Title
Non-coding RNAs of the Bithorax Complex in the Developing Drosophila Embryo

Author
Piel, Jessica Christine

Publication Date
2010

Peer reviewed|Thesis/dissertation

Non-coding RNAs of the Bithorax Complex in the Developing *Drosophila* Embryo

by

Jessica Christine Piel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Levine, Chair
Professor Iswar Hariharan
Professor Qiang Zhou
Professor Sydney Kustu

Spring 2010

Non-coding RNAs of the Bithorax Complex in the Developing

Drosophila Embryo

by Jessica Christine Piel

Abstract

Non-coding RNAs of the Bithorax Complex in the Developing *Drosophila* Embryo

by

Jessica Christine Piel

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Michael Levine, Chair

The mechanisms for the precise regulatory control of genes have long been a question in the field of molecular biology. While traditional views of gene regulation focus on the *cis*-regulatory elements and transcription factors involved, a new field of study focuses on the *trans*-regulatory functions of non-coding RNAs. In recent years, both small and long non-coding RNAs have been recognized as important elements in the regulation of many different cellular processes, however, the identification and characterization of long non-coding RNAs is still in its infancy. The Bithorax Hox Gene Complex of *Drosophila melanogaster* has long been known to contain a multitude of long non-coding RNAs of unknown function. While many of the *cis*-regulatory mechanisms of this historical complex have since been worked out, many of the non-coding RNAs remain mysterious. With the advent of new techniques and a fresh RNA-centric viewpoint, several researchers have recently returned to the Bithorax Complex to study the vast non-coding transcription that pervades its inter- and intragenic regions. These studies reveal that these non-coding RNAs act in *trans* as important regulators of the protein-coding Hox gene, *Ultrabithorax*.

This thesis is dedicated to my parents, without whose love and support I never would have made it this far.

Table of Contents

Acknowledgements

Many people contributed to the successful completion of my thesis, through both academic and emotional support and deserve my sincere gratitude. First of all, I would like to thank my thesis advisor Mike Levine for inviting me into his lab and allowing me to pursue my own research interests even as they veered away from the lab's main focus. While we did not always see eye-to-eye, my experience in his lab has made me a more confident, independent and stronger scientist than I would have become anywhere else.

I am indebted to many members of the Levine Lab, who over the years have functioned as great resources for both knowledge and support without whom I never would have made it and who deserve special recognition. Matt Ronshaugen and Fred Biemar were great sources of guidance early on and taught me not only many new biological techniques but also new modes of thinking, confidence and Levine Lab coping mechanisms. Matt especially was a constantly positive force in a generally negative environment and definitely helped buoy the spirits of all the woeful grad students. Additionally, Rob Zinzen was a wealth of knowledge both in techniques and stories of Levine lab members past. He made the late nights in the fly and injection rooms much easier with great conversations that I will never forget. Jessica Cande functioned as the mother hen of the lab and was a great repository of information both for research methods and menial questions such as "Where do we keep the screwdrivers?". Despite her complaints she was (almost) always willing to help solve whatever kind of crisis any lab member was currently facing. Alex "Wings" Mori was the lab cheerleader, always ready with a smile and an upbeat pep talk to keep your spirits up. Despite his relatively short stay he has made an indelible imprint on the Levine Lab, mostly through his nonstop caffeinated chatter and ability to get his "chicken wings flapping" about practically any subject. Our frequent coffee and rice krispie treat breaks were mini support sessions that I sorely needed. Mike "Periwinkle" Perry was always a great source for methods and techniques and an outside non-fly insect perspective: "Well, in the Patel lab we…" I always wondered what watching a chatterbox competition between Perry and Wings would be like, and who would win.

Ben Haley and Vivek Chopra were essential components of my thesis project progression. Both offered invaluable knowledge of techniques and history of their fields in their individual specialties. They were generous with materials, experimental assistance and guidance in my research planning, and it is likely that I would not have finished my project without their help. Ben additionally kept us all entertained with his seemingly endless collection of humorous and disturbing internet treasures.

Kevin Tsui and Jacques Bothma arrived at a critical time during my grad years in which I greatly benefitted from their sense of humor and willingness to be distracted by me when I needed a mental break. Additionally, their tolerance for my constant and fresh supply of kitty stories was much appreciated. Eileen "Princess" Wagner has been a great source of emotional support over the years and a much needed co-commiserator. Her command of sarcasm, sharp tongue and colorful language were an essential part of the Levine Lab experience, and our friendship has helped prevent my descent into insanity over the past few years. Finally, Jeni Beh has been an undeniably important aspect of my grad school years, both in and out of the lab. Since the first day of her arrival, we recognized each other as familiars in an unfamiliar place. We bonded over our slow and gradual command of Levine Lab science and our predilection and necessity to discuss non-science related topics throughout the day. Jeni Beh was a great grad school companion and has become an even greater friend. Thank goodness we have each other

since I couldn't imagine what the years of grad school would have been like without her company, support, and understanding. Additionally, I would like to thank the other members of the Levine Lab who have been important sources of information, suggestions and humor over the years, especially Dave Hendrix and Lionel Christiaen.

