ABSTRACT

Title of Dissertation: THE EVOLUTION AND FUNCTION OF THE PAIR-RULE GENE FUSHI TARAZU (FTZ)

Amanda L. Field, Doctor of Philosophy, 2015

Directed by: Dr. Leslie Pick, Professor
Department of Entomology
University of Maryland, College Park

The homeodomain protein Fushi tarazu (Ftz) and its obligate cofactor Ftz-F1, an orphan nuclear receptor, cooperatively bind to DNA and co-regulate the transcription of genes responsible for segmentation in the early *Drosophila* embryo. Two interesting questions have arisen about these genes. The first question concerns the evolution of Ftz, which changed in arthropods from a Hox, to a pair-rule protein in *Drosophila*. This change in function involved changes in both the expression and protein sequence of Ftz, which are being explored throughout arthropod lineages. To determine if the expression and function of *ftz* is conserved in Diptera, Ftz and related genes were examined in the mosquito *Aedes aegypti* using reverse transcriptase - PCR, in situ hybridization, and ectopic expression techniques. The second question probes the mechanisms underlying *Drosophila* Ftz/Ftz-F1 target site recognition in vivo. To date, direct computational

attempts to identify downstream target genes and their enhancers have been inadequate. Towards this end, a microarray analysis was performed comparing wild type and *ftz-f1* mutant embryos. This generated a list of Ftz/Ftz-F1 candidate target genes. To identify genes among this group that are directly regulated by Ftz/Ftz-F1, potential Ftz/Ftz-F1 binding sites around these genes were identified by combining Ftz in vivo ChIP data with a computational search for candidate Ftz-F1 binding sites. Next, to test whether these regions correspond to Ftz/Ftz-F1-dependent enhancers, *enhancer-lacZ* reporter genes were constructed and their expression was analyzed in wild type and *ftz* mutant embryos. Of 10 enhancers tested, 8 generated expression patterns that overlap with Ftz and Ftz-F1 expression in early embryos and were lost in *ftz* mutants. The enhancers found in this study, along with previously identified Ftz/Ftz-F1-dependent enhancers, were analyzed to identify binding motifs for additional transcription factors that might co-regulate gene expression with Ftz/Ftz-F1. Four transcription factors were identified that could potentially be involved in Ftz/Ftz-F1-dependent gene regulation: Deaf-1 and Zeste, both transcriptional activators, and Dichaete and GAGA factor, both repressors. Together, these studies identified five new Ftz/Ftz-F1-dependent target genes and seven new Ftz/Ftz-F1-regulated enhancers, and they suggest that other transcription factors may also play roles in the pair-rule gene regulatory system.

THE EVOLUTION AND FUNCTION OF THE PAIR-RULE GENE *FUSHI TARAZU (FTZ)*

by

Amanda L. Field

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:

Professor Leslie Pick, chair
Professor David O'Brochta
Professor Alexa Bely
Professor Taneyhill
Professor Louisa Wu

Dedication

To Mom, Dad, John, Kevin, Laura, Lisa, and Kristin

Acknowledgements

First, I want to thank my advisor, Leslie Pick, for being an outstanding mentor throughout my time in graduate school. She has always been supportive, both in my thesis work and in my personal life. She has helped me develop as a critical thinker and a scientist. I am certain I would not have enjoyed graduate school as I did without her. Thank you for your support and enthusiasm.

I would also like to thank my lab members I have enjoyed working with every day. Our lab conversations were a constant source of help and entertainment, both of which were greatly needed. Our friendships were one of the most enjoyable parts of graduate school.

Lastly, I want to thanks my family and friends, for your love and support through many years of school. I could not have made it here without you. I want to thank my parents for teaching me to be independent, goal-oriented, and hardworking. Thanks to my brothers John and Kevin for their support and understanding as we all went through our graduate programs together. My oldest friends Laura and Lisa were always a source of support and fun. And I need to thank Kristin, who having been my roommate for the last years of grad school, has listened to me complain more than anyone and helped me laugh my way through it.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Aa	*Aedes aegypti*
AEL	after egg laying
Antp	*Antennapedia*
BDGP	Berkley Drosophila Genome Project
BDTNP	Berkley Drosophila Transcriptional Network Project
CRE	cis-regulatory element
EL	egg length
en	*engrailed*
eve	*even-skipped*
Dm	*Drosophila melanogaster*
FDR	false discovery rate
ftz	*fushi tarazu*
ftz-f1	*ftz-factor1*
GAF	GAGA Binding Factor
h	*hairy*
odd	*odd-skipped*
opa	*Odd-paired*
Odz	*odd oz*
prd	*paired*
PWM	position weight matrix
RT-PCR	reverse transcriptase PCR
run	*runt*
slp	*sloppy-paired*
Ten^m	*tenascin major*
RT-PCR	reverse transcriptase PCR
UMIACS	University of Maryland Institute for Advanced Computer Science

Overview

The central issue of modern developmental biology is the molecular basis of the diverse body plans seen in nature. This question is being explored from two main angles –a genetic approach, and an evolutionary approach. Both approaches are valuable and contribute different pieces to the puzzle. The genetic approach makes it possible to understand the molecular basis of development in extant species, while the evolutionary approach allows us to understand how these molecular mechanisms arose in the context of the history of life on earth. The purpose of this work is to use both of these approaches to answer questions about the gene *fushi tarazu* (*ftz*) in the pair-rule transcriptional network found in the genetic model system *Drosophila melanogaster*.

Chapter 1 is a literature review that covers thirty years of research on the pair-rule segmentation gene network in *Drosophila*. This chapter summarizes the historical context that launched this field of study, and then reviews the discovery and characterization of each of the nine *Drosophila* pair-rule genes. This establishes the background needed to understand the questions asked in the following chapters.

Chapter 2 is a brief chapter outlining research into whether the cell-adhesion proteins, the cadherins, could play a role in the pair-rule promotion of segmentation in the embryo. It was found that the expression of the *Drosophila* cadherins suggests that they are not involved in segmentation.

Chapter 3 takes an evolutionary approach to look into the question of whether the *Drosophila* pair-rule gene *ftz* is conserved in Diptera. Within Arthropoda, *ftz* has evolved from a *Hox* gene to a pair-rule gene. To do this, both the expression and

function of *ftz* changed. Therefore, the expression and function of *ftz* in the mosquito species *Aedes aegypti* was studied to compare to that of *ftz* in *Drosophila melanogaster* to determine whether variation in either expression or function occurs within the order Diptera.

Chapter 4 takes a genetic approach to determine how Ftz and its cofactor Ftz-F1 bind their specific targets in vivo. Understanding how a set of pair-rule transcriptional regulators activate their target genes will help future attempts to identify the downstream target genes that are directly involved in the morphological formation of segments. Finding the genes involved in creating segments will ultimately give insight into how pair-rule genes direct the formation of the segmented body plan of the fly and other insects. While the core binding sites of both Ftz and Ftz/F1 are known, previous attempts to find bona fide binding sites in the *Drosophila* genome using a computational approach were not successful. My work combined experimental and computational approaches to identify first, a larger set of direct targets of Ftz/Ftz-F1, and then to identify the Ftz/Ftz-F1 responsive enhancers that regulate these targets. Once these enhancers were confirmed, they were analyzed to determine if they contained any binding site motifs that would suggest other candidate transcription factors that could have a role in Ftz/Ftz-F1 gene regulation. Several factors were identified that will be studied in the future to determine if they do have a role in the pair-rule promotion of segmentation.

The final section of the paper, Conclusions and Future Directions, summarizes the findings from this work and speculates on where this work could lead next.

Chapter 1: Introduction

1.1 The segments of the fly are established by a cascade of regulatory genes

The fruit fly *Drosophila melanogaster* has been used as a model to study development for many years, and the pathways directing early embryonic development have been well characterized. These pathways are organized in a cascade of genes that divides the embryo into monomeric units that become more and more defined throughout development (Fig. 1-1). The first step in the cascade is the set of genes whose RNA is maternally deposited, which define the basic axes of the embryos – anterior/posterior and dorsal/ventral. These egg polarity genes regulate the gap genes, which define large sections of the embryos. Large sections containing multiple segments of the embryo do not form when a gap gene is mutated. The gap genes regulate the pair-rule genes, which are expressed in striped patterns and define the borders of the segments that are later seen in the morphology of the larvae and adult. The pair-rule genes regulate the segment-polarity genes, which define the anterior and posterior sides of each newly formed segment. This definition of the borders of the segments is essential for the function of the homeotic genes, which determine the identity of each segment and direct their morphology.

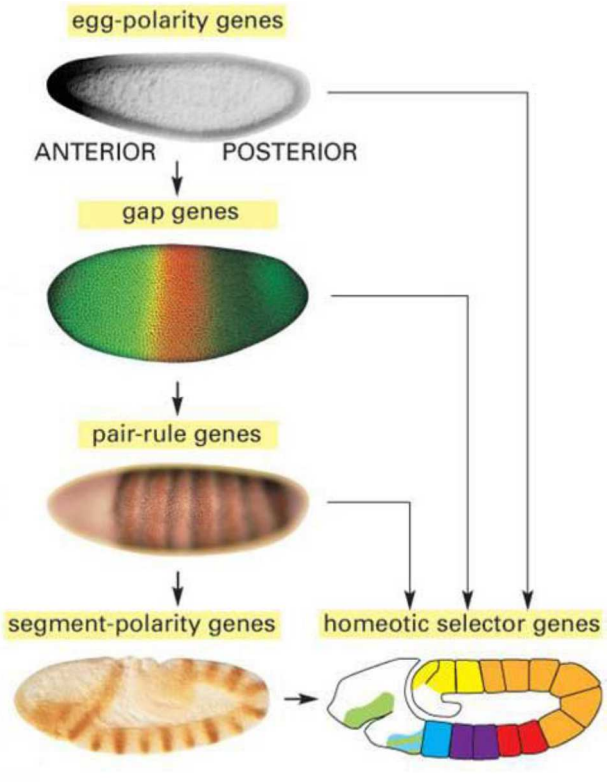


Figure 1-1. The cascade of genes that divide the *Drosophila* embryo into segments. A hierarchy of genes was determined by their mutant phenotype. The egg-polarity (maternal effect) genes regulate the gap genes, which in turn regulate the pair-rule, which then regulate the segment-polarity genes. Once the segments have been established, they are given their identity by the homeotic selector genes. Figure from Molecular Biology of the Cell. 4th edition, Alberts B, Johnson A, Lewis J, et al. 2002.

The pair-rule gene network is a complicated set of genes that regulate both each other and other genes in order to direct the formation of segments with exact borders in the embryo. These genes are essential to the proper development of embryos into larvae. When the segments are not formed correctly, first instar larvae usually die before hatching. Understanding the molecular basis of this pair-rule gene network has been a large focus in *Drosophila melanogaster* (*Dm*) research for the last thirty years, resulting in large amounts of information about the *Dm* pair-rule genes, their interactions with each other, and their regulation by and of many other genes. Here I

discuss how these genes were discovered, summarize what is known about how they interact within the segmentation gene network, and speculate about future insights.

1.2 Genetic Screens Identify Pair-rule Genes

In the early 80's and 90's, a number of genetic screens (Nusslein-Volhard et al., 1985; Nusslein-Volhard and Wieschaus, 1980; Perrimon et al., 1996; Sander et al., 1980; Wakimoto and Kaufman, 1981) were performed in order to understand the molecular basis of the metameric segments that make up the complex form of the fly. At the time, little was understood about how these segments were formed, or if indeed the segments even had a biological role or were just a structural artifact with no real function (Garcia-Bellido, 1998). The fly larva was clearly made up of a set of compartments, but it was unclear whether these compartments were truly separate from each other and whether they formed individually or simultaneously. The idea of the parasegment was also new and not fully accepted until the mid-80's - after the first of these screens were complete. The segments of the embryo were morphologically visible – there were grooves in the embryo between each segment, the longitudinal muscles attach at these grooves, and each of these segments appeared to correspond to the segments of the adult fly. The parasegments were much more difficult to detect. Parasegments are also metameric units, off register with the visible segments. Each parasegment consists of the posterior half of one segment and the anterior half of the next segment. But parasegments are delineated on a molecular level; several segmentation genes are expressed in parasegments rather than segments. Early evidence

of parasegments was drawn from multiple studies showing some of the first expression patterns of segmentation genes and early morphogenic evidence in *Drosophila* and other crustaceans (Martinez-Arias and Lawrence, 1985). So, it had yet to be determined whether the segments (and parasegments) had a genetic basis or were controlled by other means, such as widespread interactions in growth.

It was in this context that the most influential of these screens was conducted by Nusslein-Volhard and coworkers (Nusslein-Volhard et al., 1985; Nusslein-Volhard and Wieschaus, 1980), who were the first to identify classes of genes that control the basic subdivisions of the early embryo into segments. In their first screen, they looked through thousands of mutant stocks for any lines that were embryonic lethal, which ended up being 2,600 lines. The cuticles of these lines were inspected for defects, and then complementation tests were used to map the genes responsible for these mutant phenotypes. At the time, this type of large-scale screen was a newly designed approach to observing many mutant phenotypes. It proved effective if not efficient, finding 15 genes that caused the types of segmentation defects for which they were looking. This work proved invaluable and launched the study of the molecular basis of development, earning Nusslein-Volhard and Wieschaus the Nobel Prize in Physiology or Medicine in 1995.

Once the initial screen was complete, the true work of trying to understand what these defects suggested about the molecular mechanism behind the compartmental body plan of the larvae started. Despite not yet having even the concept of genetic networks formulated, Nusslein-Volhard was able to recognize that these genes worked together to direct this body plan. The authors classified the genes identified into three groups– gap,

pair-rule, and segment polarity genes – based on the phenotypes seen in the cuticles of larvae. The gap gene mutants were identified by loss of several segments in a row, so that there was a gap in the cuticle. The pair-rule mutants were identified by loss of alternating segments. The segment-polarity mutants were identified by the loss of repeating regions of each segment, often with mirror image duplications of the remaining portion of each segment. Without a proper understanding of parasegments, the function of pair-rule genes in alternating segments was interpreted as dividing the embryo into units the size of two segments at an earlier stage. It was not understood until later that these pair-rule genes worked together in overlapping expression patterns to define the borders of the segments and parasegments. The five pair-rule genes found in their first screen were *even-skipped* (*eve*), *odd-skipped* (*odd*), *paired* (*prd*), *hairy* (*h*) (originally called *barrel* in this paper), and *runt* (*run*).

eve mutants were missing the even-numbered segments - the prothoracic, metathoracic, and 2nd, 4th, 6th, and 8th abdominal segments. *Odd* mutants were missing the opposite, the odd-numbered segments. *prd* mutants were missing sections of both odd-numbered (posterior - naked cuticle) and even-numbered (anterior - denticle band) segments. *h* had the same phenotype as *prd*, just weaker. *run* mutants were missing half of their segments, and the anterior rows of the denticle belts of each segment were mirrored. Nusslein-Volhard and co-workers then conducted a larger screen, searching for any phenotypes that affect the cuticle. In this second screen, among many other cuticle defects, they found another pair-rule gene, *sloppy-paired* (*slp*), as well as confirming the *eve* and *odd* phenotypes. *slp* mutants were identified by the loss of naked cuticle in segments T2, A1, A3, A5, and A7. The *ftz* pair-rule phenotype, loss of the

even numbered parasegments, was independently discovered in a screen for larval segmentation defects with genes in the *Antp* complex (Lewis et al., 1980; Wakimoto et al., 1984). *ten*^{m} was found by two different labs in the same year, one in a search for a *Drosophila* homolog of the mammalian *tenascin* gene (Baumgartner et al., 1994), and the other after observing the expression pattern of the gene (see below), which they named *odd Oz* (Levine et al., 1994) . Ten^{m} was expressed in the seven striped expression pattern, typical for pair-rule genes, and three different mutants gave a range of phenotypes from slight segment fusions to full missing alternate segments, very similar to *odd* (Baumgartner et al., 1994). However, unlike the other pair-rule genes, *ten*^{m} is not a transcription factor, but works outside of the cell to start a signal cascade. The last pair-rule gene *ftz-f1*, not previously found as it functions in the early embryo through maternally deposited RNA, was discovered in a screen for maternal effects of zygotic lethal genes (Chou and Perrimon, 1996) and was confirmed as a pair-rule gene through its mutant phenotype (Yu et al., 1997).

1.3 Cloning and expression pattern of pair-rule genes

The pair-rule genes were first cloned and their expression patterns first determined over a decade, from the early 80's to the early 90's. The expression patterns of all pair-rule genes have a few characteristics in common. They each contain the core seven pair-rule stripes (with the exception of *ftz-f1 and opa*), which typically develop first as broader regions that are then refined into the thinner stripes during cellularization by gaps of expression forming within the region to form periodic stripes.

The stripes also generally develop anterior to posterior, and ventral to dorsal. In this section, the original cloning and determination of expression patterns of each pair-rule gene are reviewed.

1.3.1 *even-skipped*

eve was isolated from a library of bacteriophage clones containing genomic sequences (Macdonald et al., 1986). The *eve* homeobox was detected when the library was probed for homeoboxes of *Antp*, *ftz*, and *Ubx*. The sequence from the positive clones was mapped to region 46C of the polytene chromosome. Using northern blotting and RNase protection mapping of RNA from staged embryos, a 1.4 kb transcript for *eve* was found. This transcript had one 71 bp intron. The homeobox of *eve* was unique - unlike the other three homeobox-containing genes that had been characterized so far (*ftz*, *Antp, Ibx*), the homeobox was in the amino terminal region of the protein. It was also only 60% identical to other known homeoboxes. Within the homeodomain, there were two particular amino acids that contact DNA that varied in Eve, which were predicted to make it bind to different DNA sequences. Sixteen of the nineteen amino acids in Eve's homeodomain were identical to the N-term of the Engrailed homeodomain.

The spatial pattern of *eve* was first expressed as a band in the peripheral, cortical cytoplasm in a gradient anterior to posterior. This band expanded to seven stripes, each stripe three nuclei wide and separated by five nuclei. When compared to the expression pattern of *ftz*, the stripes were found to be 180 degrees out of phase with *ftz*, and shifted anteriorly from *ftz*. This means that the seven major stripes were

expressed in the progenitor cells of the odd numbered parasegments. As gastrulation proceeded, seven weaker stripes were expressed posterior to the original so that *eve* was expressed in every segment. The stripes fade as the germband extends.

1.3.2 *odd-skipped*

To isolate *odd* clones, *odd* mutant alleles containing a P-element were first made using hybrid dysgenesis (Coulter et al., 1990). The odd^{hd1} mutant had an *odd*-like phenotype, and an in vivo hybridization of a polytene chromosome showed a P-element in the chromosome interval *odd* was mapped to, making this sequence the *odd* sequence. A bacteriophage library from odd^{hd1} mutants was used to isolate the clone. The line was first crossed to M strain females to get rid of all unwanted P-elements. Then the fragment containing the P-element in the *odd* locus was subcloned and used to screen a wildtype genomic library. Four clones were obtained from this library, which together overlapped a 25 kb region. This region was confirmed using other mutants. A northern blot using genomic fragments from this region as a probe found a 2.2 kb transcript. cDNA clones from embryonic libraries were used to find the transcript for cloning because the RNA had long repeats. The transcript was confirmed by comparing it to existing mutant restriction maps. The longest cDNA contained an ORF of 392 amino acids. *odd* is the only pair-rule gene that encodes a zinc-finger DNA binding motif.

The expression pattern of *odd* was determined by performing in situ hybridization on sections of embryos, using an antisense RNA probe from a cDNA clone derived from an *odd* transcript. Transcripts were first detectable as a stripe at 62%

egg length at end of 13th nuclear division. The transcripts extended posteriorly to 24% egg length in 3-4 broad regions. Six stripes formed during cellularization, as did a domain on the very anterior of the embryo. A posterior 7th stripe appears in late cellularization, after cell elongation is complete. These seven stripes persist until gastrulation begins. The stripes are 3 cells wide and the gaps are 4-5 cells wide. The gap between 6 and 7 is slightly wider. During gastrulation, the stripes double to 14 stripes, one in each segment. Double staining with *engrailed* (which marks the posterior portion of each segment) for comparison found that the 7 *odd* stripes are slightly posterior to the *engrailed* stripes.

1.3.3 *paired*

The location of *prd* was found by chromosome walking from the gene *extra sex combs* on chromosome 2L (Kilchherr et al., 1986). The sequences from the walk were cut with ecoRI and used as probes for a southern blot of cDNA from 0-4 hour embryos. The probe hybridized to a 9.6 kb EcoRI fragment. A whole genome southern blot of wild type and *prd* mutants found the only difference was a 1.1 kb insertion within the 9.6 kb EcoRI fragment. This region was further confirmed by northern analysis, showing that the mutants had a larger allele for *prd*.

The expression pattern of *prd* begins in cellular blastoderm as a six nuclei wide band at ~70% egg length (Sander et al., 1980). This band widened after a round of nuclear division. During late syncytial blastoderm, five more bands appeared. The anterior band split, so that all seven bands are equal in width. The expression of all bands was stronger on the ventral side. The width of the six posterior bands was 5-7

cells, while the most anterior band was four cells wide. The gap between each band was 2-3 cells wide. A new band formed near the anterior pole, only on the dorsal side. A new band formed posterior to the 7th stripe. Bands 2-7 split, so that there is one anterior dorsal band, thirteen middle stripes, and one broader 14th stripe. The original seven stripes are shifted anteriorly to *ftz*. *ftz* stripes 1-7 and *prd* stripes 2-8 have their anterior border aligned, but *prd* stripes are twice as thick.

1.3.4 *hairy*

Sequences were first cloned from the *h* gene using the h^1 mutation (Holmgren, 1984). The *gypsy* transposable element that caused the mutation was used as a probe to find the *h* gene out of a library. The sequence from this clone was then used as a probe in a wild type library. Four clones covered a region of 40 kb, but the exact location of *h* was not found in this paper. *h* was previously mapped to chromosome subdivision 66D (Jeffery, 1979).

To determine the expression pattern of *h, t*wo transcripts were found to be expressed in embryogenesis - one 2.1 kb and one 2.0 kb (Ingham et al., 1985). A 1.7 kb genomic probe was made from these sequences and hybridized to sections. A staining with *ftz* was used to compare to as a reference in same the embryo in the same slide and two adjacent slides. Transcripts were first detectable at the eleventh nuclear division, like *ftz*. Transcripts are first ubiquitous, then localized in two areas. In the blastoderm there are seven stripes on the ventral side, but the most anterior stripe is wide and interpreted as two stripes. The seven most posterior stripes are also expressed on the dorsal side. These stripes are the same width as *ftz* and are mostly located in the

cytoplasm –but with some nuclear presence. Stripes one and two flank the cephalic fold during gastrulation. By the time the germ band has finished extending, *h* stripes are gone. There is *h* expression in the hindgut and foregut at this point.

1.3.5 *runt*

run was isolated using the same method as *odd.* in situ hybridization was performed on a polytene chromosome to find P-elements in a *run* mutant (Gergen and Butler, 1988). This mutant was crossed to an M strain to remove unwanted P-elements, and then a genomic library was generated from this mutant and probed for P-elements. The wild type polytene chromosome was hybridized to probes made from all of these clones, and one was found to bind to the correct region. DNA fragments from this clone were used to scan a wild type genomic library. The clones from this screen covered a 50 kb region, and were used to make a restriction map.

cDNA probe of the transcript was made from poly(A)$^+$ RNA isolated from blastoderm stage embryos. This probe was used on a southern blot of restriction digest fragments of the *run* region. Two of these fragments were hybridized to a northern blot of embryonic poly(A)$^+$ RNA and total RNA, and they detected a 2.6 kb fragment. The fragment was found to be transcribed 5' to 3' by subcloning it into a vector with RNA polymerase promoters on each side and using RNA made from this vector as probes. To confirm the *run* clone, a rescue experiment inserting the clone into a *run* mutant with P-elements was performed, but there was only partial rescue.

To determine the expression pattern, radiolabelled RNA probes were used. RNA was found in a broad domain over 10-70% egg length in nuclear cycle 13. The seven stripes became clear in nuclear cycle 14.

1.3.6 *sloppy-paired*

The enhancer detector transposon P[lArB] was used to find transposants on a CyO balancer chromosome, and two transposants mapped to the *slp* locus (Grossniklaus et al., 1992). One probe was made from each of these transposants and used to probe the genomic library. A 28 kb region was found. EcoRI fragments from this 28 kb region were used to probe northern blots of RNA from developmental timepoints. Two transcripts were found, a 1.5 kb transcript called *slp1* and a 2.2 kb transcript called *slp2*. *slp1* was expressed in 0-3 hour embryos and *slp2* appeared in 3-6 hour embryos. Both transcripts were most highly expressed in 3-6 hour embryos. Both transcripts are expressed at a steadily decreasing amount throughout embryogenesis and then are expressed strongly in the first instar larvae and weakly in the third instar larvae. The genomic sequence of the two genes showed neither had an intron. The protein contains a *forkhead* domain, a DNA binding domain also found in Forkhead proteins and other transcription factors. When both *slp1* and *slp2* were mutated, a severe pair-rule defect was found, showing functional redundancy (Cadigan et al., 1994a; Grossniklaus et al., 1992).

To determine the expression pattern, an in situ hybridization of whole mount embryos with digoxigenin probes was performed. *Slp1* is expressed first during the syncytial blastoderm in a region from 100-70% EL. The anterior most part of this

expression is lost as development continues into cellularization. This region then splits into an anterior cap and a stripe that becomes the first stripe of the seven metameric stripes. The six other stripes appear from anterior to posterior. These stripes start ventrolaterally and spread ventrally and dorsally, but never meet at the dorsal tip. The stripes are two cells wide with six cell gaps in between them. The fourth stripe is narrower in places - only one cell. The primary seven stripes are fully established during blastoderm, and towards the end of this stage, seven more secondary stripes start to form in the alternate segments. These secondary stripes also form anterior to posterior and laterally outwards, and do not connect at the dorsal tip. The anterior cap splits into two stripes at this time. A new anterior region of expression forms as well. The secondary stripes are fully established by early gastrulation, although they are slightly weaker than the secondary. Each of the stripes is two cells wide with 2-3 cells in between. Lateral cell clusters connected to each stripe by a thin strip appear during germ band extension. The stripes fade during germband retraction. *slp2* has similar expression, the main difference being a later start time. The primary stripes start expression just before cell membranes form, and the secondary stripes form just start forming during gastrulation. Double in situ hybridizations with *en* and *ftz* showed that *slp* is expressed in the posterior half of the each parasegment, just anterior to the boundary.

1.3.7 *fushi tarazu*

Antp cDNA clones were hybridized to a segment 30 kb upstream of the *Antp* gene. This segment was used as a probe for a northern blot, which hybridized to a 1.9 kb transcript (Kuroiwa et al., 1984). Another study (Scott et al., 1983) also mapped *ftz*

to this site on the chromosome. EcoRI fragments from within a clone were used to determine which region of DNA hybridized to the 1.9 kb RNA in a northern blot.

An expression profile of *ftz* was determined by hybridizing nick translated DNA to electrophoretically fractionated poly(A)+ RNA from embryos to adult. The RNA was most highly expressed at 0-3 hours, then continued at 3-6 hours, and then fainter at 6-12 hours.

To map the *ftz* transcript, the direction of transcription was determined to be left to right using strand-specific probes on clones in single stranded phage vectors. A restriction map of the clones was made by S1 nuclease mapping. Restriction mapping of a clone from a cDNA library showed an intron of 150 bp.

That same year, another study also isolated *ftz* and then determined its expression pattern (Hafen et al., 1984). A plasmid that contained gDNA homologous to *ftz* RNA transcripts was used to make a probe to hybridize to frozen sections of WT embryos at various stages, labeled with a radioactive tag. Transcripts were first barely visible after 11th nuclear division. The stripes became clearer after the 13th nuclear division. At the cellular blastoderm, there was a definite pattern of seven stripes, which continued through gastrulation. During germ band extension, the 7th stripe extended to the anterior part of the dorsal side of the embryo. No transcripts were detected at the end of germband extension so that all stripes were gone by the time visible segments appeared. *ftz* was not expressed in the anterior head segments, only in the posterior and thoracic and abdominal segments.

1.3.8 *Odd-paired*

odd-paired was first identified in a screen that used EMS to generate lines with mutations on the third chromosome that caused cuticle defects (Jürgens et al., 1984). By examining the cuticle of embryos from thousands of lines, and then performing complementation tests, they were able to find 13 mutant alleles. The mutation of *opa* caused a typical pair-rule defect, for which the odd parasegments were missing.

Unexpectedly given *opa*'s pair-rule function, it was found that *opa* was not expressed in seven pair-rule stripes, but instead was expressed ubiquitously throughout cellularization, blastoderm, and gastrulation. First, a northern blot with a probe against *opa* showed that a 3 kb mRNA was expressed starting at 2 hours after egg laying (AEL) and continues until 12 hours AEL. When the expression of this transcript was observed by in situ hybridization, it showed that it was first expressed in a band that was 10 cells at 80% egg length (EL), after which the expression expands to cover between 20% and 80% egg length. The expression is lost from the ectoderm and expression fades in between stripes to form the typical 14 stripes after germband extension begins (Benedyk et al., 1994).

1.3.9 *tenascin major/odd oz*

A genomic library was scanned with a probe for the extracellular epidermal growth factor (EGF)-like domains of ten^{a} of *Drosophila* (Baumgartner et al., 1994). Several clones were found and used to create a map of the 110kb region of ten^{m}. The region included 5 introns and a 7545 bp open reading frame. The putative protein contained two domains with tenascin-type EGF-like repeats.

ten^{m} had two transcripts, found by northern blot, of 10.5 and 11.5 kb. While the 11.5 kb transcript was found throughout development, the 10.5 kb transcript was only expressed in early embryogenesis. In situ hybridization showed that the 10.5 kb transcript was first expressed during the 14th nuclear cycle (stage 5). Expression begins as a broader region across the center of the embryo. This broad expression does not resolve into 14 stripes until germband extension is already underway, during stage 12. Transcripts were also expressed in cardiac cells and in the lymph gland at this time. After germband retraction, transcripts of *ten*^{m} continued to be expressed in various structures of the embryo throughout development. Protein expression was seen in seven stripes in the blastoderm that are 4 to 5 cells wide. There was also some weak anterior and posterior staining. Because the RNA and protein patterns do not match, it was proposed that the protein was interacting with a receptor that was expressed in the seven pair-rule stripes. At the beginning of germband extension, the seven stripes of the protein were detected on cell surfaces in the mesoderm and ectoderm, then become restricted to the mesoderm, and faded completely by the end of germband extension. Other protein expression continues throughout development. In order to determine where these stripes were located, the protein pattern was compared to the expression patterns of Ftz, Prd, and Eve. Ten^{m} was found to be localized to all odd numbered parasegments

Another study discovered *ten*^{m} at the same time, although they first named the gene *odd oz* (Levine et al., 1994). Monoclonal antibodies were made against tyrosine kinases and tyrosine kinase substrates. Performing immunohistochemistry on whole mount embryos, seven stripes located on the periphery of cells were found in cellular

blastoderm embryos with one of these antibodies. During gastrulation and germ band extraction, the seven stripes expanded to fourteen stripes. This protein continued to be expressed in various structures throughout embryonic development, most notably the CNS. The antibody was used to isolate a cDNA clone from a library. They found 10 and 11 kb transcripts and used in situ hybridization to observe the RNA pattern, which they also found not be expressed in seven stripes in the blastoderm. They used the transcripts to hybridize to polytene chromosomes to isolate the gene location and clone it out from a genomic library. They used P-element insertions to verify that mutations in the cloned genes generated the *ten*^{m} phenotype.

1.3.10 *Ftz-Factor1*

Ftz-F1 was first identified in a search for proteins binding an enhancer element of *ftz* that directs expression in a striped pattern, called the zebra element. First several factors were found in vitro using an exonuclease III experiment, and then a gel mobility shift assay was used to show a factor binding to the zebra element. This factor was purified and named Ftz-F1 (Ueda et al., 1990).

ftz-f1 was later shown to be maternally deposited (Guichet et al., 1997) and expressed ubiquitously through blastoderm by in situ hybridization (Yu et al., 1997; Yussa et al., 2001), which were unusual characteristics for a gene that gives the mutant phenotype of missing alternate parasegments typical of a pair-rule gene (Yu et al., 1997). This was explained by many studies that show that Ftz and Ftz-F1 are obligate cofactors that interact directly to promote segmentation in the early *Drosophila* embryo. Ftz and Ftz-F1 physically interact both in vitro and in vivo (Guichet et al., 1997; Yu et

al., 1997) in wild type *Drosophila* embryos; they bind synergistically to DNA containing composite Ftz/Ftz-F1 binding sites (Yu et al., 1997), they interact through the Ftz-F1 AF2 domain (Guichet et al., 1997; Schwartz et al., 2001) and the Ftz LXXLL motif (Yussa et al., 2001), and coordinately activate the transcription of target genes (Yussa et al., 2001).

1.4 Pair-rule genes interact in a complicated network of activation and repression

The pair-rule genes interact in a complicated network to ensure the correct formation of the parasegments. Several layers of repression and activation occur to define the exact borders of expression of the segment polarity gene *engrailed*. In order to help organize our understanding of this network, scientists have classified the pair-rule genes somewhat arbitrarily into a hierarchy of primary, secondary, and tertiary pair-rule genes. The primary pair-rule genes are mostly activated by gap genes, are expressed early in the pair-rule cascade in the blastoderm, regulate each other to define their borders, and then help to activate or repress the later expressed pair-rule genes. Secondary pair-rule genes are activated and maintained by primary pair-rule genes, and also regulate each other to define their borders, but these genes do not regulate primary pair-rule genes. Pair-rule genes that might be categorized as tertiary would be activated by secondary pair-rule genes, and then work together to define the precise expression of *en*. However, the hierarchy of genes does not work as simply as defined by this order. After an examination of the interaction of pair-rule genes, summarized in Fig. 1-2, the accuracy of ordering them in this hierarchy will be reexamined below.

