements and transcription factors control gene expression in a context dependent manner. Extensive evidence from the *slp1* gene reveals the relatively simple combination of pair-rule transcription factors acting at PESE and DESE. Moreover, the transcription factors acting at these CRMs mediate repression in different ways, even in the same cell context. As found for endogenous *slp1*, Runt and Ftz repress DESE by preventing the transition of the Pol II complex that has initiated transcription into an actively elongating complex. In contrast, the Runt-dependent repression of PESE involves blocking PESE-dependent recruitment of Pol II and the initiation of transcription. It was found that the mechanism of Eve-dependent repression in cell type I is the consequence of Eve blocking PESE-dependent release of a paused Pol II complex.

Experiments were designed to investigate Runt-dependent changes in enhancer promoter contacts for *slp1* and other segment polarity genes. Based on the model, shifting the percentage of cells of a particular context by manipulating Runt, should result in a change in DNA contacts. Preliminary data shows that interactions can be detected, but results among different experimental viewpoint were inconsistent. A slightly modified version of the Gheldof, et al. is being used to generate additional 4C-seq libraries.

Molecular signatures, such as H3K3Me1, Rad 21 association, CBP association, and eRNA output have been used to predict enhancer location in the genome. Determining whether these signatures are altered *in vivo* in a Runt dependent manner should provide additional insight into the molecular mechanisms underlying target gene regulation. Preliminary experiments show

that Rad21 association and H3K4Me1 can be detected, but that a higher resolution method is required to detect eRNA output. CBP binding is another signature indicated by modENCODE data, and future experiments will include ChIP with CBP/Nejire antibody.

When considering enhancer activity, the models of tracking and looping focus on enhancer contacts with the promoter and how transcription factors can influence the interactions. It is becoming recognized that long distance inter-chromosomal contacts, or loops, are common in transcription regulation (Dekker, 2008; Sanyal et al., 2013). Much work has been done regarding chromosome looping at the ß-globin locus, and other instances have been found in Drosophila (Pink et al., 2010; Splinter, 2006; Tolhuis et al., 2002) A great deal of evidence for trans-action of enhancers, or inter-chromosomal contacts, has also come from studies on the Bithorax complex (Drewell et al., 2002; Hopmann et al., 1995; Sexton et al., 2012). The development of chromosome conformation capture technology (Dekker et al., 2002) was key in demonstrating looping in the contexts mentioned above. More recently genome-wide studies such as ENCODE and modENCODE have shown that long-range looping contacts are quite common throughout both the human genome as well as Drosophila (Sanyal et al., 2013; Sexton et al., 2012).

Transcription factors binding can mediate these long range and looping interactions (Deng et al., 2012; Nolis et al., 2009), and the binding site architecture of a CRM may determine its activity. In contrast to the view of the enhancer as an information-processing unit, which has either an on of off state, a single, compact enhancer may serve as an information display, representing on and off states, at the same time and in the same nucleus. Multiple transcription factors can bind at the same time to influence enhancer positioning and contact with the promoter. The billboard model (Kulkarni and Arnosti, 2003; Arnosti and Kulkarni, 2005) is

favorable in light of both our observation that *slp1* enhancers PESE and DESE are repressed by different mechanisms in the same cell, reflecting enhancer competition with the promoter.

Non-additive interactions have been observed among primary and shadow enhancers, (Hong et al., 2008) and function in the regulation of the key patterning genes *knirps*, *hunchback*, and *snail* in developing Drosophila embryos. Additive interactions are observed for *knirps* enhancers, but non-additive, or sub-additive interactions are seen for *hunchback* and *snail*. Quantitative modeling of enhancer-promoter interactions suggests that weakly active enhancers function additively while strong enhancers behave sub-additively due to competition with the target promoter. (Bothma et al., 2015). This suggests that enhancer strength is one factor that can determine if it interacts with the promoter in a particular cell context.

Binding site arrangement can also influences enhancer activity and strength. The Otx-a enhancer, mediates gene expression in the neural plate of Ciona embryos. High-throughput analysis provides evidence that enhancer specificity depends on submaximal recognition motifs having reduced binding affinities or "suboptimization". Observed binding sites contain imperfect matches to consensus motifs and depending on their spacing in a particular enhancer, perfect matches mediate robust but ectopic expression. Changes in spacing of the sites alters activity as well (Farley et al., 2015). Mutations in the Runt binding sites of PESE and DESE affect enhancer activity (data not shown) and may also for *wg3911* and *wg2946.*

Taken together, one can consider that conferring phenotypic robustness in gene expression is a major consequence of having multiple enhancers. Enhancers for *shavenbaby (svb)* demonstrate this idea. At optimal temperatures for embryonic development, *svb* enhancer deficiency causes only minor defects in patterning of a subset of sensory bristles, referred to as trichomes. In embryos that develop at extreme temperatures, however, absence of these

secondary enhancers leads to extensive loss of trichomes. Finally, removal of one copy of *wg*, required for normal trichome patterning, causes a similar loss of trichomes, but only in flies lacking the secondary enhancers. These results support the hypothesis that secondary enhancers contribute to phenotypic robustness in the face of environmental and genetic variability. (Frankel et al., 2010). These single enhancer studies do not address how the different enhancers interact with the promoter and what influences those interactions.

It remains to be seen if differences in enhancer strength or binding site architecture contribute to the differential activities of PESE and DESE or *wg3911* and *wg2946.* The results described from the single enhancer studies above are only a part of the full scope of enhancer activity early in development. Similar to observations for *slp1* and *wg,* functional analysis of enhancers in other species finds provide additional examples of complex interactions between multiple *cis*-regulatory elements in development. In zebrafish, the *wnt8s* gene is expressed in phases just before gastrulation. Two CRMs were identified that respond to at least 3 known factors, as well as unidentified factors. It is observed that the presence of one CRM activity can inhibit the activity of the other (Narayanan & Lekven, 2012). Functional analysis of limb enhancers carried out by enhancer deletion, concludes that the observations seen from the deletions cannot be explained by just the actions of the 2 enhancers alone (Nolte et al., 2014).