People outside the Levine Lab were also important supporters of my graduate work. My Cornell friends Hania Kronfol and Katie McMillan punctuated some of the dreary days with inside jokes, stories and memories of our fond (yet stressful) years in icy Ithaca. No matter how many miles (or continents) or months apart, talking with them makes it seem like we had just seen each other yesterday. Indeed, no matter how down I was, Hania would constantly one-up me with descriptions of "crazy, hectic and depressing" recent events in her life to make me feel better. I would also like to thank my friend Jen Burgess who was always there for me one way or another. She would listen to me complain even though she had most likely spent the night before as a medical student without sleep sewing patients back together.

Kate Chase has been an invaluable necessity to me ever since I first arrived as a Levine Lab member. She was always there to counsel, console, advise and listen to me and has helped me tackle a multitude of problems ranging from filling out forms to life-changing issues. She has been not only a great asset but also a great friend.

My family has been a source of unending support and encouragement. My brother Bill kept alive his college tradition of coming out to visit every year, and took my car on more crazy adventures than I have ever been. His enjoyment of the Berkeley atmosphere and the California weather made me stop and appreciate the simple joys that surround me every day. Often having to rip me out of my apartment, Bill would force me to go out and have fun, giving me much needed perspective on the immense lab-related stress with which my life is often consumed. My parents have been behind me every step of the way, supporting me in all aspects of my life. They have always been my number one fans, and I am so grateful knowing that I could always count on them.

Jesse Patterson is the main reason that I have been able to successfully complete my thesis and earn my degree. Without his understanding, enthusiasm for science of all types, and kindness I would not have endured the past few years. He has been a constant pillar of support and encouragement for all of my endeavors and has always been there for me whenever I needed him, even flipping fly vials and collecting virgins when I could not. I simply could not imagine my life without him.

Finally, I would like to acknowledge my kitty Emma for just being her wonderful self. She was my constant companion during my thesis writing, staying up late with me until I finished what I needed to do. She was always willing to join me for a nap or outside time whenever I needed a break and her contented purring while sleeping beside me has been a consistent source of comfort to me at the end of the day.

Chapter I:

Introduction

The idea of the RNA World hypothesizes that the early world was once dominated by RNAs, both active and informative, that were the main source of organic and evolutionary processes on Earth (Walter 1986). This worldview was sparked by the discovery of self-catalyzing RNA molecules, called ribozymes, that were present in *E.coli*. The knowledge that RNA could function not only as a strand of data or a structural motif but as an enzymatic agent shaped the notion that RNA was once the original and only molecular component of life. Prior to these revelations, however, the early days of molecular biology and the genesis of the Central Dogma relegated RNA to the role of messenger, a mere conduit between DNA, the blueprint of life, and proteins, the building blocks. Despite the known existence of non-messenger RNAs, such as tRNAs and rRNAs, the idea that RNA could be actively and independently useful beyond the occasional structural "glue" of RNA-Protein complexes was not popular. The standard view was that DNA encoded proteins, which were the main and active components of life, and the bigger genome, and thus the more proteins, the more complex the organism.

The completion of the human genome sequence turned this notion on its head. Excited to quantify the superiority of the human race through sheer numbers of coding genes, researchers were quite shocked to find that only ~2% of the genome coded for proteins (2004; Frith et al. 2005). These ~20,000 protein-coding genes put humans on about equal footing with the nematode worm, *Caenorhabitis elegans* in terms of protein output (Frith et al. 2005). Scrambling for explanations, several theories became popular that suggested the superior complexity of the human form was determined not by its proteins but by the intricate regulation of them by the complex *cis*-regulatory elements that surely composed a majority of the remaining 98% of our genome (Levine and Tjian 2003). While this is a valid concept, further analyses indicated that a vast extent (at least 60-70%) of the human (and other species) genome is transcribed (Carninci et al. 2005; Frith et al. 2005). Some were quick to dismiss this phenomenon as transcriptional "noise" or irrelevant debris of evolution and transposons (Wang et al. 2004; Schwabish and Struhl 2007; Ebisuya et al. 2008). However, many believed that these mysterious transcripts held functions that were yet unknown (Mattick 2001).

Indeed, further evidence showed that a majority of these non-coding transcripts are regulated in similar ways to normal protein-coding genes. For instance, many studies have shown non-coding RNAs that exhibit developmentally regulated expression patterns (Blackshaw et al. 2004; Dinger et al. 2008a), cell-type specific expression (Ravasi et al. 2006; Dinger et al. 2008b), and are differentially spliced (Willingham et al. 2005). This suggests that these transcripts are purposely transcribed and are not mere products of wayward transcription machinery. Additionally, studies into genome complexity show a better conservation between the amount of non-coding transgenic and intronic sequences and organism complexity than the number of protein coding sequences (Mattick 2001; Frith et al. 2005). If the model that higher complexity of an organism is dictated not by the number of protein coding genes but by the more complex and higher-order regulation of them, then these data suggest that this mysterious genome transcription is actually involved in unique forms of regulatory mechanisms.