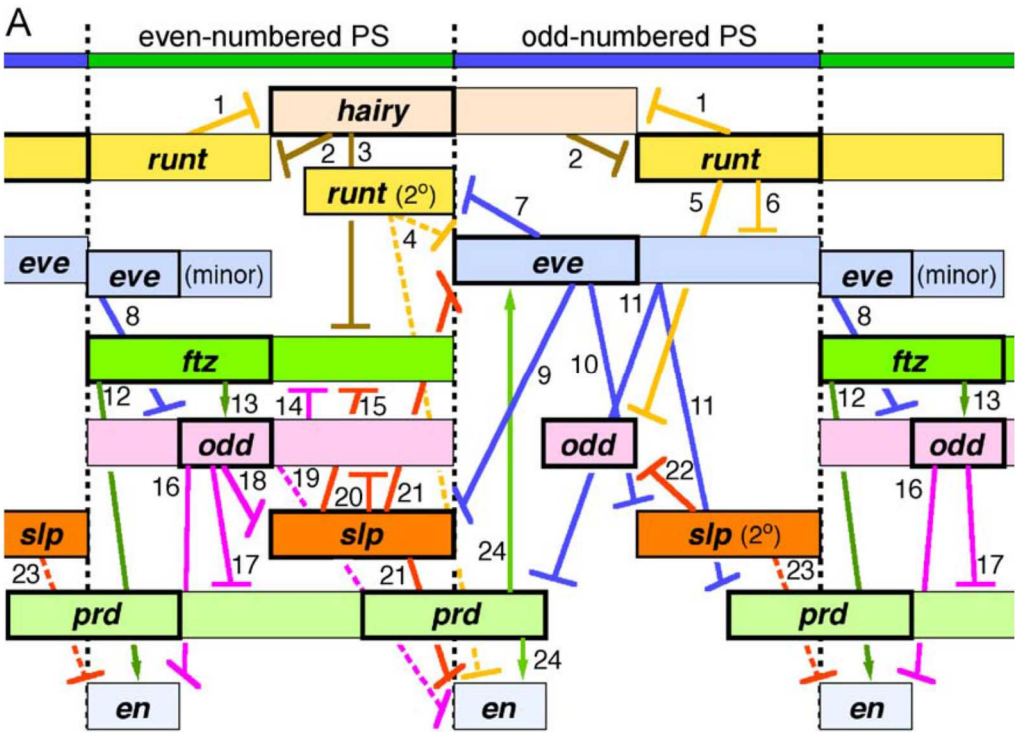


Figure 1-2. The network of pair-rule interactions in even and odd-numbered parasegments. The pair-rule genes repress (bar) and activate (arrow) each other in order to define their borders and ultimately to define the borders of *en* in the anterior of each parasegment. Dotted lines represent interactions that are not confirmed, as discuss below. Figure from (Jaynes and Fujioka, 2004).

1.4.1 Primary pair-rule genes

Because of its role early in the pair-rule hierarchy, *h* is considered one of the 'primary' pair-rule genes, along with *run* and *eve.* Hairy represses both *run* and *ftz. run* was shown to be mildly affected in *h* mutants, with the seven stripes still appearing but broader and less defined. The *run* transcripts then increase between the stripes so that the expression is uniform by gastrulation (Ingham and Gergen, 1988). *run* and *ftz* were also activated by a *h* derivative containing heterologous transcriptional activation domains, showing that *h* represses *run* by binding the promoter sites of *run* (Jimenez et

al., 1996). *ftz* stripes do not resolve properly in *h* mutant, they are expanded but still show a periodicity (Howard and Ingham, 1986; Ingham and Gergen, 1988).

run is also considered one of the 'primary' pair-rule genes. r*un's* initial broad expression begins before gap genes fade, characteristic of activation by gap genes, and is resolved into stripes by the other early acting pair-rule genes *eve*, *run*, and to a lesser degree, *h* (Klinger and Gergen, 1993). It is expressed at the beginning of the pair-rule network and controls the expression of other early expressing as well as later expressing genes. *run* represses *h*, as shown by an expansion of *h* stripes in *run* mutants (Hartmann et al., 1994; Ingham and Gergen, 1988). By the onset of gastrulation, the expression has become uniform between the first and last stripes (Ingham and Gergen, 1988). *eve* is expanded in *run* mutants, suggesting *run* represses *eve* (the opposite of what was seen in *h* mutants) (Howard and Ingham, 1986; Ingham and Gergen, 1988). *eve* was rapidly repressed when *run* was expressed throughout the embryo with a heatshock promoter (Manoukian and Krause, 1993). *eve* was also activated by a *run* derivative containing heterologous transcriptional activation domains, showing that *run* represses *eve* by binding *run* CRE(s) (Jimenez et al., 1996)**.** Once *run* has helped establish the early pair-rule genes *eve*, *h*, and itself through autoregulation, it then has a later role in maintenance and the repression of some secondary pair-rule genes. *ftz* expression is lost in *run* mutants (Carroll and Scott, 1986; Howard and Ingham, 1986). These effects occur during cellularization for *ftz* expression (Carroll and Scott, 1986) and gastrulation for *eve* expression. Run regulates the expression of *odd* in the odd parasegments by repressing the posterior border of *odd* expression, which expands in *run* mutants (Jaynes and Fujioka, 2004). Run along with Opa are necessary and sufficient to activate

slp1 in odd parasegments, while Run represses *slp1* with the help of Ftz in even parasegments (Swantek and Gergen, 2004). Run controls *slp1* independently of its co-repressor Groucho, shown by repressor function independent of the Runt binding motif VWRPY (Walrad et al., 2010) and a Runt and Ftz coexpression with reduced Groucho expression (Walrad et al., 2011). Later, *run* is involved in both the repression and maintenance of *en* stripes, although it is possible this role is fulfilled by Run repressing the anterior end of *eve* at the anterior border of the odd parasegment (see section 1.5).

A large amount of evidence points to *eve* also being a primary pair-rule gene. The strongest evidence comes from a series of studies that have shown that elements in the cis-regulatory elements (CREs) of individual stripes of *eve* are regulated by different maternal and gap genes (Fujioka et al., 1999; Goto et al., 1989; Harding et al., 1989; Small et al., 1992, 1996). The next line of evidence comes from the *eve*'s role early in the hierarchy. A promoter fusion experiment found the promoter region of *eve* that early expressed Eve binds to autoactivate and refine later expression of *eve* (Harding et al., 1989) along with a deletion analysis of the cis-regulatory region of *eve* (Goto et al., 1989). *eve* helps establish/maintain the posterior border of the later *run* stripes in even parasegments, and so *run* expands into the odd parasegments in *eve* mutants (Fujioka et al., 1995). This was further shown when *eve* was expressed under a heatshock promoter, which strongly repressed late *run* stripes (Manoukian and Krause, 1992). As a primary pair-rule gene, *eve* also regulates several later pair-rule genes through repression – *odd, slp*, and *prd* – which then directly regulate the formation of *en* stripes to set up the borders of the parasegments (see section 1.5). *eve* repression of the anterior third of late *odd* expression, in the even parasegments, is lost in *eve* mutants, which allows for the

expression of the even *en* stripe to be expressed in this space (Coulter and Wieschaus, 1988; DiNardo and O'Farrell, 1987; Fujioka et al., 1995). When *eve* was overexpressed under a heatshock promoter, *odd* was rapidly repressed (Manoukian and Krause, 1992). Mutation of *eve* also causes *slp* to completely expand throughout the odd parasegments, meaning that *eve* must repress the posterior of *slp* (Fujioka et al., 1995; Jaynes and Fujioka, 2004). *eve* also represses the anterior portion of late *slp* in the odd parasegments, as shown when *eve* is expressed abnormally in *run* mutants and so *odd* expression, normally repressed by *slp*, expands to the domain of *slp* expression (Jaynes and Fujioka, 2004). Similar to the effect on *slp*, loss of *eve* also causes *prd* expansion in odd parasegments (Baumgartner and Noll, 1990; Fujioka et al., 1995), and overexpression of *eve* under a heatshock promoter represses *prd* (Manoukian and Krause, 1992).

1.4.2 Secondary pair-rule genes

ftz has been defined as a primary pair-rule gene by some because it is activated by maternal and gap genes (Carroll and Scott, 1986; Carroll and Vavra, 1989; Vavra and Carroll, 1989), but it has also been considered a secondary pair-rule gene because of its role slightly later in the network and its direct interaction with *en*. A number of studies have shown that *ftz* autoregulates, and the enhancers mediating this direct regulation have been identified (Hiromi and Gehring, 1987; Hiromi et al., 1985; Pick et al., 1990b; Schier and Gehring, 1992, 1993). In addition, anterior expansion of *ftz* stripes was observed after ectopic expression of *ftz* throughout the embryo (Ish-Horowicz et al., 1989). Ftz activates *odd*, as well as activating *en* (see section 1.5). *ftz*

overexpression by a heatshock promoter activated *odd*, and *ftz* mutant embryos have reduced *odd* expression (Nasiadka and Krause, 1999). The same study showed that *ftz* mutants also have broadened *slp* stripes, but this is not due to Ftz repression of *slp* and *prd*, but is instead due to *ftz* activation of *odd*, which then represses *slp* and *prd* (Jaynes and Fujioka, 2004).

Once Ftz activates *odd*, Odd then represses the posterior half of *ftz* early stripes, refining the stripes, as seen in the failure of *ftz* stripes to narrow *odd* mutants (Mullen and DiNardo, 1995) and in the repression of *ftz* when *odd* is overexpressed by heatshock (Saulier-Le Drean et al., 1998). Odd also represses *prd*, seen in *odd* mutants where *prd* expression continues throughout the even parasegment, instead of narrowing into *prd* stripes that overlap the even and odd parasegments (Baumgartner and Noll, 1990), and in overexpression of *odd* by heatshock repressing *prd* (Saulier-Le Drean et al., 1998). Odd represses *slp*, with late Odd repressing the anterior end of late *slp* in the even parasegments (Jaynes and Fujioka, 2004). This makes *odd* a repression intermediate between *eve* and *slp*, meaning that *eve* does not directly repress *slp*, but first represses *odd* which then represses *slp*. A complicated combination of mutant and expression analysis led Jaynes et al. to this conclusion. Perhaps a simpler and certainly a more definite way of understanding these pair-rule direct/indirect interactions would be to determine whether Odd binds directly to *slp* enhancers.

Although the Ftz cofactor Ftz-F1 has not traditionally been studied in the pair-rule network, it has been shown that for every one of the 16 non-pair-rule targets of *ftz* found so far, *ftz* and *ftz-f1* have both been necessary to activate them, including *en*, in the segmentation pathway (Bowler et al., 2006; Hou et al., 2009)(this work). It seems

likely then, that if studies were done of *ftz-f1* in the pair-rule network, it would show that *ftz-f1* is necessary for *ftz* regulation of *odd* as well as *en*.

Opa has been shown to be involved in the activation of *slp* (cooperatively with Run) (Swantek and Gergen, 2004) and in the regulation of *prd*, repressing the middle of *prd* stripes to split into their even and odd parasegment stripes and activating the anterior of the *prd* stripes to maintain them (Baumgartner and Noll, 1990). However, there is no evidence that these interactions are direct, and so the role of *opa* in the pair-rule network needs to be explored further. Opa has also been shown to up-regulate *en* expression through repression of *odd* (Benedyk et al., 1994). The presence of a putative Opa-binding site in the *en* intron enhancer suggests that Opa may also exert a direct effect on *en* (Florence et al., 1997).

1.4.3 Tertiary pair-rule genes

The final pair-rule genes are activated by secondary pair-rule genes and interact directly with *en*. *Slp* represses the posterior half of early *odd* stripes, so that s*lp* mutants cause *prd* stripes to expand throughout the even parasegment (Jaynes and Fujioka, 2004). This *odd* expansion has an effect on *en* stripe formation more thoroughly explained in section 1.5. The posterior half of early *ftz* expression is repressed by *slp*, shown by the failure of *ftz* stripes to narrow in *slp* mutants (Cadigan et al., 1994b). Slp also represses the late *odd* stripes in the odd parasegments. It was shown that late *odd* stripes are lost in *eve* mutants (Fujioka et al., 1995), but this was shown to be caused by the Slp repression of *odd* once *slp* stripes were expanded in *eve* mutants (Jaynes and Fujioka, 2004). Jaynes et al. (2004) also showed that *slp* represses *eve* in the odd

parasegments by demonstrating that *en* stripes, usually lost in *eve* mutants, are returned to expression in *eve, slp* double mutants. In vitro binding and transgenic analysis shows that Prd activates late *eve* (Fujioka et al., 1996).

The terms 'primary, secondary, and tertiary' pair-rule genes are well established in the literature. However, after examining the pair-rule interactions, this assignment is ambiguous at best. Different genes have been defined as primary – *h*, *run*, and *eve* are often considered the primary pair-rule genes (Akam, 1989a). However, one of the defining features of a primary pair-rule gene is activation by gap genes, which is also true of *ftz* (Carroll and Scott, 1986; Yu and Pick, 1995) and *prd* (Gutjahr et al., 1993). A defining feature of a tertiary pair-rule gene is the direct regulation of *en*, which both *ftz* and *prd* also do. Another factor that adds confusion to the order is the fact that the expression of most of pair-rule genes is split into early and late expression. This leads to differences in establishment and maintenance of genes. For example, *eve* is first established by gap genes, but later expression is maintained by Slp, a tertiary pair-rule gene. *Run* is also regulated by secondary and tertiary pair-rule genes *ftz* and *prd* (Klinger and Gergen, 1993).

Enough exceptions to the rule have now been found to question whether the original rule should be applied any longer. It would be convenient if the pair-rule network did follow a set cascade of genes, as the segmentation network does, but the evidence no longer supports this way of thinking. Trying to insert pair-rule genes into categories that do not exist only makes it more difficult understand the complicated interactions in this network.

1.5 The pair-rule genes work together to establish *en* expression

The segment polarity gene *en* is a highly conserved segmentation gene important in the establishment of parasegments. *eve* and *ftz* are two major controls that define the expression of *en* and thus the parasegment borders. They have a complementary expression pattern, with *eve* stripes being expressed in the odd parasegment primordia and *ftz* stripes expressed in the even parasegment primordia (Harding et al., 1986). It has been shown that the odd parasegments narrow (Frasch and Levine, 1987; Fujioka et al., 1995; Fujioka et al., 2002) and expand (Fujioka et al., 1999; Hughes and Krause, 2001) with the reduced and extended expression of *eve*. The odd parasegments also expand when the repressive function of *eve* is increased (Fujioka et al., 2002; Kobayashi et al., 2001). Similarly, when Ftz is overexpressed (Hughes and Krause, 2001) or its stability is increased (Kellerman et al., 1990), the even parasegments expand. Ftz also binds to an enhancer located in the first intron of *en* (Florence et al., 1997), showing direct regulation by Ftz. The expansions of one parasegment always come at the cost of the other, which become smaller. *en* is expressed at the anterior end of the *eve* and *ftz* stripes, and so the anterior of each parasegment (Ingham et al., 1988; Lawrence et al., 1987). While *eve* and *ftz* are the key players, seven of the pair-rule genes work together in a complicated network of activation and repression to ensure that the control of *en* expression is tightly regulated. Once *en* stripes have been established, they must be maintained. *en* stripes are in part maintained by late *eve* expression (Fujioka et al., 2002) and continued *slp* expression. *en* itself represses *eve* to end its expression (Harding et al., 1986), although it is not known if this repression is direct.

Early research identified the genes that directly activated or repressed *en*, as well as the genes that regulated those direct genes. It was shown that even *en* stripes were lost in *ftz* mutant embryos and odd *en* stripes were lost in *prd* mutants (DiNardo and O'Farrell, 1987) and that *en* expression was expanded with *ftz* overexpression by heatshock (Ish-Horowicz et al., 1989), indicating *ftz* and *prd* activate *en* even and odd stripes, respectively. *odd* and *slp* are direct repressors of *en. odd* defines the posterior border of even *en* stripes, as can be seen in the posterior expansion of *en* stripes both in *odd* mutants (DiNardo and O'Farrell, 1987), in overexpression of *eve* leading to *odd* repression (Manoukian and Krause, 1992), and in the restoration of cuticle defects, normally seen in *odd* mutants, in *odd, en* double mutants (Coulter and Wieschaus, 1988), and the posterior expression of even *en* stripes ending sharply at the expression of the *odd* stripes (Fujioka et al., 1995). *slp* was shown to define the anterior border of *en* stripes through repression as overexpression led to *en* repression (Cadigan et al., 1994a), through anterior expansion of *en* stripes in *slp* mutants (Cadigan et al., 1994b), and lost *en* stripes in *eve* mutants returning in *eve, slp* double mutants (Jaynes and Fujioka, 2004).

Along with these known regulators of *en*, there is speculation that *run* could be a direct repressor of *en* as well. There was an idea that *run* has a redundant function with *slp*, directly repressing odd *en* stripes. Odd *en* stripes were lost in *run* ectopic expression by heat shock (Aronson et al., 1997), even though expression of early *eve* remained unaffected (Manoukian and Krause, 1993). Additionally, overexpression of *run* both

under the control of a heatshock promoter (Tsai and Gergen, 1994) and using the GAL4 system (Tracey et al., 2000) resulted in the repression of *en*. And it was shown that the establishment of *en* stripes is dependent on *run* and *tramtrack* expression (Wheeler et al., 2002) but not dependent on *run*'s repressive function with its cofactor Groucho (Aronson et al., 1997; Wheeler et al., 2002) or Hairless (Walrad et al., 2011). This model is consistent with the fact that *en* expression expands throughout the *ftz* domain in *run,slp* mutants (Jaynes and Fujioka, 2004). However, this redundancy is not the only possible explanation. Instead, this expansion could result from loss of both *odd* and *slp* expression in these double mutants. It was shown that *slp* is an intermediate between *eve* and *en*, and so a change in *eve* expression could affect *slp* and then *en* expression. It was also shown that the anterior border of *en* may be under the control of *odd* when *slp* is null (Jaynes and Fujioka, 2004), because *slp* is no longer present to repress *en*, and *odd* expression overlaps with *slp* in the even parasegment and also represses *en*, it takes on that role. In *run* mutants, early *eve* repression is lost and early *eve* expression extends into the even parasegment (Fujioka et al., 1995). If *runt* mutants also caused the expansion of late *eve* expression, then late *eve* would repress *slp* in the even parasegment, leading to the expansion of *en* stripes anteriorly. To summarize the possible results from this model: 1) in a single *slp* mutant, *odd* would still be present and so would repress *en* normally (Jaynes and Fujioka, 2004) 2) in a single *run* mutant, *slp* would be repressed and so *en* would extend anteriorally (Aronson et al., 1997; Manoukian and Krause, 1993) in a *run*,*slp* double mutant, expression of both *odd* and *slp* would be lost and *en* would expand anteriorally (Jaynes and Fujioka, 2004). Thus, these results could have been caused either by a redundancy in *slp* and *run* direct

repression of the anterior border of the even *en* stripe, or it could have been caused by *run* directly repressing *eve* expression, keeping *eve* expression from expanding into the even parasegment.

There is a lack of definite evidence for either of these possibilities. This is an excellent example of how the effects on expression in loss of function or overexpression mutants is not enough to determine direct regulation, as the effects on the rest of the network may confuse the results about the gene of interest. Only so much can be learned by comparing different expression patterns in different sets of mutants. To truly test for direct regulation, the enhancers for the pair-rule genes must be found and the binding sites for the suspected activators and repressors must be mutated to confirm a direct effect.

Indirect interactions are just as important as the direct interactions in ensuring well defined *en* stripes. The pair-rule gene *eve* is an important control point in the network of *en* stripe formation even though it does not regulate *en* directly. *Eve* is necessary for *en* stripe expression in both odd and even parasegments, as shown by the failure of the trunk region to develop in *eve* mutants (Nusslein-Volhard et al., 1985) and the lack of both sets of *en* stripes in *eve* null mutants (Harding et al., 1986; Macdonald et al., 1986). Beyond just the presence of some of these indirect and direct regulators, the concentration of the regulators may affect which targets are being regulated in each parasegment. The concentrations of Eve and Ftz may be important in establishing the expression of *en*. *eve* regulates *en* expression by controlling both *ftz* and *prd* expression

through repression. *eve* was shown to work in a gradient dependent fashion on several pair-rule genes (Manoukian and Krause, 1992). The expression pattern of pair-rule genes in *eve* embryos showed that early activation of *eve* was essential for the establishment of all *en* stripes, while later expression of *eve* was important for the continued strong expression of the odd *en* stripes (Fujioka et al., 1995). This led to the first model of a gradient-dependent regulation of *en* stripes, which suggested that *en* could only be expressed in odd parasegments when *eve* expression was strong enough to repress *en* repressors *slp* and *runt*, but weak enough to allow expression of the *en* activator *prd*. This odd parasegment network was next to the even parasegments, in which *ftz* expression activated the even *en* stripes. *eve* repressed *ftz* expression, guaranteeing it would not activate *en* expression in the odd parasegment.

While this model held after years of further study, a review of pair-rule literature was used to make an expanded model explaining the effects of concentration changes of both of *eve* and *ftz* on how each *en* stripe is set (Fig. 1.3) (Jaynes and Fujioka, 2004). In this model, it was proposed that reduced expression of *eve*, away from the anterior border of the odd parasegment, and the increased expression of *ftz*, into the posterior of the odd parasegment, could change the downstream network to narrow the odd parasegment. Reduced *eve* expression would change the border of repression for *slp* and *prd*. *slp* only requires weak expression of *eve* to be repressed, and so *slp* expression would expand until the beginning of the reduced *eve* expression. p*rd* requires strong *eve* expression, and so would continue to be expressed until it reached the domain of strong *eve* expression. Thus, the odd *en* stripe would be activated by *prd* more posteriorly into the region of the odd parasegment and would be repressed by *slp* also more posteriorly.

The entire *en* stripe would be moved posteriorly from its normal position. Since the *en* stripe defines the border of the parasegment, the parasegment itself would also move to posterior from its normal position, and the even parasegment on its anterior would expand with the expansion of *wg* expression. In the same vein, an increased *ftz* concentration would shift the odd parasegment posterior border anteriorly, narrowing the parasegment on the posterior side. In this case, increased *ftz* would lead to a shift in activation of both *en*, which would more anteriorly, and *odd*, which would repress the posterior border of *en* more anteriorly. This model suggests that small changes in the gradients of either of these two key genes would have a large effect on the placement of the parasegments, and the formation of the embryo.

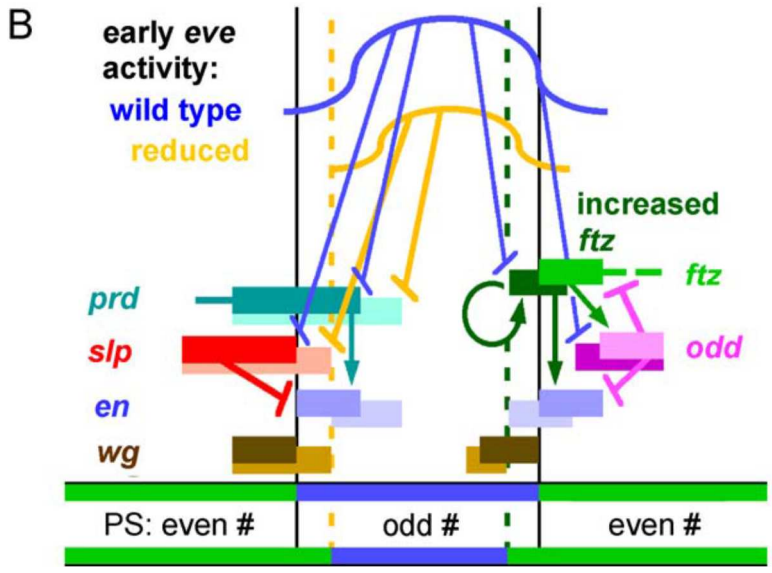


Figure 1-3. A model suggesting small changes in the expression of *eve* and *ftz* could have large effects on the borders of the parasegments. A change in the wild type expression of *eve* (blue wave and darker expression bars) to a reduced expression (yellow wave) creates a shift of its targets' expression to the posterior (lighter expression bars). Similarly, an increase in *ftz* expression (dark green expression bar) causes an anterior shift in the expression of its targets. These shifts in expression cause a change in the position of the border of the odd numbered parasegment. Figure from (Jaynes and Fujioka, 2004).

1.6 Pair-rule Conclusions

The network of pair-rule interaction to establish *en* expression has largely been elucidated through a large number of experiments that examined the effects of expression patterns of pair-rule genes in other pair-rule mutants. These mutant and expression pattern analyses are useful and relatively simple, but they cannot determine if control of pair-rule expression is direct or indirect. For example, it was thought that Ftz repressed *slp* until it was shown that this repression occurs through the activation of *odd* by Ftz, which then represses *slp*. Another difficult and unresolved case is the question of whether Runt directly represses *en* or does so through a cascade of repression through *eve* and *slp* (discussed in section 1.5). The only way to fully determine these network interactions is to determine the binding of the pair-rule transcription factor to the enhancer of its candidate target. However, enhancer analyses are much more difficult, as searching for enhancers is a time-consuming process. Even so, to have a complete understanding of the pair-rule network, it will be necessary to find enhancers for pair-rule genes, and after doing so find all functional binding sites within the enhancers.

To date, the body of work elucidating Ftz direct binding of targets is the best example of a complete set of experiments showing direct interaction. It has been shown that *ftz* autoactivates beyond doubt. First, the *ftz* enhancer was identified through a deletion analysis of the genome upstream from the transcription start site of *ftz*, and then it was shown that Ftz was necessary for proper *ftz* expression (Hiromi and Gehring, 1987; Hiromi et al., 1985; Ish-Horowicz et al., 1989; Pick et al., 1990b; Schier and Gehring, 1993). Finally, it was shown that Ftz binds directly to the *ftz* enhancer by

mutating Ftz binding sites, which resulted in a loss of enhancer-directed expression. This loss of expression was rescued by expressing a Ftz protein with altered DNA specificity that gained the ability to bind these altered Ftz sites (Schier and Gehring, 1992). Additionally, the direct interaction of Ftz with its target *en* has been thoroughly established. Again, first the expression of *en* was shown to be dependent on *ftz* (DiNardo and O'Farrell, 1987; Howard and Ingham, 1986), and then an *en* enhancer was found in the first intron of *en* that contains Ftz (and Ftz-F1) binding sites. When the Ftz binding sites were mutated, a loss of *lacZ* expression was observed in an *en-lacZ* reporter gene (Florence et al., 1997).

In this work, I set out to further understand how Ftz works with its cofactor Ftz-F1 to identify and activate its targets. To do this, we first identified Ftz/Ftz-F1 targets through a microarray experiment. We next identified candidate enhancers of these target genes and tested their function in vivo. Once bona fide enhancers were found, they were analyzed to learn more about what motifs they contain and which other transcription factors could be involved in Ftz/Ftz-F1 target binding. Learning more about how Ftz/Ftz-F1 bind their targets will give us the ability to more fully understand their place in the pair-rule gene hierarchy, as well as giving more insight into how other pair-rule genes may regulate both their pair-rule and segmentation targets.

Chapter 2: Cadherin Expression in the early *Drosophila* embryo

2.1 Introduction

In order to understand how pair-rule genes promote segmentation, the genes they regulate, which encode products that actually carry out this process, must be determined. In *Drosophila*, segments get established simultaneously during germband extension, but the morphological and molecular basis for this process is not understood. It is known that in order to expand, the cells of the germband intercalate, causing the germband to thicken and lengthen. Once intercalation has occurred, the cells that have moved into position next to each other remain together as the germband expands and retracts. This suggests that cell adhesion may be essential for this process, and so for the formation of segments.

As cell surface proteins that bind like proteins and can initiate signaling within the cell, cadherins have many functions in development, including cell-cell adhesion, cell sorting, boundary formation, and cell movement. Cadherins have an extracellular domain and an intracellular domain. The extra-cellular domain is often dependent on calcium and involved in binding other cadherin molecules of the same type. The intracellular sequence binds the actin cytoskeleton, both anchoring the cadherin to the cell and allowing the cadherin to signal cell shape changes and movement (reviewed in (Halbleib and Nelson, 2006). There are 17 cadherins in the *Drosophila* genome (Hill et al., 2001; Hynes and Zhao, 2000). Three of them, (Shotgun/DE-Cad, Cadherin-N/DN-Cad, CadN2), fall into the category of classical cadherins, which includes E-cadherin,

which functions in the epithelia, and D-cadherin, which functions in the nervous system. Seven of the *Drosophila* non-classical cadherins, which do not contain a catenin-binding domain (Fung et al., 2008; Hill et al., 2001; Tepass et al., 2000) can be categorized as Fat-like (Fat, Fat2, Dachsous), protein kinase cadherins (Cad96Ca and Ret), seven-pass transmembrane cadherins (Starry Night/Flamingo), and the calsyntenin cadherin Calsyntenin-1 (Zartman et al., 2009) . The remaining seven cadherins (Cad74A, Cad86C, Cad87A, Cad88C, Cad89D, Cad96Cb, Cad99C) are not categorized (Fung et al., 2008; Hill et al., 2001). Because of their function as cell-cell adhesion molecules, cadherins are candidates for the role of cell adhesion that holds cells together germband extension. Identifying such a role would be a first step towards understanding how pair-rule genes promote segmentation. To determine if cadherins are involved in segmentation, in situ hybridizations against the RNA for all *Drosophila* cadherins were performed in the embryo. It was determined that they did not have a striped expression pattern that could suggest a role in segmentation, and so are unlikely to be involved in segmentation.

2.2 Methods and Materials

Whole mount in situ hybridization

Standard protocols were followed for *in vivo* hybridization (Kosman and Small, 1997; Tautz and Pfeifle, 1989) with one modification - in place of Proteinase K treatment, embryos were heated at 95°C for 5 minutes. Digoxigenin-labeled RNA probes were made by isolating coding and/or 3'UTR sequences by PCR using a reverse primer with the T7 polymerase promoter sequence (TAATACGACTCACTATAGGG).

Sheep anti-digoxigenin (1:2000, Roche) was used for RNA probe detection. Stainings were visualized using DIC on a Leica DMRB microscope.

2.3 Results

With such functions as cell adhesion, sorting, and movement, it is not surprising that cadherins play a vital role in development. With so many roles and functions already known in development, it stands to reason cadherins may have important functions in the early *Drosophila* embryo. The first step to discovering this possible function was to examine the expression of the 17 *Drosophila* cadherins in the early embryo, from blastoderm formation to germband extension, when pair-rule genes are expressed in their peak patterns of seven stripes and when segmentation is determined A target of a pair-rule gene would be expressed in stripes corresponding to the pair-rule gene regulating it. Expression of the 17 cadherin-like genes was examined by in situ hybridization to early and late embryos. Pools of embryos, that were collected for both 0-6 hours and overnight, were fixed, and RNA probes were used against individual cadherin mRNAs to see where they were being expressed. Cadherin expression in the early *D. melanogaster* embryo for seventeen cadherins was examined (Fig. 2-1). Only one cadherin, *fat*, showed a reliable striped pattern, but it did not appear until the end of germ band elongation, after gastrulation was complete. With such late expression, it is unlikely *fat* is directly regulated by a pair-rule gene. Staining of an older embryo showed that *fat* was not a Ftz/Ftz-F1 target, as it is expressed on the anterior of each alternate segment at late stages, while Ftz expression is reduced to the posterior of each

segment. Because no other cadherin had a striped expression pattern, the cadherins do not appear to be targets of Ftz/Ftz-F1 or any pair rule gene.

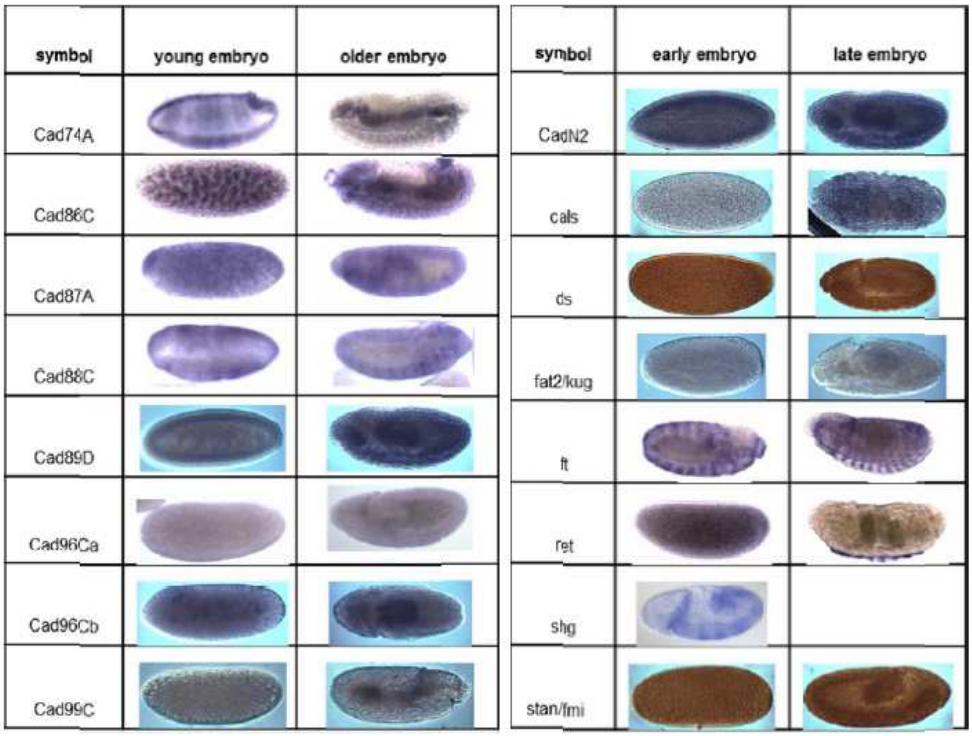


Figure 2-1. Expression of cadherins in D. melanogaster embryos. Each cadherin expression pattern is shown in both early and late embryos. *Cad74* showed faint striped staining at stage 8. The *shg* image, taken from the Berkely Drosophila Genome Project expression database, showed strong striped expression in embryos as early as the end of germband elongation.