The *wg* gene provided an additional developmental context to investigate if similar mechanisms are underlying transcription regulation of another segment polarity gene, and moreover to potentially identify the individual contributions of two separate enhancers to the overall *wg* pattern and the combinatorial codes acting at each. I have identified two CRMs capable of driving *wg* expression in the segmented region of the blastoderm embryo. Each of the CRMs, *wg3911* and *wg2946* can activate expression in all 14 stripes. Ectopic expression of

wg2946, and misbehaving activity of both CRMs with pair-rule gene manipulation are greatly reduced when both CRMs are present. These two CRMs are sufficient to modulate *wg* expression, but may be regulated by different mechanisms. It is proposed that *wg2946* is a secondary enhancer, but able to maintain *wg* expression in certain genetics backgrounds.

At the blastoderm stage, transcription regulation is quite dynamic. As shown in Figure 1-5, in just 3 hours post fertilization transcription has been taken over by the embryo, entire sets of patterning genes are activated and regulated in a dynamic manner. The establishment of the initial patterns of these genes is dependent solely on the transcription factors present in the early embryo, as epigenetic modifications are still being established. Just a few hours later, regulation is more stable and patterning is largely taken over by the trithorax and polycomb groups. These 2 classes of factors act to stably repress or activate targets, respectively (reviewed in Geisler & Paro, 2015). Polycomb response elements (PREs) are located throughout the Drosophila genome and in addition to enhancers, modulate stable repression (Maeda & Karch, 2011). As the embryo transitions from this highly dynamic induction of transcription to more stable regulation, it stands to reason that multiple enhancers acting in the same cell provide an intermediate and flexible level of regulation.

Unlike pair-rule genes, which are expressed throughout parasegments, segment polarity genes have two sets of stripes, one in the odd parasegment and one in the even parasegment. For pair-rule genes, such as *eve* stripe 2, there is a stripe specific enhancer- just 1 enhancer for that stripe, and the factors acting upon it are clearly identified. For *slp1* it is observed that overlapping transcription factors are acting at both PESE and DESE, which each contribute differently to the odd and even stripes. It was determined that the same set of transcription factors comprise two distinct combinatorial codes regulating PESE or DESE. Observations of *wg*

suggest a similar mode of regulation. A single CRM would not be able to provide the same integration of context dependent combinatorial cues that is possible with multiple CRMs. With the addition of complexity in expression and body pattern from the pair-rule to segment polarity transition, multiple enhancers are necessary so that the correct combinatorial code can be integrated at the promoter to drive accurate gene expression.

Runt likely regulates more target genes than we know and interacts with as yet unidentified factors. The Drosophila segmentation pathway provides us with a rich context to study gene expression, from basic genetics to genome-wide, high resolution molecular biology. Substantial evidence from *slp1* gene regulation and new insight into *wg* regulation, provide *in vivo* confirmation of predicted CRMs and insights into the mechanisms by which Runt may regulate an entire class of target genes. Genetic screens to uncover new Runt interacting factors will lead way to more biochemical and high-throughput experiments to fully understand regulation of target gene expression, particularly the molecular mechanisms and physical basis for the observed effects on gene expression. The more we know about how transcription regulation works early in development, the better we can understand disease states resulting from aberrant regulation.

References

Andrioli, L. P., L. A. Digiampietri, L. P. de Barros and A. Machado-Lima (2012). "Huckebein is part of a combinatorial repression code in the anterior blastoderm." Dev Biol 361(1): 177-185.

Akbari, OS, Bousum, A, Bae, E, Drewell, RA (2006). "Unraveling *cis*-regulatory mechanisms at the abdominal-A and Abdominal-B genes in the Drosophila bithorax complex." Dev Biol **293**(2): 294-304.

Banerjee S, Smallwood A, Lamond S, Campbell S, Nargund G. "Igf2/H19 imprinting control region (ICR): an insulator or a position-dependent silencer?" Scientific World Journal **1**:218-24.

Banerji J, Rusconi S, Schaffner W. (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27(**2 Pt 1): 299-308.

Ben-Tabou de-Leon S1, Davidson EH (2007). "Gene regulation: gene control network in development." Annu Rev Biophys Biomol Struct. **36**:191-212.

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 & Development **8**(1): 105-117.

Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-12.

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., Levine, M. (2015) "Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo." Elife 4:e07956. DOI: 10.7554/eLife.07956

Cadigan KM, Grossniklaus U, Gehring WJ. "Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments." Genes Dev. **8**(8):899-913.

Core, L. J., Lis, J. T (2008). "Transcription regulation through promoter-proximal pausing of RNA polymerase II." Science **319**(5781): 1791-92.

Dekker, J. (2008). "Gene regulation in the third dimension". Science **319**(5781):1793-94.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A, Blobel, G. A. (2012). "Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor". Cell **149**(6):1233-44.

Drewell, R. A., Bae, E., Burr, J., Lewis, E. B. (2002) "Transcription defines the embryonic domains of *cis*-regulatory activity at the Drosophila bithorax complex." Proc Natl Acad Sci U S A **99**(26):16853-58.

Duncan IW. (2002). "Transvection effects in Drosophila." Annu Rev Genet. **36**:521-56.

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

Engel N, Raval AK, Thorvaldsen JL, Bartolomei SM. (2008). "Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking." Hum Mol Genet. **17**(19):3021-9.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., Stern, D. L. (2010). "Phenotypic robustness conferred by apparently redundant transcriptional enhancers" Nature 466(7305): 490-3.

Fujioka, M., Jaynes, J. B. (2012). "Regulation of a duplicated locus: Drosophila sloppy paired is replete with functionally overlapping enhancers." Dev Biol **362**(2):309-19.

Gallo, S. M., Li, L., Hu, Z.,Halfon, M. S. (2006). "REDfly: a Regulatory Element Database for Drosophila." Bioinformatics **22**(3):381-3.