Relative to eukaryotes, prokaryotes rely much more heavily on increasing their complexity through regulatory proteins. In fact, some studies have shown that prokaryotic regulatory proteins scale almost quadratically with genome size, but do not show the same trend

of increased non-coding transcription as well (Croft 2003; van Nimwegen 2003; Mattick and Makunin 2006). This indicates that eukaryotes have increased their diversity capacity by relying on the adaptability and nimbleness of RNA for increased regulatory complexity. The fact that eukaryotes also contain a much more sophisticated and developed set of systems for RNA processing, binding, and signaling (such as RNAi) supports this idea (Mattick and Gagen 2001; Mattick 2003; Mattick and Makunin 2005). The discoveries of functions for both small and long non-coding RNAs over the past several years have proven that RNA is far more involved in the regulatory process then ever imagined.

Small Non-coding RNAs in Regulation

Non-coding RNAs receive the arbitrary distinction of being "small" if they are below ~200 bps in length. There are numerous types of small RNAs, including snoRNAs, miRNAs, piRNAs, snRNAs, etc. In the past several years, this class of non-coding RNAs has gone from relative obscurity to become a major field of research of gene regulation. This is true most of all for microRNAs, or miRNAs.

miRNAs and siRNAs are non-coding RNAs of approximately 22 nucleotides that are derived from either a hairpin or double-stranded RNA precursor transcript. They are both translational silencers that work by binding to transcripts through complementary base pairing and either interfering with the translational process or degrading the transcript through interaction with the RISC complex. miRNAs are often found in introns and exons of genes that are transcribed by RNA Polymerase II (reviewed in (Mattick and Makunin 2005). The protein Drosha recognizes the hairpin and cleaves the pre-miRNA from the primary transcript. The hairpin is then exported to the cytoplasm where Dicer cleaves it into a ~21 nt miRNA duplex at which point it is free to target other RNA transcripts for repression.

miRNAs were first discovered by the Victor Ambros group in 1993, when they found that a gene called *lin-4*, which was known to regulate *C. elegans* larval development by repressing another gene called *lin-14*, did not code for protein. They were stunned to find that the seemingly functional product of this gene was a tiny RNA only ~22 nt long. Further study showed that this RNA had complementary sequence to a region within the 3' UTR of the *lin-14* gene, which had already been shown to be important for its repression. This suggested that *lin-4* mediated its repression of *lin-14* through binding of this transcript sequence (Wightman et al. 1991; Lee et al. 1993; Wightman et al. 1993). Initially thought of as a *C. elegans* peculiarity, miRNAs began to be discovered in many different species, from *Drosophila* to mammals, and were shown to be involved in all forms of cellular functions requiring gene regulation, including developmental timing, cell proliferation, and apoptosis (Bartel 2005; Berezikov and Plasterk 2005; Zamore and Haley 2005; Naguibneva et al. 2006). In fact, miRNAs are such good indicators of the current regulatory environment of the cell that miRNA profiling has become a very accurate diagnostic tool for cancer typing (Calin et al. 2005; Lu et al. 2005).

miRNAs have become a formidable aspect of cellular gene regulation research as both a field of study and as a tool to artificially degrade transcripts of interest (Haley et al. 2008). Additionally, studies have also shown them to be a key target of evolution. Due to their low complexity and non-catalytic nature, miRNA target sequences can be modified with a single basepair substitution, allowing them to be a dynamic and rapid target of evolutionary forces (Dai and Chen; Mattick and Makunin 2006). This suggests that the prevalence and manipulation of miRNAs in the eukaryotic genome allows for an additional level of organism complexity due to

their significant regulatory abilities of coding and non-coding genes alike. In fact, genome studies of organism complexity corroborate this idea by showing that genomes of higher organisms correlate with longer UTRs, key targets of miRNA action (Frith et al. 2005).

Additional classes of functional small RNAs have received much attention for their regulatory roles in human disease, such as the snoRNAs. snoRNAs are a type of small non-coding RNA that have significant functions within the cell's nucleolus during tRNA and rRNA processing. Before leaving for the cytoplasm, some tRNAs and rRNAs need additional chemical modifications, such as 2'O-ribose-methylations and pseudouridylations. snoRNAs facilitate these modifications by associating with other modification proteins to form a complex in which the RNAs act as a targeting guide to the appropriate location for modification of the tRNA or rRNA (Maden and Hughes 1997). The full effect of these modifications has not yet been established, however, they seem to subtly enhance the folding of the RNA and its subsequent interactions with proteins.