2.4 Conclusion

Because the cadherins were not expressed in stripes patterns during blastoderm and gastrulation stages, it is unlikely they have a role in segmentation. Therefore, we did not continue with this project but instead decided to focus on finding targets of the pair-rule genes Ftz/Ftz-F1.

Chapter 3: Conservation of *ftz* in Diptera

3.1 Introduction

The diverse body plans found in the animal kingdom are established early in embryonic development by a small set of regulatory conserved genes (Carroll et al., 2005). How this small, conserved group of genes could control such different forms is a difficult question. The regulation of expression of these genes is one key feature influencing their ability to create diversity. A single gene can direct the formation of two very different structures in different contexts. For example, the human Pax6 gene, which controls the formation of a human eye, drives the formation of a fly eye when inserted into a fly (Halder et al., 1995). This experiment and many others have suggested that changes in the regulation of developmental genes allowed for their evolution. Because regulatory genes are often pleiotropic, meaning a gene is used at different times and for different purposes throughout development, even a small change to the coding region could have catastrophic consequences to the organism (Stern and Orgogozo, 2008). This reasoning suggested that a major driving force behind the evolution of developmental strategies has been change in the cis-regulatory elements (CREs) of these pleiotropic regulatory genes (Carroll, 2008; Prud'homme et al., 2007). Mutations in CREs allow changes in the timing and location of gene expression, giving new function to already existing genes. For example, the loss of limbs in snakes was not due to a loss of gene that controls the development of limbs, but instead to the expansion of the expression of a *Hox* gene that controls development of thoracic vertebrae into the forelimb region (Cohn and Tickle, 1999). Similarly, the *Hox* genes expressed in broad overlapping domains in the trunk of the ancestral brachiopod

crustacean *Artemia franciscana* evolved through changes in their CREs in at least two lineages, leading to compartmentally defined expression of *Ubx* and *Abd-A* in both the insects and the malacostracan crustacean *Porcellio scaber*, while the actual expression patterns evolved differently in both (Abzhanov and Kaufman, 2000). By changing the CREs, the proteins themselves were left unaffected, allowing them to continue their function in their often necessary ancestral roles while gaining new functions in other cell types.

However, changes to protein coding sequences are also essential during evolution, and several examples have been discovered to show this as well. For example, while CRE changes allowed for expression differences of Ubx to lead to a change in its role in trunk formation, changes in its protein sequences allowed it to gain a function - the repression of the limb-forming gene *Distal-less* (Grenier and Carroll, 2000). Ubx is expressed in both the thoracic and abdominal segments in crustaceans without affecting limb formation, allowing for varying numbers of limbs throughout. With the acquisition of the repression domain, Ubx became restricted to abdominal segments in hexapods, repressing limb formation in these segments but leaving limbs in the thoracic segment untouched (Galant and Carroll, 2002; Ronshaugen et al., 2002). Another example of a protein that changed function is the homeodomain transcription factor Ftz (Lohr and Pick, 2005; Lohr et al., 2001; Telford, 2000). In *Drosophila,* Ftz binds DNA with its cofactor Ftz-F1 (Florence et al., 1997; Guichet et al., 1997; Schwartz et al., 2001; Yu et al., 1997; Yussa et al., 2001) to activate genes involved in segmentation (Bowler et al., 2006; Hou et al., 2009) and the segment polarity gene *en* (DiNardo and O'Farrell, 1987; Florence et al., 1997; Harding et al., 1986; Hughes and

Krause, 2001; Kellerman et al., 1990). In *Drosophila*, *ftz* is essential for establishing the even numbered parasegments. When *Dm-ftz* expression is lost, the primordia of the even parasegments do not develop (Schwartz et al., 2001; Yu et al., 1997), and reciprocally, when it is ectopically expressed throughout the embryo, the primordia of the odd parasegments do not develop, a phenotype which has been dubbed the 'anti-ftz' phenotype (Struhl, 1985).

The function of *ftz* completely changed during arthropod evolution from an ancestral *Hox* gene to a pair-rule segmentation gene in *Drosophila* (Alonso et al., 2001; Gibson, 2000; Lohr et al., 2001). *ftz* first arose as a duplication of the *Hox* gene *Antp* (Telford, 2000). With a similar sequence and function to *Antp* and the other adjacent *Hox* gene *Scr* (Lohr et al., 2001), the *ftz*-ancestor had no constraint on its function and was free to evolve (Force et al., 2005; Ohno, 1970). Looking into the history of *ftz*, when and where it changed during the evolution of arthropods, gives insight into how genes are able to gain new roles. *ftz* changed both its expression and function to gain its new role as a segmentation gene. Homeotic genes are expressed in a single region of the embryo, where they function to define the identity of that body region (McGinnis and Krumlauf, 1992). In basally branching arthropods, *ftz* is expressed in a single broad region expected for a homeotic *Hox* gene, but in *Drosophila*, *ftz* is expressed in the typical seven stripe pattern of a pair-rule segmentation gene (Hughes and Kaufman, 2002; Janssen and Damen, 2006; Papillon and Telford, 2007; Telford, 2000). In addition to this change in expression pattern, the *ftz* protein sequence changed, allowing *ftz* to gain a role in segmentation. Specifically, Hox proteins find their specific target genes by binding DNA with a cofactor Extradenticle, and they bind Exd through a

YPWM motif upstream of the homeodomain (Johnson et al., 1995; Mann and Chan, 1996; Passner et al., 1999; Zhao et al., 1996). *Drosophila* Ftz lost this YPWM motif and so lost its potential to function as a homeotic gene (Lohr et al., 2001). Instead, Ftz gained an LRALL motif, which allowed it to bind to a new cofactor, Ftz-F1, giving it specificity to bind new target genes (Schwartz et al., 2001; Yu et al., 1997; Yussa et al., 2001). With this new ability, *ftz* gained the potential to regulate segmentation (Lohr et al., 2001) and was incorporated into the segmentation pathway (Lohr and Pick, 2005), where it became necessary for body plan formation. It still needs to be determined when in evolution these changes took place. Did the change in expression pattern come first, or the change in protein sequence? Could an animal survive a change in expression pattern before a change in function, or vice versa? Answering these questions will help clarify how it is possible for genes to gain new functions without harming the complex organism in which the change occurred.

To study the change in expression and function of *ftz*, the *ftz* expression pattern and DNA sequence has been determined in several species spanning Arthropoda (Heffer et al., 2010). It was shown that the first step was the change from a *Hox*-like expression pattern to stripes. *ftz* lost its *Hox*-like expression in Crustacea and then gained striped expression at some point within Insecta, although the striped expression was lost again within this clade as *ftz* is not expressed in stripes in all derived insects. Along with this, the YPWM motif degraded at least eight separate times, first seen in Pancrustacea. Finally, the LXXLL motif was stably gained at one point during evolution, at the base of holometabolous insects. This allowed Ftz to interact with its new cofactor Ftz-F1. These changes allowed for the loss of homeotic potential and the

gain of segmentation potential. However, these studies did not show the actual function of Ftz in the species harboring different forms of Ftz, rather they used ectopic expression in *Drosophila* to assess functional potential. Functional studies outside of model species like *Drosophila melanogaster* is challenging. It is now possible to use RNAi to knock down gene expression through feeding or injection in many species, and this technique is extremely useful in functional studies. Still, it is better to create stable lines with a permanent gene alteration to assess gain and loss of function phenotypes.

In this study, I generated transgenic *Aedes aeypti* (*Aa*), the mosquito vector for the viruses that cause Dengue and Yellow Fever, to determine the expression and function of *ftz* within Diptera. *Aedes* diverged from *Drosophila* 210 million years ago (Yeates and Wiegmann, 1999), at the base of Diptera, such that *Aedes* and *Drosophila* represent diverse branches. I reasoned that if the expression and function of *ftz* is conserved between *Drosophila* and *Aedes*, it is likely conserved throughout all of Diptera. The *Aedes* genome was already sequenced and annotated (Nene et al., 2007), in situ hybridization techniques were worked out (although are still unreliable), and ability to make transgenic mosquitoes already was well honed, making *Aedes* a good candidate for this study. However, in situ techniques proved to be quite difficult. To determine the expression of *ftz* and its cofactor *ftz-f1*, reverse transcriptase (RT)-PCR was performed for several time points in the early embryo, showing a similar expression progression to *Drosophila*. To assess the function of *Aa-ftz,* a transgene was generated in which *Aa-ftz* was placed downstream of a heat shock promoter. Transgenic mosquitoes were generated in order to make animals that overexpress *ftz*. We attempted to determine the function of *ftz* by overexpressing *ftz* by heat shock and then observing any segmentation

abnormalities, either by observing developed embryos or observing the expression pattern of a known Ftz target in early embryos. Unfortunately, technical difficulties with in situ hybridization precluded an analysis of embryonic phenotypes.

3.2 Methods and Materials

Mosquito rearing

Mosquitoes were kept at 28º C and 68% humidity with a 12 hr. light: 12 hr. dark cycle. Adults were fed with cotton balls soaked in 10% glucose and laid on the mesh top of a cage. To induce egg laying, adults were starved overnight, and were then fed bovine blood. The blood was first heated in an incubator and then aliquoted into tubes with parafilm stretched over one side to simulate a skin membrane. Eggs were collected by placing cups with a small amount of water in the bottom and filter paper on the sides, constantly wet from the water, into the adult cages. To induce larger amounts of egglaying in a short amount of time, bloodfed females were removed from the large cage and placed in a small collection cup, with water covering the bottom and filter paper on the sides, to encourage them to lay more eggs. Filter paper with eggs were placed in larger trays filled with water, where larvae would live when they hatched. Larvae were fed by pouring a small amount of ground dog food on the surface of the water, after which it would sink to the bottom. Pupae were selected and moved to cages to eclose as adults.

Generation of *Aedes* transgenic line

Aedes aegypti ftz was isolated from genomic DNA by PCR, using the primers 5' CTAGAA GCTTATGGCAGCCGCATATCCATAT and 5' GATCAAGCTTGATCAGAGGTCAATAAA AAAATGTTTGCTGATAT. The gene was inserted into pCaSpeR-hs in order to gain the *Hsp70* heatshock promoter. The heatshock promoter+*ftz* was cut from the plasmid and inserted into pBac[3xP3-EGFafm], a plasmid used for mosquito transformation as it has a high rate of insertion into the genome (Kokoza et al., 2001). This construct was designated *hs-ftz*.

Transgenic mosquito lines were generated by the University of Maryland IBBR Insect Transformation Facility. Two transgenic lines were established from this transformation. The 3xP3-EGFP marker in the piggyback plasmid allowed for mutant animals to be selected under a fluorescent microscope, as it expresses EGFP in the eyes the larval stage. Selection of fluorescing larvae continued with each generation until all larvae were fluorescing for three generations in a row. Once this occurred, the lines were thought to be homozygous. Lines were occasionally tested for homozygosity by back crossing with wild type mosquitoes.

Reverse Transcriptase-PCR expression profile

Aedes aegypti embryos were collected for 3 hours and aged to five different ages – 0-3 hours, 3-6 hours, 6-9 hours, 9-12 hours, and 12-15 hours. The embryos were ground in RNA*later* Solution with mortar and pestle and stored overnight at -20^{o}C. RNA was extracted using the Omega Microelute Total RNA Kit, then converted to cDNA using the Qiagen Quantitect Reverse Transcription Kit. Primers were designed to amplify

three genes out of this cDNA. The primers were designed to span introns so that any remaining gDNA would not be amplified. The sequences were: *actin* – 5' TGACTCAGATCATGTTTGAGACCT and 5' GTTGCCAATGGTGATGACCT, *ftz* – 5' GTTAATACTCCGCCACTTTCAC and 5' GCCAAATTGTAATCCAACTGAG, and *ftz-f1* – 5' GGACAAGATCTCCGGCTTCC and 5' TGCGATCCTCACGAATTGCT. These primers were used to amplify *actin, ftz*, *and ftz-f1* from the cDNA of each timed collection, to examine the expression during the first fifteen hours of *Aedes aegypti* embryonic development. The PCR conditions for all four primer sets were: 25 cycles, annealing temperature of 55^{o} C with a one minute elongation time.

In situ hybridization and antibody staining

Embryos were fixed and dissected as described by others (Clemons et al., 2010b). Established protocols were followed for in situ hybridization (Haugen et al., 2010) and antibody staining (Clemons et al., 2010a). RNA probes were transcribed using template generated by PCR amplification of genomic DNA, using primers with T7 polymerase sequence added to the 5'end of the reverse primer. For antibody staining, *engrailed* antibody 4F11 (Patel et al., 1989) was used.

Heatshock

To express *Aa-ftz* ectopically, embryos were collected for three hours and then aged to the appropriate time. To heatshock, embryos were removed from filter paper using a paintbrush and placed in a scintillation vial with water. The vials were placed in a 37^{o} C water bath for the appropriate time. To remove the embryos from the vial, the water was

poured over a new filter paper and placed back into a cup with shallow water to allow aging to continue. Aging progressed either until germband extension (9-12 hours), at which point embryos were fixed for *en* antibody staining, or until they were fully developed (72 hours), at which point embryos were dissected to observe phenotypes.

3.3 Results

3.3.1 *Drosophila* and *Aedes* Ftz have similar protein sequences

In order to study the *Aa*-Ftz sequence, the genomic sequence was obtained from the published *Aedes* genome (Nene et al., 2007). An alignment between the protein sequences of *Dm*-Ftz and *Aa*-Ftz shows that the sequences essential to Ftz function are highly conserved between the two species (Fig. 3-1). The first nine nucleotides of the homeodomain, which confer binding specificity to Ftz (Heffer et al., 2010; Telford, 2000) are identical. Comparing the entire homeodomain, the amino acid sequences are 75% identical and 95% similar. In addition, both have lost the YPWM, which binds the cofactor Exd and gives homeotic potential, and gained the LRALL motif, which confers the ability to bind the cofactor Ftz-F1 that is necessary for Ftz segmentation potential. Around the LRALL motif, an additional 3 amino acids are shared downstream – TNP, and the two amino acids upstream are the same but reversed – PS in *Drosophila* and SP in *Aedes*. This shows that conservation remained high in the area immediately surrounding the functional LXXLL motif. The remainder of the protein sequence shows little sequence conservation, consistent with the notion that these sequences are not necessary to conserve Ftz function (Heffer et al., 2010). This similarity between the

sequences of the two Ftz proteins suggested that *Aa*-Ftz would have the same segmentation function as *Drosophila* Ftz. To test this hypothesis, we set out to analyze the expression and function of *Aa-ftz*.

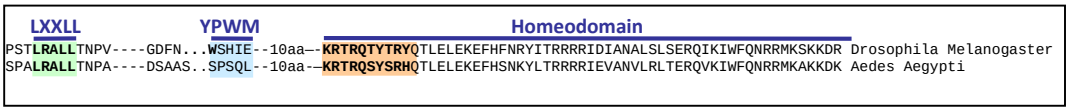


Figure 3-1. Amino Acid alignment of *Dm*-Ftz and *Aa*-Ftz. A protein alignment between *Dm*-Ftz and *Aa*-Ftz is shown. Strong conservation of functional sequences is evident. The homeodomains are 75% similar and 95% identical. Both contain the Ftz-F1 binding motif LRALL, while both have lost the Exd binding motif YPWM.

3.3.2 *ftz* and *ftz-f1* expression in *Aedes* is consistent with expression in *Drosophila melanogaster*

RNA was extracted from *Aa* embryos at five timepoints covering development through germband extension (15 hrs.) and then converted to cDNA. From this cDNA, RT-PCR was performed to detect expression of *Aa-ftz* and *Aa-ftz-f1*, using *Aa-actin* as a control to determine overall RNA level from each sample. Then the samples were run on an agarose gel (Fig. 3-2). *Aa-ftz* expression began and was strongest during 3-6 hours AEL and continued weaker 6-15 hours AEL. *Aa- ftz-f1* was also expressed during the entire 15 hours, with highest expression between 3-9 hours AEL. *Aa-ftz-f1* expression decreased during 9-15 hours AEL. The timeline for *Aa-ftz* expression began in blastoderm and then decreased and continuted through germband extension, which matches the expression timeline for *Dm*-ftz. *Aa-ftz-f1* expression also matches the *Dm-ftz-f1* expression timeline, which is maternally deposited and continues to be ubiquitously expressed in the embryo until it begins to fade during germband extension.

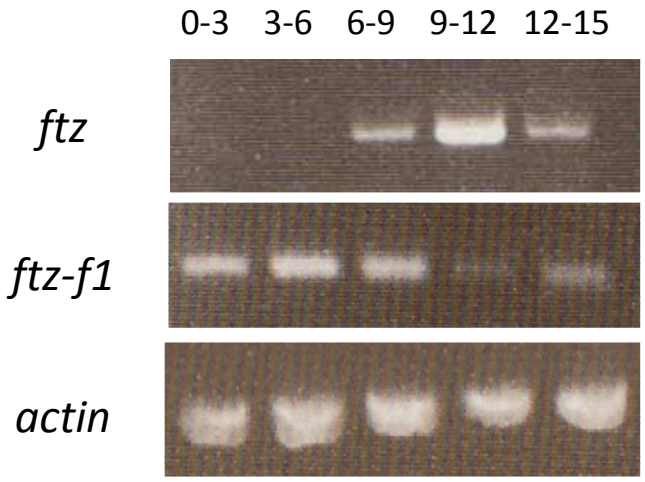


Figure 3-2. Expression profile of *ftz* and *ftz-f1* in early *Aedes* embryos. cDNA was collected from *Aedes aegypti* embryos aged to 5 different time points and RT-PCR was performed to examine the expression levels across the first 15 hours of embryonic development. a*ctin* was used as a control across all time points. *ftz* was most highly expressed at 9-12 hours AEL and *ftz-f1* was most highly expressed at 0-9 hours AEL.

3.3.3 *engrailed* is likely expressed in stripes in *Aedes*

The *engrailed* (*en*) gene is expressed in segmental stripes in developing embryos of all insects examined (Patel et al., 1989), including the mosquito *Anopheles gambiae* (Yoder and Carroll, 2006), and thus serves as a good control for establishing in situ hybridization and antibody staining in *Aedes* embryos. In addition, in *Drosophila*, Ftz and Ftz-F1 directly regulate *en* gene expression (Florence et al., 1997). After ectopic expression of *Dm-ftz*, *en* expression is altered (Ish-Horowicz et al., 1989; Struhl, 1985). Thus, En expression also serves as a marker for a potential anti-ftz phenotype in *Aedes*.

In previous studies, the commercially available anti-EN antibody 4D9 was tested on *Aedes* embryos. None of these stainings were successful (personal communication with Molly Duman-Scheel). Therefore, I tested the anti-EN antibody 4F11 (Patel et al., 1989), which had been used successfully in other insect species.

Antibody staining proved to be difficult, giving unreliable results. Figure 3 shows an attempted staining with the 4F11 antibody. Arrows point to potential stripes, although no stainings were clear enough to give definite results.

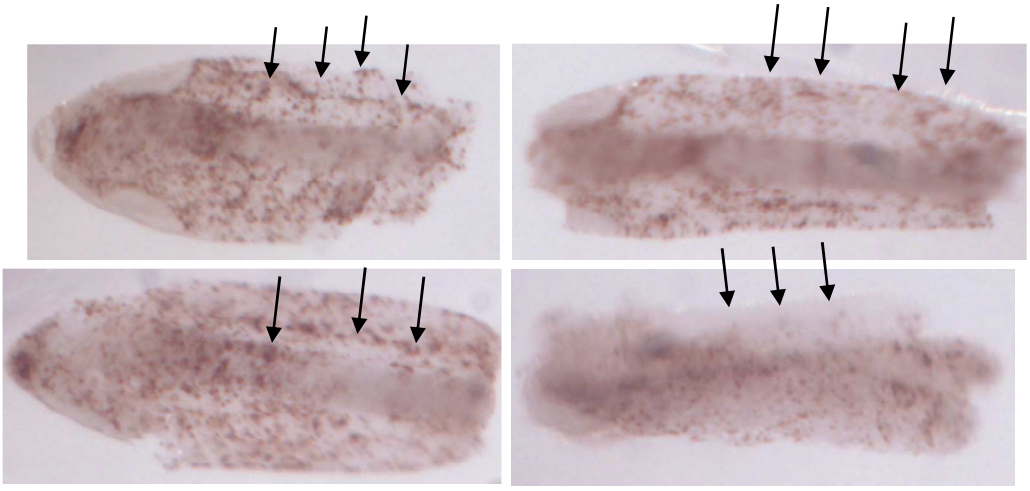


Figure 3-3. Antibody staining of *Aedes* embryos with anti-*en* 4F11. Anti-*en* staining of embryos during germband extension shows potential stripes, indicated by arrows. High amounts of background make it difficult to determine whether stripes are present or an artifact of staining.

Embryos were then stained with the 4F11 antibody along with the fluorescent nuclear dye Sytox green, to allow visualization of the embryo and germband within the yolk (Fig. 3-4). Color and fluorescence pictures were taken before (Fig. 3-4A and 3-4C) and after (Fig. 3-4B and 3-4D) dissection of two embryos, one of which is shown in Fig. 3-4A and 3-4B and the other is shown in Fig. 3-4A and 3-4C. Because patterns were difficult to assess on whole embryos due to a spotty background, embryos were dissected under GFP fluorescence to try to separate the embryo from the yolk. Again, arrows point to potential, but unclear, stripes (Fig. 3-4C and D). Interestingly, it appears that these potential stripes are not only expressed in the germband, but in some embryos

seem to spread across the entire width of the embryo (Fig. 3-4C). However, despite repeated attempts, and some promising preliminary results, neither in situ hybridization (not shown) nor antibody staining (Fig. 3-3) worked well enough in early embryos to be useful for assessing the effects of ectopic *Aa-ftz* expression.

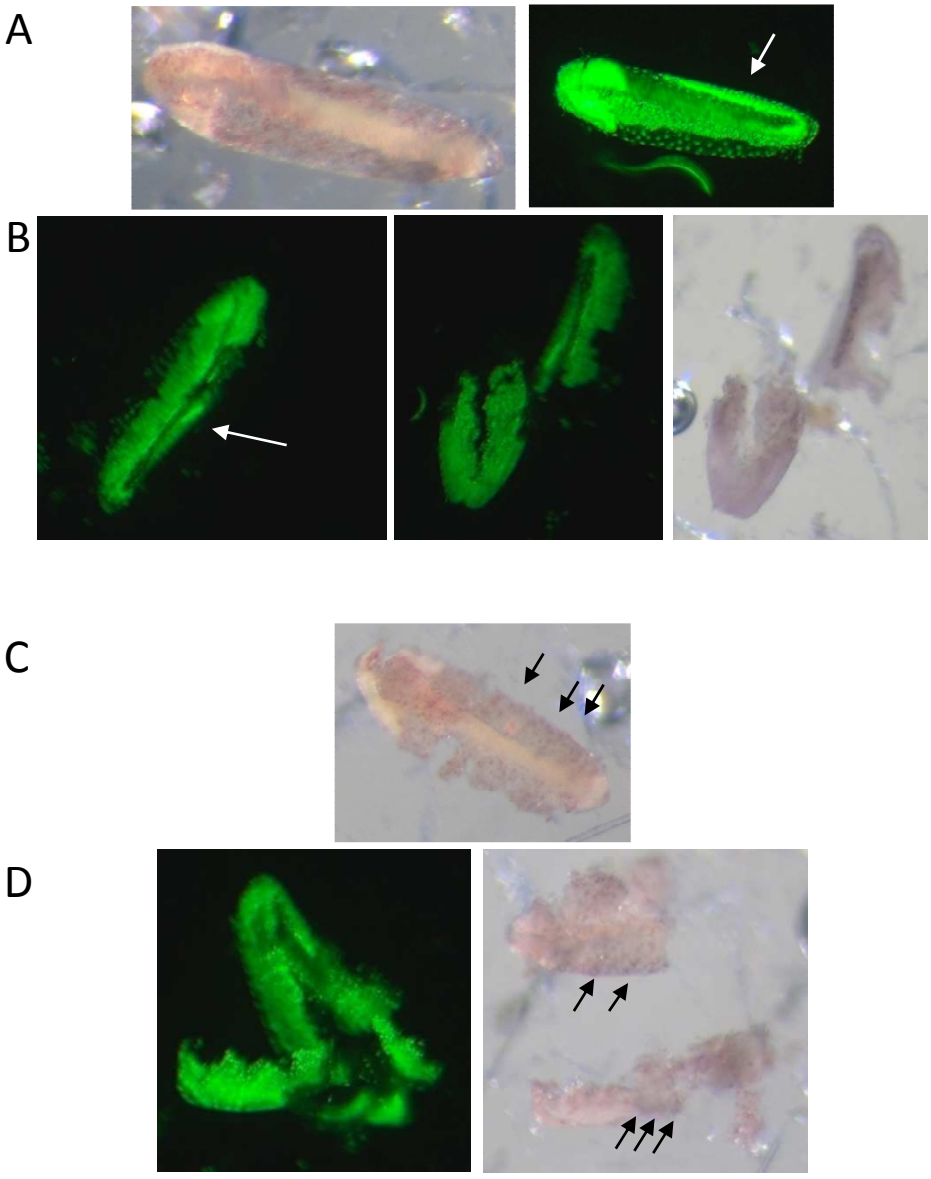


Figure 3-4. Embryo dissection minimally improves visualization of staining. To aid in visualization of staining of the embryo germband, dissection of the germband out of the yolk were performed. Two dissections of anti-*en* and Sytox Green stained wild type embryos are shown. Nuclear stain Sytox Green was used to visualize germband, indicated by white arrow in A) and B). Potential *en* stripes are indicated by black arrows in C) and D).

3.3.4 Overexpression of *ftz* by heatshock indicates some lethal effect

Both wildtype and *hs-Aa-ftz* embryos were collected and subjected to heatshock for various times. Embryos were then either allowed to hatch to determine effects on survival, or they were fixed at blastoderm to germband stages for analysis of gene expression patterns or when nearly fully developed as larvae for segmentation analysis by dissection from the chorion. Table 3-1 shows survival of embryos after heat shock. There was a large difference in the survival rate of wildtype embryos compared to those in which *Aa-ftz* was overexpressed. When heat shock was performed after 9-12 hours AEL (germband extension), when endogenous Aa-*ftz* expression is strongest (Fig. 3-2), wildtype embryos had a 44% survival rate, while Aa-*ftz* transgenic embryos had a 0% survival rate. However, dissection of heatshocked embryos that were fixed after 70+ hours of development did not show any segmentation defects (Fig. 3-5). Next, heatshocks were performed during blastoderm and gastrulation (6-9 hr), to determine if a segmentation phenotype would be observed if *ftz* was overexpressed at the onset of *ftz* expression (Fig. 3-2). In this case, 17% of wildtype embryos died before developing segments, while 11% of Aa-*ftz* transgenic embryos died early. These numbers are too similar to be meaningful.

Table 3-1. Overexpression of *Aa-ftz* suggests a lethal effect

	WT			FTZ		
age	total	% survival	% developed	total	% survival	% developed
9-12 hr	54	44		8	0	
6-9 hr	29		83	28		89

Again, dissections of embryos that developed in a range from blastoderm to the beginning of germband extension (6-9 hrs AEL) and were then subjected to heatshock did show some developmental defects, but no segmentation defects (Fig. 3-5). Fig. 3-5A and B shows wild type and *Aa-ftz* embryos respectively that were heatshocked for 15 minutes. The wild type embryos were overall healthier and more whole upon dissection, while the *Aa-ftz* embryos looked less well formed. This pattern continues in embryos that were heatshocked for 40 min (Fig. 3-5C and D). While both wild type and *Aa-ftz* embryos look to be more affected by this longer heatshock, the *Aa-ftz* appear to have more defects than wild type. Note that *Aa-ftz* embryos that were heatshocked for 40 min are severed due to dissection, and not as an effect of heatshock. However, these embryos were much more delicate and so difficult to dissect, which was an effect of heatshock found much stronger in *Aa-ftz* embryos compared to wild type. No definite segmentation defects, which would be indicative of a pair-rule phenotype, were observed.

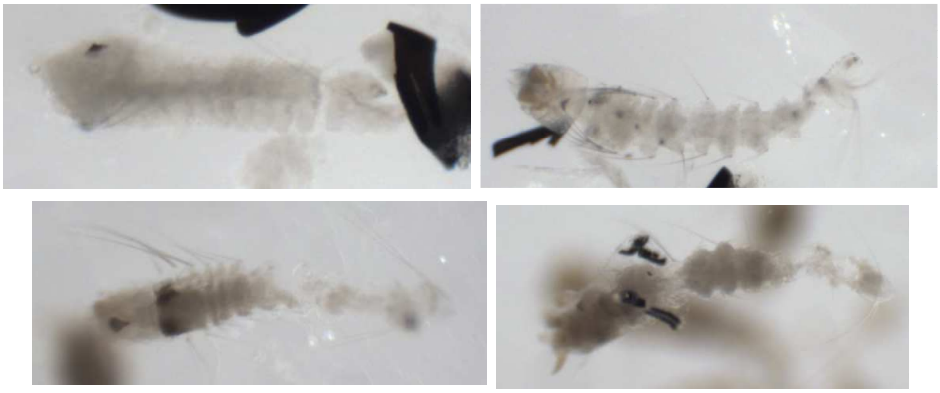
Figure 3-5. Overexpression of *Aa-ftz* results in some developmental defects. Heatshock was performed on wild type and *Aa-ftz* embryos after aging them to blastoderm and gastrulation stages (6-9 hrs). Heatshock was performed for A,B) 15 min and C,D) 40 min. Wildtype embryos heatshocked for 15 minutes were largely unaffected (A) and for 40 min had mild defects (C). *Aa-ftz* embryos heatshocked for 15 minutes had mild defects (B) and for 40 min had stronger defects (D).

A

WT
Hs 15 min

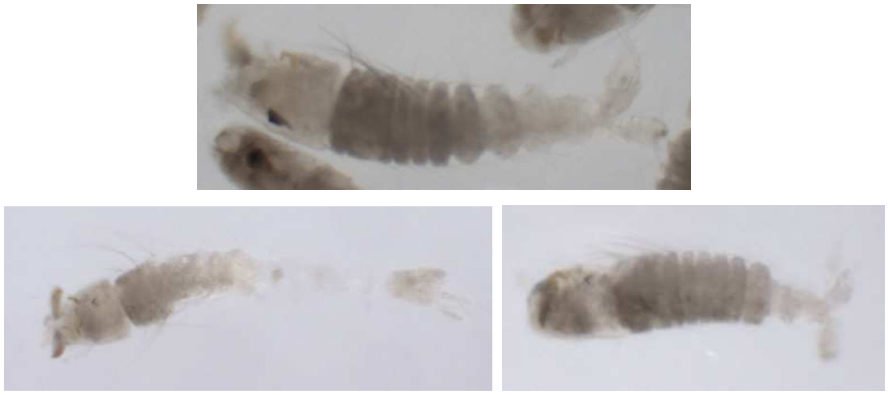
B

Aa-ftz
Hs 15 min



C

WT
hs 40 min



D

Aa-ftz
hs 40 min

3.4 Discussion and Future Directions

The mosquito embryo develops approximately three times slower than the *Drosophila melanogaster* embryo. In *Aedes aegypti,* blastoderm stage ends at 9 hours after egg laying (AEL), germband extension begins around 9 hours and is finished around 14 hours (Rezende et al., 2008). In *D. melanogaster*, the blastoderm stage ends around 3 hours AEL, and germband extension goes from 3 hours to 5 hours AEL. The expression of *ftz* and *ftz-f1* also follow this same pattern, being expressed about 3 times later and during the same stages in *Aedes* as in *Drosophila* (Fig. 3-2). In *Aedes aegypti*. *Drosophila ftz* is expressed from the blastoderm stage to the beginning of germband extension, 2-4 AEL (Hafen et al., 1984). Mosquito *ftz* expression also begins during the early blastoderm stage (3-6 hours), decreased some through gastrulation (6-9 hours), and continued through germband extension, which matches the expression timeline for *Dm*-ftz. In flies, *ftz-f1* is maternally deposited, so expression begins immediately. In *Aedes*, *ftz-f1* expression is strong at 0-9 hours and then begins to fade. The weaker *Aa-ftz-f1* band for 9-12 hours can be explained by a lower RNA concentration for the collection at this time point, as shown by weaker band for the actin control. The expression profile of *ftz* and its cofactor *ftz-f1* in *Aa* from RT-PCR matches the expression file of these genes known for *Drosophila* from in situ hybridization techniques (Yu and Pick, 1995), suggesting that Aa *ftz* would have the same expression as *Dm-ftz*. Attempts at performing an in situ hybridization with a *ftz* probe were unsuccessful. However, knowing the expression pattern of *Aa-ftz* in the early embryo could be indicative of potential segmentation function, so it would be valuable to continue this experimental line.

The sequence of *Aa*-Ftz suggests that it will have the same function as *Dm-ftz* (Fig. 3-1), but experiments must be performed to assess its functionality. The evidence collected in this project does indicate that *Aa-ftz* has a segmentation function, but experiments were not conclusive. The lethality of *ftz* overexpression indicates that overexpressing this gene does have a serious effect on development (Table 3-1). This is encouraging, but without knowledge of what those effects are, it is impossible to conclude that *ftz* has the same segmentation function in mosquitoes that it has in flies. In sum, functional studies using heatshock to overexpress *ftz* have been promising but inconclusive**.**

A segmentation function for *Aa-ftz* does appear likely, and the question is certainly worth pursuing. The next step should be to continue varying heat shock times in both wild type and Aa-*ftz* transgenic embryos, and then to determine what phenotypes result from *Aa-ftz* overexpression. While analyzing *en* expression to determine effects on segmentation earlier in the development of the embryo would be the most effective way of determining roles in segmentation, the difficulty of staining protocols in *Aedes aegypti* may make this method too difficult to accomplish. The embryos could instead be allowed to fully develop and then dissected to analyze morphological defects. Alternatively, in situ hybridization and antibody stainings are much easier to perform in the closely related mosquito species, *Anopheles gambiae* (personal communication with and unpublished protocol from J. Juhn (Juhn and James, 2006), due to differences in their chorions. Attempting this same experiment in *Anopheles* may be a more feasible experimental approach.