Geisler, S. J., Paro, R. (2015). “Trithorax and Polycomb group-dependent regulation: a tale of opposing activities.” Development **142**(17):2876-87.
Gergen, J. P. Butler, B. A. (1988) “Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis.” Genes & Development **2**(9):1179-1193.
Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., Furlong, E. E. (2014). “Enhancer loops appear stable during development and are associated with paused polymerase.” Nature **512**(7512):96-100.
Gheldof, N., Leleu, M., Noordermeer, D., Rougemont, J., Reymond, A. (2012). “Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method.” Methods Mol Biol **786**:211-25.
Groth AC, Fish M, Nusse R, Calos MP. (2004). “Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31.” Genetics **166**(4):1775-82.
Hang S, Higgins M Gergen JP, submitted. “Control of enhancer-promoter interactions by regulation of Pol II pausing.”
Hopmann R, Duncan D, Duncan I (1995).” Transvection in the iab-5,6,7 region of the bithorax complex of Drosophila: homology independent interactions in trans.” Genetics**139**(2):815-33.
Horsfield, J. A., Anagnostou, S. H. Hu, J. K., Cho, K. H., Geisler, R., Lieschke, G., Crosier, K. E., Crosier, P. S. (2007). “Cohesin-dependent regulation of Runx genes” Development **134**(14):2639-49.
Ingham PW. (1988) “The molecular genetics of embryonic pattern formation in Drosophila.” Nature **335**(6185):25-34.
Ingham PW, Hidalgo A. (1993). “Regulation of wingless transcription in the Drosophila embryo.” Development.**117**(1):283-91.”
Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M., Gergen, P. (1993). “The Runt domain identifies a new family of heteromeric transcriptional regulators.” Trends Genet **9**(10):338-41.
Kim, A., Dean, A. (2012). “Chromatin loop formation in the beta-globin locus and its role in globin gene transcription.” Mol Cells **34**(1):1-5
Klingler M1, Soong J, Butler B, Gergen JP. (1996). “Disperse versus compact elements for the regulation of runt stripes in Drosophila.” Dev Biol. **177**(1):73-84.
Klug A. (2005). “Gene regulatory proteins and their interaction with DNA.” Ann N Y Acad Sci. **758**:143-60.
Koch, F., Jourquin, F., Ferrier, P., Andrau, J. C. (2008). "Genome-wide RNA polymerase II: not genes only!" Trends Biochem Sci **33**(6): 265-73.
Kuhn, D. T., Chaverri, J. M., Persaud, D. A., Madjidi, A. (2000). "Pair-rule genes cooperate to activate en stripe 15 and refine its margins during germ band elongation in the D. melanogaster embryo." Mech Dev **95**(1-2):297-300.
La Rosée A1, Häder T, Taubert H, Rivera-Pomar R, Jäckle H. (1997). "Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the Drosophila embryo." EMBO J. **16**(14):4403-11.
Lessing D1, Nusse R. (1998). "Expression of wingless in the Drosophila embryo: a conserved *cis*-acting element lacking conserved Ci-binding sites is required for patched-mediated repression." Development **125**(8):1469-76.
Li, Q., Barkess, G., Qian, H. (2006). "Chromatin looping and the probability of transcription." Trends Genet **22**(4): 197-202.

Li QL, Ito K, Sakakura C, Fukamachi H, Inoue Ki, Chi XZ, Lee KY, Nomura S, Lee CW, Han SB, Kim HM, Kim WJ, Yamamoto H, Yamashita N, Yano T, Ikeda T,Itohara S, Inazawa J, Abe T, Hagiwara A, Yamagishi H, Ooe A, Kaneda A, Sugimura T, Ushijima T, Bae SC, Ito Y. (2002). “Causal relationship between the loss of RUNX3 expression and gastric cancer.” Cell 109(1):113-24.
MacArthur, S., Li, X. Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., Grondona, B. P.Hechmer, A., Simirenko, L., Keranen, S. V., Knowles, D. W., Stapleton, M. Bickel, P.Biggin, M. D., Eisen, M. B. (2009). "Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions." Genome Biol **10**(7): R80.
Maeda, R. K., Karch, F. (2011). “Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions.” Curr Opin Genet Dev **21**(2):187-93.
Manoukian, A. S., Krause, H. M. (1993). "Concentration-dependent activities of the even-skipped protein in Drosophila embryos". Genes Dev **6**(9):1740-51.
Miskiewicz, P., Morrissey, D., Lan, Y., Raj, L., Kessler, S., Fujioka, M., Goto, T.Weir, M. (1996). " "Both the paired domain and homeodomain are required for in vivo function of Drosophila Paired". Development **122**(9):2709-18.
Miyazono K, Maeda S, Imamura T. (2004). "Coordinate regulation of cell growth and differentiation by TGF-beta superfamily and Runx proteins." Oncogene. **23**(24):4232-7.
Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., Ohki, M. (1991). "t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1." Proc Natl Acad Sci U S A **88**(23):10431-4.
Morrissey, D., Askew, D., Raj, L., Weir, M. (1991). "Functional dissection of the paired segmentation gene in Drosophila embryos". Genes Dev **5**(9):1684-96.
Mousavi, K., Zare, H., Koulnis, M., Sartorelli, V. (2014). “The emerging roles of eRNAs in transcriptional regulatory networks.” RNA Biol **11**(2):106-110.
Mullen, J. R., DiNardo, S. (1995). " Establishing parasegments in Drosophila embryos: roles of the odd-skipped and naked genes." Dev Biol **169**(1): 295-308.
Narayanan, A, Lekven, C. (2012). “Biphasic wnt8a Expression Is Achieved Through Interactions of Multiple Regulatory Inputs.” Dev Dyn **241**(6):1062-75.
Natoli G, Andrau JC. (2012). "Noncoding transcription at enhancers: general principles and functional models." Annu Rev Genet. **46**:1-19
Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J. R. Rushlow, C. (2011). “Temporal coordination of gene networks by Zelda in the early Drosophila embryo.” PLoS Genet **7**(10):e1002339.
Nolis, I. K., McKay, D. J., Mantouvalou, E., Lomvardas, S., Merika, M., Thanos, D. (2009). "Transcription factors mediate long-range enhancer-promoter interactions." Proc Natl Acad Sci U S A **106**(48):20222-7.
Nolte, M. J., Wang, Y., Deng, J. M., Swinton, P. G., Wei, C., Guindani, M., Schwartz, R. J., Behringer, R. R. (2014). “Functional analysis of limb transcriptional enhancers in the mouse.” Evol Dev **16**(4):207-23.
Noordermeer D, Branco MR, Splinter E, Klous P, van Ijcken W, Swagemakers S, Koutsourakis M, van der Spek P, Pombo A, de Laat W. (2008). "Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region."PLoS Genet. 4(3):e1000016.