Recently, additional functions, including a role in alternative splicing, have been shown for various snoRNAs (Kishore and Stamm 2006). The fact that the snoRNAs exhibit conservation between distant eukaryotes suggests that they perform an important cellular function (Bachellerie et al. 2002). Additionally, many of the snoRNA clusters are tissue-specific and developmentally regulated (Cavaille et al. 2000; Cavaille et al. 2001; Rogelj and Giese 2004). Some, such as the snoRNA *5-HTBII-52*, are even epigenetically imprinted. This RNA regulates the splicing pattern of the transcripts for the 5-HT(2C)R serotonin receptor gene. Deletion of the region encapsulating this snoRNA cluster on the paternal chromosome (the maternal locus is epigenetically silenced) leads to improper splicing of the 5-HT(2C)R transcripts and contributes to development of Prader-Willi Syndrome (Sahoo et al. 2008). This disastrous developmental outcome shows the magnitude that small RNAs' effects can have on gene regulation. As scientists further probe the genome for small RNAs, they are finding that this class of regulatory RNAs is pervasive throughout more and more regulatory pathways and mechanisms. Recent studies suggest that approximately 30% of all human protein-coding genes are regulated through miRNA targeting (Du and Zamore 2005).

Long Non-coding RNAs and Regulation

While small RNAs have generally been celebrated in their newfound rise to notoriety within the scientific community, long non-coding RNAs have received a mixed welcome. Despite their much longer history, the field of long non-coding RNAs is shrouded in mystery and contention. Small RNAs are generally united through their use of stem loops and similar targeting mechanism of consensus sequence binding, however, long non-coding RNAs have no unifying feature and are extremely diverse in not only their structural characteristics and domains but in their purported functions as well. Research abounds with descriptions of various non-coding RNAs that have exons and introns, are differentially spliced, have no introns, contain multiple transcripts, are polyadenylated or not, are capped or not, are confined within a gene region, span several genes, and so on (Wang et al. 2004; Mattick and Makunin 2006; Struhl 2007). Their inherent resistance to being categorized and defined makes them a difficult topic to tackle and gives outsiders the perspective that they are not concrete objects with real function but merely ephemeral wisps of shadowy expression (Wang et al. 2004; Struhl 2007). Additionally, the diversity of this group makes it difficult to imagine long non-coding RNAs with any consistent activity of function. Indeed, long non-coding RNAs have been associated with a

menagerie of functions including recruitment of transcription factors or repressors, chromatin remodeling, imprinting, chromosome stability, promoter interference and more (Plath et al. 2003; Willingham et al. 2005; Rinn et al. 2007; Zhao et al. 2008). The very lack of a pattern in this incohesive field has led many to abandon even the most well-characterized RNAs in favor of more tractable protein-coding genes (Hogness et al. 1985).

Much of the prejudice against long non-coding RNAs stems from a perpetuation of the protein-centric view of biology. Small RNAs are so different from normal coding genes in terms of processing and function that they are easily thought of as a separate entity from the rest of the coding genome. However, many long non-coding RNAs (lncRNAs) so closely resemble protein coding genes in terms of their transcript splicing, polyadenylation and regulation that it is understandable to want to view and judge these elements through the lens of a protein researcher. Many of the criteria by which a protein-coding gene would be judged are not consistent amongst non-coding RNAs, such as conservation, lack of comparable "functional domains", appropriate transcript modifications such as polyadenylation, fuzzy borders of a defined gene domain, transient expression times and so on.

However, one overarching theme that does seem to develop throughout this field is that RNA is not constrained by the same forces of functionality that govern the protein world. Insertion or deletion of just a few nucleotides within a coding region can have disastrous effects on the resultant protein's activity. However, lncRNAs likely have much more flexibility in their sequence structure since they are not restrained by the rules of the codon reading frame. Recent studies have shown that lncRNAs whose sequences show potential for significant secondary structures have much more sequence constraint overall than other lncRNAs (Marques and Ponting 2009). This implies that the RNA's function might be entwined with its resultant structure and that as long as the structure were preserved the nucleotide sequence could change dramatically. This suggests that lncRNAs have much smaller "functional domains" than protein-coding genes and that these short stretches of highly conserved sequence could be linked together with significant amounts of nucleotides with very low selective pressures. In fact, stem loops are formed through base-pairing of pair-able sequences; as long as reciprocal changes maintained the pairing, the sequence could change significantly over time. An extremely important lncRNA involved in dosage compensation in mammals, *Xist*, only exhibits high conservation over short sections of its length (Pang et al. 2006). Thus, high levels of sequence conservation do not seem to be a necessary characteristic of lncRNAs to impart functionality.