Another informative experiment that could be performed in the future would be to use RNAi to knockdown *ftz* expression in mosquito. Because the phenotypic effect would occur in the embryo, the dsRNA would have to be injected in the very early embryo, before *ftz* expression begins during blastoderm. This protocol has been established in mosquitoes (Caplen et al., 2002; Clemons et al., 2010c) and so should be possible. I hypothesize that this would give the opposite phenotype from overexpression of *ftz*. That is, the primordia of the even numbered segments would be lost in these knock down embryos, as they are in *ftz* mutant *Drosophila* (Schwartz et al., 2001; Yu et al., 1997). Ideally, both the overexpression of *ftz* through transgenic lines and the loss of *ftz* through RNAi would be performed, giving a complete picture of *ftz* function in mosquitoes.

Chapter 4: The Systematic Identification of *Drosophila* Ftz/Ftz-F1 Target Genes

4. 1 Introduction

Highly conserved cascades of regulatory genes control embryonic development of diverse animal species. The regulatory genes are often members of large families of DNA binding transcription factors that control the expression of larger sets of so-called downstream or target genes that are involved in cell growth and differentiation. These regulatory transcription factors activate or repress expression of these sets of downstream targets via sequence-specific DNA binding in the genome. Understanding the mechanisms used by embryonic transcription factors to identify genomic binding sites is the key to understanding their function.

The Hox family of transcription factors regulates segmental identity but controls diverse processes in all metazoans (McGinnis and Krumlauf, 1992). Hox proteins share a highly conserved DNA binding domain, the homeodomain, a sequence-specific DNA binding domain (Akam, 1989b; McGinnis and Krumlauf, 1992). However, these homeodomains in different proteins do not have specific binding sites that are easily distinguished, all binding very similar DNA sequences (Berger et al., 2008; Mann et al., 2009; Noyes et al., 2008). This contrasts with the unique biological activities seen for individual Hox proteins in vivo. This 'Hox paradox' was thought to have been resolved by the identification of binding partners for Hox proteins (Mann et al., 2009; Moens and Selleri, 2006). Many Hox proteins were found to form heterodimers with a divergent homeodomain-containing protein, Extradenticle (Exd) through the protein interaction motif YPWM upstream of the homeodomain, which can be thought of as 'homeotic

Hox genes' (Chan et al., 1994; Mann et al., 2009; Passner et al., 1999; Phelan et al., 1995). Hox-Exd complexes bind to specific ten base pair DNA sequence TGATNNATNN (Chan and Mann, 1996; Lu et al., 1995; Mann and Chan, 1996; Piper et al., 1999)**.** While these findings explain some of the specificity of Hox regulation, it is surprising that so many Hox proteins can heterodimerize with the same partner and maintain such diverse specificity. Recent reports suggest that Hox-Exd specificity is achieved through the different protein complexes' ability to distinguish between different low affinity binding site (Crocker et al., 2015), but Ftz/Ftz-F1 use a completely different method to gain specificity (Yu et al., 1997). This may be because having a unique cofactor confers all the specificity needed, so low affinity binding sites that the HOX-Exd uses to distinguish different target sites is not needed. Thus, the *Hox* gene *ftz* has evolved its own method for achieving specificity.

In contrast to the *Hox* genes in *Drosophila* that contain this homeodomain, the *Drosophila fushi tarazu* (*ftz*) shares the chromosomal location and homeodomain of *Hox* genes but is expressed in stripes (Hafen et al., 1984) and functions as a pair-rule segmentation gene. It appears to have lost the ability to functionally interact with Exd in vivo, and instead *Drosophila* Ftz interacts with the orphan nuclear receptor Ftz-F1 (Florence et al., 1997; Guichet et al., 1997; Schwartz et al., 2001; Yu et al., 1997; Yussa et al., 2001). Many studies have shown that Ftz and Ftz-F1 are obligate cofactors that interact directly to promote segmentation in the early *Drosophila* embryo. Ftz and Ftz-F1 physically interact both in vitro and in vivo in wild type *Drosophila* embryos; they bind synergistically to DNA containing composite Ftz/Ftz-F1 binding sites, and coordinately activate the transcription of target genes. Importantly, *ftz-f1* mutants

display the same pair-rule phenotype as *ftz* mutants, with loss of alternate body segments, despite the fact that Ftz-F1 is expressed ubiquitously, present in all somatic nuclei of blastoderm embryos (Guichet et al., 1997; Yu et al., 1997; Yussa et al., 2001).

One of the best characterized targets of Ftz and Ftz-F1 is *ftz* itself, which autoregulates gene expression through independent enhancers in the upstream element (Hiromi 1987, Pick 1990). Mutation of the three Ftz-F1 binding sites in a core region of the upstream element resulted in loss of expression of a *lacZ*-reporter gene, as did mutation of the Ftz binding sites (Schier and Gehring, 1993). Similarly, Ftz and Ftz-F1 coordinately regulate the expression of *engrailed* (*en*) through binding to composite sites in an intronic cis-regulatory element (Florence et al., 1997). *en* is expressed in the even numbered parasegments (Florence et al., 1997), and the pair-rule gene *even-skipped* establishes *en* expression in odd parasegments by regulating the activator *paired* and the repressors *runt* and *sloppy-paired* (Fujioka et al., 1995). *en* and another segment polarity gene, *wingless*, are expressed in alternating stripes, which establish and define the borders of the parasegments (DiNardo et al., 1988; Heemskerk et al., 1991; Martizez Arias et al., 1988; Vincent and O'Farrell, 1992). Together, these studies demonstrated that the function of Ftz and Ftz-F1 in the early *Drosophila* embryo are tightly linked, both being dependent on each other to create the pair-rule segment formation.

The findings summarized above suggest that the unique regulatory specificity of Ftz can be explained by its interaction with a unique co-factor, Ftz-F1, which is distinct from the homeotic Hox protein interaction with Exd. Ftz-F1 has a longer and more specific DNA binding site than do Hox proteins, increasing the specificity of the Ftz-

F1/Ftz heterodimer in vivo. A computational screen identified 30,000 potential Ftz-F1 binding sites in the *D. melanogaster* genome while over 14 million candidate Ftz binding sites were found (Bowler 2006). Thus, the requirement for Ftz-F1 binding site would restrict Ftz binding to a much smaller set of potential sequences in the genome. Further, Ftz-F1 alone is a strong transcriptional activator in vitro(Yussa et al., 2001), and it was shown that a Ftz protein lacking its homeodomain can carry out many of the functions of full length Ftz protein, likely due to its interaction with Ftz-F1 (Copeland et al., 1996; Fitzpatrick et al., 1992). These studies all support the notion that Ftz-F1 plays the major role in the function of the Ftz-F1/Ftz heterodimer. However, because Ftz-F1 is expressed ubiquitously while Ftz expression is restricted to seven pair-rule stripes, it is possible that Ftz-F1 is able to regulate some target genes in Ftz- cells.

In order to characterize how Ftz/Ftz-F1 regulate their targets, previous attempts to find targets and their Ftz-F1 responsive enhancers have been attempted, with some success. Bowler (2006) tried a computational screen, using two known configurations of composite Ftz/Ftz-F1 sites - binding sites for *ftz* itself (the F1F-site) (Yu 1997, Yussa 2001) and *engrailed* (called the en-site) (Florence et al., 1997). Ftz uses the same core binding site as other Hox genes, ATTA, while the core binding sequence found in all Ftz-F1 binding sites is AAGG (Florence et al., 1997; Han et al., 1993; Ueda et al., 1992). By searching for the Ftz/Ftz-F1 binding site that matched these configurations in the entire genome, the authors hoped to identify bona fide Ftz/Ftz-F1 targets. The screen did find two new Ftz/Ftz-F1 targets, but only two of the thirty candidate targets tested were actual targets.

In another study, in situ hybridization data from the Berkley Drosophila Genome Project was used to identify candidate Ftz/Ftz-F1 targets on the basis of their striped expression patterns (Hou et al., 2009). Next, the genomic regions surrounding the genes expressed were examined for potential Ftz and Ftz-F1 binding sites. Three targets were identified in this study: *drm*, *noc*, and *5-HT2*. Two Ftz/Ftz-F1 responsive enhancers of *drm* were found that directed expression in the striped expression during blastoderm, late gastrulation, and early germband extension. Further, expression of *drm-lacZ* fusion genes was lost in *ftz* or *ftz-f1* mutant embryos, and point mutations of Ftz-F1 sites within these two enhancers constructs abolished reporter gene expression, showing *drm* to be a direct target of Ftz/Ftz-F1 (Hou et al., 2009). Together, these studies identified seven targets of Ftz/Ftz-F1: *ftz* itself (Florence et al., 1997; Yu et al., 1997), *en* (Florence et al., 1997),*apt*, *Sulf1* (Bowler et al., 2006), *drm*, *noc*, and *5-HT2* (Hou et al., 2009). For only two of these, Ftz/Ftz-F1-responsive enhancer regions have been identified. However, it is likely there are many more targets as yet undiscovered. The methods used in both of these studies proved to be inefficient, and their results indicate that there is more information about the in vivo target site recognition for Ftz/Ftz-F1 than could be predicted by their binding sequences alone.

Here we have combined computational and experimental approaches to find an efficient method for identifying Ftz/Ftz-F1 targets and their Ftz/Ftz-F1 responsive enhancers in the *Drosophila* genome. A microarray analysis was used to identify genes regulated by *ftz-f1* in early embryos. Expression of nine of the top ten genes overlapped with that of Ftz, suggesting they could be coordinately regulated by Ftz and Ftz-F1. The expression of these nine candidate Ftz-F1 targets was lost in *ftz* and *ftz-f1* mutant

embryos, also indicating that they are coordinately regulated by Ftz and Ftz-F1. To determine whether these are direct Ftz/Ftz-F1 targets, the DNA surrounding these targets was examined for Ftz binding and Ftz-F1 binding sites. These enhancer regions were analyzed for additional motifs and transcription factor binding sites that could provide additional information on how Ftz/Ftz-F1 bind to their enhancer sequences. The DNA- binding proteins Deaf-1 and Zeste were identified as potentially necessary for Ftz/Ftz-F1 binding, while Dichaete and GAGA Factor were identified as potential repressors. This work has identified new Ftz/Ftz-F1 targets and enhancers, which will help give insight into how these transcription factors work together to activate their targets and elucidate a complicated system of regulation.

4.2 Materials and Methods

Fly line and genetics

Flies were maintained at 25 °C on a standard diet. The *ftz* mutant was *ftz*^{9H34}/*TM3Ser, hb-lacZ,* with expression of β-galactosidase used to identify mutant embryos. Embryos derived from *ftz-f1* germline clones (referred to as *ftz-f1* mutants) were generated with the autosomal FLP-DFS technique (Chou and Perrimon, 1992, 1996; Chou et al, 1993) using *ftz-f1*^{19} (Broadus et al., 1999; Fortier et al., 2003; Pick et al., 2006). Transgenic fly lines were generated by Rainbow Transgenic Flies, CA and BestGene, CA. The PhiC31 integration system was used to insert transgenes into the genomic attP site VK00022. Transgenic lines were maintained as homozygotes. One

transgenic line, containing the *blot* enhancer construct, was homozygous lethal but was homozygous viable when crossed into a *ftz* mutant background. In this *ftz* background, a rare phenotype was observed in which half of the thoracic segment was missing and only one wing grew.

To examine transgene expression in *ftz* mutant embryos, virgin females w^{-}*; ftz*^{9H34}*/Tm3 Sb* were crossed with w^{-}*; P[enhancer-lacZ]*/*P[enhancer-lacZ]*; *Dr*/*Tm3 Sb* males. From this cross, w^{-}*; P[enhancer-lacZ]*/+*; ftz*^{9H34}*/Tm3 Sb* males and females were crossed to generate w^{-}; *P[enhancer-lacZ]*/*P[enhancer-lacZ]*; *ftz*^{9H34}*/Tm3 Sb* offspring, which were then self-crossed to maintain the stable transgenic enhancer lines in *ftz* mutant background for examining β-galactosidase expression.

Germline clones homozygous for the *ftz-f1*^{19} were generated using the 'autosomal FLP-DFS' technique (Chou and Perrimon 1996). Briefly, *yw hsFLP;FRT*^{2A}*ftz-f1*^{19}*/TM3Sb* virgin females were crossed with *w; FRT*^{2A} *ovo*^{D}*/TM3Sb* males. Females were allowed to lay eggs for 1 day in vials and their progeny were heat-shocked for 2 hours at 37ºC in a circulating water bath on the third and fourth days after egg laying. Subsequently, embryos derived from the females of genotype of *yw hsFLP;FRT*^{2A}*ftz-f1*^{19}*/ FRT*^{2A} *ovo*^{D} (identified as non-*Sb* females) were analyzed. All of the *FRT*^{2A} *ovo*^{D} recombinant chromosomes were associated with a fully penetrant DFS phenotype such that all eggs laid by these females are derived from germline recombination events. For the control for the microarray experiment, *yw hsFLP;FRT*^{2A}*/ FRT*^{2A} virgins were crossed to *w; FRT*^{2A} *ovo*^{D}*/TM3Sb* males and subjected to the same heat shock and selection protocol in parallel.

Whole mount in situ hybridization and Immunochemistry

For in situ hybridization, standard protocols were followed (Kosman and Small, 1997; Tautz and Pfeifle, 1989) with one modification - in place of Proteinase K treatment, embryos were heated at 95°C for 5 minutes. Digoxigenin-labeled RNA probes were made by isolating coding and/or 3'UTR sequences by PCR using a reverse primer with the T7 polymerase promoter sequence (TAATACGACTCACTATAGGG). Primer sequences for probes available upon request. Sheep anti-digoxigenin (1:2000, Roche), followed by incubation with NBT/BCIP, was used for for RNA probe detection. Standard protocols were followed for antibody staining (Gutjahr et al., 1994) For reporter constructs, anti-β-galactosidase antibody (Cappel, 1:2000; Gutjahr et al., 1993) was used with a secondary biotinylated - anti-rabbit IgG (Vector Laboratories, 1:500), followed by amplification with an ABC kit (Vector Labratories, Inc) and detection with 3,3'-Diamino-benzidine tablets (Sigma). Protein/RNA double staining followed standard in situ hybridization protocols, using an RNA probe against *engrailed*. After several washes in PBST, and rinses with PBS, embryos were mounted in 90% glycerol, 0.1M Tris–HCl, pH 7.9. Stainings were visualized using DIC on a Leica DMRB microscope.

Microarray

ftz-f1 heterozygous or control females were mated to w^{1118} males in collection cages at 25°C and allowed to lay eggs for up to 2 hours. Eggs were dechorionated in 3% sodium hypochlorite for 3 minutes and then covered with halocarbon oil and aged. Embryos were visualized under phase contrast optics at 100-200x magnification. Individual embryos were pooled into groups of roughly equivalent developmental

stages and kept out of the light path of the microscope as much as possible. To generate pools of embryos at specific stages, each pool was monitored closely and selected by visual inspection at late cellularization, stage 5 (when the invaginating cell membrane had reached approximately 80-90% of its maximal depth and before any sign of the cephalic furrow invagination was visible), at the onset of gastrulation, stage 6 (when the cephalic furrow had invaginated approximately one full cell depth), or at mid germband extension, stage 8 (when the posterior end of the germband had extended approximately half way towards the cephalic furrow.

Embryos were manipulated using a small needle. As each embryo reached the desired stage of development, it was transferred to 100 ul of TRIZOL on ice. Approximately 100 embryos were collected per time point. Individual collections were stored at -80^{o}C. Each experiment was done in triplicate. Total RNA was extracted using the Qiagen RNEasy kit according to standard protocols, and samples were hybridized to Affymetrix Drosophila 2.0 expression arrays by the UMBI microarray facility.

Affymetrix drosophila2 genechip CEL files were imported into BioConductor/R(Gentleman et al., 2004) using the ReadAffy function of the affy package (Gautier et al., 2004) and assigned to developmental stage (5,6,8), wild type or *ftz-f1* mutant condition (0,1) and batch number (processing and hybridization batch). Using 3 replicates of 3 stages and 2 states of *ftz-f1* gave a total of 18 genechip arrays. All arrays were normalized by the expresso function using quantiles normalization, only perfect match, and median polish summary method. This generated the normalized expression set used for all further data analysis. The normalization procedure produced $\log_2$ expression results, and the fold change between the average of any two data sets s1

and s2 was calculated as $\frac{s2-s1}{|s2-s1|}2^{|s2-s1|}$. To identify differentially expressed genes showing a response to the presence or absence of functional Ftz-F1, the microarray analysis of variance package, maanova (Wu et al., 2011), was run in oneColor mode fitting a mixed effect ANOVA model using the formula ŷ=Stage+Ftz.F1+Batch. Batch represents the groups of RNA that were processed and hybridized on the same day and was treated as a random or non-repeatable term in the model. The function matest was run with a permutation count of 100 to compute the p-value for the Ftz.F1 model term. p-values were further controlled for the N discovery rate using the Q-value method of the adjPval function (Dabney et al., 2011). Probesets having an FDR adjusted p-value < 0.05 were considered potential target genes. We restricted the analysis to genes showing at least a 1.1 fold increase between stage 5 and stage 6 of our wildtype microarrays. This eliminated genes with solely maternal expression that could mask responses to Ftz/Ftz-F1.

Analysis of BDNTP Chip Data to identify candidate enhancers

Data files containing the 1% FDR bound regions for all 22 transcription factors having ChIP-chip data, excluding polymerases (MacArthur et al., 2009), available at http://bdtnp.lbl.gov/Fly-Net/browseChipper.jsp were downloaded and all interval coordinates along with the peak binding positions were remapped from release 4 to release 6 of the *Drosophila* genome using the Coordinate Converter provided by Flybase.

These ChIP-chip data were used to identify Ftz binding sites within 70 kb each of the top 9 Ftz-F1 targets identified from the microarray by uploading the published

BDTNP data onto Flybase (supplementary table 4). Genomic regions of ~1 kb surrounding each finding Ftz binding interval were queried for consensus Ftz-F1 binding sites (BSAAGGHYRHH). For each gene, the region that contained the strongest Ftz binding and at least one candidate Ftz-F1 binding site was selected to be tested functionally. A candidate enhancer region was found for all but two targets at 1% FDR. No Ftz binding was detected at this level within 70kb of the target gene *astray*, so the Ftz binding peak found in the dataset using 25% FDR was used. No Ftz binding was found with 70kb surrounding the target gene *5-HT2*. However, because *5-HT2* was already known to be *ftz/ftz-f1* target (Hou et al., 2009), a candidate enhancer was selected based on a candidate *ftz-f1* binding site near the gene. In order to increase the strength of this candidate enhancer, an exception for the size of the region was made, making it 2 kb to include both the strongest potential Ftz-F1 binding sites and Zelda binding sites, which are important in the activation of transcription in many developmental genes (Liang et al., 2008). This provided a good test of the importance of Ftz binding in choosing an enhancer. Using these criteria, one enhancer-reporter construct was made for each candidate target by isolating the 1 kb region from gDNA by PCR and fusing each enhancer to generate *ken-lacZ, aay-lacZ, mid-lacZ, 5HT2-lacZ, trn-lacZ, hh-lacZ, Antp-lacZ,* and *blot-lacZ*. Transgenic flies were generated and analyzed as described above.

Motif Analysis

401bp windows surrounding the peak binding position for all intervals were merged and the underlying genomic sequences were extracted from a repeat masked copy of the *Drosophila melanogaster* genome. This dataset was processed with the meme application fasta-get-markov, to generate a 5th order hidden markov model of the available background enhancers found in the *Drosophila* blastoderm (Bailey and Elkan, 1994). This model or its associated fasta file were used as the background model for all relevant processes of the meme suite of applications.

A subset of the 401bp sequences, defined by the ftz_3_032707-sym-1 dataset, was used as the positive set for Ftz binding. This set contains 403 intervals identified as bound by Ftz in stage 5 *Drosophila melanogaster* embryos, along with the location of maximal binding within each interval. Genomic sequence for each of these regions was extracted from the repeat masked genome. This dataset was processed with MEME and DREME applications to identify the core binding motifs of Ftz. For both MEME and DREME, a q-value threshold of 0.05 was used with the previously computed background file searching both forward and reverse strands. MEME was restricted to a maximum motif width of 10, while DREME's default of 8 was used (Bailey, 2011).The results of both MEME and DREME runs were passed to TOMTOM with the previously computed background model, q-value threshold of 0.05, and scanning was performed on all of the *Drosophila* databases distributed with the MEME suite (flyreg.v2.meme, idmmpmm2009.meme, and dmmpmm2009.meme) (Gupta et al., 2007).

To search for additional motifs within the enhancer regions, then Ftz/Ftz-F1 enhancer regions were analyzed using the MelinaII software and University of Maryland Institute for Advanced Computer Science (UMIACS) position weight matrix (pwm) program. For MelinaII, 4 algorithms were used – Consensus, MEME, MDScan, and Gibbs. These algorithms search for motifs de novo. For UMIACS pwm, the *Drosophila* JASPAR database was used to scan enhancer sequences, with a confidence of 95% in all 10 sequences. The sequences used for these enhancers were *ftz* upstream element*, en* intron (Florence et al., 1997)*, ken-lacZ, trn-lacZ, mid-lacZ, hh-lacZ, Antp-lacZ, blot-lacZ, drm2* and *drm34* (Hou et al., 2009).

4.3 Results

4.3.1 Overall gene expression determined more by developmental stage than genotype.

In order to identify candidate targets of Ftz/Ftz-F1, gene expression was compared between wild type and *ftz-f1* germline clones embryos (referred to throughout as *ftz-f1* mutants) at three time points during development: stage 5, when the blastoderm forms and *ftz* expression begins and is strongest, stage 6, when gastrulation occurs, and stage 8, when germband extension occurs and *ftz* expression continues to be strong (Fig. 4-1). Gene expression was compared in whole embryos (wild type vs. *ftz-f1* mutants). Previous analysis of *ftz-f1* mutants suggested that gene expression would only be affected in the Ftz+ cells (see above); these cells represent a maximum of 25% of the cells in the whole embryo, thereby potentially diluting the overall fold change observed for genes regulated by Ftz/Ftz-F1 and other transcription factors. However, this method also allowed us to identify potential Ftz-F1 targets that are not co-regulated by Ftz.

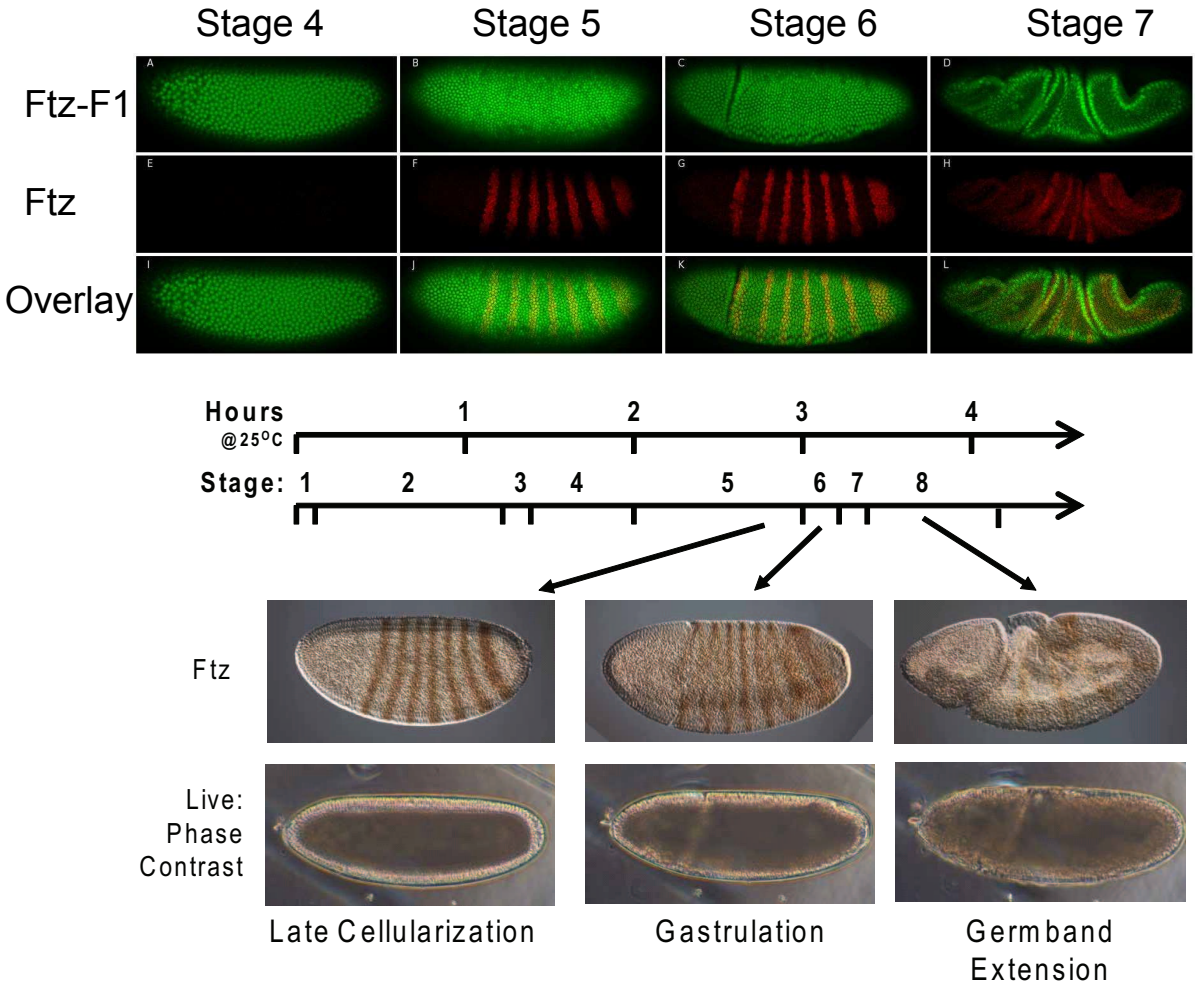


Figure 4-1. Ftz-F1 and Ftz protein expression and staging of embryos. A) Stage 4 (A,E,I), stage 5 (B,F,J), stage 6 (C,G,K), and stage 7 (D,H,L) embryos showing Ftz-F1(green) is maternally expressed and localizes to the nuclei. Ftz (red) reaches its peak level during late cellularization (stage 5), where it is expressed in seven stripes. At the onset of gastrulation (stage 6), the most anterior stripe of Ftz is immediately posterior to the cephalic furrow. The Ftz stripes weaken throughout germband extension (stage 8). Both proteins are coexpressed (yellow) in nuclei of the primordia of even numbered parasegments. B) Hand staging of embryos at stages 5, 6, and 8 was performed using phase contrast microscopy of live, dechorionated embryos in halocarbon oil as shown.

Of the 18952 probesets on the genechips, 735 (3.9%) showed detectible alterations in response to Ftz-F1. 3944 (20.8%) of the probesets showed a 1.1 fold increase in expression level between stage 5 and stage 6 (note: this was meant to be an inclusive cutoff, so no statistical threshold was applied to this set). The intersection of these two data sets produced a list of 379 potential targets for validation. 314 (82.8%) of these potential targets showed upregulation in response to the loss of Ftz-F1. As both Ftz and Ftz-F1 have been shown to function *in vivo* as transcriptional activators (Bowler et al., 2006; Dubrovsky et al., 2011; Florence et al., 1997; Hou et al., 2009), many of these may be indirectly regulated targets of Ftz-F1. Overall, 65 probesets showed at least a 1.1 fold increase in expression level between stages 5 and 6 and a loss of expression in *ftz-f1* mutants, marking these as potential Ftz-F1 targets. These 65 probesets map to 63 unique genes. Figure 4-2 shows a heatmap of the changes in expression level for potential target genes and Table 4-1 shows the top 12 candidate target genes, ordered by their fold change in expression level in *ftz-f1* mutants. Each showed an average of at least -1.49 fold change in expression levels between wildtype and *ftz-f1* mutants. Mutant fold change shows the average ratio of expression levels between *ftz-f1* mutant embryos and wildtype embryos at stages 6 and 8. Activation fold change shows the ratio of expression levels between wildtype stages 5 and 6. And the q-value shows the FDR adjusted p-value for a change in expression with respect to the state of Ftz-F1 in the embryos. Interestingly, pairwise comparison of Pearson correlation coefficients showed that individual genechips clustered varied more strongly with developmental stage and then by the presence or absence of a functional Ftz-F1 protein (Tables 4-2 and 4-3).

Table 4-1. Top eleven targets identified by the microarray.

Gene	Affymetrix Probset	Mutant Fold Change	Activation Fold Change	q-value
ken	1628840_at	-6.1	3.43	0.00E+00
en	1627445_s_at	-4.16	4.22	0.00E+00
aay	1633488_at	-1.89	1.22	0.00E+00
tektin-C	1628238_at	-1.87	1.23	0.00E+00
mid	1637867_at	-1.88	5.28	2.22E-04
tal-1A	1625897_s_at	-1.84	3.21	0.00E+00
5-HT2	1633198_a_at	-1.83	2.05	6.66E-04
trn	1639235_at	-1.69	3.32	0.00E+00
hh	1626527_at	-1.56	2.39	2.15E-02
Antp	1624759_s_at	-1.59	2.61	6.94E-03
blot	1626839_s_at	-1.56	2.41	0.00E+00

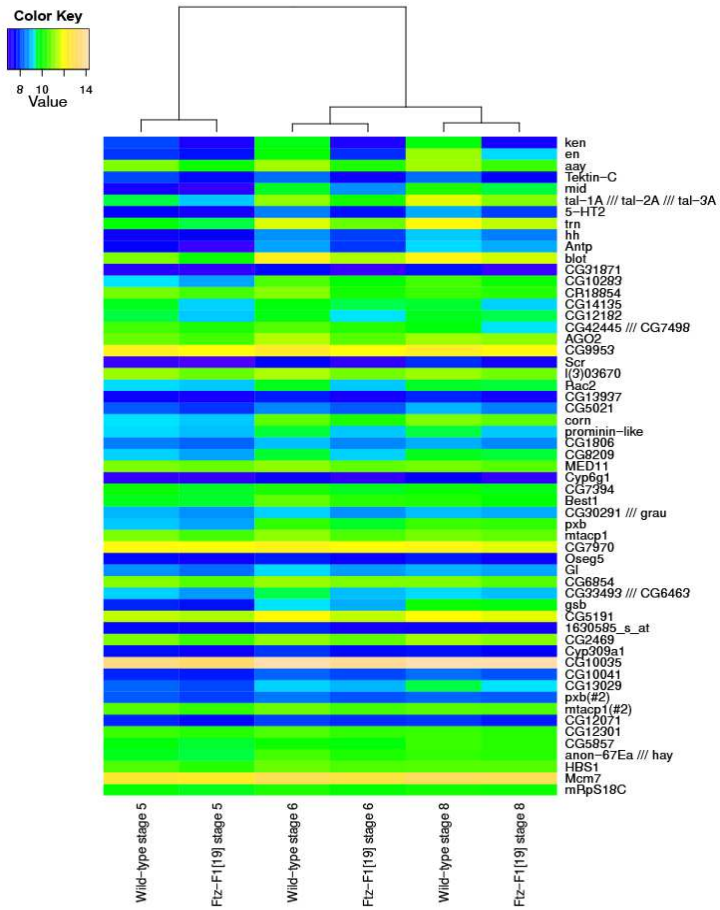


Figure 4-2. Microarray identified genes differentially expressed in ftz-f1 mutant embryos. RNA was extracted from wildtype or *ftz-f1* mutant embryos at stages 5,6, and 8, as indicated, and were hybridized to Affymetrix Drosophila 2.0 expression arrays The dendrogram shows Pearson correlation coefficients of mean expression levels across all probesets under assayed conditions. The heatmap shows expression levels of the 60 probesets activated at gastrulation in wildtype embryos, and showing statistically significant loss of expression in *ftz-f1* mutant embryos. Genes were sorted from highest to lowest average fold change in expression between wildtype and *ftz-f1* mutants after cellularization. This experiment was carried out by W. Ray Anderson.

Table 4-2. Overview of the upregulation and downregulation of RNA transcripts identified using the microarray at different different p-values.

p-value	0.25	0.10	0.05	0.01
Total # Differential expressed	897	500	345	159
Total % regulated	4.7%	2.6%	1.8%	0.8%
# down-regulated	116	78	60	31
% down-regulated	0.6%	0.4%	0.3%	0.2%
# up-regulated	781	422	285	128
% up-regulated	4.1%	2.2%	1.5%	0.6%

Table 4-3. A comparison of the upregulation of RNA in different stages in wild type and *ftz-f1* mutant embryos.