Okuda, T., Nishimura, M., Nakao, M., Fujita, Y. (2001)." RUNX1/AML1: a central player in hematopoiesis." Int J Hematol **74**(3):252-7.
Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R. (1998). "Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis." EMBO J. **17**(17):5141-50.
Paris, M., Kaplan, T.,Li, X. Y., Villalta, J. E., Lott, S. E. Eisen, M. B. (2013). "Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression." PLoS Genet **9**(9):e1003748.
Pereira PS, Pinho S, Johnson K, Couso JP, Casares F. (2008). "A 3' cis-regulatory region controls wingless expression in the Drosophila eye and leg primordial." Dev Dyn. **235**(1):225-34.
Perry, M, Boettiger ,AN and Levine, M. (2011). "Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo" Proc Natl Acad Sci U S A **108**(33): 13570–13575.
Prazak, L., M. Fujioka and J. P. Gergen (2010). "Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors." Dev Biol **344**(2): 1048-59.
Sanyal, A., Lajoie, B. R., Jain, G., Dekker, J. (2013). "The long-range interaction landscape of gene promoters". Nature **489**(7414):109-13.
Schaaf, C. A., Kwak, H., Koenig, A., Misulovin, Z., Gohara, D. W., Watson, A., Zhou, Y., Lis, J. T., Dorsett, D.(2013). "Genome-wide control of RNA polymerase II activity by cohesin." PLoS Genet **9**(3):e1003382.
Schaukowitch, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., Kim, T. K. (2014). "Enhancer RNA facilitates NELF release from immediate early genes." Mol Cell **56**(1):29-42.
Sexton T, Kurukuti S, Mitchell JA, Umlauf D, Nagano T, Fraser P. (2012). "Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip." Nat Protoc. **7**(7):1335-50
Small S, Blair A, Levine M. (1992). "Regulation of even-skipped stripe 2 in the Drosophila embryo." EMBO J. **11**(11):4047-57.
Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N.de Laat, W. (2006). "CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus" Genes Dev. **20**(17):2349-54.
Swantek, D. Gergen, J. P. (2004). "Ftz modulates Runt-dependent activation and repression of segment-polarity gene transcription." Development **131**(10):2281-90.
Tamkun, J. W. (2007). "Stalled polymerases and transcriptional regulation." Nat Genet **39**(12): 1421-22.
Tolhuis B[1], Palstra RJ, Splinter E, Grosveld F, de Laat W. (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell.**10**(6):1453-65.
Tracey, W. D., Jr., 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., Gergen, J. P. (1994)."Gap gene properties of the pair-rule gene runt during Drosophila segmentation." Development **120**(6):1671-83.
Tugrul, M., Paixao, T., Barton, N. H., Tkacik, G. (2015). "Dynamics of Transcription Factor Binding Site Evolution." PLoS Genet **11**(11):e1005639.
Wang, X., Lee, C., Gilmour, D. S., Gergen, J. P. (2007). "Transcription elongation controls cell fate specification in the Drosophila embryo." Genes Dev **21**(9):1031-6.

Wang, X., Hang, S., Prazak, L., Gergen, J. P. (2010). “NELF potentiates gene transcription in the Drosophila embryo.” PLoS One **5**(7):e11498.
Zhang YW, Yasui N, Ito K, Huang G, Fujii M, Hanai J, Nogami H, Ochi T, Miyazono K, Ito Y (2000). "A RUNX2/PEBP2alpha A/CBFA1 mutation displaying impaired transactivation and Smad interaction in cleidocranial dysplasia". PNAS **97**(19):10549–54.

Appendix

Primers used in 4C-seq experiments

IPCRFORADAPT	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGAT CTGAAAAATGGTGAAAAGTGAAATGG
IPCRREVADAPT	CAAGCAGAAGACGGCATACGAGCATCAGGTAGTTGGGCAGT
BAITREV4CLIG	CTAGAATATTGGATTGGTGCTTTG
5kbnewF	CAAAGGTGTCAACCAAACTGTG
5kbnewR	GCAGCAACAGGATATGAAAGG
8kbnewF	GAGAGGGAGACGACTTTTCATTC
8kbnewR	GGGGCGAGAGTACTTTACTAATCA
20kbnewF	CCTCCATTTTCCATGAGC
20kbnewR	CGAAGACCACCTTGATTTTCTG