Additionally, one of the other more challenging hurdles that scientists have in assigning functionality and significance to these RNAs is that many of them do not exhibit the normal characteristics one would use to define something as a gene. Generally, a gene is defined by processed transcripts that emerge from a unique and discrete locus that result in some sort of gene product. lncRNAs break most of these rules, making it difficult for scientists to even determine whether some of these RNAs are or are not even genes. Various genome tiling array studies have revealed many non-coding transcripts that exist within or transcribe through documented loci of known genes (Bertone et al. 2004; Cheng et al. 2005; Jongeneel et al. 2005). Additionally, some lncRNAs seem to exist and function in some arrangement of sense/anti-sense pairing (Chen et al. 2004; Dahary et al. 2005; Katayama et al. 2005). Our current terminology for gene definitions is not set up for defining characteristics of that type. Also, protein-centric customs insist upon gene transcription from one direction, since obviously mRNAs cannot be translated from two different directions. Our current understanding of gene transcription does not make it easy for us to imagine how a locus can sustain, and even require, transcription from

opposing sides and function. Indeed, the *Xist* locus requires this competing transcriptional activity to function properly.

Xist is one of the most studied and highly characterized long non-coding RNAs. It is a 17 kb non-coding RNA located on the X inactivation center of the mouse X chromosome (Brown et al. 1992; Penny et al. 1996). Upon transcription it accumulates on and apparently coats the entire inactivated X chromosome (Clemson et al. 1996; Lucchesi et al. 2005). Additionally, this RNA has an antisense partner called *Tsix* that remains active only on the active X chromosome (Lee and Lu 1999). Recently, a smaller non-coding RNA, called *RepA*, was found to be expressed within the *Xist* locus in the same direction (Yue et al. 2009). It is proposed that *RepA* and *Tsix* compete for binding of the Polycomb complex, PRC2 through chromosome pairing. Somehow, a decision is made on each chromosome to inactivate one of the opposing RNAs. If *RepA* successfully binds the PRC2 group, then *Tsix* expression is turned off, full length *Xist* is activated, and PRC2 spreads throughout the chromosome and maintains its inactivation (Yue et al. 2009). This locus functions through recruitment of PRC2 through non-coding RNA signals and is significantly important for proper X inactivation and dosage compensation. Interruption of these non-coding RNA activities leads to improper X inactivation and severe developmental defects. Obviously, this example proves that sense/anti-sense transcriptional activity within the genome can have very important regulatory functions.

Some characteristics of lncRNAs make them not only difficult to characterize but also difficult just to locate. Many lncRNAs are expressed at levels significantly lower than most protein-coding genes and for much shorter lengths of time (Bertone et al. 2004; Carninci et al. 2005). These shadowy transcripts are difficult to capture through both *in situs* or current expression methods such as genome tiling arrays since they would appear barely above background. Many researchers would point to the low and transient expression as a reason why these lncRNAs would be merely artifacts (Huttenhofer et al. 2005; Werner and Berdal 2005). On the contrary, it supports the idea that these lncRNAs are involved mainly in the regulation of other genes. Recruitment or obstruction of regulatory proteins would likely take far fewer transcripts than those needed for reasonable expression of a protein product. Additionally, most of these studies are performed using polyA+ cytoplasmic RNA in order to eliminate high backgrounds of rRNA and tRNA that compose the majority of an organism's total RNA. However, It is becoming more apparent that a large part of the previously unstudied genomic transcription is in fact nuclear-bound polyA- RNA (Cheng et al. 2005; Kiyosawa et al. 2005). There is a very good chance that many of the more dynamic and interesting lncRNAs have not yet been discovered due to these research techniques designed mainly to detect regular coding genes.

New methods need to be designed with the idea of capturing all non-coding RNAs in mind. The advent of the Solexa sequencing system is allowing for significantly deeper coverage of transcriptomes than previous technology, allowing researchers to identify short-lived RNAs more easily than before. Additionally, most RNA purification methods aimed at reducing background of tRNA and rRNA usually work by selecting for polyA+ RNA. However, this eliminates any polyA- non-coding RNAs that may be playing important roles in the cell. A new method of using beads coated with sequences specific to rRNA sequences instead depletes the offending RNAs from the pool, leaving the rest of the polyA- transcriptome behind. To perform new screens for functional non-coding RNAs, some groups are using RNAi screens against putative non-coding RNAs found through genome tiling arrays. In fact, the non-coding RNA NRON was discovered using this technique.

NRON (non-coding repressor of NFAT) is a non-coding RNA that was discovered to have repressor activity of the human transcription factor NFAT (Willingham et al. 2005). This lncRNA was discovered by performing a large-scale RNAi screen against 454 mouse ncRNAs previously identified and logged in the Functional Annotation of the Mouse (FANTOM) Consortium. The NRON locus is composed of three exons which are alternatively spliced, yielding transcripts from 0.8 to 3.7 kb in size. RT-PCR experiments showed that, like many non-coding RNAs, NRON is expressed at very low levels, suggesting that it is involved in some sort of regulatory function. When shRNAs were generated and expressed against the NRON region, significant increase in NFAT activity resulted. This suggests that somehow the NRON RNA is involved in repressing NFAT activity. Additionally, overexpression of the NRON transcript resulted in decreased NFAT activity. Further characterization showed that NRON physically associates with three members of the importin-beta superfamily which directly mediate the nucleocytoplasmic transport of NFAT proteins. This example shows the concentrated activity that exists within only a few transcripts of a lncRNA. Also, it demonstrates the effectiveness of new RNA-centered techniques for identifying putative functional long non-coding RNAs.