	FF1.5	FF1.6	FF1.8	FRT.5	FRT.6	FRT.8
FF1.5	1	0.97114	0.947377	0.995691	0.968568	0.946157
FF1.6		1	0.991669	0.96222	0.997102	0.989425
FF1.8			1	0.937397	0.990501	0.996754
FRT.5				1	0.962672	0.939053
FRT.6					1	0.992129
FRT.8						1

4.3.2 Candidate Ftz-F1 target genes require Ftz and Ftz-F1 for expression in embryos

In order to determine whether the microarray correctly identified Ftz-F1-responsive genes that are also regulated by Ftz, the expression patterns of ten of the top eleven genes from the microarray were examined and compared to the expression pattern of *ftz* RNA (Fig. 4-3). *tektin-C* was not examined because it does not show a detectable level of expression at this time. For 10 of the 11 remaining candidate target genes, expression of the target gene overlapped with *ftz* expression. A) two stripes of *ken* expression overlapped with the anterior and posterior stripes of *ftz*. B) *en* is expressed in 14 stripes, every other stripe overlapping with a *ftz* stripe. C) the 7 stripes of both *aay* and *ftz* overlap. D) 7 alternating stripes of the 14 stripes of *mid* overlap with *ftz* expression. E) stripes 2 and 3 of *tal-1a* are more strongly expressed than the rest of the 7, and the stripes of *tal-1a* do not overlap with *ftz* stripes. F) *5HT2* expression is broader than *ftz* expression, but all 7 *5HT2* stripes cover the 7 *ftz* stripes. G) the 7 posterior stripes of *trn* overlap the 7 stripes of *ftz*, while the most anterior *trn* stripe does not. H) 7 alternating stripes of the 14 stripes of *hh* overlap with *ftz* expression. I) The one *Antp* stripe overlaps with the second *ftz* stripe. J) *blot* expression, while expressed in a striped pattern, is not as precise as other *ftz/ftz-f1* targets. The pattern is not formed in tight stripes, but in more blurred lines with soft edges. This expression does overlap with *ftz* stripes. Because *tal-1a* stripes did not overlap with *ftz* expression, *tal-1a* could not be a direct target of *ftz*, and so was removed as a possible target.

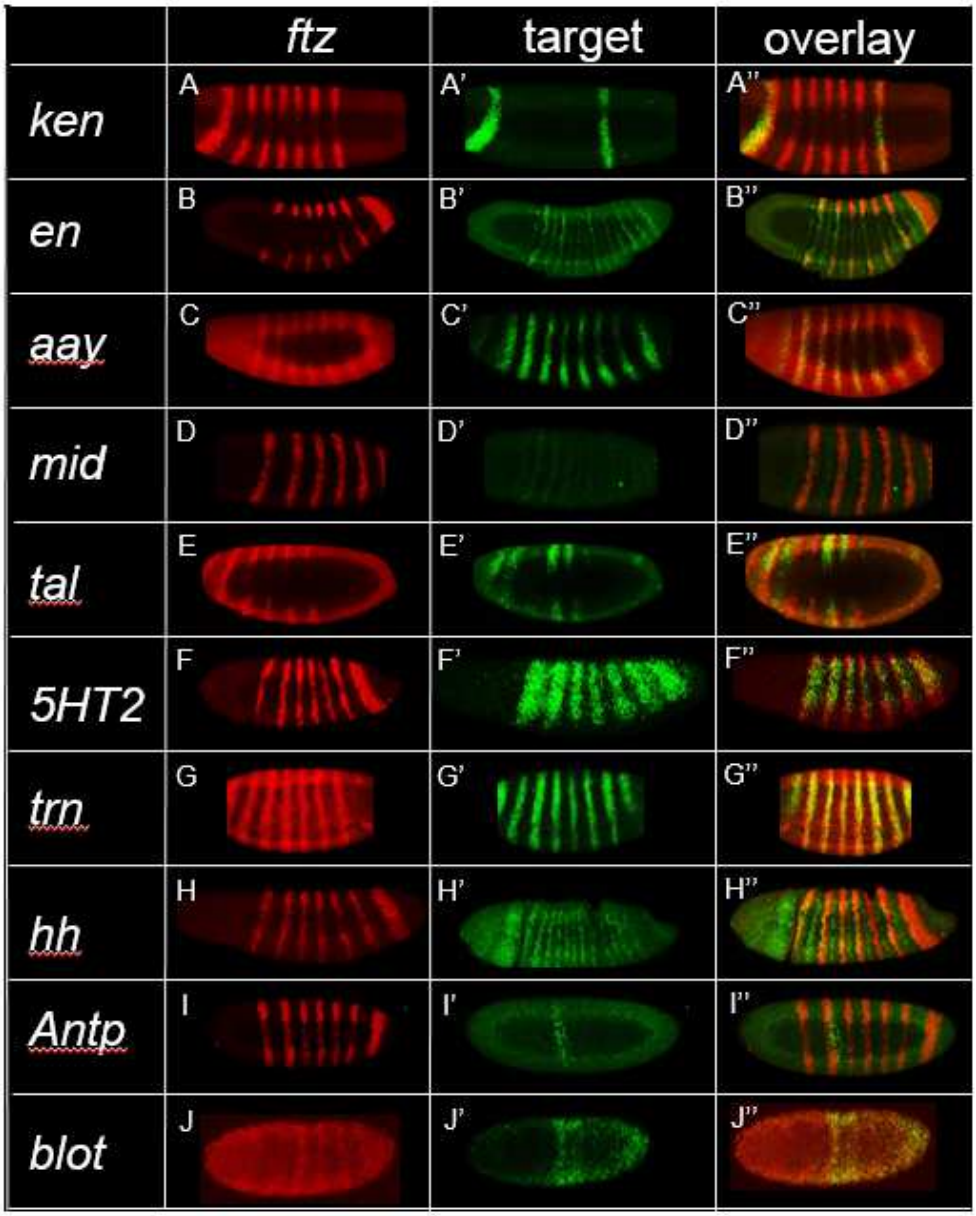


Figure 4-3. Nine of the ten candidate target expression patterns overlap with *ftz* stripes. A fluorescent double in situ was performed against *ftz* and each target gene, as indicated. *ftz* stripes, red, and target gene stripes, green, overlap, yellow. Some or all of each of the target gene expression patterns overlap *ftz* expression during the blastoderm stage for all candidate targets but *tal-1A*. A) *ken* B) *en* C) *aay* D) *mid* E) *tal-1A* F) *5HT2* G) *trn* H) *hh* I) *Antp* (J) *blot*

To further test whether these targets are controlled by Ftz/Ftz-F1, expression patterns of nine of the top ten genes from the microarray were examined in control, *ftz*, and *ftz-f1* mutant embryos (Fig. 4-4). A striped expression in early stage embryos (A-I) was altered or undetectable in both *ftz* (A'-I') and *ftz-f1* (A''-I'') mutants. A) *ken* was expressed in 2 stripes, which were lost in *ftz* (A') and *ftz-f1* (A'') mutants. B) *en* is well known to be expressed in 14 stripes, and 7 alternating stripes were lost in both B') *ftz* and B'') *ftz-f1*. C) *aay*, was expressed in seven stripes in control embryos, which were lost in C') *ftz* and C'') *ftz-f1* mutants. D) *mid* was normally expressed in 14 stripes. Like *en*, in the D') *ftz* and D'') *ftz-f1* mutants, 7 of the 14 stripes were lost, presumably those controlled by Ftz/Ftz-F1. *tal-1a* expression was lost in both E') *ftz* and E'') *ftz-f1* mutants, and although it was shown that it could not be a direct target (Fig. 4-2E), this indicates it is an indirect target of Ftz/Ftz-F1. F) *5HT2* was expressed in seven stripes beginning in early gastrulation in control embryos, which were lost in F) *ftz* and F'') *ftz-f1* mutants. G) *trn* was expressed in 8 stripes, the 7 most posterior overlapping with Ftz expression (data not shown). In *ftz* (G') and *ftz-f1* (G'') mutant embryos, the expression of these 7 posterior stripes is lost while the Ftz-independent stripe remains. H) *hh* was expressed in 14 stripes, 7 of which were lost in H') *ftz* and H'') *ftz-f1* mutants. H) *Antp* was expressed in one broad band posterior to the cephalic furrow of the embryo, as shown in (Carroll et al., 1986). Expression was also lost in I') *ftz* and I'') *ftz-f1* mutants. *blot* was weakly expressed in seven stripes in (J), but unlike the other candidate targets from the microarray, no loss or alteration of expression was observed in either *ftz* (J') or *ftz-f1* (J") embryos.

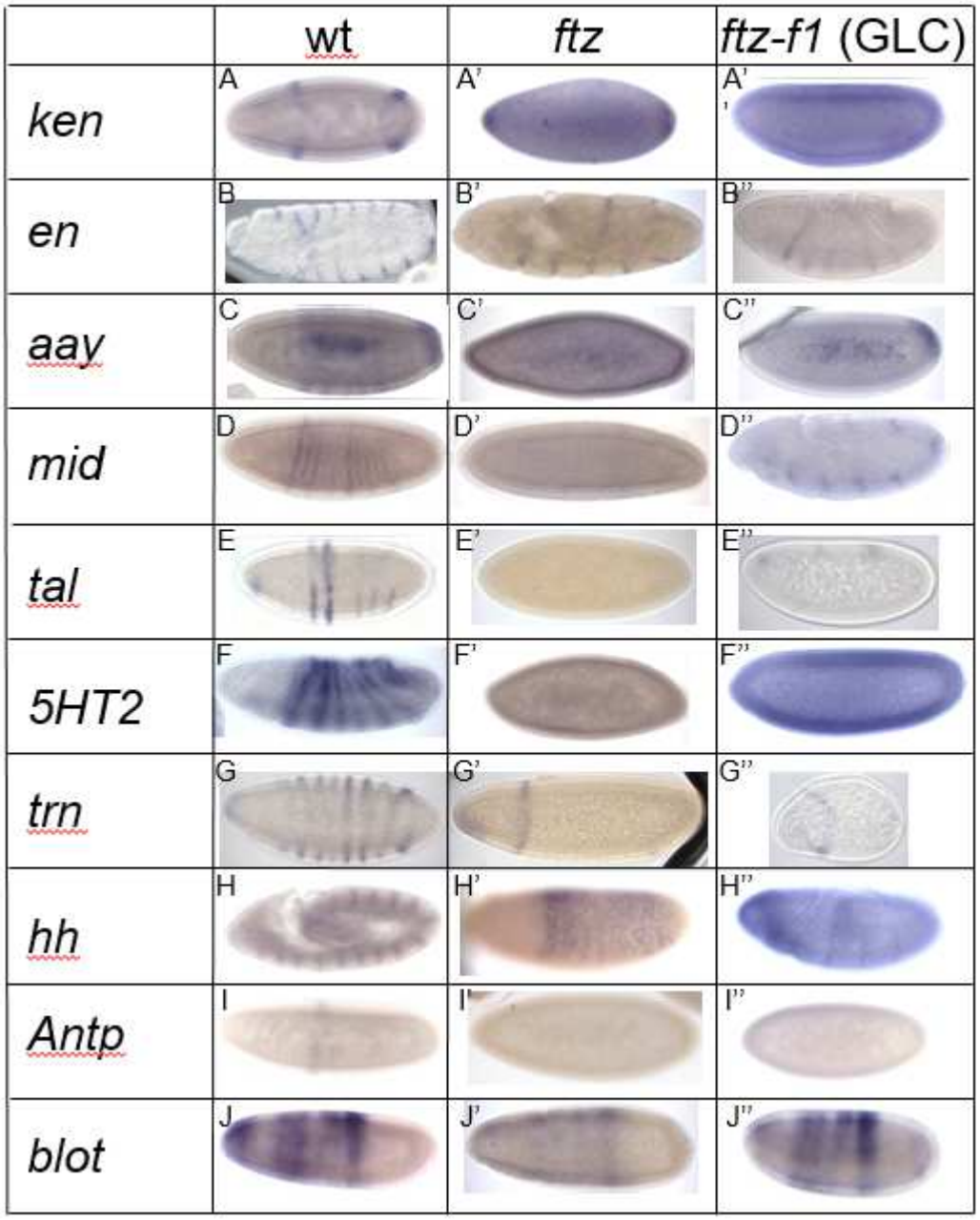


Figure 4-4. Expression patterns of nine of the top ten candidate targets are lost in *ftz* and *ftz-f1* mutants. 10 of the top 11 candidate target genes from the microarray are expressed in stripes at the blastula/gastrula stage of embryogenesis, *tektin-C* being the exception. All but one of those 10 also lost expression of at least a subset of those stripes in embryos lacking functional Ftz or Ftz-F1 protein, *blot* being the exception.

In sum, of the top 9 Ftz-F1 targets from the microarray experiment, all showed striped expression patterns in the early embryo that suggests regulation by a pair-rule gene. Interestingly, all 10 of these responsive targets responded in the same way to loss of either Ftz or Ftz-F1. Thus, of the 13 Ftz-F1-responsive genes identified to date, all require Ftz for expression in embryos. These 13 genes are the seven targets found in the microarray (*ken*, *aay*, *mid*, *hh*, *trn*, A*ntp*, *blot*), *en* (Florence et al., 1997) *ftz* itself, *apt*, *Sulf1* (Bowler 2006), *drm*, *noc*, and *5HT2* (Hou et al., 2009). This finding is consistent with the fact that *ftz* and *ftz-f1* pair-rule phenotypes are indistinguishable (see above) and strengthens the conclusion that Ftz-F1 absolutely requires Ftz for its activity in early embryos.

4.3.3 Ftz ChIP-chip peaks overlap many known cis-regulatory modules

The Ftz-F1-responsive target genes analyzed above could be either directly or indirectly regulated by Ftz and/or Ftz-F1. To address this, we made use of BDNTP published ChIP-chip data to identify candidate Ftz/Ftz-F1 responsive candidate enhancers for each of these targets. BDNTP published ChIP-chip data on a large set of *Drosophila* transcription factors, including Ftz (Fig. 4-5).

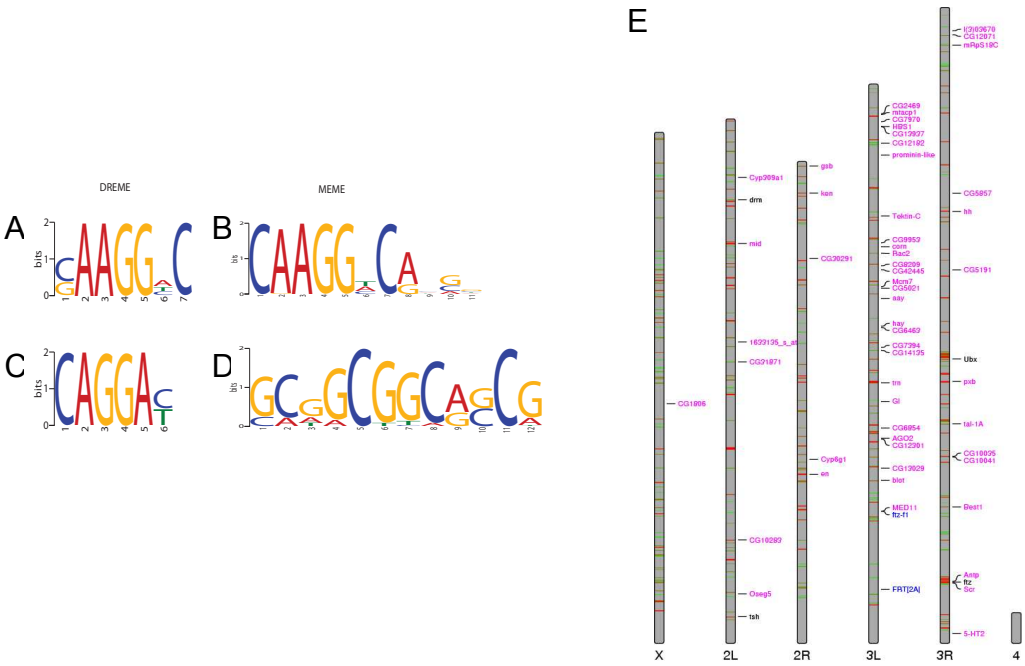


Figure 4-5 ChIP-chip data shows Ftz binding over all chromosomes. Over-represented sequences in the Ftz ChIP-chip binding data, identified by A,B) DREME and C,D) MEME. Ftz-F1 candidate binding sites are overrepresented where Ftz binds DNA. E) The red stripes represent strong Ftz binding, and the green stripes represent weak Ftz binding. Candidate targets from the microarray are shown near strong Ftz binding in pink.

Ftz binding was located within 70 kb each of the 9 Ftz-F1 targets by uploading the published BDTNP data onto Flybase (Table 4-4). By searching for candidate Ftz-F1 sites within the strongest Ftz binding surrounding each target, the strongest candidate Ftz/Ftz-F1 responsive enhancer for each Ftz/Ftz-F1 target was chosen for further analysis (Fig. 4-6). In addition to the criteria stated here, these figures show accessible chromatin data also published by BDTNP. This data shows that all candidate enhancers were in accessible chromatin at stage 5.

Table 4-4. All enhancers found within 70kb of the Ftz/Ftz-F1 targets and related data.

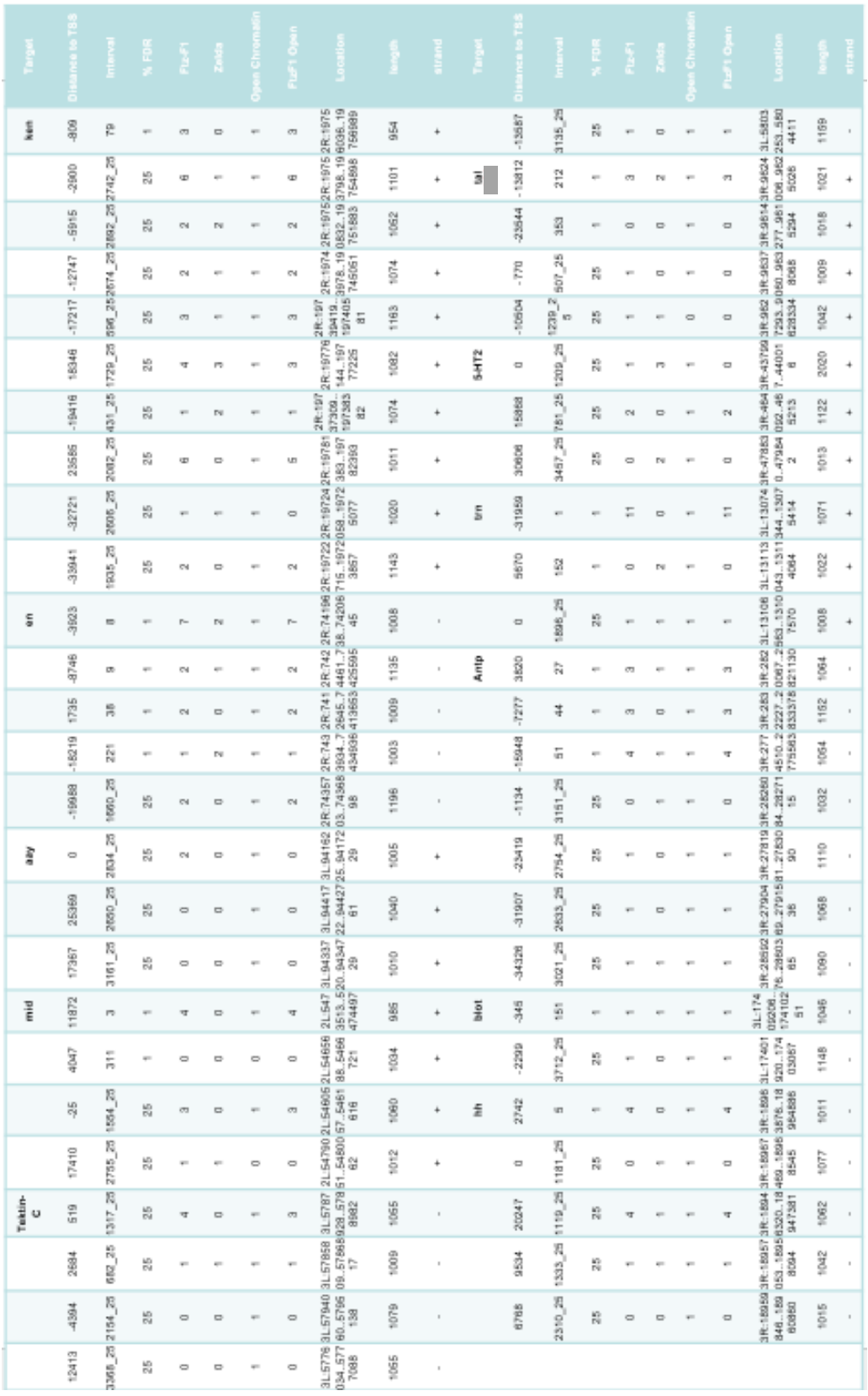
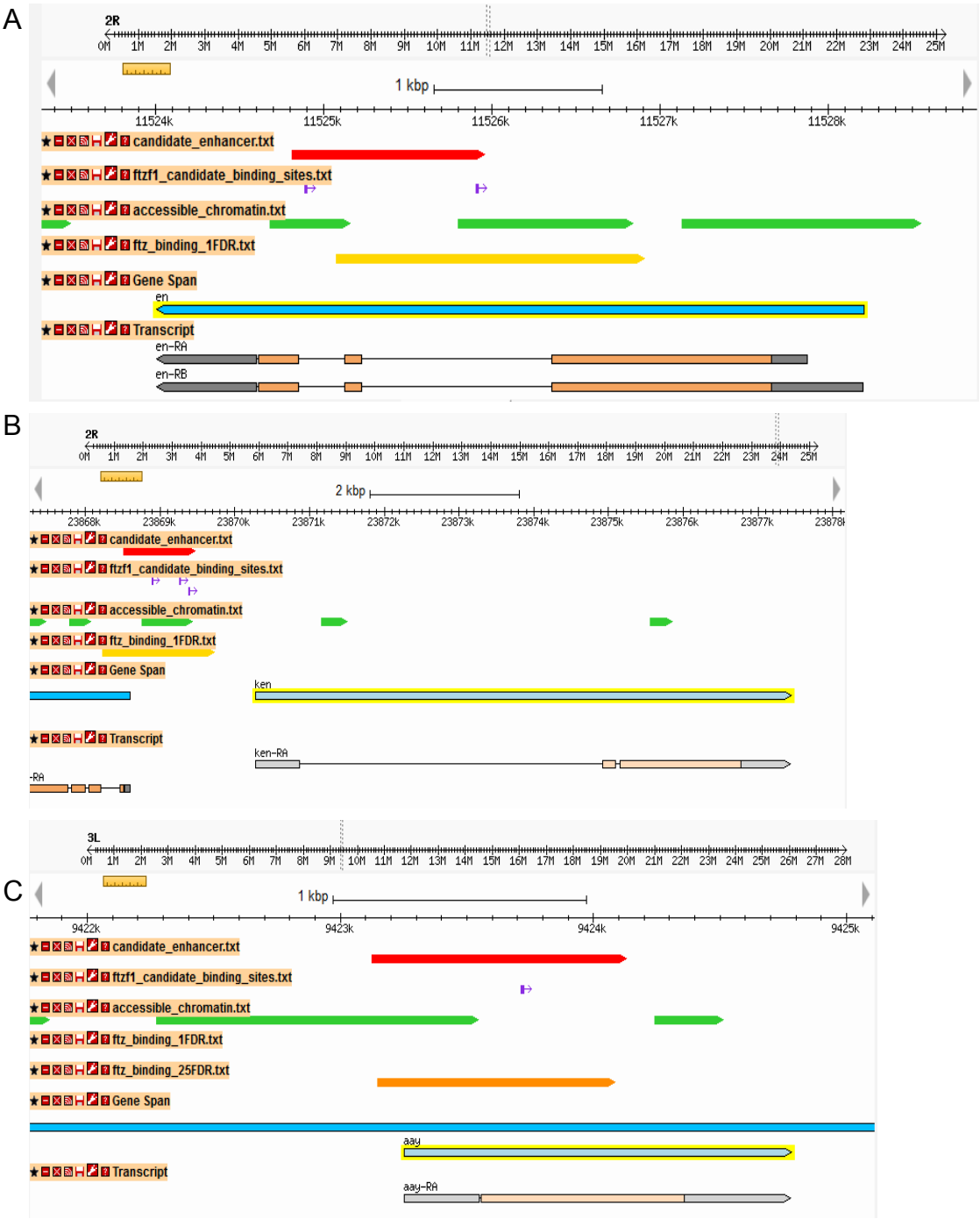
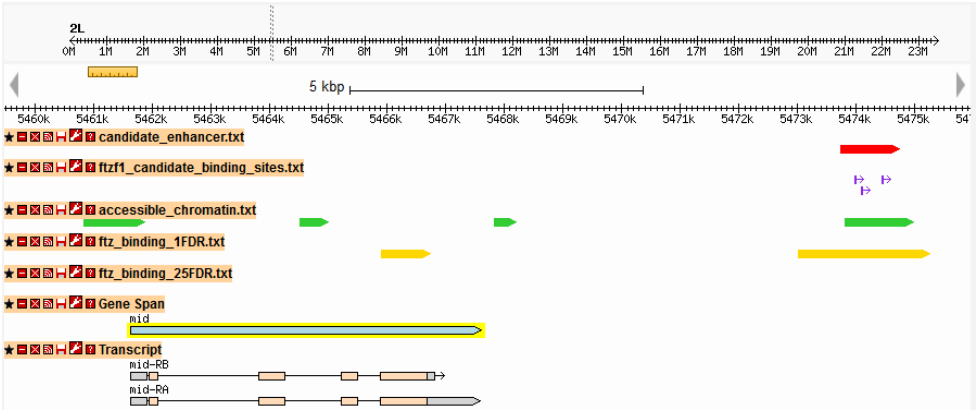


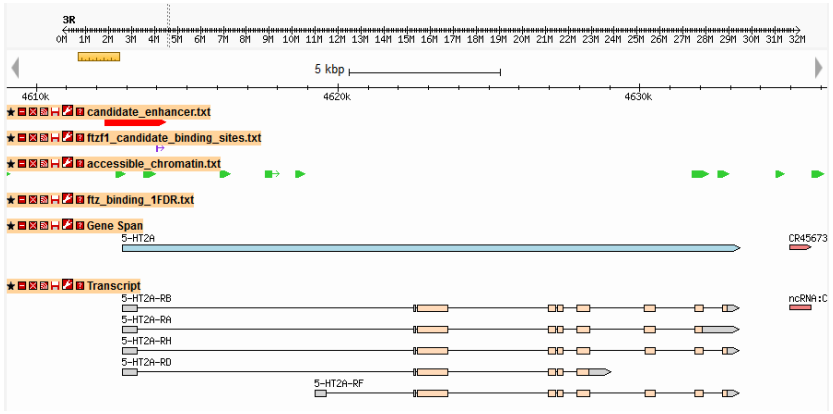
Table 4 also shows the association of these Ftz binding intervals with 1) overlapping open chromatin, 2) candidate Zelda sites and 3) candidate Ftz-F1 sites. Not all enhancers had a Zelda site within the binding interval, indicating that candidate Ftz-F1 sites are more predictive than Zelda binding sites for Ftz ChIP-chip peaks near potential target genes.



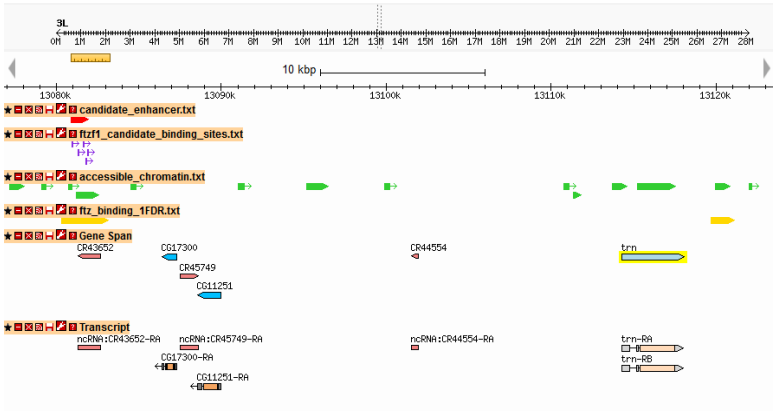
D


2L
5 kbp
candidate_enhancer.txt
ftzf1_candidate_binding_sites.txt
accessible_chromatin.txt
ftz_binding_1FDR.txt
ftz_binding_25FDR.txt
Gene Span
mid
Transcript
mid-RB
mid-RA

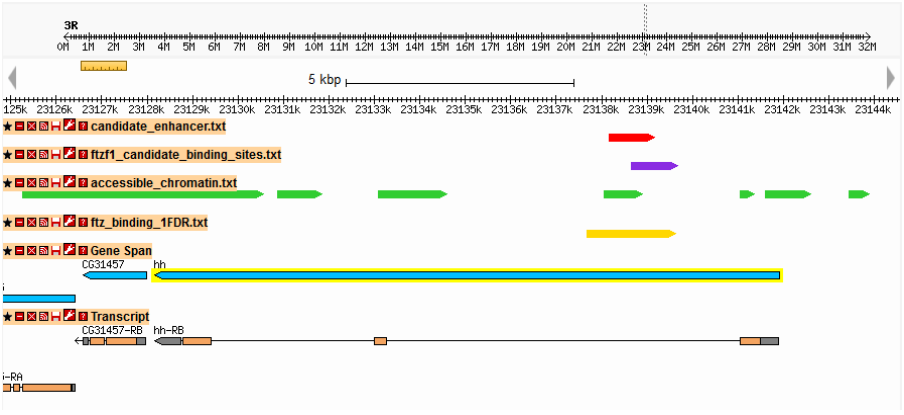
E


3R
5 kbp
candidate_enhancer.txt
ftzf1_candidate_binding_sites.txt
accessible_chromatin.txt
ftz_binding_1FDR.txt
Gene Span
5-HT2A
CR45673
Transcript
5-HT2A-RB
5-HT2A-RA
5-HT2A-RH
5-HT2A-RD
5-HT2A-RF

F


3L
10 kbp
candidate_enhancer.txt
ftzf1_candidate_binding_sites.txt
accessible_chromatin.txt
ftz_binding_1FDR.txt
Gene Span
CR43652
CG17300
CR45749
CG11251
CR44554
trn
Transcript
ncRNA:CR43652-RA
ncRNA:CR45749-RA
CG17300-RA
CG11251-RA
ncRNA:CR44554-RA
trn-RA
trn-RB

G


3R
5 kbp
candidate_enhancer.txt
ftzf1_candidate_binding_sites.txt
accessible_chromatin.txt
ftz_binding_1FDR.txt
Gene Span
CG31457
hh
Transcript
CG31457-RB
hh-RB

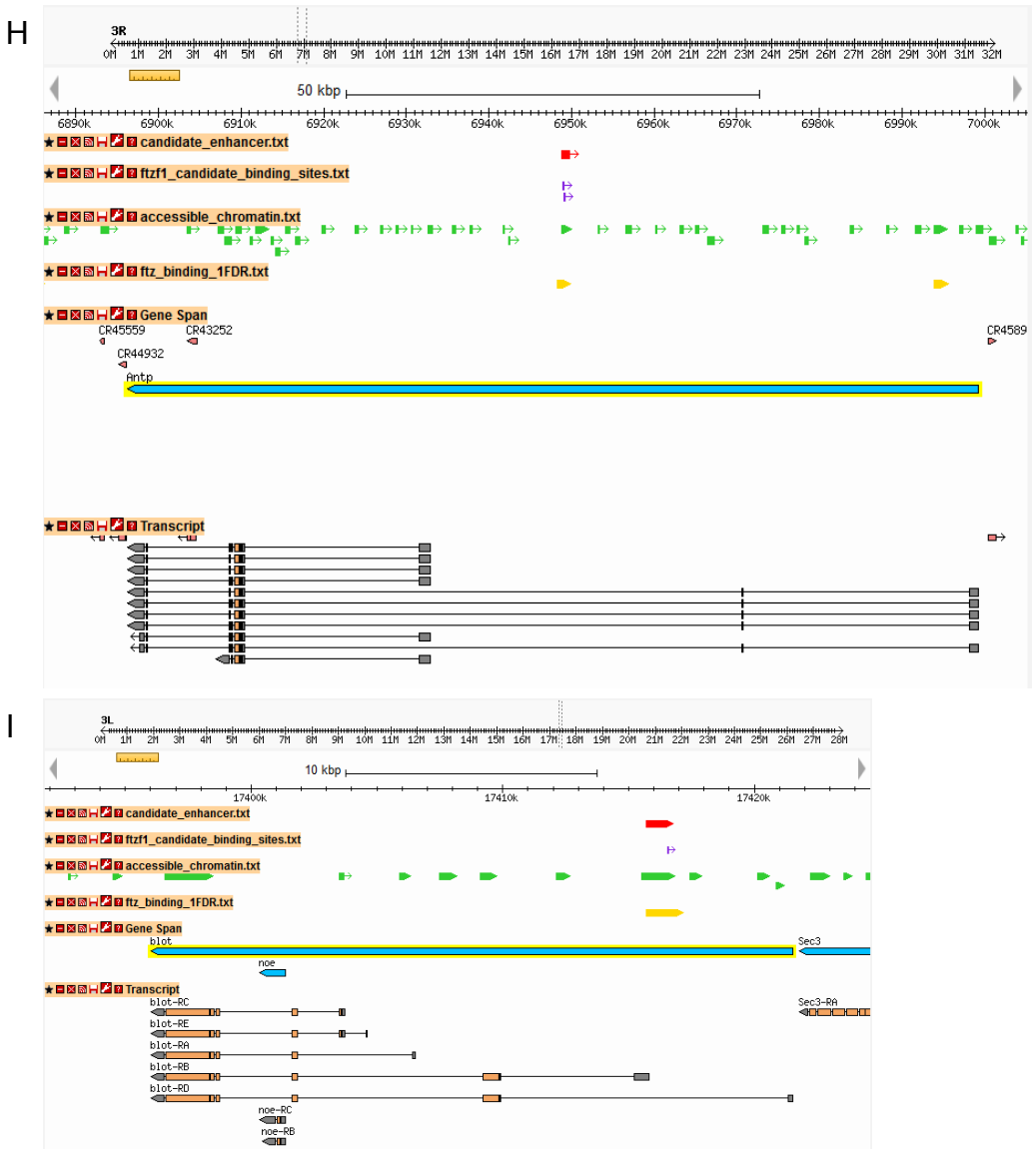


Figure 4-6. Candidate enhancers were identified by combining experimental and computational experiments. Candidate enhancers were identified from three major criteria, shown here. 1) Within 70kb of the of the target gene 2) Ftz binding was observed (with the exception of 5-HT2), and 3) candidate Ftz-F1 binding sites were found within the bound region. Open chromatin was also found in blastoderm stage at all enhancers. A) *ken* B) *en* C) *aay* D) *mid* E) *5HT2* F) *trn* G) *hh* H) *Antp* I) *blot*

4.3.4 Expression of the enhancer region reporter genes in vivo

For functional analysis, ~1 kb genomic regions containing candidate Ftz-F1 binding sites within the strongest binding peak of Ftz near each target were isolated (Methods). These candidate enhancers were fused upstream of a basal promoter and *lacZ* in a *lacZattB* vector and transgenic flies were generated. As shown in Figure 4-7, expression of reporter genes was analyzed by anti-β galactosidase antibody staining of transgenic embryos.