DESEEcoRI5199F	GCAAGACGCTTTCCAAAATC
DESEEcoRI5199R	GGTCCTTGCACTTGGAAAAA
DESECsp6I5657F	CGTGTCTGCCCTTTCATTTT
DESECsp6I5657R	CTTTCAGTGCAGCCAGGATT
PESEEcoRI9537F	ACCCACAGTGGAACGAAAAC
PESEEcoRI9537R	TACGCCTGCCTCATTAGCTC
PESEMseI9855F	CGATTATGCTCAAGGTGTGC
PESEMseI9855R	TCAACTGCAAGGTGTGCCTA
PESE2Csp6I9190F	TGACAGAACTCAGCGTTTCG
PESE2Csp6I9190R	CATCCTCATCCTCGCACTTT
Slp1uncut8100F	GATTTCACACGCCCAGAAAG
Slp1uncut8100R	ACGATGCTAGGCTGTGTGTG
Slp2proMseI23575F	CGACTGCGATTGGGAATTAT
Slp2proMseI23575R	GCTATTCAAATTGGGGCTCA
Slp2proEcoRI23499F	GGAAAGAGAGTGCGCTGAGT
Slp2proEcoRI23499R	CGACGCTCGACGTTAATACA
Slp2pro2Csp623221F	TCAAAAGAGCAAGGGGAAAA
Slp2pro2Csp623221R	CCGACATGAAACACGAAAAA
Slp2uncut24200F	TTTCCGCCCAACCATAATAA
Slp2uncut24200R	GGTTGTCTCCCCATCACATT
TwiproMseI1583F	AATCCTCGCACAGCAAAGTT
TwiproMseI1583R	TGTTGTTGTTTGCAGCGTTAG
TwiproBamHI2207F	CGCTCGAGAAAATCGAAATC
TwiproBamHI2207R	ATCTGCTCGCACGCACTTAC
Twiuncut1900F	ATCCCAACACGCATACTTCC
Twiuncut1900R	GGGGATTTCGATTTTCTCG
Wg3911BamHI7234F	ACAAAGACACGACACGCTCA
Wg3911BamHI7234R	GCAGGGTGTGAAAATTGCTT

Wg3911MseI8646F	CCTACAGCAGCGAGAGGTTT
Wg3911MseI8326R	CCTGAAAAATTAGCGCGTTC
Wg2946MseI13977F	GTGTCCCTGAGTCCCAGTGT
Wg2946MseI13977R	GCAGGACGAAAATCAGAAGC
Wg2946EcoRI14699F	AAAACGAAACCACCAACGAC
Wg2946EcoRI14699R	GGTGTTGATTGACGCATCTG
Wguncut12750F	TAACCCAGCTCGAATCCAGT
Wguncut12750R	GCCCCTTCAAAAGGGTAGTC
Enpro/regRI7927F	GCCCAGCAATTTTGAGGTAA
Enpro/regRI7927R	CACATTACGCTGCATGAAAA
Enpro/regCsp11818F	GGAATGGAGTCCTCGGATG
Enpro/regCsp11818R	ACGAACCTTTGGGACTTCCT
EnregEcoRI12879F	TCGCCGGAGTTCTACTCTTC
EnregEcoRI12879R	CGTCGTTGGTCTTGTCCTTT
Enuncut9900F	AGCCATTTTCCTGGCCTACT
Enuncut9900R	TCCCTCTCGCTCTCACTCTC
GsbproCsp6I5831F	CACTGCTCTCGCTCACACTT
GsbproCsp6I5831R	AGCCGCAAGACTAAAACGAA
GsbproEcoRI6069F	CAGTTTCGAGCTGTCAAGCA
GsbproEcoRI6069R	TCCTTGCGGTAATCCTTTTG
Gsbuncut6800F	GCATCTGCTGCTCTTATGGA
Gsbuncut6800R	TTGAAGATTGGCACAAACGA
PrdregEcoRI5014F	TTGCTTGCCACGGTGTATTA
PrdregEcoRI5014R	GGCCTTGATTGTTTCACGTT
PrdregMseI5586F	TGAATGAATGCGCATGGTTA
PrdregMseI5586R	TGCTGTGTTCATTTGACCTTTC
Prduncut6000F	GTAGGGTCGCAGGTAAACGA
Prduncut6000R	CCATAAACGCTGCAAGTCAG
OddregBamHI516F	CCAAATTTCGAGTTGGCTGT
OddregBamHI516R	TTCCCCTCCCTACCAAAAAC
OddregMseI1241F	AGTCAGCCGAAGGTTGAGAA
OddregMseI1241R	TCGCTAATGGTCACAGCAAG
Odduncut4950F	AGCCCTAAGCTCGTTGGATT
Odduncut4950R	AATCCAGGAAGGGTTGATCC
WgCsp6I1009F	ACGCTCTCGGAGATGAAATG
WgCsp6I1009R	AGACGGAGGCCCATAACTTT
Slp1BamHI11855F	TGTCATCCCCATTGAAGTGA
Slp1BamHI11855R	CCGCTCGCTGTTTTGACTAT
Slp1Csp6I3505F	GCGACATACAAGTGGCGATA
Slp1Csp6I3505R	AACACTAGAGGCGCCAAGAA
Slp2Csp6I25002F	GACCACCGGGACCTCAAG
Slp2Csp6I25002R	ACCGAGAGTTGCCTTTGATG
Slp2BamHI24636F	GGCTGGCAGAACTCCATAAG
Slp2BamHI24636R	CTGTGGTTCTCCTCCTCAGC

Slp1BamHI4554F	CAACGCCCAAACCAAAATAC
Slp1BamHI4554R	CCTGGTTAACCGTACGCAGT
slp1proPEfor	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGAAAAATGGTGAAAAGTGAAATGG
slp1proPErev	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGCATCAGGTAGTTGGGCAGT
slp1prorightnew	AATGGTGAAAAGTGAAATGGAAT
PESEleft	CTATGGAGATGCATAAGCGAAT
PESEright	GAATCGCTAATCAATTCGTTTTAC
PESE2left	GCCTTTCAGTCATTGTCCTGT
PESE2right	GCCACGGTGTCTACTTGGAAT
DESEleft	TCCACGTGCATCCTTTAGAA
DESEright	TTAACCGACTTTCGAACACATC
slp2proleft	TTGAATCCGGCAGCTGTTA
slp2proright	GGATTGGGAATTATTGGGAATT
slp2pro2left	ACGCACACAAATAAACACAAACTC
twiproleft	ATGTTGTTGTTTGCAGCGTTA
twiproright	CAATGGCCGGAAGGAT
wg3911left	GGGATCGGTATCGGGATC
wg3911right1	CTTTTATGTCCTACTATTTCTTTGATTTG
wg2946left3	CGGGAGTACAGCTCATCTCG
wg2946right	CCAAATTATTTCGTGGTCGAAT
enproregleft	TGCGATCATATCAACGGAAT
enproregright	GAGACGGGATCCACCAC
enregleft	TGACCCACTTAAACGCTAGGTAA
enregright	CTCCGGCCGAGGAAT
gsbproleft	GAGAGGCTGCGATCGGTA
gsbproright	CAGATACACACTCCAAAAATAATTTAGAAT
prdregleft	CATGATTTCCAGAAATTAGGGAAT
prdregright	AACTGCATGAAAGGTGTTGC
oddregleft	GAATCAAGTAATGGACAAACAGGAT
oddregright4	GAACCTTCTGTATTTTCAAATTTCC
Slp1pronewPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAATGGTGAAAAGTGAAATGGAAT
PESEPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCTATGGAGATGCATAAGCGAAT
PESEPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGAATCGCTAATCAATTCGTTTTAC
PESE2PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGCCTTTCAGTCATTGTCCTGT
PESE2PE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGCCACGGTGTCTACTTGGAAT
DESEPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTCCACGTGCATCCTTTAGAA
DESEPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTTAACCGACTTTCGAACACATC
Slp2proPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTTGAATCCGGCAGCTGTTA

Slp2proPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGGATTGGGAATTATTGGGAATT
Slp2pro2PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTACGCACACAAATAAACACAAACTC
Slp2pro2PE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCGATTGGGAATTATTGGGAATT
TwiproPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTATGTTGTTGTTTGCAGCGTTA
TwiproPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCAATGGCCGGAAGGAT
wg3911PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGGGATCGGTATCGGGATC
wg3911PE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCTTTTATGTCCTACTATTTCTTTGATTTG
wg2946PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCGGGAGTACAGCTCATCTCG
wg2946PE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCCAAATTATTTCGTGGTCGAAT
EnproregPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTGCGATCATATCAACGGAAT
EnproregPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGAGACGGGATCCACCAC
EnregPE1	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCTCCGGCCGAGGAAT
GsbproPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGAGAGGCTGCGATCGGTA
GsbproPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCAGATACACACTCCAAAAATAATTTAGAAT
PrdregPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCATGATTTCCAGAAATTAGGGAAT
PrdregPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAACTGCATGAAAGGTGTTGC
OddregPE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGAATCAAGTAATGGACAAACAGGAT
OddregPE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGAACCTTCTGTATTTTCAAATTTCC
slp1pronewPE1l	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGCATCAGGTAGTTGGGCAGT
Slp1proPE1old	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGAT CTGAAAAATGGTGAAAAGTGAAATGG
slp1proLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTAACACCTAGCATCAGGTAGTTGGGCAGT
slp1proLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTACTAACTGGCATCAGGTAGTTGGGCAGT
slp1proLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTATCGCCAGGCATCAGGTAGTTGGGCAGT
slp1proRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCATTCCAAAATGGTGAAAAGTGAAATGGAAT
slp1proRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCGCATTAAAATGGTGAAAAGTGAAATGGAAT
slp1proRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCTTCGCGCAATGGTGAAAAGTGAAATGGAAT
PESELyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTACGTAGCTCTATGGAGATGCATAAGCGAAT
PESELRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTATCCTATTCTATGGAGATGCATAAGCGAAT
PESELRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCAGGAGGCCTATGGAGATGCATAAGCGAAT

PESERyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCGACTGGGGAATCGCTAATCAATTCGTTTTAC
PESERRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCTTAAGATGAATCGCTAATCAATTCGTTTTAC
PESERRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGCAAGTAGGAATCGCTAATCAATTCGTTTTAC
PESE2Lyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTATATAGGAGCCTTTCAGTCATTGTCCTGT
PESE2LRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCACGTGTTGCCTTTCAGTCATTGTCCTGT
PESE2LRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCGAACTGTGCCTTTCAGTCATTGTCCTGT
PESE2Ryw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCTCTGTCTGCCACGGTGTCTACTTGGAAT
PESE2RRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGATGGAATGCCACGGTGTCTACTTGGAAT
PESE2RRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGGCAGACGGCCACGGTGTCTACTTGGAAT
DESELyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCACAGTTGTCCACGTGCATCCTTTAGAA
DESELRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCCTTTACATCCACGTGCATCCTTTAGAA
DESELRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCTAGTCATTCCACGTGCATCCTTTAGAA
DESERyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGATCCAGCTTAACCGACTTTCGAACACATC
DESERRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGGATATGGTTAACCGACTTTCGAACACATC
DESERRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGTGACTACTTAACCGACTTTCGAACACATC
slp2proLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCCTACAACTTGAATCCGGCAGCTG
slp2proLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCTAGATTCTTGAATCCGGCAGCTG
slp2proLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGAGTTAACTTGAATCCGGCAGCTG
slp2proRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGGACGAGAGGATTGGGAATTATTGGGAATT
slp2proRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGTCTACATGGATTGGGAATTATTGGGAATT
slp2proRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTATACCGTGGATTGGGAATTATTGGGAATT
twiproLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTCGTCGGCTATGTTGTTGTTTGCAGCGTTA
twiproLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGAGAACTCATGTTGTTGTTTGCAGCGTTA
twiproLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGCTGGCGAATGTTGTTGTTTGCAGCGTTA
twiproRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTGTCCATTACAATGGCCGGAAGGAT
twiproRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTAGTCACACAATGGCCGGAAGGAT
twiproRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTGACGCATCAATGGCCGGAAGGAT
wg3911Lyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGACGTCAAGGGATCGGTATCGGGATC
wg3911LRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGCTCAGTTGGGATCGGTATCGGGATC