The world of non-coding RNAs is full of new genes and activities that have yet to be discovered. Until now, the full extent of the role of non-coding RNAs in the regulatory mechanisms of genes has never been imagined. The discovery that both short and long non-coding RNAs have important functions related to regulation, development and human diseases is an important eye opener to the scientific community. Further studies into the detection and characterization of these things will hopefully begin to fill in the holes of our knowledge of the complex web of gene regulation.

Organization of the Thesis

This thesis describes my studies of various long non-coding RNAs in the fruit fly *Drosophila melanogaster*. This organism is an excellent choice for the study of non-coding RNAs for several reasons. As described above, both long and short non-coding RNAs have been shown to be key players in a wide variety of regulatory mechanisms. For decades, *Drosophila melanogaster* has been one of the main model organisms used to study the regulation of genes involved in development. Groundbreaking work in terms of enhancer function and Hox gene regulation was preferentially done in this animal. Its genome has been vastly sequenced and annotated, which makes combing it for anomalies and novel transcripts much simpler. Additionally, years of *Drosophila* research have developed well-established tools for husbandry, imaging, and the production of transgenic animals. In fact, in recent years many groups have identified important functions for *Drosophila* microRNAs in early development, showing that this organism also relies heavily on non-coding RNAs as regulatory elements (Biemar et al. 2005; Ronshaugen et al. 2005). Finally, *Drosophila* is steeped in history of non-coding transcription. Since the sequencing of the *Drosophila* Bithorax Complex, more and more mysterious non-coding transcripts have been revealed throughout this complex, some of whose functions remain uncharacterized even today. In this thesis, I describe my efforts in revisiting the Bithorax Complex for the purpose of discovering new functions for these RNAs. I particularly focus on two: *bithoraxoid*, a complicated and well-established transcribed non-coding locus, and *750*, a non-coding RNA that I discovered in the course of this work.

Anterior-Posterior Axis Specification in the Early *Drosophila* Embryo

The early *Drosophila* embryo is a syncitium of nuclei in which little to no transcription of zygotic genes occurs. Instead, development is driven by the interplay of maternally deposited transcripts and proteins within the egg. Early distinguishers of the A-P axis are mRNAs encoding for the transcriptional and translational regulatory proteins Bicoid, Hunchback, Caudal and Nanos. *bicoid* and *nanos* transcripts are tethered to opposite poles of the embryo (Ingham 1988; Hoch and Jackle 1993; Kornberg and Tabata 1993; DiNardo et al. 1994). Upon translation of these transcripts, they form opposing protein gradients within the egg since the protein products are able to diffuse throughout the embryo. Interaction with these proteins and others present within the cell also set up opposing gradients of Hunchback and Caudal expression. This results in an embryo with high concentrations of Bicoid and Hunchback at the future anterior pole and high concentrations of Nanos and Caudal at the future posterior pole (Fig. 1). Regulatory interactions between these and other maternally-deposited transcription factors lead to the expression of the three gap genes, *giant*, *Kruppel*, and *knirps*.

The gap genes are transcription factors that are expressed in broad domains, dividing the embryo into three large segments: the head section, the tail section and the middle section. The gap genes interact with each other to express the next wave of genes, the pair-rule genes, such as *ftz*. These genes are expressed in seven evenly spaced stripes across the embryo and are the first indicators of segmentation (Fig. 1). These pair-rule genes divide the embryo up into 14 parasegments across the A-P axis of the embryo through their alternating on-off patterns. These parasegments will persist throughout development and each will become a particular segment in the adult fly, each with its own characteristic structures. The patterning of the structures for each individual segment is accomplished through the expression of the Hox genes.

Hox genes are transcription factors of the homeobox family of proteins. They activate and repress multitudes of downstream targets in order to properly pattern each segment in which they are expressed. They are highly regulated genes that display interesting regulatory characteristics such as colinearity and posterior dominance. Colinearity refers to the fact that the Hox genes are located within the genome in the same order in which they are expressed in the developing embryo (Lewis 1978). For instance, the gene *Abdominal-B* is the most posteriorly expressed Hox gene and it is also the most 3' located Hox gene within the genome. Posterior dominance refers to the fact that more posterior Hox genes are capable of repressing other more anteriorly expressed Hox genes (Sanchez-Herrero et al. 1994). This means that *Ultrabithorax* (*Ubx*) is capable of being repressed by both *abdominal-A* and *Abd-B*. If *abd-A* is deleted, *Ubx* expression will expand posteriorly to fill the segments normally patterned by *abd-A* all the way up to the *Abd-B* border.