Seven of nine of the candidate enhancers directed *lacZ* reporter gene expression in striped patterns similar to those of the endogenous genes (Fig. 4-7). The *ken-lacZ* reporter gene was expressed in two weak stripes, anterior and posterior, in blastoderm (A). The enhancer for *en* was identified previously (Florence et al., 1997), and so was not tested (B). *mid-lacZ* was expressed in 14 stripes, similar to *mid* wild type expression (C). *trn-lacZ* was expressed in 7 of the 8 stripes of *trn* (D). This finding clearly indicates that a Ftz-F1/Ftz-responsive enhancer was identified for *trn*, as only seven of the eight endogenous *trn* stripes are regulated by *ftz* (Chang et al. 1993 and Fig. 4-3). *Antp-lacZ* was expressed in one strong stripe just posterior of the cephalic furrow at the blastoderm stage (E), while two weaker bands anterior and posterior of the first band appear later, during gastrulation and germband extension (F). These two weaker bands are not expressed by the endogenous gene. *blot-lacZ* was expressed in 7 stripes, similar to the endogenous expression (I). *hh –lacZ* was expressed in an interesting pattern, in that 14 stripes were detected but alternating stripes were strong and weak (G). The endogenous *hh* gene is expressed in all 14 stripes at similar levels. One fusion gene, *5-HT2-lacZ,* was expressed in a weak pattern, but this pattern did not correspond to wild

type expression of *5HT2* and the expression was spotty (E). *5-HT2-lacZ* was expressed in three thick bands evenly spaced, between engrailed stripes 2 and 3, 4 and 5, and 6 and 7 (Fig. 8 E'). For one of the candidate enhancers, *aay-lacZ* (C) no expression of *β-Galactosidase* was detected.

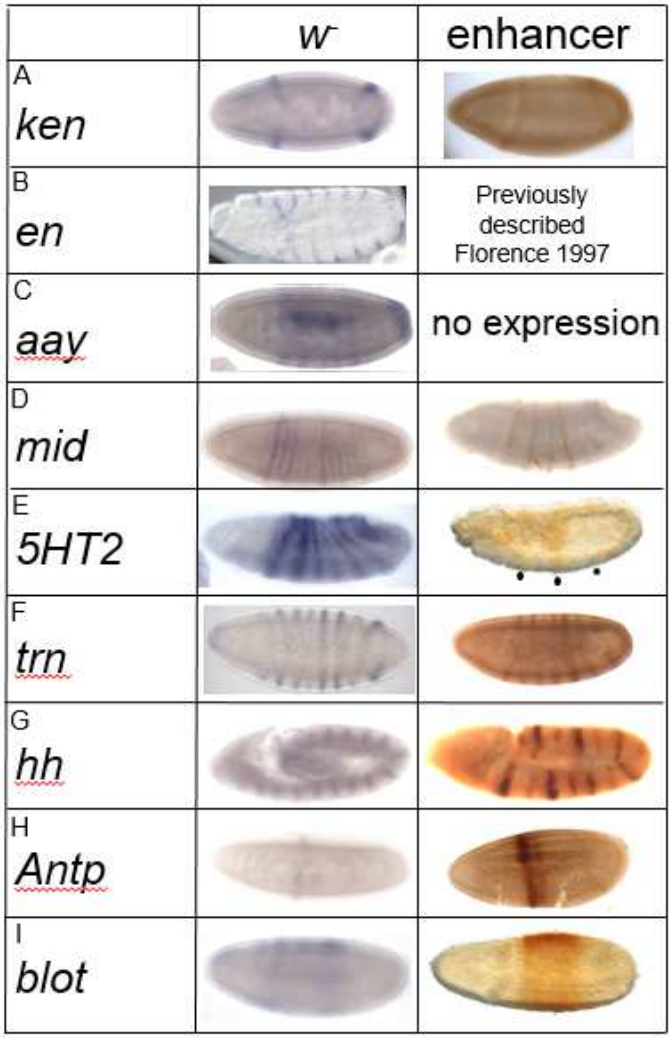


Figure 4-7. Candidate enhancer regions for Ftz show *ftz*-like stripy expression pattern. Expression of enhancer-lacZ reporter constructs in transgenic embryos is shown. Eight of the nine candidate enhancer regions tested directed expression of lacZ in patterns similar to the endogenous genes. These expression patterns match the wild type expression pattern seen in Fig. 4-2 (shown again here for easy comparison), in the regions that overlap with *ftz.* No expression was detected in transgenic flies carrying an empty *attB-lacZ* vector (not shown). A) *ken* B) *en* C) *aay* D) *mid* E) *5HT2* F) *trn* G) *hh* H) *Antp* I) *blot*.

In sum, seven of the nine reporter transgenes – *ken-lacZ*, *mid-lacZ*, *trn-lacZ*, *hh-lacZ, Antp-lacZ*, and *blot-lacZ* - were expressed in striped patterns, in part or wholly similar to that of the endogenous gene. These were categorized as "strong" enhancers. One of the nine –*5HT2-lacZ* – was expressed in a weak, banded pattern that did not align with the expression pattern of its target. This was categorized as a "weak" enhancer. For one of the nine transgenes – *aay* - no expression was detectable, suggesting that this region does not function as an enhancer. Overall, 88% of the candidate enhancers were confirmed as either strong or weak, making this method of finding enhancers highly efficient.

4.3.5 Candidate Ftz/Ftz-F1 target gene enhancers are Ftz-responsive

To test whether these candidate enhancers are *ftz*-responsive, transgenic flies carrying the eight confirmed enhancer-*lacZ* lines (*ken-lacZ*, *en-lacZ*, *mid-lacZ*, *trn-lacZ*, *Antp-lacZ*, *blot-lacZ*, *hh-lacZ,* and *5HT2-lacZ*), and the empty *attB-lacZ* vector as control, were each crossed to ftz^{9H34} /*TM3Sb* flies. If expression is dependent upon *ftz*, it should be lost in ftz^{9H34} homozygotes. The embryos from these crosses were double stained with anti-β-galactosidase antibody (brown) to detect enhancer expression and *en* (blue) to identify the *ftz* mutants (Fig. 4-8). *en* staining in *ftz* mutants would only show expression in 7 stripes, rather than its normal 14 stripes. The enhancer-*lacZ* reporter gene expression pattern (A-G) was partially or completely lost (A'-G') for all eight enhancers, shown during germband extension when expression was strongest. The two stripes of *ken-lacZ* expression (A) were completely lost in the mutant (A'). While some bands were still visible for *trn-lacZ* in *ftz* mutants (B'), these were much weaker and spotty. The seven strong stripes for *mid-lacZ* (C) were completely lost in *ftz* background

(C'). All 14 stripes of *hh-lacZ* (D) were lost (D'), which was surprising, as only the 7 stripes corresponding the *ftz* expression should be lost if the enhancer is under Ftz control. For *5HT2-lacZ* (E), the weak enhancer with 3 spotty, thick bands, the weak bands were completely lost in a *ftz* background (E'). *Antp-lacZ* developed two weaker bands during germband extension, posterior to its original stripe. (F). These weaker bands were completely lost and the stronger, original stripe was much weaker in the *ftz* background (F'). The seven stripes of expression driven by *blot-lacZ* (G) were lost in the *ftz* background (G'). Figure 6H, H' show that empty *attB-lacZ* vector did not show any expression pattern in wild type and *en* was expressed in 14 stripes in wild type embryos and seven stripes in *ftz* mutant embryos. Overall, each enhancer for all eight targets had weakened or complete loss of expression in a *ftz* mutant background, indicating that these enhancers are Ftz-responsive.

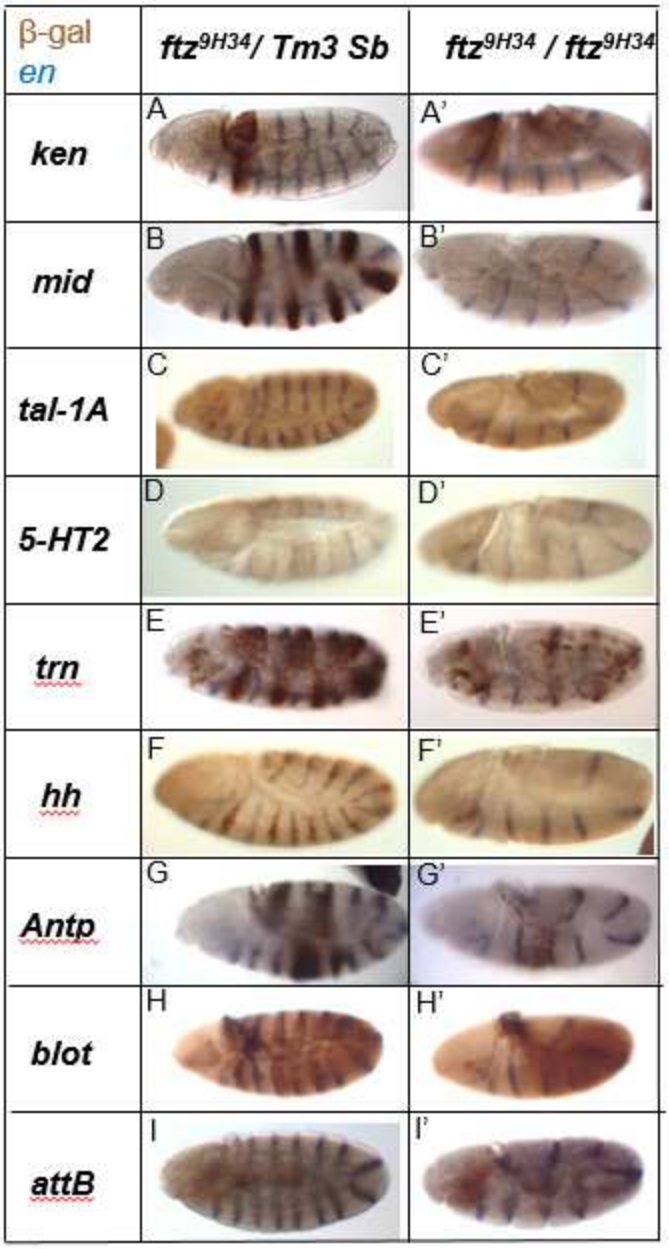


Figure 4-8. Expression of candidate enhancer regions is lost in *ftz* mutants. Anti-β-galactodsidase staining was used to detect expression of the *enhancer-lacZ* transgenes (brown), and in situ hybridization against *en* (blue). *en* is expressed in 14 stripes in *w* embryos and 7 stripes in *ftz* embryos. The expression of all candidate enhancer are shown in wild type and in *ftz* mutants A) *ken* B) *mid* C) *tal-1A* D) *5HT2* E) *trn* F) *hh* G) *Antp* H) *blot* and I) empty attB-lacZ vector.

4.3.6 Motif analysis finds Ftz-F1 sites along with new candidate sites

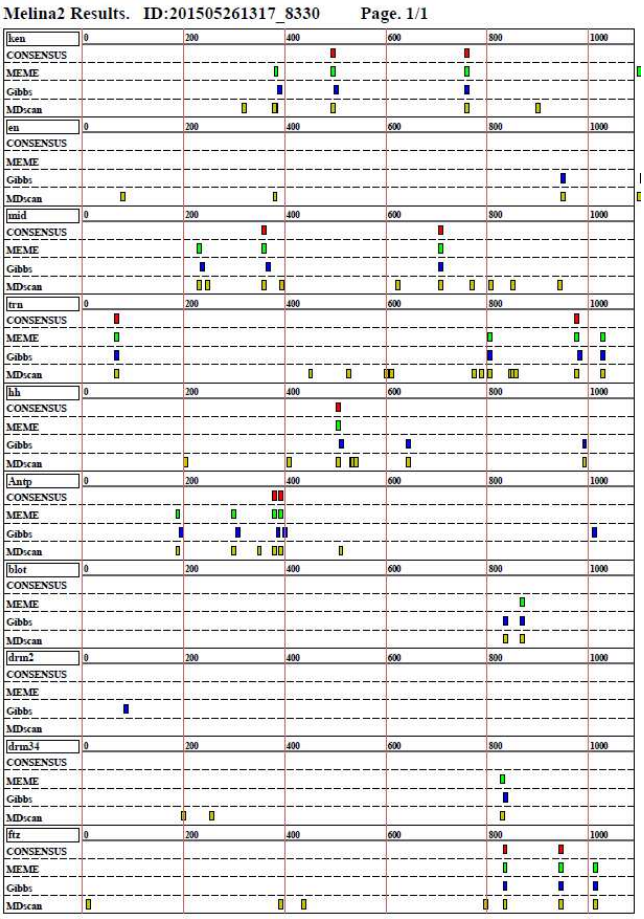
MEME and DREME analysis of the candidate enhancer regions each produced 2 overrepresented sequence motifs (Fig. 4-5). A). meme_motif #1 was built from 95 sites, and had a q-value of 5.5e-018. meme_motif#2 was built from 36 sites with a q-value of 1.2e-015. Dreme_motif#1 was built from 314 sites with a q-value of 4.7e-25. Dreme_motif#2 was built from 290 sequences with a q-value of 0.006. In each case, TOMTOM analysis identifies the strongest sequence as a Ftz-F1 binding site. When searched over the entire genome, the Ftz-F1 like binding motif is over-represented in Ftz ChIP-chip data (Fig. 5A-D).

Once the enhancer regions were shown to be controlled by Ftz/Ftz-F1, they were examined for additional motifs that might be important for Ftz/Ftz-F1 regulation. For this analysis, the seven strong enhancers found in this study (enhancers for *ken*, *mid*, *trn*, *hh*, *Antp*, and *blot*) were combined with three previously confirmed Ftz/Ftz-F1 responsive enhancers (*ftz* proximal enhancer, *en*, *drm2*, and *drm34*) to determine if there were any binding sites common to all ten. All four of the *de novo* algorithms used by the MelinaII program -Consensus, Meme, MDScanner, and Gibbs - found that the binding site for Ftz-F1 was the most common motif (Fig. 4-9A,B) which was also found in the PWM analysis of the negative enhancers (Table 4-4). This is expected, as all enhancers were chosen based on whether they had at least one candidate Ftz-F1 sequence.

Enhancers that fit the criteria used to search for enhancers in these studies (they contain at least both Ftz and Ftz-F1 binding sites), but were then shown to not be

enhancers, were designated as ‘negative enhancers’. Three negative enhancers have been found so far, the enhancers found in this study (*aay-lacZ*) and a previous study (*drm1 and drm5*) (Hou et al., 2009). The same analysis was run with the three negative enhancers, and the most common binding motif found was for the Forkhead (Fig 4-9C,D), which was also found in the PWM analysis of the negative enhancers below (Table 4-4).

A



B

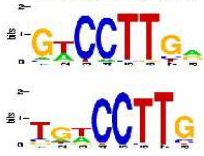
Consensus: GTCCGA

GACCTCG

Meme: TGTCCTTG

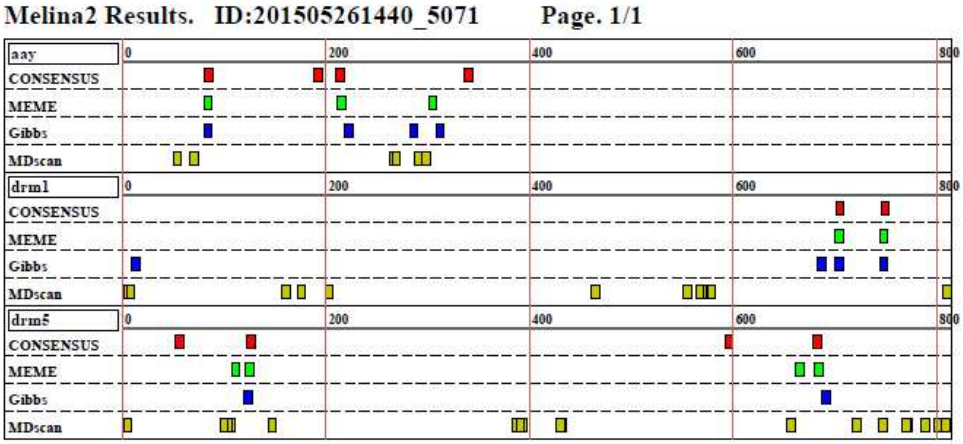
MDScan: GTCCTTGA

TGTCCTTG



Gibbs: TGTCCTTG

C



D

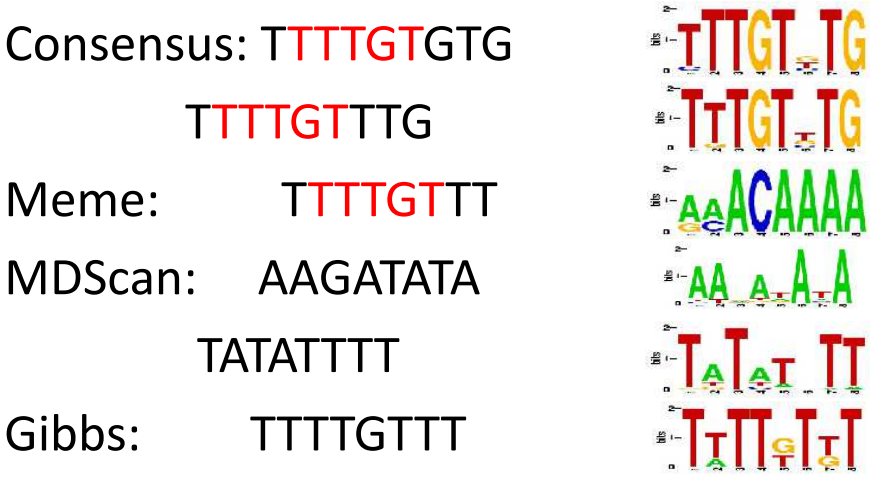


Figure 4-9. Examination of motifs in candidate Ftz enhancers shows Ftz-F1 as the prominent binding motif. Melina II software was used to compare A) the 10 positive enhancers to search for common motifs using 4 algorithms – CONSENSUS, MEME, Gibbs, and MDScan. The boxes represent motifs found by each algorithm common to all 10 enhancers across each enhancer. B) All 5 algorithms showed the most common motif to be the binding site for Ftz-F1. The software was also used to compare C) the negative enhancers *aay-lacZ, drm1,* and *drm5*. D) The strongest motif found between these three sequences was the binding site for Forkhead.

When the enhancers were analyzed using the UMIACS pwm programs, which scans sequences for known transcription binding sites from the JASPAR database, binding sites were found for 67 known transcription factors (Table 4-5). 57 of the 67 transcription factors contain a homeodomain, as does Ftz. These homeodomain-

containing transcription factors were eliminated, as they were likely identified by the program because their binding site is similar to the binding site of Ftz. One transcription factor that did not contain a homeodomain, DEAF-1, had a particularly high PWM score. The three negative enhancers were also analyzed using the UMIACS PWM program (Table 4-5). This pwm analysis also found many homeodomain binding sites, 64 out of 77, and only 13 transcription factors that did not contain a homeodomain. When the lists of transcription factors that do not contain a homeodomain for the positive enhancers and the negative enhancers were compared, only one transcription factor in the positive enhancer group was identified, while five were found in the negative enhancer group. For the positive enhancers, this transcription factor was Zeste (z), and for the negative enhancers, they were Trithorax-like (Trl), which codes for GAGA factor, Scalloped, Dichaete, Forkhead, and Sloppy-paired1 (Table 4-5).

Table 4-5. A position weight matrix program was used to search for transcription factor binding motifs within the positive and negative enhancer groups.

10 positive enhancers		3 negative enhancers	
protein	hox domain	protein	hox domain
br_Z2	n		
br_Z3	n	br_Z3	n
br_Z4	n	br_Z4	n
hb	n	hb	n
Ubx	y		
ovo	n	ovo	n
Abd-B	y	Abd-B	y
Antp	y		
Awh	y	awh	y
B-H1	y	B-H1	y
B-H2	y	B-H2	y
C15	y	C15	y
		CG11085	y
dbx	y	dbx	y
lms	y		
		CG13424	y
CG15696	y	CG15696	y
CG18599	y	CG18599	y
CG32105	y	CG32105	y
CG32532	y	CG32532	y
Vsx1	y	Vsx1	y
CG4328	y	CG4328	y
CG7056	y	CG7056	y
CG9876	y	CG9876	y
Deaf1	n	Deaf1	n
Dfd	y		
Dll	y	Dll	y
E5	y	E5	y
		Gsc	y
HGTX	y	HGTX	y
		Hmx	y
Lag1	y	Lag1	y
		Lim1	y
Lim3	y	Lim3	y
NK7.1	y	NK7.1	y
Oct	n	Oct	n
		OdsH	y
Optix	y	Optix	y
Pph13	y	Pph13	y
		Rx	y
Scr	y	Scr	y

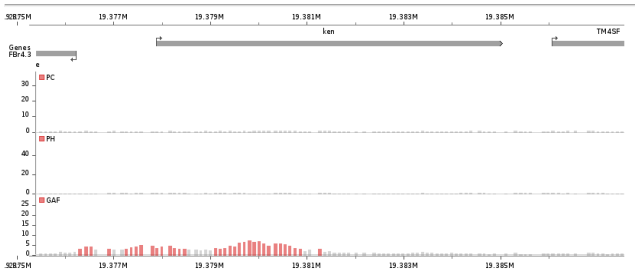
10 positive enhancers		3 negative enhancers	
protein	hox domain	protein	hox domain
Six4	y	Six4	y
		Trl	n
abd-A	y	abd-A	y
ap	y	ap	y
ara	y	ara	y
		bcd	y
bsh	y	bsh	y
btn	y		
cad	y	cad	y
caup	y	caup	y
ct	y	ct	y
ems	y	ems	y
en	y	en	y
eve	y	eve	y
		exd	y
		exex	y
ftz	y	ftz	y
		hbn	y
ind	y	ind	y
lab	y		
lbe	y	lbe	y
lbl	y	lbl	y
mirr	y	mirr	y
		oc	y
otp	y	otp	y
prd	y	prd	y
repo	y	repo	y
ro	y	ro	y
		sd	n
slbo	n	slbo	n
slou	y	slou	y
tup	y	tup	y
		unc-4	y
unpg	y	unpg	y
vvl	y	vvl	y
z	n		
zen2	y	zen2	y
CG34031	y	CG34031	y
		D	n
		fkh	n
H2.0	y	H2.0	y
PHDP	y	PHDP	y
		slp1	n
BEAF-32	n	BEAF-32	n

4.3.7 ChIP-chip data identifies GAF binding near negative enhancers

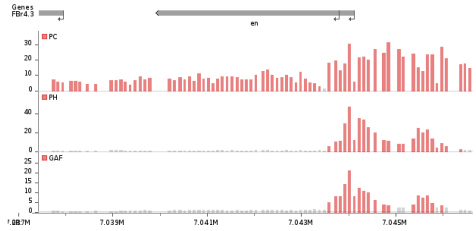
A ChIP-chip analysis was performed on pooled embryos 4-12 hours AEL for several chromatin markers, including methylation markers K4 and K27, as well as GAF, PC, and Pho (Schuettengruber et al., 2009). These data were used to analyze binding surrounding Ftz/Ftz-F1 targets and enhancers. For the methylation markers, no clear pattern was detected, and in fact, some targets that are known to be active during the blastoderm stage had strong K27 binding (data not shown). This could be explained

by the broad collection time of the embryos. These genes may be both activated and repressed over this long time period and in different cells, and so repression markers would be observed that are bound in some cells of the whole embryo at any point in time during the collection. When GAF binding, along with PC and Pho binding, was analysed surrounding the Ftz-F1 responsive targets and enhancers, an interesting pattern emerged (Fig. 4-10). While PC and Pho showed no distinct difference between positive and negative enhancers, strong GAF binding was found only near the three negative enhancers. All but one of the positive enhancers contained no GAF binding, while the *ken* enhancer had weak GAF binding.

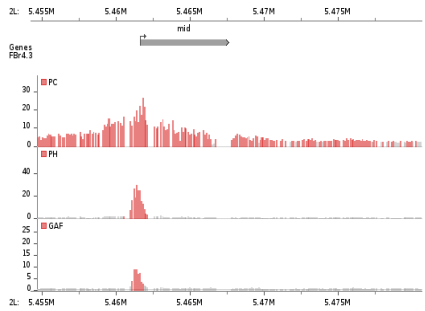
ken-lacZ



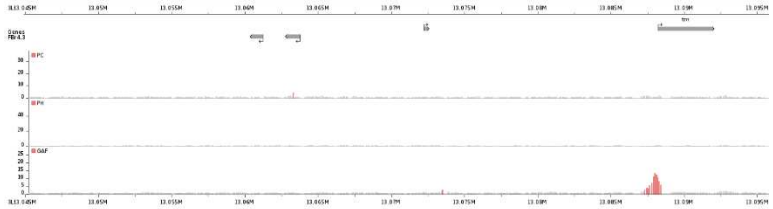
en-lacZ



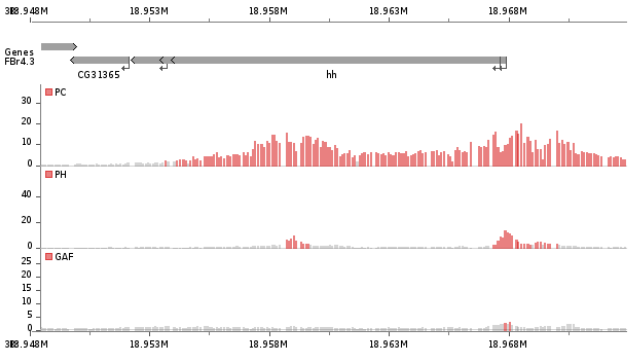
mid-lacZ



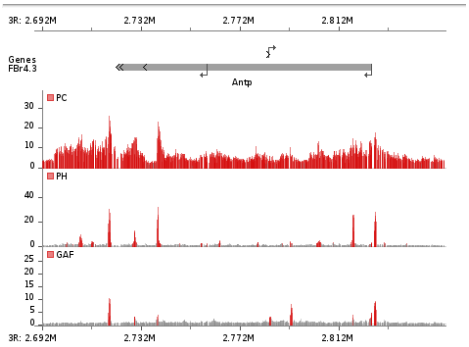
trn-lacZ



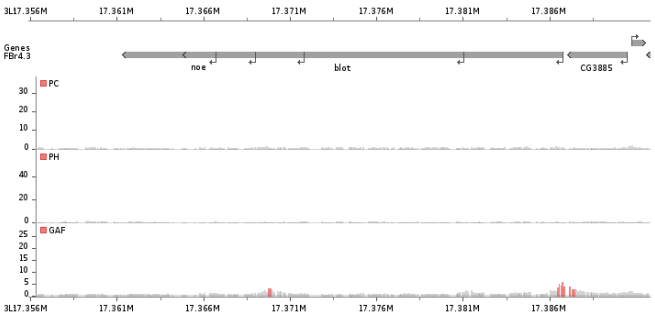
hh-lacZ



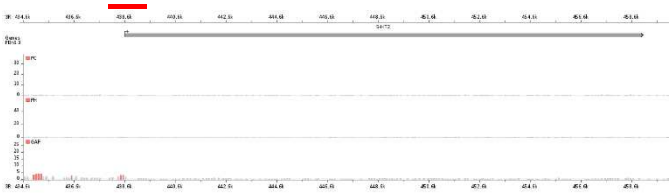
Antp-lacZ



blot-lacZ



5-HT2-lacZ



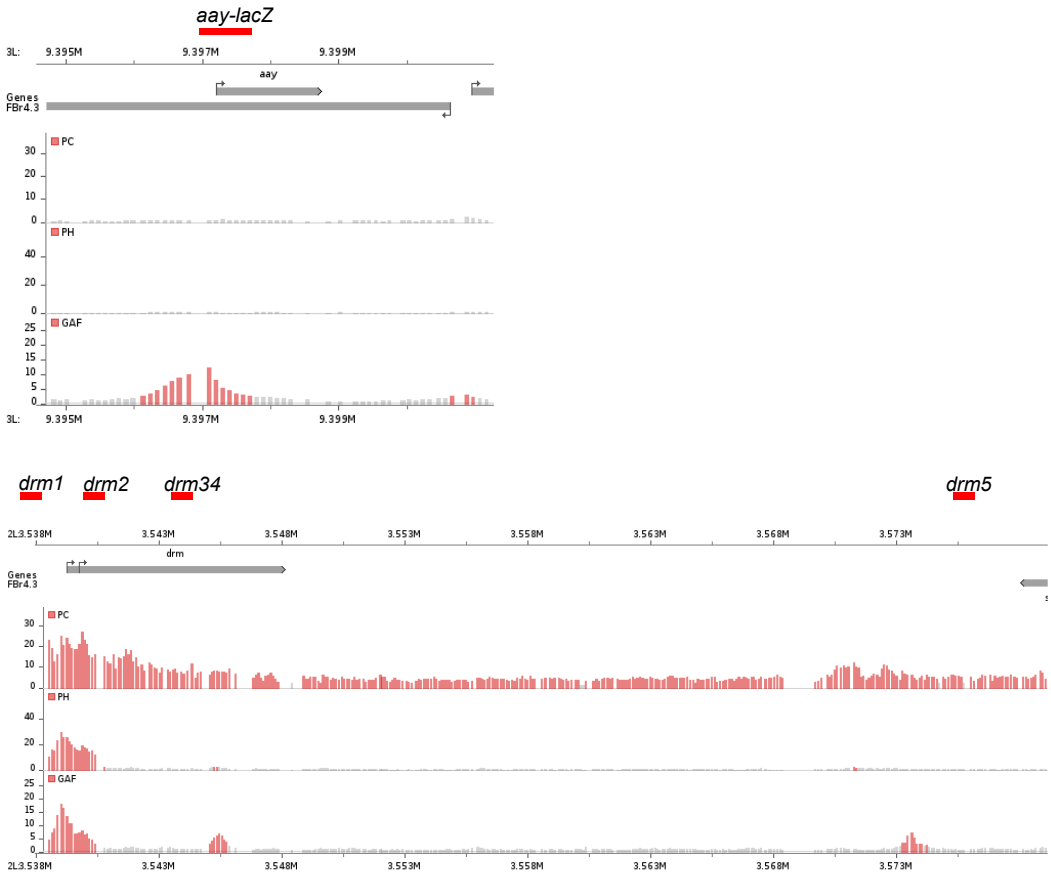


Figure 4-10 ChIP-chip data for Pho, PC, and GAF surrounding Ftz-Ftz-F1 enhancers. ChIP-chip data published by Schuettengruber et al., 2009 shows binding by fold change (y-axis) of transcription factors GAF, PC, and Pho along the DNA (x-axis). GAF binds at all three of the negative enhancers, while GAF does not bind near nine of the positive enhancers, and binds weakly near the *ken* enhancer. No discernable pattern could be seen for Pho and PC binding.

4.4 Discussion

This study identified target genes coordinately regulated by *Drosophila* Ftz and Ftz-F1 using microarray data, identified and confirmed the enhancers of those genes based on ChIP-chip data and enhancer reporter constructs respectively, and then analyzed these newly identified enhancers for interesting motifs. The microarray was used to compare RNA expression during blastoderm and gastrulation from control and

ftz-f1 mutant embryos, identifying 63 candidate Ftz-F1 targets genes that had at least a 1.1 fold increase between the two stages and a downregulation in *ftz-f1* mutants (Fig. 4-2). Nine of the top ten positive targets from this list were shown to be regulated by Ftz-F1 and Ftz in that their expression overlapped with that of *ftz* (Fig. 4-3), and target gene expression was lost in both *ftz* and *ftz-f1* mutants (Fig. 4-4). In order to determine whether Ftz/Ftz-F1 directly regulate expression of these target genes, Ftz/Ftz-F1 responsive enhancers were found by combining ChIP-chip data of Ftz binding (Fig. 4-5) and a computational search of candidate Ftz-F1 binding sites (Fig. 4-6). Analysis of the expression of these enhancers using *lacZ* reporter constructs in transgenic *Drosophila* embryos showed that eight of the nine the enhancers direct expression in a striped pattern (Fig. 4-7). These patterns were similar to those of the endogenous Ftz-F1/Ftz target genes they were expected to regulate, indicating that they were in fact enhancers of these genes, and the enhancer expression was lost in a *ftz* mutant background (Fig. 4-8). Analysis of these new enhancers, along with previously identified Ftz/Ftz-F1-regulated enhancers (Fig. 4-9), identified Deaf1, Dichaete, Zeste, and GAGA factor (Table 4-5) as possibly necessary for correct Ftz/Ftz-F1 regulation.