wg3911LRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGTAGAGCTGGGATCGGTATCGGGATC
wg3911Ryw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTAGCTAGTCTTTTATGTCCTACTATTTCTTTGATTTG
wg3911RRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTGAATTCGCTTTTATGTCCTACTATTTCTTTGATTTG
wg3911RRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTTCCTCACCTTTTATGTCCTACTATTTCTTTGATTTG
wg2946Lyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGCGTTTCGCGGGAGTACAGCTCATCTCG
wg2946LRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGTACTTGCCGGGAGTACAGCTCATCTCG
wg2946LRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTACTGCGCCGGGAGTACAGCTCATCTCG
wg2946Ryw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTCGGTACCCCAAATTATTTCGTGGTCGAAT
wg2946RRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTTAAACAGCCAAATTATTTCGTGGTCGAAT
wg2946RRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAATGCTGACCAAATTATTTCGTGGTCGAAT
enproLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGGTCTGACTGCGATCATATCAACGGAAT
enproLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTACGAATCTGCGATCATATCAACGGAAT
enproLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTCGCGTACTGCGATCATATCAACGGAAT
enproRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTTGTGCTATGAGACGGGATCCACCAC
enproRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAATCACACGAGACGGGATCCACCAC
enproRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAGGTCAGTGAGACGGGATCCACCAC
gsbproLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTGTTTCACTGAGAGGCTGCGATCGGTA
gsbproLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTCCTACTAGAGAGGCTGCGATCGGTA
gsbproLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTGTAGGTCGAGAGGCTGCGATCGGTA
gsbproRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAAGATTGCCAGATACACACTCCAAAAATAATTTAGAAT
gsbproRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAGGCAATGCAGATACACACTCCAAAAATAATTTAGAAT
gsbproRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTATTGCATCCAGATACACACTCCAAAAATAATTTAGAAT
prdregLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTCCAGCCTCATGATTTCCAGAAATTAGGGAAT
prdregLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTGGTCTTCCATGATTTCCAGAAATTAGGGAAT
prdregLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTAAGCGGTCCATGATTTCCAGAAATTAGGGAAT
prdregRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTAGAACACCAACTGCATGAAAGGTGTTGC
prdregRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTATGCATCCAACTGCATGAAAGGTGTTGC
prdregRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCCATACACAACTGCATGAAAGGTGTTGC
oddregLyw	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTTGCGGTTAGAATCAAGTAATGGACAAACAGGAT

oddregLRO	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTAACCGTGTGAATCAAGTAATGGACAAACAGGAT
oddregLRF	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTACTCTAAGGAATCAAGTAATGGACAAACAGGAT
oddregRyw	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTATGAGGAAGAACCTTCTGTATTTTCAAATTTCC
oddregRRO	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCCAGCACGGAACCTTCTGTATTTTCAAATTTCC
oddregRRF	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC TCTTCCGATCTCGCTTCTGGAACCTTCTGTATTTTCAAATTTCC

slp1pro1L	CTGGGTTTTTTTTTTCGATC
slp1pro1R	GTTTGCTCACCAACTTTTGTAC
slp1pro2L	TTTTCAAGTGTCGATAACTCGATC
slp1pro2R	CTGGGCACCGCAGGTAC
PESE1L	GATTTATGCCTCGGCGATC
PESE1R	CGAATTTCTTTTATTTCGTCCATG
PESE2L	ATTGCAGTGCCGTGGATC
PESE2R	CAAATATTTGCCAGAGTCCATG
DESE1/2L1	GGACGGCAGCGGAGAT
DESE1/2L2	AAATTCCTTGGACGGCAG
DESE1R	GCAAGGACCAATTATACCCTGT
DESE2R	TACTTTTTGTTTCCTTTTCCGTAC
slp2pro1L	GCTTGTTTGTCCAACGGATC
slp2pro1R	GCCTGCTATAAATCTGGCTGATTAG
slp2pro2L	CTTCCTGTATAGCAACATTAAGATC
slp2pro2R	AAGGGTGACCCACTACATG
wgpro1L	GGTCATAGGTCAGCACCGATC
wgpro1R	TCTGCCTGCTGTCCAGGTAC
wgpro2L	TCTTTCCACTTGACTTGAGATC
wgpro2R	ACATTACGCCCATCATGTAC
wg3911-1L	AGGTCCGCGTTGGGATC
wg3911-2R	GGCTTATTTTCAAATCTCACCCA
wg3911-2L	CCAGGTCGCGACTCGATC
wg3911-2R	GCTACCCTTACGAATCCCCG
wg2946-1L	TGTTCCTCCGACCCGATC
wg2946-1R	TGTAAACAAAAGTCGCGCATG
wg2946-2L	GTGGTTCGCTTGTGGAGATC
wg2946-2R	CATATGCTCTTTTCGCAGTGTAC
slp1pro1DpnL	ATTGAACGCCGAGTTGTTTC
slp1pro1DpnR	CAATGCGATGATGACCAAAG
slp1pro1CspL	AGTCGATGGTTAACGGCAAT
slp1pro1CspR	TGTCGATAACTCGATCCAAAAA
slp1pro2DpnL	CGAGAGCCCTTGTAGTTTGC
slp1pro2DpnR	ACGGAAAAAGAAGCTGCTGA
slp1pro2CspL	AGAGCGGCAAACATGAAAAT
slp1pro2CspR	AGCCGGGGCTATATGGGATA