Mutations in Hox genes lead to improper segment patterning. This means that segmental identities will be transformed into other improper segmental identities (Lewis 1978). The most famous of these is regulatory mutations for the *Ubx* gene. *Ubx* is normally expressed in the T3 segment and patterns the formation of the balancing organ called the haltere (Fig. 1). When *Ubx* activity in this segment is lost, a duplication of the T2 segment forms, causing the fly to develop two sets of wings. In the opposite situation, if *Ubx* is upregulated in the T2 segment, duplication of the T3 segment will result, leading to flies with no wings. These genes are obviously extremely potent and important for proper development and the fly employs a complex system of regulatory mechanisms, all of which are not yet understood, to properly express these genes. The Hox genes are split into two different genomic complexes: the Antennapedia Complex,

which patterns the anterior most structures, and the Bithorax Complex, which patterns the posterior half of the fly. This thesis deals mainly with the Bithorax Complex.

The Bithorax Complex

The Bithorax Complex exists as a massive 330 kb of sequence on the third chromosome of *Drosophila melanogaster*. Detailed studies of this complex have been going on for over fifty years, and its research has provided invaluable insights into our understanding of general gene regulation. Since homeotic genes are conserved throughout the animal kingdom, the Bithorax Complex has in many ways acted as a guide to deciphering even our own mammalian regulatory mechanisms. Despite its enormous size, this complex contains only three known protein coding genes, *Abdominal-B (Abd-B), abdominal-A (abd-A)* and *Ultrabithorax* (*Ubx*) (Fig. 2) (Sanchez-Herrero et al. 1985; Martin et al. 1995). All three are important Hox genes involved in segmental patterning of the *Drosophila* embryo. However, protein coding regions make up less than 2% of the total sequence in the Bithorax Complex. It has been shown that a majority of this remaining sequence is composed of a complex web of *cis*-regulatory elements that maintains a tight regulation of the Hox genes (Irish et al. 1989; Casares and Sanchez-Herrero 1995). It was found that spaced throughout the Bithorax Complex were regulatory elements and that each was responsible for appropriate expression of the Hox proteins within a single parasegment. The regulatory regions governing expression of the *Abd-B* and *abd-A* genes were named *infraabdominal-* (or *iab-*) *2* through *8* and are expected to host at least one enhancer element within each region (Beachy et al. 1985; Celniker et al. 1990; Karch et al. 1990; Macias et al. 1990; Sanchez-Herrero 1991)(Fig. 2). Additionally, regulatory regions that controlled *Ubx* expression within parasegments 5 and 6 were also characterized and were called the *abx/bx* and *bxd/pbx* regions (Beachy et al. 1985; White and Wilcox 1985; Little et al. 1990) (Fig. 2).

These regulatory elements function by specifying expression patterns for a single parasegment. If that element is mutated, then the specific segment it was controlling would produce an improper identity. For instance, flies that are homozygous for a mutation affecting only the *iab-7* element would have their seventh abdominal segment transformed into segment 6 fate, resulting in a duplication of this identity (Galloni et al. 1993).

Bithorax Complex Regulatory Elements Produce Non-coding Transcripts

The discovery of the functions of the *iab* regulatory domains in the Bithorax Complex was the answer to one of the most interesting questions in the field of *Drosophila* gene regulation. However, further study of this complex revealed that additional mysteries were yet unsolved. For instance, several groups discovered that the protein-coding Hox genes were not the only source of expressed transcripts from this region (Sanchez-Herrero and Akam 1989; Casares and Sanchez-Herrero 1995; Zhou et al. 1999). They showed that non-coding transcripts were emanating from the *iab* regulatory regions and were preceding expression of *abd-A* and *Abd-B* genes. A more detailed study showed that transcription from these segments mirrored the enhancer-driven patterns that each domain drives as a *cis*-regulatory element (Bae et al. 2002). This means that each *iab* domain is driving an enhancer pattern that has a similar expression pattern to its non-coding transcript. The majority of this mysterious transcription still has no known function. Since the transcription occurs slightly before activation of the enhancer function, it was suggested that the transcription is needed to open the chromatin domain in order

for regulatory elements to properly access the enhancer. However, this would suggest that after the initial activation of the enhancer there would no longer be a need to continue the transcription, but several of these transcripts do persist until later in development. More recent models theorize that these transcripts are necessary not only for the initiation of the enhancer but also for the continued activity and that they function to recruit regulatory proteins to the enhancer sites (Akbari et al. 2006).