4.4.1 An RNA expression-based search for Ftz/Ftz-F1 targets was more efficient than those based on Ftz and Ftz-F1 binding sites

Two approaches were used previously to identify Ftz/Ftz-F1 target genes (Bowler et al., 2006; Hou et al., 2009) but both proved to be inefficient in identifying targets. The first (Bowler et al., 2006) used a computational approach to scan the genome for known Ftz and Ftz-F1 binding sites, based on the sequence configurations

of Ftz and Ftz-F1 binding sites in the two previously verified target genes, *engrailed* (Florence et al., 1997) and *ftz* itself (Yu et al., 1997). Genes that had at least 5 composite Ftz and Ftz-F1 binding sites in these orientations within 40 kb of their transcription start site were considered to be candidate targets. Of the 111 candidate targets found this way, 30 had known expression patterns published by BDGP. Their expression patterns were analyzed to determine whether they overlap with Ftz and thus could be directly regulated. Of these 30 candidate targets, only 2 candidate genes were shown to be new targets of Ftz/Ftz-F1. While the expression of most of the candidate targets was not affected in *ftz* mutants, 3 of those 30 candidate targets had a strong striped pattern that was altered in *ftz* mutants, and one of those was *en* itself. The two new genes, *Sulf1* and *apt*, both co-localize with *ftz* expression, and this expression was lost in *ftz* and *ftz-f1* mutants. To further determine if they were targets, *ftz* was ubiquitously expressed using an NGT40-GAL4 driver. This ectopic expression of Ftz altered *apt* and *Sulf1* expression similarly to that of *en,* a known direct target, with fused stripes of expression, indicating that they were regulated by Ftz/Ftz-F1. *Sulf1* was shown to be a direct target of Ftz/Ftz-F1. Two candidate Ftz-F1 binding sites, *Sulf2* and *Sulf34*, were found within 8.5 kb upstream of the TSS for *Sulf1*. Direct interaction between the enhancer *Sulf34* with Ftz and Ftz-F1 in an electromobility shift assays was observed, and this interaction was lost when the Ftz-F1 site was mutated in the *Sulf34* enhancer. Enhancer-reporter transgenes demonstrated that *Sulf34* was sufficient to generate a *Sulf1* pattern, while *Sulf2* was not sufficient on in its own but worked with *Sulf34* to generate a cleaner pattern.

In the second study (Hou et al., 2009), the opposite approach was used. Known expression patterns published by BDGP were visually examined for any striped expression patterns suggestive of pair-rule targets, which led to 95 candidate targets. The genomic regions (~20 kb) around these targets were scanned for Ftz and Ftz-F1 candidate binding sites in the *ftz* or *en* configuration, similar to method used in Bowler 2006, which narrowed the candidate targets to 30. The expression pattern of these 30 genes was examined in both wildtype and *ftz-f1* mutants, and only 10 of the genes showed an aberration in the mutants. *5HT2*, *noc*, and *drm* were further investigated. For all three, 7 of their stripes overlapped with Ftz expression. Ectopic expression of Ftz altered the expression of all three genes, with several stripes becoming fused or paired, and expression became expanded or unclear in all. One of these candidate targets*, drm* was further analyzed. Ftz-F1 binding sites were identified in the genomic region surrounding the *drm* coding region and four enhancers were identified. Two of these enhancers, *drm2* and *drm34*, directed striped expression in the early embryo. These were shown to be directly regulated by Ftz/Ftz-F1 by mutating the core AAGG of the Ftz-F1 sites to AGAT and fusing these mutated enhancers to a lacZ reporter gene. The mutated enhancers gave reduced or no expression, showing the Ftz-F1 site was necessary for the enhancers to function, indicating direct control by Ftz/Ftz-F1. Interestingly, two other candidate enhancers for *drm*, *drm1 and drm5*, which were identified by Ftz and Ftz/F1 binding sites, did not give a striped pattern and so were not enhancers at all. This showed the presence of candidate Ftz and Ftz/F1 binding sites was not enough to identify a Ftz/Ftz-F1 responsive enhancer. Overall, this approach identified 10 candidate Ftz/Ftz-F1 target genes, of which only only 3 targets were

further investigated and shown to be actual targets. Together, these two approaches showed that a search for Ftz/Ftz-F1 targets based on expression patterns and Ftz and Ftz-F1 binding sites was not enough to identify targets efficiently. This indicates that Ftz/Ftz-F1 use more information to locate their target sites in the genome than just the core Ftz and Ftz-F1 binding sites.

In the current study, a microarray comparing wildtype and *ftz-f1* at blastoderm stage identified 63 candidate targets, 10 of which were investigated further and shown to overlap with *ftz* expression. The expression of each of the ten genes was altered in *ftz* and *ftz-f1* mutants, indicating that they were indeed Ftz/Ftz-F1 targets. This method may have been more successful in finding targets because it did not depend on computationally identifying Ftz and Ftz-F1 sites in the genome, but was instead based on RNA expression. This again suggests that more information is needed to determine how Ftz/Ftz-F1 choose their targets than just the presence of Ftz and Ftz-F1 binding sites.

4.4.2 Ftz/Ftz-F1 target genes have known roles in development

The new Ftz/Ftz-F1 target genes generated by this study all have roles in development. *En* and *hh* have many roles in development, the most significant one as pair-rule targets being segment polarity genes (Peel, 2004). *hh* is also the ligand in the hedgehog signaling pathway (Briscoe and Therond, 2013). *mid* is also a segment polarity gene (Tearle and Nusslein-Volhard, 1987), along with being involved in axon guidance (Liu et al., 2009), heart development (Miskolczi-McCallum et al., 2005; Qian et al., 2005; Qian et al., 2008; Reim et al., 2005), and nervous system development (Leal et al., 2009) (Gaziova and Bhat, 2009) (Buescher et al., 2006). *Antp* is a *Hox* gene

that specifies segments in the thorax (Schneuwly et al., 1987), as well as later roles in lymph gland development (Mandal et al., 2007), specification of the muscle cell fate (Enriquez et al., 2010), and neuroblast development (Tsuji et al., 2008). All of the other genes have known roles in development in the later embryo. *ken* works in genitalia development (Lukacsovich et al., 2003) and is a regulator of JAK/STAT pathway (Arbouzova et al., 2006). *aay* has a role in axon guidance/nervous system development (Prokopenko et al., 2000). *trn* encodes a cell surface protein that has many roles - in cell movement and migration in the imaginal discs (Milan et al., 2001), tracheal branch outgrowth "joining of tracheal branches over segment boundaries" (Krause et al., 2006), salivary gland invagination (Maybeck and Roper, 2009), retinal epithelial integrity (Mao et al., 2008), and mediate homophilic cell adhesion (Shinza-Kameda et al., 2006). *Blot* is a putative neurotransmitter with a role in morphogenesis of epithelium (Johnson et al., 1999). *5HT2* is a serotonin receptor with specific ligand binding (Colas et al., 1995). It helps control morphogenetic movements in gastrulation - mutants had defective germband extension and mislocalization of Armadillo, suggesting abnormalities in adherens junctions (Colas et al., 1995; Colas et al., 1999a; Colas et al., 1999b; Schaerlinger et al., 2007). It was already shown to be a target of Ftz/Ftz-F1 (Hou et al., 2009), as discussed in the previous section.

4.4.3 Nine Ftz/Ftz-F1 enhancers direct expression in striped patterns

Nine of the ten Ftz/Ftz-F1 enhancers found in this study directed reporter gene expression in patterns in the early embryo. The one exception, the Ftz/Ftz-F1 enhancer for *aay*, was the only candidate enhancer found using the ChIP-chip data at 25% FDR,

and so was likely a false positive. If *aay* is a Ftz/Ftz-F1 target, either the Ftz binding within 100 kb of the gene was missed by the ChIP-chip analysis, or the enhancer is outside of the 100 kb range used.

Seven of the ten newly found Ftz/Ftz-F1 enhancers were categorized as "strong" enhancers, as they directed *lacZ* expression in patterns similar to the endogenous genes they were expected to regulate (Fig. 4-7). That is, *ken-lacZ* was expressed in two stripes, one anterior and one posterior, similar to endogenous *ken*. *en-lacZ*, *mid-lacZ*, and *hh-lacZ* were expressed in 14 stripes. *tal-1A-lacZ*, *5-HT2-lacZ*, and *blot-lacZ* were expressed in seven stripes. *Antp-lacZ* was expressed in one broad stripe. *trn-lacZ* was particularly interesting, because while *trn* is expressed in eight stripes, the Ftz/Ftz-F1 enhancer directed expression in only seven stripes. The most anterior *trn* stripe does not overlap with *ftz*, and so was unlikely to be controlled by Ftz. The expression pattern of *trn-lacZ* indicates that the seven posterior stripes are indeed regulated by Ftz/Ftz-F1, consistent with the overlap of these seven stripes.

One of the Ftz/Ftz-F1 enhancers, *5HT2-lacZ*, was expressed in a pattern that differed from the endogenous expression of *5HT2*. For this "weak" enhancer, rather than the 7 stripes expected, the reporter genes were expressed in 3 broad, spotty bands in the middle of the embryo. This was the only enhancer chosen that did not entirely fit the criteria – there was no Ftz binding at this location based on the ChIP-chip data. Instead, this enhancer was chosen because *5HT2* was a known *ftz/ftz-f1* target, but no Ftz binding was detected within 70kb of the gene. Instead, *5HT2-lacZ* was selected based on a candidate *ftz-f1* site found near the gene. This result could have several explanations: 1) the 1 kb chosen for the candidate enhancer includes only a small part of

the functional enhancer; 2) there are other enhancer(s) working with this enhancer to direct the correct expression pattern; and/or 3) the candidate enhancer is not a true enhancer for the proposed target, although it could still be enhancer for a different gene. Combining the strong and weak enhancers, nine out of the ten enhancers chosen (with the criteria of Ftz binding within 100 kb of a Ftz/Ftz-F1 target gene and containing at least one candidate Ftz-F1 site), were shown to have enhancer activity at least partially overlapping the expression of the endogenous gene.

aay-lacZ was the only candidate enhancer that directed no β-Gal expression, indicated that it is not an enhancer. This enhancer was chosen at a region that only had Ftz binding at 25% FDR, so this may have been a false positive.

4.4.4 Ftz binding in the *Drosophila* genome

The ChIP-chip data published by BDNTP identified 403 Ftz binding sites in the genome of blastoderm stage embryos using a 1% FDR. An additional 3,318 sites were identified when a 25% FDR was used as a cutoff. It is likely that the true number of Ftz targets is between these numbers, as 1% FDR is a stringent cutoff that would miss many sites while 25% FDR would allow many false positives. In contrast, 14.5 million candidate Ftz binding sites were found in the *Drosophila* genome based on sequence alone (Bowler et al., 2006). This begs the question of how there can be so few actual Ftz binding sites. One possibility is that Ftz is only bound when Ftz-F1 is also bound. Bowler searched for two possible Ftz/Ftz-F1 cooperative configurations of binding sites in the genome, for *en* and *ftz* enhancers, and found 3,310 candidate sites. However, the number of genes that are differentially expressed between wild type and *ftz-f1* mutants at different p-value cut offs (Table 4-1) suggests that the number of Ftz/Ftz-F1 targets is

between 159 and 897. There are likely more than only these two cooperative configurations for Ftz and Ftz-F1 binding sites, and so there are likely more than the 403-3,318 candidate cooperative binding sites. If Ftz binding is restricted to areas that are also bound by Ftz-F1, what explains the difference between candidate cooperative binding sites (403-3,318) and the likely number Ftz/Ftz-F1 targets (159-897)? It is possible that another factor besides Ftz-F1 is also restricting the binding of Ftz.

Of the 403 Ftz bound regions, 56 were CRMs that have previously been described, including nearly all published enhancer elements previously shown to be Ftz responsive. Some examples are enhancers for *teashirt* (Coré et al., 1997), *en* (Florence et al., 1997), gooseberry (Bouchard et al., 2000), *Ubx* (*pbx*) (Müller and Bienz, 1992; Pirrotta et al., 1995), *Ubx* (*bx*) (Qian et al., 1993),and the *ftz* upstream element (Pick et al., 1990a). Several known and candidate Ftz/Ftz-F1 responsive elements were not identified by either the ChIP-chip data or the analysis in this study, including *drm34* (Hou et al., 2009), the *ftz* zebra element (Ueda et al. 1990), the 5th *hairy* stripe (Langeland and Carroll, 1993). As mentioned above, two Ftz/Ftz-F1 responsive enhancers had been identified for *drm – drm2* and *drm34* (Hou et al., 2009). Only *drm2* was bound by Ftz in the ChIP-chip analysis - *drm34* was not identified at either 1% or 25% FDR. While the *drm34* site was originally selected based on the presence of a consensus Ftz-F1 binding site (Hou et al., 2009), that Ftz-F1 binding site was not detected by this study using the Ftz-F1 dreme position weight at the threshold (Fig. 4-5A). However, because *drm34* was confirmed as a Ftz/Ftz-F1 responsive enhancer, this indicates that bona fide targets were missed in the ChIP-chip analysis. Additionally, the *hairy* stripe 5 element has been suggested to be regulated by Ftz-F1 due to binding site

conservation (Langeland and Carroll, 1993). Since these sites were missed by this analysis, other areas with Ftz binding may have been missed in the analysis, so it is likely there are more Ftz/Ftz-F1 enhancers that were not found with the current data. If Ftz's role in segmentation is to be understood, these other Ftz bound regions need to be investigated to see which control targets are involved in segmentation.

4.4.5 Searching for motifs in Ftz/Ftz-F1 enhancers

As mentioned above, the number of actual Ftz/Ftz-F1 target genes appears to be much smaller than those predicted based upon predicted or verified Ftz genomic binding at the blastoderm stage. Analysis of verified Ftz/Ftz-F1-regulated enhancers was carried out to identify candidate co-regulators that may limit the action of Ftz and/or Ftz-F1 to specific genes (Fig. 4-9, Table 4-2). The MelinaII analysis was only able to find Ftz-F1 binding sites (Fig. 4-7B in red) using *de novo* algorithms. This is likely because each enhancer was chosen on the criteria that they must contain at least one Ftz-F1 site, and most contained multiple (Table. 4-4), making the Ftz-F1 site the strongest site. Examining the ten enhancers using the UMIACS pwm program gave a list of strong transcription factor binding sites found in each sequence (Table 4-5). Most of these binding sites were for transcription factors with homedomains. Given the similarity of homeodomain binding sites, it is likely that these binding site consensus sequences were similar enough to the Ftz binding site to be found in this search. However, several transcription factors did not contain a homeodomain and so could not have been found from a site similar to the Ftz binding site. Similar results were found when the three negative enhancers *drm1, drm5* and *aay-lacZ* were analyzed by the UMIACS pwm program.

The transcription factors that did not contain homeodomains were further analyzed to determine if they might have a significance for Ftz/Ftz-F1 binding. DEAF-1 stood out as a possible co-regulator of Ftz/Ftz-F1 target genes, as it had the strongest pwm score for the binding sites in the positive enhancers. However, DEAF-1 binding sites were also found in the three negative enhancers, with a lower pwm score. This does not preclude DEAF-1 as a possible co-regulator, but it does mean that a lack of DEAF-1 binding sites is not what kept the negative enhancers from functioning as enhancers. DEAF1 contains two motifs, SAND and MYND. The 3D structure of the SAND domain has been characterized and appears to encode a DNA binding motif involved in transcriptional regulation (Bottomley et al., 2001). The MYND motif in *Drosophila* DEAF1 may have a role in protein-protein interaction (Gross and McGinnis, 1996). Mutant maternal germline clones of *Deaf1* caused early embryonic death in 70% of embryos, before gastrulation, but some completed embryogenesis. The embryos that survived to first instar larvae or the end of embryogenesis had segmentation defects including loss and fusion of all segments across the anterior/posterior axis - that could suggest a role in pair-rule regulation (Veraksa et al., 2002). *ftz* expression was also lost or altered in *Deaf1* maternal mutants, as was the expression of gap genes *knirps* and *kruppel*. Thus, *Deaf1* likely plays multiple roles in early embryonic development making specific interactions difficult to unravel. However, the consistent presence of strong DEAF1 binding sites in enhancers that are regulated by Ftz/Ftz-F1 makes DEAF1 a good candidate for further study as a possible co-regulator of Ftz/Ftz-F1.

When the lists of transcription factors that did not contain homeodomains for both the positive and negative enhancers were compared, one transcription factor was identified in the positive enhancer group that was not in the negative enhancer group. All ten positive enhancers contained binding sites for the transcription factor Zeste, which were not found in the three negative enhancers. Zeste can act as both an activator and repressor (Hur et al., 2002; Mulholland et al., 2003; van Steensel et al., 2003), working on the chromatin level. Z acts as an activator by recruiting the *Drosophila* Brahma (BRM) chromatin remodeling complex, which opens chromatin (Dejardin and Cavalli, 2004; Kal et al., 2000).

When the reverse search was done, five transcription factors lacking homeodomains were identified in the negative enhancer group that were not found in the positive enhancer group. Three of these can easily be dismissed as necessary for Ftz/Ftz-F1 function – *sd* and *fkh* do not have a pattern consistent with Ftz/Ftz-F1 function, while *slp1* is a known pair-rule gene that does not affect Ftz/Ftz-F1 function. The functions of the two other transcription factors identified, Dichaete and Trl, are known and examined further here.

Dichaete is a member of the protein superfamily High Mobility Group (HMG). This group of proteins affects regulation on the chromatin level (Grosschedl et al., 1994; Landsman and Bustin, 1993), and is capable of bending DNA when it binds (Ferrari et al., 1992; Giese et al., 1992). *Dichaete* is expressed in the blastoderm in a striped pair-rule pattern and contains one HMG DNA-binding domain (Nambu and Nambu, 1996; Russell et al., 1996). Embryos that are null mutants for *Dichaete* show defects in *eve, h, run*, and *ftz* expression. The expression of *ftz* in these mutants have

stripes 3 and 4 fused, suggesting a repressive function on *ftz*. *Dichaete* does not appear to be necessary for the establishment of any pair-rule genes, but may be involved in the refinement and maintenance (Nambu and Nambu, 1996). When *Dichaete* was overexpressed, it caused the expansion of stripes and an expansion throughout the posterior for the expression of *eve*, *h*, and *run*. This overexpression had a much smaller effect on *ftz* expression, causing a shift of the most posterior stripe 3 or 4 cells more towards the posterior, although all stripes maintained their normal shape and size (Russell et al., 1996). Like DEAF-1, the segmentation defects caused by a *Dichaete* mutant make it a possible that it is necessary for Ftz/Ftz-F1 function, although it is also possible that these segmentation defects are caused by some other function of Dichaete in segmentation. Furthermore, Dichaete is a particularly interesting candidate for Ftz/Ftz-F1 function because it was previously found to interact with Ftz-F1 in a whole genome yeast two-hybrid experiment.

Trl, like Z, can also act as an activator or repressor on the chromatin level (Hur et al., 2002; Mulholland et al., 2003; van Steensel et al., 2003). Trl is in the trithorax family, commonly known to activate genes (Lehmann, 2004; Simon and Tamkun, 2002). *Trl* codes for GAGA Factor (GAF) (Farkas et al., 1994), which contains three domains -a zinc finger DNA-binding domain (Pedone et al., 1996) and a BTB/POZ protein-protein interaction domain (Bardwell and Treisman, 1994) are connected by a Q domain, essential for transcriptional activation (Greenberg and Schedl, 2001). GAF assists in gene activation through a mechanism of anti-repression (Croston et al., 1991; Lehmann, 2004). However, it has also been shown to repress genes. GAF binding sites have been found to be essential sequences in polycomb response elements (PREs)

(Busturia et al., 2001; Hagstrom et al., 1997; Horard et al., 2000), where Polycomb Group proteins (PcG) bind to repress homeotic genes. Further, both GAF and Pleiohomeotic (Pho) are involved in recruiting PcG (Fujioka et al., 2008; Mishra et al., 2001). It has been shown that GAF binds chromatin and facilitates the binding of Pho as well, and together these two proteins recruit PcG to the chromatin to repress genes (Mahmoudi et al., 2003). ChIP-chip binding data of GAF surrounding both the positive and negative the Ftz/Ftz-F1 enhancers, it was determined that GAF bound only at the negative enhancers, with the exception of weak binding near *ken-lacZ* (Fig. 4-10). The collection time for this experiment was broad so it is impossible to determine from this data if GAF was binding during blastoderm and germband extension, but this strong binding difference between positive and negative enhancers should be further investigated. The GAF binding sites found in the three negative enhancers, along with the GAF binding data, provide a possible explanation as to why these candidate enhancers did not function as Ftz/Ftz-F1 enhancers – they may have been repressed at the chromatin level by GAF, using a mechanism of repression that does not require PC and Pho. Both Z and GAF should be further studied as possible Ftz/Ftz-F1 co-activators or repressors, respectively.

Conclusions and Future Directions

Here I have studied the expression and function of the developmental gene *ftz* in the mosquito *Aedes aegypti* and compared it with what is known in *Drosophila*. RNA from *Aa* embryos was isolated and converted to cDNA. RT-PCR was used to examine the timeline of expression for *Aa-ftz* and its cofactor *Aa-ftz-f1*. These temporal expression patterns were consistent with *Dm-ftz* and *Dm-ftzf1*. Along with this, an alignment of the sequences of *Dm*-Ftz and *Aa*-Ftz showed that the functional sequences were nearly identical, indicating that the two genes will have the same pair-rule function. However, further attempts to determine the expression and function of *Aa-ftz* were unsuccessful. In situ hybridization using an *Aa-ftz* RNA probe and antibody staining with anti-*en* 4F11 were difficult due to the tough chorion of this species. This inability to observe the expression pattern of *en* also made it impossible to determine any defects in the early embryos that may have been caused by a functional experiment in which *Aa-ftz* was overexpressed in a transgenic line of *Aedes*. However, the low survival rate of heatshocked *Aa-ftz* embryos compared to wild type embryos does indicate that overexpression of *Aa-ftz* has a lethal effect.

Further experiments need to be carried out to determine the expression pattern of *ftz* in the blastoderm through germband extension in mosquitoes. This experiment could continue to be attempted in *Aedes*, or it could be tried in the mosquito *Anopheles gambiae*, now known to be easier to stain because of differences in the chorion. Functional studies might also be more successful in *Anopheles*, as the *en* pattern might be easier to obtain and so segmental defects could be observed. A knockdown of *ftz* and

ftz-f1 by RNAi could also be used to determine the loss of function phenotype of *ftz* in mosquitoes.

In addition to the conservation of *ftz* in Diptera, the function of *ftz* was studied in *Drosophila*. Specifically, I wanted to find a larger set of Ftz/Ftz-F1 targets and enhancers so that we could determine any other inputs that are necessary for this function. The question of how transcription factors find their targets in vivo is complex and varied across different factors. Before this study, the core binding site of both Ftz and Ftz-F1 was known, but studies trying to determine more genomic Ftz/Ftz-F1 binding sites based on a computational search for the in vitro DNA binding sites only found a few new target genes. In my study, I combined the experimental data from a microarray and a ChIP-chip experiment with a computational search for candidate *ftz-f1* binding sites, and using this approach, nine of the top eleven targets were confirmed to be bona fide targets. Five of these were new target genes. Interestingly, every target from the microarray comparing wild type and *ftz-f1* mutant embryos proved to also require *ftz* to be activated. So far, all known targets of Ftz and Ftz-F1 in their segmentation role have needed both factors to be activated. This provides more evidence that Ftz and Ftz-F1 do not function separately in the early embryo, despite the differences in their expression patterns and their roles later in development.

I was also able to identify seven new Ftz/Ftz-F1 responsive enhancers in the *Drosophila* genome – enhancers for *ken*, *mid*, *5HT2*, *trn, hh, Antp*, and *blot* (the *en* enhancer was already known). Combined with previously known enhancers (enhancer for *ftz* and *drm*), there are now ten confirmed enhancers of Ftz/Ftz-F1. This collection of enhancers gave us the opportunity to examine them for any binding motifs that were

previously unknown to be necessary for Ftz/Ftz-F1 function. There are also three enhancers (for *aay* and *drm*) that fit previous search criteria, including Ftz binding and candidate *ftz-f1* sites, which proved to not function as enhancers. This group allowed us to also search for any binding motifs that may explain why Ftz/Ftz-F1 did not activate these enhancers. Using a position weight matrix that searched the two groups of enhancers for known binding sites of *Drosophila* transcription factors stored in the Jaspar database, we found three factors that may play a role in how Ftz/Ftz-F1 find their targets. DEAF-1 and Zeste should be further explored as possible activators of Ftz/Ftz-F1 targets, while GAGA Factor should be studied as a possible repressor of candidate Ftz/Ftz-F1 targets. A simple first step would be to observe the expression pattern of all known Ftz/Ftz-F1 targets in embryos carrying mutations in the genes encoding these factors, although this would be made difficult made by the universal effects all three factors have on development. In addition, the binding sites for the factor should be mutated in some or all of the known Ftz/Ftz-F1 enhancers to see if the expression patterns directed by these enhancers are affected.

Another interesting line of study from this work would be to explore if and how the new set of Ftz-F1 targets are involved in segmentation. While the fact that Ftz/Ftz-F1 has a role in segmentation has been well documented, it is not known how they promote segmentation. More information on how their targets affect segmentation would go a long way towards answering this question. Beyond Ftz/Ftz-F1, this is still a question for all of the pair-rule genes, and any work to clarify this question for Ftz/Ftz-F1 would give insight into how all pair-rule genes contribute to segmentation.

Finally, having a greater understanding of how Ftz/Ftz find their enhancers, and any other factors that may be involved, will only help to expand our understanding of how Ftz evolved into its role as a segmentation gene. If other factors are involved, it would interesting to know at which point in the evolution of arthropods they joined in the regulation controlled by Ftz/Ftz-F1, or if instead they had to be co-opted into development or segmentation before Ftz/Ftz-F1 gained their segmentation role.

Appendix: *Antp* data in *Aedes Aegypti.*

Antp and *ftz* are closely related and have many similarities. *ftz* likely arose as a duplication of *Antp* (Telford, 2000), and Ftz has a nearly identical homeodomain to Antp . Ectopic expression of basally branching *ftz* gene also cause an antennae-to-leg transformation similar to ectopic expression of *Antp* (Lohr et al., 2001), indicating they have the ability to regulate some of the same targets. Because of this, originally I was going to use *Aa-Antp* as a control when studying *Aa-ftz* expression and function. *Aa-Antp* would have served as a control for heatshock experiments, hypothesizing that the overexpression of *Antp* in *Aedes* would cause the same antennae-to-leg transformation seen in *Drosophila*. At the same time, it would have been the first homeotic transformation performed in a mosquito. However, due to the difficulties working with *Aedes* described in Chapter 3, this line of experimentation was not followed. Here, I am presenting the preliminary results I discovered about *Aa-Antp*. First, an alignment between *Dm-Antp* and *Aa-Antp* shows that they both contain the YPWM motif and that their homeodomains are 100% identical (Fig. A-1), indicating that they will have the same function in both *Drosophila* and *Aedes*.

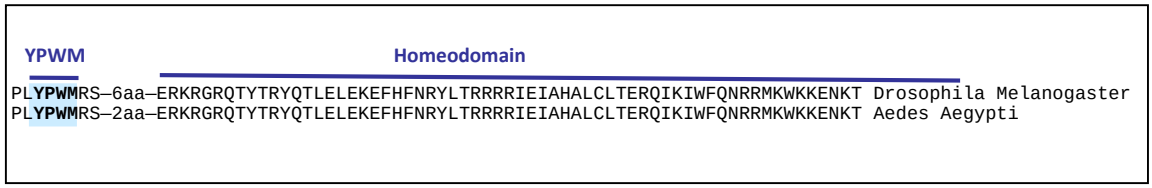


Figure A-1. An alignment between the homeodomain of *Dm-Antp* and *Aa-Antp*. The alignment shows that both *Dm-Antp* and *Aa-Antp* have a YPWM motif, and that the homeodomains of the two proteins are 100% identical.

The RT-PCR experiment shown in chapter 3 (Fig. 3-2) originally also included the expression profile of *Antp*. This figure is presented in full here (Fig. A-2). In *D. melanogaster*, the blastoderm stage ends around 3 hours AEL, and germband extension goes from 3 hours to 5 hours AEL. The expression of *Antp*, *ftz*, and *ftz-f1* also follow this same pattern, being expressed about 3 times later and during the same stages in *Aedes* as in *Drosophila*. In *Drosophila*, *Antp* is expressed strongly starting during blastoderm and lasts through germband extension, approximately 2 -5 hours AEL (Bermingham et al., 1990). In *Aedes aegypti*, *Antp* began expression at 6-9 hours, in which blastoderm ends and gastrulation begins, and continues throughout germband extension. *Aa- ftz-f1* was expressed during the entire 15 hours, with highest expression between 3-9 hours AEL. *Aa-ftz-f1* expression decreased during 9-15 hours AEL. The weaker *Aa-ftz-f1* band for 9-12 hours can be explained by a lower RNA concentration for the collection at this time point, as shown by weaker band for the actin control. *Aa-ftz-f1* expression matches the *Dm-ftz-f1* expression timeline, which is maternally deposited and continues to be ubiquitously expressed in the embryo until it begins to fade during germband extension. *Drosophila ftz* is expressed from the blastoderm stage to the beginning of germband extension, 2-4 AEL (Hafen et al., 1984). The timeline for *Aa-ftz* expression began in the early blastoderm (3-6 hours), decreased some through gastrulation (6-9 hours), and continued through germband extension, which also matches the expression timeline for *Dm*-ftz.

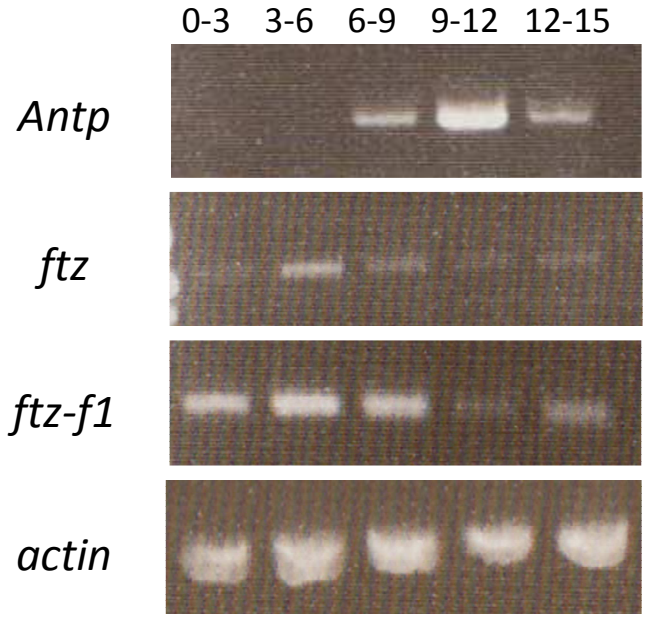


Figure A-2. Expression profile of *Antp*, *ftz* and *ftz-f1* in early *Aedes* embryos. cDNA was collected from *Aedes aegypti* embryos aged to 5 different time points and RT-PCR was performed to examine the expression levels across the first 15 hours of embryonic development. a*ctin* was used as a control across all time points. *Antp* was most highly expressed at 3-6 hours AEL, *ftz* was most highly expressed at 9-12 hours AEL, and *ftz-f1* was most highly expressed at 0-9 hours AEL.

When *Aa-Antp* was isolated from *Aedes* in order to make a *hsp70-Aa-Antp* construct (originally to be transformed into *Aedes* for heatshock overexpression experiments), two isoforms of *Antp* were isolated (Fig. A-3). One contained an exon that was not found in the other. It was never determined which one was expressed in the early *Aedes* embryo. To do this, RT-PCR should be performed on RNA taken from a pool of 0-8hr embryos, using isoform specific primers.

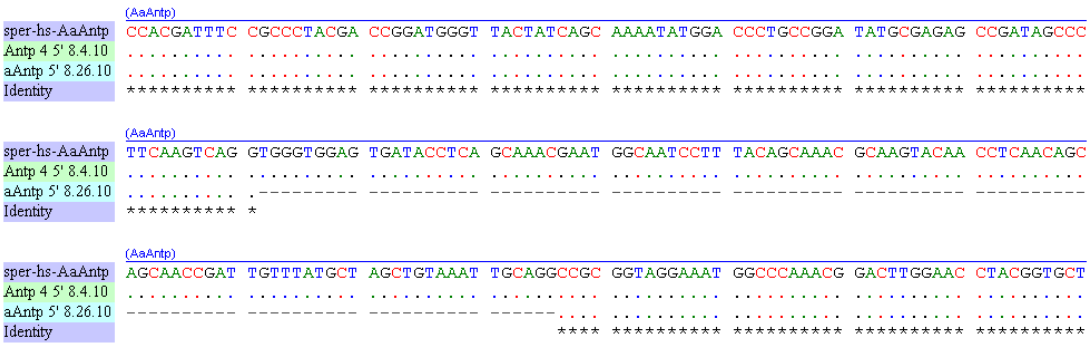


Figure A-3. Two isoforms of *Antp* were isolated from *Aedes aegypti* gDNA.

References

Abzhanov, A., and Kaufman, T.C. (2000). Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. Development *127*, 2239-2249.

Akam, M. (1989a). Drosophila development: making stripes inelegantly. Nature *341*, 282-283.

Akam, M. (1989b). Hox and HOM: homologous gene clusters in insects and vertebrates. In Cell (United States), pp. 347-349.

Alonso, C.R., Maxton-Kuechenmeister, J., and Akam, M. (2001). Evolution of Ftz protein function in insects. In Curr Biol (England), pp. 1473-1478.

Arbouzova, N.I., Bach, E.A., and Zeidler, M.P. (2006). Ken & barbie selectively regulates the expression of a subset of Jak/STAT pathway target genes. Curr Biol *16*, 80-88.

Aronson, B.D., Fisher, A.L., Blechman, K., Caudy, M., and Gergen, J.P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. Mol Cell Biol *17*, 5581-5587.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653-1659.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Bardwell, V.J., and Treisman, R. (1994). The POZ domain: a conserved protein-protein interaction motif. Genes Dev *8*, 1664-1677.

Baumgartner, S., Martin, D., Hagios, C., and Chiquet-Ehrismann, R. (1994). Tenm, a Drosophila gene related to tenascin, is a new pair-rule gene. EMBO J *13*, 3728-3740.

Baumgartner, S., and Noll, M. (1990). Network of interactions among pair-rule genes regulating paired expression during primordial segmentation of Drosophila. Mechanisms of Development *33*, 1-18.