PESE1NlaL	GCGGCTGTTTCTTTTTGGTA
PESE1NlaR	GTGGCCGCATATATCCTTTG
PESE2NlaL	AAAATCTTAAACACACACACACAAA
PESE2NlaR	GAGCACCGGGATTTGTTCTA
DESE1DpnL	ATCCTGCCGCGTGGATAC
DESE1DpnR	CCCAAAACTCAGGTGCTGTT
DESE1NlaL	GGCTGGTCCTTTAGGCAAA
DESE1NlaR	TTCAAATTGGAAAATGTCTTCG
DESE2DpnL	ATCCTGCCGCGTGGATAC
DESE2DpnR	CCCAAAACTCAGGTGCTGTT
DESE2CspL	TGATGTTGCACGTTCATCTG
DESE2CspR	TCTTTCAGTGCAGCCAGGAT
slp2pro1DpnL	CGACTGCGACTGCATGTAAA
slp2pro1DpnR	GTTCTTGGCCGTAAAGTTGC
slp2pro1CspL	GCAACTTTACGGCCAAGAAC
slp2pro1CspR	ATTGATCAAAACGCCGAAAA
slp2pro2NlaL	ACGGTCCGGGACACAATTA
slp2pro2NlaR	GCCACGCACACAAATAAACA
wgpro1DpnL	TGAGAAACAGGAGGCCCATA
wgpro1DpnR	ATCTCCCAATTTTCGCCTTT
wgpro1CspL	GGAGATGGAGTCGGAGTCAC
wgpro1CspR	TCCACATCCATTTCCTAGCAC
wgpro2DpnL	CGCCGGCAATTGACGTAT
wgpro2DpnR	ATGCTTGGCTGAGTTTTTCG
wgpro2CspL	ATCGTTCAGCTGGTGGAAAT
wgpro2CspR	CGTCGCTGTTTCCTTCTCA
wg3911-1DpnL	CGGACAACCATTTTCGGTAG
wg3911-1DpnR	CCACCCCTCCCACTACCTC
wg3911-1NlaL	GACCACCCGACCACCATC
wg3911-1NlaR	CACTCCCACTCCGGCTAAG
wg3911-2DpnL	GCTGGATTTCCTGCCACTAC
wg3911-2DpnR	GGATTTCGATTGGATGGATG
wg3911-2NlaL	CCCCCATTCGCTATGGTC
wg3911-2NlaR	GGATTTCCAGGAAGATTACACAA
wg2946-1NlaL	TTGTCACTCACTCAGTCATTCAA
wg2946-1NlaR	TCAGTTTCAGTTGCAGGATCTC
wg2946-2DpnL	GTGGACATTCCACTCGCTTA
wg2946-2DpnR	CCTCGTTGTTGTGCAGATTC
wg2946-2CspL	CGAGCGGTAAGTTTAGATACCC
wg2946-2CspR	AGGGGGAGCGAAAGAGAGA
slp1proforL	CGAGAGCCCTTGTAGTTTGC
slp1proF2kbRR	CTCAGTGTTGGCTGGCAGT
slp1proR2kbFL	AGCCTACAAGTATTATCGCATTCT
slp1prorevR	GCGATGATGACCAAAGGAAT

slp1proF10kbRR	GTGTCCTGGCCATTTTCATT
slp1proR10KbFL	CCAGCTTTTCCATTAATTCCA
PESEforL	TCGGTTTCCGAATTTCTTTT
PESEF2kbRR	AGTCGTCTTTTCGCCCTCAT
PESER2kbFL	AGTGCGAGGATGAGGATGAG
PESErevR	GCTGCCAACAAATGTCTCAA
PESEF10kbRR	TTTTTCCAAAGAAACATCTGAAG
PESER10kbFL	TGTGGCAGAGCTCTATTTACCA
DESEforL	TTTCCCCGATTCAAAAGACA
DESEF2kbRR	TTAAGTTGTCCATGCGGTTG
DESER2kbFL	CCGCATGGACAACTTAATCA
DESErevR	TGCAGTGAGTCCAGAAGGAG
DESEF10kbRR	ACCACAAGGACCCGCAAG
DESER10kbFL	GGTCCTTGTGGTCCTCTGAA
wgproforL	CATAATTAGGCGAAAAACTCAGC
wgproF2kbRR	TTGTATCGTAAATTTGGCTGTTT
wgproR2kbF	CGAAACAGCCAAATTTACGA
wgprorevR	GCAATCTTCGGAAAGAGGTT
wgproF10kbRR	AAACAAAACTAATTAGGCATAAAGTGC
wgproR10kbFL	CACGCACTTTATGCCTAATTAGTTT
wg3911forL	AAGTTGGCCATTGGTGGTTA
wg3911F2kbR	CGCGGAAAATCCAAATGA
wg3911R2kbFL	ATCGATTACACCGAAAATGC
wg3911revR	CCAATAAGGTGAGAGGTTTGG
wg3911F10kbRR	TGTTGTTGCAACTTGCTGTG
wg3911R10kbFL	GCGAGTCGATACGCAAGAAG
wg2946forL	CAGCAAAATCCCAGGAGAAC
wg2946F2kbRR	TTTTTAAGCGAGTGGAATGTC
wg2946R2kbFL	CCGAGAGACGAGCTTCATTT
wg2946revR	GGTTCAGTGGTTCCAGTGGT
wg2946F10kbRR	GGCCAACAAAAACAACGAG
wg2946R10kbFL	GGTGGGTGCGATTTGATAAT

Primers used in ChIP and eRNA experiments

DESEeRNAF0	TTGCACGTTCATCTGCGACT
DESEeRNAF2	GCACGTTCATCTGCGACTT
DESEeRNAF18	ACTTGCCTCGCTCGACTC
DESEeRNART5	TTTTTCGTACGGAAAAGGAAAC
DESEeRNAR0	CGTACGGAAAAGGAAACAAAA
DESEeRNAR34	GCCGATGTGTTCGAAAGTC
PESEeRNAF0	AGCAAAAGAATAAAAAGAGAGAAAAAGG
PESEeRNAF2	CAAAAGAATAAAAAGAGAGAAAAAGG
PESEeRNAF21	AAAAAGGGAAGCCGCAAA

PESEeRNART3	TTTCAATCGAGCGATAGAGAAAGA
PESEeRNAR0	CAATCGAGCGATAGAGAAAGAAAG
PESEeRNAR2	TCGAGCGATAGAGAAAGAAAGG
PESEeRNAR20	AAAGGTATAGGCAGTCGAAAAAAGTA
Rp49F	TACAGGCCCAAGATCGTGAA
Rp49R	TCTCCTTGCGCTTCTTGGA
slp1-prmtr_F(20007)	GGGCTCTCTTCGTGTAGACTTCGT
slp1-prmtr_R(20010)	GGAGAAGTTGCTCTTGAATTCCATT
PESE 9372F	TTGACAGCGAGGTTCCTCAA
PESE 9372R	CCTGCCTCATTAGCTCACAAA
PESE 9537F	GAATTCGCTTATGCATCTCCATA
PESE 9537R	GAGCATAATCGAGCGTGGAT
DESE 5093F	GGCAAGACGCTTTCCAAA
DESE 5093R	GGTATAATTGCTCCTTGCACTTG