Benedyk, M.J., Mullen, J.R., and DiNardo, S. (1994). odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos. Genes Dev *8*, 105-117.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. In Cell (United States), pp. 1266-1276.

Bermingham, J.R., Jr., Martinez-Arias, A., Petitt, M.G., and Scott, M.P. (1990). Different patterns of transcription from the two Antennapedia promoters during Drosophila embryogenesis. Development *109*, 553-566.

Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., and Sattler, M. (2001). The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol *8*, 626-633.

Bouchard, M., St-Amand, J., and Côté, S. (2000). Combinatorial activity of pair-rule proteins on the Drosophila gooseberry early enhancer. Dev Biol *222*, 135-146.

Bowler, T., Kosman, D., Licht, J.D., and Pick, L. (2006). Computational identification of Ftz/Ftz-F1 downstream target genes. Dev Biol *299*, 78-90.

Briscoe, J., and Therond, P.P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. Nat Rev Mol Cell Biol *14*, 416-429.

Buescher, M., Tio, M., Tear, G., Overton, P.M., Brook, W.J., and Chia, W. (2006). Functions of the segment polarity genes midline and H15 in Drosophila melanogaster neurogenesis. Dev Biol *292*, 418-429.
Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development *128*, 2163-2173.
Cadigan, K.M., Grossniklaus, U., and Gehring, W.J. (1994a). Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. Proc Natl Acad Sci U S A *91*, 6324-6328.
Cadigan, K.M., Grossniklaus, U., and Gehring, W.J. (1994b). Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. Genes Dev *8*, 899-913.
Caplen, N.J., Zheng, Z., Falgout, B., and Morgan, R.A. (2002). Inhibition of viral gene expression and replication in mosquito cells by dsRNA-triggered RNA interference. In Mol Ther (United States), pp. 243-251.
Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. In Cell (United States), pp. 25-36.
Carroll, S.B., Gernier, J.K., and Weatherbee, S.D. (2005). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd edn (Oxford: Blackwell Science Ltd.).
Carroll, S.B., Laymon, R.A., McCutcheon, M.A., Riley, P.D., and Scott, M.P. (1986). The localization and regulation of Antennapedia protein expression in Drosophila embryos. Cell *47*, 113-122.
Carroll, S.B., and Scott, M.P. (1986). Zygotically active genes that affect the spatial expression of the fushi tarazu segmentation gene during early Drosophila embryogenesis. Cell *45*, 113-126.
Carroll, S.B., and Vavra, S.H. (1989). The zygotic control of Drosophila pair-rule gene expression. II. Spatial repression by gap and pair-rule gene products. Development *107*, 673-683.
Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. In Cell (United States), pp. 603-615.
Chan, S.K., and Mann, R.S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proc Natl Acad Sci U S A *93*, 5223-5228.
Chou, T.B., and Perrimon, N. (1996). The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. Genetics *144*, 1673-1679.
Clemons, A., Flannery, E., Kast, K., Severson, D., and Duman-Scheel, M. (2010a). Immunohistochemical analysis of protein expression during Aedes aegypti development. Cold Spring Harb Protoc *2010*, pdb prot5510.
Clemons, A., Haugen, M., Flannery, E., Kast, K., Jacowski, C., Severson, D., and Duman-Scheel, M. (2010b). Fixation and preparation of developing tissues from Aedes aegypti. Cold Spring Harb Protoc *2010*, pdb prot5508.
Clemons, A., Haugen, M., Severson, D., and Duman-Scheel, M. (2010c). Functional analysis of genes in Aedes aegypti embryos. Cold Spring Harb Protoc *2010*, pdb prot5511.

Cohn, M.J., and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. Nature *399*, 474-479.
Colas, J.F., Launay, J.M., Kellermann, O., Rosay, P., and Maroteaux, L. (1995). Drosophila 5-HT2 serotonin receptor: coexpression with fushi-tarazu during segmentation. Proc Natl Acad Sci U S A *92*, 5441-5445.
Colas, J.F., Launay, J.M., and Maroteaux, L. (1999a). Maternal and zygotic control of serotonin biosynthesis are both necessary for Drosophila germband extension. Mech Dev *87*, 67-76.
Colas, J.F., Launay, J.M., Vonesch, J.L., Hickel, P., and Maroteaux, L. (1999b). Serotonin synchronises convergent extension of ectoderm with morphogenetic gastrulation movements in Drosophila. Mech Dev *87*, 77-91.
Copeland, J.W., Nasiadka, A., Dietrich, B.H., and Krause, H.M. (1996). Patterning of the Drosophila embryo by a homeodomain-deleted Ftz polypeptide. Nature *379*, 162-165.
Coré, N., Charroux, B., McCormick, A., Vola, C., Fasano, L., Scott, M.P., and Kerridge, S. (1997). Transcriptional regulation of the Drosophila homeotic gene teashirt by the homeodomain protein Fushi tarazu. Mech Dev *68*, 157-172.
Coulter, D.E., Swaykus, E.A., Beran-Koehn, M.A., Goldberg, D., Wieschaus, E., and Schedl, P. (1990). Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. EMBO J *9*, 3795-3804.
Coulter, D.E., and Wieschaus, E. (1988). Gene activities and segmental patterning in Drosophila: analysis of odd-skipped and pair-rule double mutants. Genes Dev *2*, 1812-1823.
Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F.*, et al.* (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell *160*, 191-203.
Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R., and Kadonaga, J.T. (1991). Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science *251*, 643-649.
Dabney, A., Storey, J.D., and Warnes, w.a.f.G.R. (2011). qvalue: Q-value estimation for false discovery rate control.
Dejardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. In EMBO J (England), pp. 857-868.
DiNardo, S., and O'Farrell, P.H. (1987). Establishment and refinement of segmental pattern in the Drosophila embryo: spatial control of engrailed expression by pair-rule genes. Genes Dev *1*, 1212-1225.
DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J.A., and O'Farrell, P.H. (1988). Two-tiered regulation of spatially patterned engrailed gene expression during Drosophila embryogenesis. Nature *332*, 604-609.
Dubrovsky, E.B., Dubrovskaya, V.A., Bernardo, T., Otte, V., DiFilippo, R., and Bryan, H. (2011). The Drosophila FTZ-F1 nuclear receptor mediates juvenile hormone activation of E75A gene expression through an intracellular pathway. In J Biol Chem (United States), pp. 33689-33700.

Enriquez, J., Boukhatmi, H., Dubois, L., Philippakis, A.A., Bulyk, M.L., Michelson, A.M., Crozatier, M., and Vincent, A. (2010). Multi-step control of muscle diversity by Hox proteins in the Drosophila embryo. Development *137*, 457-466.
Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994). The Trithorax-like gene encodes the Drosophila GAGA factor. Nature *371*, 806-808.
Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R., and Bianchi, M.E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. EMBO J *11*, 4497-4506.
Fitzpatrick, V.D., Percival-Smith, A., Ingles, C.J., and Krause, H.M. (1992). Homeodomain-independent activity of the fushi tarazu polypeptide in Drosophila embryos. Nature *356*, 610-612.
Florence, B., Guichet, A., Ephrussi, A., and Laughon, A. (1997). Ftz-F1 is a cofactor in Ftz activation of the Drosophila engrailed gene. Development *124*, 839-847.
Force, A., Cresko, W.A., Pickett, F.B., Proulx, S.R., Amemiya, C., and Lynch, M. (2005). The origin of subfunctions and modular gene regulation. In Genetics (United States), pp. 433-446.
Frasch, M., and Levine, M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. Genes Dev *1*, 981-995.
Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., Goto, T., and Jaynes, J.B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. Development *126*, 2527-2538.
Fujioka, M., Jaynes, J.B., and Goto, T. (1995). Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. Development *121*, 4371-4382.
Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A.A., Weir, M., and Goto, T. (1996). Drosophila Paired regulates late even-skipped expression through a composite binding site for the paired domain and the homeodomain. Development *122*, 2697-2707.
Fujioka, M., Yusibova, G.L., Patel, N.H., Brown, S.J., and Jaynes, J.B. (2002). The repressor activity of Even-skipped is highly conserved, and is sufficient to activate engrailed and to regulate both the spacing and stability of parasegment boundaries. Development *129*, 4411-4421.
Fujioka, M., Yusibova, G.L., Zhou, J., and Jaynes, J.B. (2008). The DNA-binding Polycomb-group protein Pleiohomeotic maintains both active and repressed transcriptional states through a single site. In Development (England), pp. 4131-4139.
Fung, S., Wang, F., Chase, M., Godt, D., and Hartenstein, V. (2008). Expression profile of the cadherin family in the developing Drosophila brain. J Comp Neurol *506*, 469-488.
Galant, R., and Carroll, S.B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. In Nature (England), pp. 910-913.
Garcia-Bellido, A. (1998). The engrailed story. Genetics *148*, 539-544.
Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy---analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Gaziova, I., and Bhat, K.M. (2009). Ancestry-independent fate specification and plasticity in the developmental timing of a typical Drosophila neuronal lineage. Development *136*, 263-274.
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology *5*, R80.
Gergen, J.P., and Butler, B.A. (1988). Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev *2*, 1179-1193.
Gibson, G. (2000). Evolution: hox genes and the cellared wine principle. In Curr Biol (England), pp. R452-455.
Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. In Cell (United States), pp. 185-195.
Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. In Cell (United States), pp. 413-422.
Greenberg, A.J., and Schedl, P. (2001). GAGA factor isoforms have distinct but overlapping functions in vivo. Mol Cell Biol *21*, 8565-8574.
Grenier, J.K., and Carroll, S.B. (2000). Functional evolution of the Ultrabithorax protein. Proc Natl Acad Sci U S A *97*, 704-709.
Gross, C.T., and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. Embo j *15*, 1961-1970.
Grosschedl, R., Giese, K., and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet *10*, 94-100.
Grossniklaus, U., Pearson, R.K., and Gehring, W.J. (1992). The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev *6*, 1030-1051.
Guichet, A., Copeland, J.W., Erdelyi, M., Hlousek, D., Zavorszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H.M., and Ephrussi, A. (1997). The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. Nature *385*, 548-552.
Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol *8*, R24.
Gutjahr, T., Frei, E., and Noll, M. (1993). Complex regulation of early paired expression: initial activation by gap genes and pattern modulation by pair-rule genes. Development *117*, 609-623.
Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984). Spatial distribution of transcripts from the segmentation gene fushi tarazu during Drosophila embryonic development. In Cell (United States), pp. 833-841.
Hagstrom, K., Muller, M., and Schedl, P. (1997). A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. Genetics *146*, 1365-1380.
Halbleib, J.M., and Nelson, W.J. (2006). Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. In Genes Dev (United States), pp. 3199-3214.
Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788-1792.

Han, W., Yu, Y., Altan, N., and Pick, L. (1993). Multiple proteins interact with the fushi tarazu proximal enhancer. Mol Cell Biol *13*, 5549-5559.
Harding, K., Hoey, T., Warrior, R., and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. EMBO J *8*, 1205-1212.
Harding, K., Rushlow, C., Doyle, H.J., Hoey, T., and Levine, M. (1986). Cross-regulatory interactions among pair-rule genes in Drosophila. Science *233*, 953-959.
Hartmann, C., Taubert, H., Jackle, H., and Pankratz, M.J. (1994). A two-step mode of stripe formation in the Drosophila blastoderm requires interactions among primary pair rule genes. In Mech Dev (Ireland), pp. 3-13.
Haugen, M., Tomchaney, M., Kast, K., Flannery, E., Clemons, A., Jacowski, C., Simanton Holland, W., Le, C., Severson, D., and Duman-Scheel, M. (2010). Whole-mount in situ hybridization for analysis of gene expression during Aedes aegypti development. Cold Spring Harb Protoc *2010*, pdb prot5509.
Heemskerk, J., DiNardo, S., Kostriken, R., and O'Farrell, P.H. (1991). Multiple modes of engrailed regulation in the progression towards cell fate determination. Nature *352*, 404-410.
Heffer, A., Shultz, J.W., and Pick, L. (2010). Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution. In Proc Natl Acad Sci U S A (United States), pp. 18040-18045.
Hill, E., Broadbent, I.D., Chothia, C., and Pettitt, J. (2001). Cadherin superfamily proteins in Caenorhabditis elegans and Drosophila melanogaster1. Journal of Molecular Biology *305*, 1011-1024.
Hiromi, Y., and Gehring, W.J. (1987). Regulation and function of the Drosophila segmentation gene fushi tarazu. In Cell (United States), pp. 963-974.
Hiromi, Y., Kuroiwa, A., and Gehring, W.J. (1985). Control elements of the Drosophila segmentation gene fushi tarazu. In Cell (United States), pp. 603-613.
Holmgren, R. (1984). Cloning sequences from the hairy gene of Drosophila. EMBO J *3*, 569-573.
Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. Mol Cell Biol *20*, 3187-3197.
Hou, H.Y., Heffer, A., Anderson, W.R., Liu, J., Bowler, T., and Pick, L. (2009). Stripy Ftz target genes are coordinately regulated by Ftz-F1. Dev Biol *335*, 442-453.
Howard, K., and Ingham, P. (1986). Regulatory interactions between the segmentation genes fushi tarazu, hairy, and engrailed in the Drosophila blastoderm. Cell *44*, 949-957.
Hughes, C.L., and Kaufman, T.C. (2002). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. Development *129*, 1225-1238.
Hughes, S.C., and Krause, H.M. (2001). Establishment and maintenance of parasegmental compartments. Development *128*, 1109-1118.
Hur, M.W., Laney, J.D., Jeon, S.H., Ali, J., and Biggin, M.D. (2002). Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. Development *129*, 1339-1343.
Hynes, R.O., and Zhao, Q. (2000). The Evolution of Cell Adhesion. The Journal of Cell Biology *150*, 89-96.

Ingham, P., and Gergen, P. (1988). Interactions between the pair-rule genes runt, hairy, even-skipped and fushi tarazu and the establishment of periodic pattern in the Drosophila embryo. Development *104*, 51-60.
Ingham, P.W., Baker, N.E., and Martinez-Arias, A. (1988). Regulation of segment polarity genes in the Drosophila blastoderm by fushi tarazu and even skipped. Nature *331*, 73-75.
Ingham, P.W., Howard, K.R., and Ish-Horowicz, D. (1985). Transcription pattern of the Drosophila segmentation gene hairy. Nature *318*, 439-445.
Ish-Horowicz, D., Pinchin, S.M., Ingham, P.W., and Gyurkovics, H.G. (1989). Autocatalytic ftz activation and metameric instability induced by ectopic ftz expression. In Cell (United States), pp. 223-232.
Janssen, R., and Damen, W.G. (2006). The ten Hox genes of the millipede Glomeris marginata. Dev Genes Evol *216*, 451-465.
Jaynes, J.B., and Fujioka, M. (2004). Drawing lines in the sand: even skipped et al. and parasegment boundaries. In Dev Biol (United States), pp. 609-622.
Jeffery, D.E. (1979). Position Effects at the Hairy Locus in DROSOPHILA MELANOGASTER. Genetics *91*, 105-125.
Jimenez, G., Pinchin, S.M., and Ish-Horowicz, D. (1996). In vivo interactions of the Drosophila Hairy and Runt transcriptional repressors with target promoters. EMBO J *15*, 7088-7098.
Johnson, F.B., Parker, E., and Krasnow, M.A. (1995). Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proc Natl Acad Sci U S A *92*, 739-743.
Johnson, K., Knust, E., and Skaer, H. (1999). bloated tubules (blot) encodes a Drosophila member of the neurotransmitter transporter family required for organisation of the apical cytocortex. Dev Biol *212*, 440-454.
In.
Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. Wilhelm Roux's archives of developmental biology *193*, 283-295.
Kal, A.J., Mahmoudi, T., Zak, N.B., and Verrijzer, C.P. (2000). The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. Genes Dev *14*, 1058-1071.
Kellerman, K.A., Mattson, D.M., and Duncan, I. (1990). Mutations affecting the stability of the fushi tarazu protein of Drosophila. Genes Dev *4*, 1936-1950.
Kilchherr, F., Baumgartener, S., Bopp, D., Frei, E., and Noll, M. (1986). Isolation of the *paired* gene of *Drosophila* and its spatial expression during early embryogenesis. Nature *321*, 493-499.
Klinger, M., and Gergen, J.P. (1993). Regulation of runt transcription by Drosophila segmentation genes. Mechanisms of Development *43*, 3-19.
Kobayashi, M., Goldstein, R.E., Fujioka, M., Paroush, Z., and Jaynes, J.B. (2001). Groucho augments the repression of multiple Even skipped target genes in establishing parasegment boundaries. Development *128*, 1805-1815.
Kokoza, V., Ahmed, A., Wimmer, E.A., and Raikhel, A.S. (2001). Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac

transposable element vector pBac[3xP3-EGFP afm]. In Insect Biochem Mol Biol (England), pp. 1137-1143.
Krause, C., Wolf, C., Hemphala, J., Samakovlis, C., and Schuh, R. (2006). Distinct functions of the leucine-rich repeat transmembrane proteins capricious and tartan in the Drosophila tracheal morphogenesis. Dev Biol *296*, 253-264.
Kuroiwa, A., Hafen, E., and Gehring, W.J. (1984). Cloning and transcriptional analysis of the segmentation gene fushi tarazu of Drosophila. Cell *37*, 825-831.
Landsman, D., and Bustin, M. (1993). A signature for the HMG-1 box DNA-binding proteins. Bioessays *15*, 539-546.
Langeland, J.A., and Carroll, S.B. (1993). Conservation of regulatory elements controlling hairy pair-rule stripe formation. Development *117*, 585-596.
Lawrence, P.A., Johnston, P., Macdonald, P., and Struhl, G. (1987). Borders of parasegments in Drosophila embryos are delimited by the fushi tarazu and even-skipped genes. Nature *328*, 440-442.
Leal, S.M., Qian, L., Lacin, H., Bodmer, R., and Skeath, J.B. (2009). Neuromancer1 and Neuromancer2 regulate cell fate specification in the developing embryonic CNS of Drosophila melanogaster. Dev Biol *325*, 138-150.
Lehmann, M. (2004). Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. In Trends Genet (England), pp. 15-22.
Levine, A., Bashan-Ahrend, A., Budai-Hadrian, O., Gartenberg, D., Menasherow, S., and Wides, R. (1994). odd Oz: A novel Drosophila pair rule gene. Cell *77*, 587-598.
Lewis, R.A., Wakimoto, B.T., Denell, R.E., and Kaufman, T.C. (1980). Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of DROSOPHILA MELANOGASTER. II. Polytene Chromosome Segments 84A-84B1,2. Genetics *95*, 383-397.
Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. In Nature (England), pp. 400-403.
Liu, Q.X., Hiramoto, M., Ueda, H., Gojobori, T., Hiromi, Y., and Hirose, S. (2009). Midline governs axon pathfinding by coordinating expression of two major guidance systems. Genes Dev *23*, 1165-1170.
Lohr, U., and Pick, L. (2005). Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of Drosophila melanogaster. In Curr Biol (England), pp. 643-649.
Lohr, U., Yussa, M., and Pick, L. (2001). Drosophila fushi tarazu. a gene on the border of homeotic function. In Curr Biol (England), pp. 1403-1412.
Lu, Q., Knoepfler, P.S., Scheele, J., Wright, D.D., and Kamps, M.P. (1995). Both Pbx1 and E2A-Pbx1 bind the DNA motif ATCAATCAA cooperatively with the products of multiple murine Hox genes, some of which are themselves oncogenes. Mol Cell Biol *15*, 3786-3795.
Lukacsovich, T., Yuge, K., Awano, W., Asztalos, Z., Kondo, S., Juni, N., and Yamamoto, D. (2003). The ken and barbie gene encoding a putative transcription factor with a BTB domain and three zinc finger motifs functions in terminalia development of Drosophila. Arch Insect Biochem Physiol *54*, 77-94.
MacArthur, S., Li, X.-Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keränen, S.V.E.*, et al.* (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*, R80.
Macdonald, P.M., Ingham, P., and Struhl, G. (1986). Isolation, structure, and expression of even-skipped: a second pair-rule gene of Drosophila containing a homeo box. In Cell (United States), pp. 721-734.
Mahmoudi, T., Zuijderduijn, L.M., Mohd-Sarip, A., and Verrijzer, C.P. (2003). GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. Nucleic Acids Res *31*, 4147-4156.
Mandal, L., Martinez-Agosto, J.A., Evans, C.J., Hartenstein, V., and Banerjee, U. (2007). A Hedgehog- and Antennapedia-dependent niche maintains Drosophila haematopoietic precursors. Nature *446*, 320-324.
Mann, R.S., and Chan, S.K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. In Trends Genet (England), pp. 258-262.
Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. In Curr Top Dev Biol (United States), pp. 63-101.
Manoukian, A.S., and Krause, H.M. (1992). Concentration-dependent activities of the even-skipped protein in Drosophila embryos. Genes Dev *6*, 1740-1751.
Manoukian, A.S., and Krause, H.M. (1993). Control of segmental asymmetry in Drosophila embryos. Development *118*, 785-796.
Mao, Y., Kerr, M., and Freeman, M. (2008). Modulation of Drosophila retinal epithelial integrity by the adhesion proteins capricious and tartan. PLoS One *3*, e1827.
Martinez-Arias, A., and Lawrence, P.A. (1985). Parasegments and compartments in the Drosophila embryo. Nature *313*, 639-642.
Martizez Arias, A., Baker, N.E., and Ingham, P.W. (1988). Role of segment polarity genes in the definition and maintenance of cell states in the Drosophila embryo. Development *103*, 157-170.
Maybeck, V., and Roper, K. (2009). A targeted gain-of-function screen identifies genes affecting salivary gland morphogenesis/tubulogenesis in Drosophila. Genetics *181*, 543-565.
McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283-302.
Milan, M., Weihe, U., Perez, L., and Cohen, S.M. (2001). The LRR proteins capricious and Tartan mediate cell interactions during DV boundary formation in the Drosophila wing. Cell *106*, 785-794.
Mishra, R.K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S.E., and Schedl, P. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. Mol Cell Biol *21*, 1311-1318.
Miskolczi-McCallum, C.M., Scavetta, R.J., Svendsen, P.C., Soanes, K.H., and Brook, W.J. (2005). The Drosophila melanogaster T-box genes midline and H15 are conserved regulators of heart development. Dev Biol *278*, 459-472.
Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. In Dev Biol (United States), pp. 193-206.

Mulholland, N.M., King, I.F., and Kingston, R.E. (2003). Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. In Genes Dev (United States), pp. 2741-2746.
Mullen, J.R., and DiNardo, S. (1995). Establishing parasegments in Drosophila embryos: roles of the odd-skipped and naked genes. In Dev Biol (United States), pp. 295-308.
Müller, J., and Bienz, M. (1992). Sharp anterior boundary of homeotic gene expression conferred by the fushi tarazu protein. EMBO J *11*, 3653-3661.
Nambu, P.A., and Nambu, J.R. (1996). The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development *122*, 3467-3475.
Nasiadka, A., and Krause, H.M. (1999). Kinetic analysis of segmentation gene interactions in Drosophila embryos. Development *126*, 1515-1526.
Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., Xi, Z., Megy, K., Grabherr, M.*, et al.* (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. In Science (United States), pp. 1718-1723.
Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. In Cell (United States), pp. 1277-1289.
Nusslein-Volhard, C., Kluding, H., and Jurgens, G. (1985). Genes affecting the segmental subdivision of the Drosophila embryo. Cold Spring Harb Symp Quant Biol *50*, 145-154.
Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.
Ohno, S. (1970). Evolution by Gene Duplication (Berlin: Springer-Verlag).
Papillon, D., and Telford, M.J. (2007). Evolution of Hox3 and ftz in arthropods: insights from the crustacean Daphnia pulex. Dev Genes Evol *217*, 315-322.
Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., and Aggarwal, A.K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714-719.
Patel, N.H., Martin-Blanco, E., Coleman, K.G., Poole, S.J., Ellis, M.C., Kornberg, T.B., and Goodman, C.S. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. In Cell (United States), pp. 955-968.
Pedone, P.V., Ghirlando, R., Clore, G.M., Gronenborn, A.M., Felsenfeld, G., and Omichinski, J.G. (1996). The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. Proc Natl Acad Sci U S A *93*, 2822-2826.
Peel, A. (2004). The evolution of arthropod segmentation mechanisms. Bioessays *26*, 1108-1116.
Perrimon, N., Lanjuin, A., Arnold, C., and Noll, E. (1996). Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by P-element-induced mutations. Genetics *144*, 1681-1692.
Phelan, M.L., Rambaldi, I., and Featherstone, M.S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Mol Cell Biol *15*, 3989-3997.

Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., and Gehring, W.J. (1990a). Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. Genes Dev *4*, 1224-1239.
Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., and Gehring, W.J. (1990b). Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. Genes Dev *4*, 1224-1239.
Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. In Cell (United States), pp. 587-597.
Pirrotta, V., Chan, C.S., McCabe, D., and Qian, S. (1995). Distinct parasegmental and imaginal enhancers and the establishment of the expression pattern of the Ubx gene. Genetics *141*, 1439-1450.
Prokopenko, S.N., He, Y., Lu, Y., and Bellen, H.J. (2000). Mutations affecting the development of the peripheral nervous system in Drosophila: a molecular screen for novel proteins. Genetics *156*, 1691-1715.
Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. In Proc Natl Acad Sci U S A (United States), pp. 8605-8612.
Qian, L., Liu, J., and Bodmer, R. (2005). Neuromancer Tbx20-related genes (H15/midline) promote cell fate specification and morphogenesis of the Drosophila heart. Dev Biol *279*, 509-524.
Qian, L., Mohapatra, B., Akasaka, T., Liu, J., Ocorr, K., Towbin, J.A., and Bodmer, R. (2008). Transcription factor neuromancer/TBX20 is required for cardiac function in Drosophila with implications for human heart disease. Proc Natl Acad Sci U S A *105*, 19833-19838.
Qian, S., Capovilla, M., and Pirrotta, V. (1993). Molecular mechanisms of pattern formation by the BRE enhancer of the Ubx gene. EMBO J *12*, 3865-3877.
Reim, I., Mohler, J.P., and Frasch, M. (2005). Tbx20-related genes, mid and H15, are required for tinman expression, proper patterning, and normal differentiation of cardioblasts in Drosophila. Mech Dev *122*, 1056-1069.
Rezende, G.L., Martins, A.J., Gentile, C., Farnesi, L.C., Pelajo-Machado, M., Peixoto, A.A., and Valle, D. (2008). Embryonic desiccation resistance in Aedes aegypti: presumptive role of the chitinized serosal cuticle. BMC Dev Biol *8*, 82.
Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. In Nature (England), pp. 914-917.
Russell, S.R., Sanchez-Soriano, N., Wright, C.R., and Ashburner, M. (1996). The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. Development *122*, 3669-3676.
Sander, K., Lohs-Schardin, M., and Baumann, M. (1980). Embryogenesis in a Drosophila mutant expressing half the normal segment number. Nature *287*, 841-843.
Saulier-Le Drean, B., Nasiadka, A., Dong, J., and Krause, H.M. (1998). Dynamic changes in the functions of Odd-skipped during early Drosophila embryogenesis. Development *125*, 4851-4861.
Schaerlinger, B., Launay, J.M., Vonesch, J.L., and Maroteaux, L. (2007). Gain of affinity point mutation in the serotonin receptor gene 5-HT2Dro accelerates germband extension movements during Drosophila gastrulation. Dev Dyn *236*, 991-999.

Schier, A.F., and Gehring, W.J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. Nature *356*, 804-807.
Schier, A.F., and Gehring, W.J. (1993). Functional specificity of the homeodomain protein fushi tarazu: the role of DNA-binding specificity in vivo. Proc Natl Acad Sci U S A *90*, 1450-1454.
Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816-818.
Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. In PLoS Biol (United States), p. e13.
Schwartz, C.J., Sampson, H.M., Hlousek, D., Percival-Smith, A., Copeland, J.W., Simmonds, A.J., and Krause, H.M. (2001). FTZ-Factor1 and Fushi tarazu interact via conserved nuclear receptor and coactivator motifs. EMBO J *20*, 510-519.
Scott, M.P., Weiner, A.J., Hazelrigg, T.I., Polisky, B.A., Pirrotta, V., Scalenghe, F., and Kaufman, T.C. (1983). The molecular organization of the Antennapedia locus of Drosophila. In Cell (United States), pp. 763-776.
Shinza-Kameda, M., Takasu, E., Sakurai, K., Hayashi, S., and Nose, A. (2006). Regulation of layer-specific targeting by reciprocal expression of a cell adhesion molecule, capricious. Neuron *49*, 205-213.
Simon, J.A., and Tamkun, J.W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. In Curr Opin Genet Dev (England), pp. 210-218.
Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J *11*, 4047-4057.
Small, S., Blair, A., and Levine, M. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. In Dev Biol (United States), pp. 314-324.
Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? In Evolution (United States), pp. 2155-2177.
Struhl, G. (1985). Near-reciprocal phenotypes caused by inactivation or indiscriminate expression of the Drosophila segmentation gene ftz. Nature *318*, 677-680.
Swantek, D., and Gergen, J.P. (2004). Ftz modulates Runt-dependent activation and repression of segment-polarity gene transcription. In Development (England), pp. 2281-2290.
Tearle, R.G., and Nusslein-Volhard, C. (1987). Tubingen mutants and stock list, pp. 209-269.
Telford, M.J. (2000). Evidence for the derivation of the Drosophila fushi tarazu gene from a Hox gene orthologous to lophotrochozoan Lox5. In Curr Biol (England), pp. 349-352.
Tepass, U., Truong, K., Godt, D., Ikura, M., and Peifer, M. (2000). Cadherins in embryonic and neural morphogenesis. Nat Rev Mol Cell Biol *1*, 91-100.
Tracey, W.D., Jr., Ning, X., Klingler, M., Kramer, S.G., and Gergen, J.P. (2000). Quantitative analysis of gene function in the Drosophila embryo. Genetics *154*, 273-284.

Tsai, C., and Gergen, J.P. (1994). Gap gene properties of the pair-rule gene runt during Drosophila segmentation. Development *120*, 1671-1683.
Tsuji, T., Hasegawa, E., and Isshiki, T. (2008). Neuroblast entry into quiescence is regulated intrinsically by the combined action of spatial Hox proteins and temporal identity factors. Development *135*, 3859-3869.
Ueda, H., Sonoda, S., Brown, J.L., Scott, M.P., and Wu, C. (1990). A sequence-specific DNA-binding protein that activates fushi tarazu segmentation gene expression. Genes Dev *4*, 624-635.
Ueda, H., Sun, G.C., Murata, T., and Hirose, S. (1992). A novel DNA-binding motif abuts the zinc finger domain of insect nuclear hormone receptor FTZ-F1 and mouse embryonal long terminal repeat-binding protein. Mol Cell Biol *12*, 5667-5672.
van Steensel, B., Delrow, J., and Bussemaker, H.J. (2003). Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. In Proc Natl Acad Sci U S A (United States), pp. 2580-2585.
Vavra, S.H., and Carroll, S.B. (1989). The zygotic control of Drosophila pair-rule gene expression. I. A search for new pair-rule regulatory loci. Development *107*, 663-672.
Veraksa, A., Kennison, J., and McGinnis, W. (2002). DEAF-1 function is essential for the early embryonic development of Drosophila. Genesis *33*, 67-76.
Vincent, J.P., and O'Farrell, P.H. (1992). The state of engrailed expression is not clonally transmitted during early Drosophila development. Cell *68*, 923-931.
Wakimoto, B.T., and Kaufman, T.C. (1981). Analysis of larval segmentation in lethal genotypes associated with the antennapedia gene complex in Drosophila melanogaster. In Dev Biol (United States), pp. 51-64.
Wakimoto, B.T., Turner, F.R., and Kaufman, T.C. (1984). Defects in embryogenesis in mutants associated with the antennapedia gene complex of Drosophila melanogaster. In Dev Biol (United States), pp. 147-172.
Walrad, P.B., Hang, S., and Gergen, J.P. (2011). Hairless is a cofactor for Runt-dependent transcriptional regulation. In Mol Biol Cell (United States), pp. 1364-1374.
Walrad, P.B., Hang, S., Joseph, G.S., Salas, J., and Gergen, J.P. (2010). Distinct contributions of conserved modules to Runt transcription factor activity. In Mol Biol Cell (United States), pp. 2315-2326.
Wheeler, J.C., VanderZwan, C., Xu, X., Swantek, D., Tracey, W.D., and Gergen, J.P. (2002). Distinct in vivo requirements for establishment versus maintenance of transcriptional repression. In Nat Genet (United States), pp. 206-210.
Wu, H., Yang, m.b.H., with ideas from Gary Churchill, K.S., Kerr, K., and Cui., X. (2011). maanova: Tools for analyzing Micro Array experiments.
Yeates, D.K., and Wiegmann, B.M. (1999). Congruence and controversy: toward a higher-level phylogeny of Diptera. Annu Rev Entomol *44*, 397-428.
Yoder, J.H., and Carroll, S.B. (2006). The evolution of abdominal reduction and the recent origin of distinct Abdominal-B transcript classes in Diptera. In Evol Dev (United States), pp. 241-251.
Yu, Y., Li, W., Su, K., Yussa, M., Han, W., Perrimon, N., and Pick, L. (1997). The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. Nature *385*, 552-555.
Yu, Y., and Pick, L. (1995). Non-periodic cues generate seven ftz stripes in the Drosophila embryo. Mech Dev *50*, 163-175.

Yussa, M., Lohr, U., Su, K., and Pick, L. (2001). The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains. In Mech Dev (Ireland), pp. 39-53.
Zartman, J.J., Kanodia, J.S., Yakoby, N., Schafer, X., Watson, C., Schlichting, K., Dahmann, C., and Shvartsman, S.Y. (2009). Expression patterns of cadherin genes in Drosophila oogenesis. In Gene Expr Patterns (Netherlands), pp. 31-36.
Zhao, J.J., Lazzarini, R.A., and Pick, L. (1996). Functional dissection of the mouse Hox-a5 gene. EMBO J *15*, 1313-1322.