These studies of the *iab* region also revealed additional mysterious non-coding transcription throughout this region, including anti-sense transcripts within the *iab-4* and *iab-6* domains. These transcripts are expressed at the same time but in mutually-exclusive patterns as their sense counterparts (Bae et al. 2002). Recently, various groups identified microRNAs that target *Ubx* transcripts within both the sense and anti-sense transcripts of *iab-4* and showed that they have regulatory effects on *Ubx* expression (Cumberledge et al. 1990; Aravin et al. 2003; Ronshaugen et al. 2005). However, the reason for the sense and anti-sense pairing of both the *iab-4* and *iab-6* transcripts still remains a mystery.

These studies of the *iab* do not even encompass all of the non-coding transcription within the Bithorax Complex. Non-coding transcripts are now thought to play a role in proper function of insulator regions within this complex (Ho et al. 2009). While the previous examples mainly dealt with non-coding transcription involving regulation of *Abd-B* and *abd-A*, this thesis will describe new discoveries of non-coding RNA regulation of the gene *Ubx*.

Organization of the *Ubx* Locus

Ultrabithorax (*Ubx*) is one the earliest sequenced genes of the *Drosophila* genome and the first ever recorded Hox gene (Lewis 1978). It is a protein-coding gene consisting of a 5' exon containing the 5' UTR, 2 microexons that are alternatively spliced, and a 3' exon that contains the homeobox domain characteristic of all Hox genes. The gene produces six alternatively spliced transcripts expressed at differing times during development. Five produce different DNA-binding transcription factors and one produces a *4.7* kb spliced non-coding transcript of unknown function (Akam and Martinez-Arias 1985; Akam et al. 1985; Beachy et al. 1985; Hogness et al. 1985). *Ubx* transcription factor proteins act as key regulators of downstream developmental genes necessary for proper patterning of structural organs within parasegments 5 and 6, including the T3 leg and the haltere.

Ubx is expressed early in embryogenesis beginning in Stage 5 within the posterior half of the embryo. Despite being expressed in more posterior segments, *Ubx*'s main region of function is within the parasegments 5 and 6. Posterior to this, its activity is repressed by the actions of *abd-A* and *Abd-B*, due to the effect of posterior dominance within Hox genes. *Ubx* expression is maintained later in development in thoracic segment 3 discs, haltere and third leg, and more weakly in posterior parts of the thoracic segment 2 (Irvine et al. 1991). Its most visible function is the repression of T2 fate (wings and T2 leg) within the T3 segment. Loss of *Ubx* function within the T3 segment results in duplication of T2 fate and a second set of wings and T2 legs. Misexpression of *Ubx* within the T2 segment leads to T3 fate duplication and repression of wings to form wingless flies with two pairs of halteres (Castelli-Gair et al. 1990). *Ubx* has a complex set of *cis*-regulatory elements located both upstream of its promoter and within its own introns (Fig. 2). These were characterized by Ed Lewis through deletion, insertion and translocation mutants exhibiting different degrees of *Ubx* phenotypes (Lewis 1978).

Within the upstream regulatory region of the *Ubx* promoter exists a non-coding gene called *bithoraxoid* (*bxd*) that produces several different transcripts of unknown function (Fig. 2). Although its transcription has been previously characterized (Hogness et al. 1985) its main function is still unclear.

Organization of the *bxd* Locus

bithoraxoid (*bxd*) is a gene located in the *Drosophila* Bithorax Complex only 8 kb upstream of the Hox gene *Ubx*. ~20 kb in length, the gene encodes 5-6 differentially-spliced polyadenylated transcripts from 2-3 separate promoters (Fig. 4). These various transcripts are differentially expressed both in time and intensity (Fig). While several models suggest a function for these transcripts, currently there is no reasonably acceptable model for their activity. Additionally, the locus encodes well-characterized *cis*-regulatory regions, including various forms of two distinct enhancer loci termed *pbx* and *bxd* (Fig. 7). The *pbx* and *bxd* elements function as important enhancers for the proper regulation of the downstream protein coding Hox gene, *Ubx*. In general, these enhancers are thought to pattern parasegment 6, including the posterior compartment of the haltere (T3 segment) and the anterior compartment of the A1 abdominal segment. Enhancer expression analysis using lacZ constructs and enhancer trap lines indeed show that *pbx* elements give a *Ubx*-like pattern with an anterior boundary at parasegment 6 (Simon et al. 1990; Muller and Bienz 1991) (Fig. 24). *bxd* mutations tend towards a transformation of the first abdominal compartment with thoracic cuticle (Lewis 1963) suggesting that in its absence, the *pbx* enhancers takes over the role of patterning that segment. *pbx* mutations tend toward a transformation of the posterior compartment of the haltere into a wing suggesting that in absence of *pbx*, no significant *Ubx* activity exists within this compartment, leading to a haltere-to-wing phenotype. Obviously, with such disastrous phenotypes, *pbx* and *bxd* enhancer elements are significantly important for the proper regulation of *Ubx* expression throughout development.

In addition to the characteristics already discussed, the *bxd* locus also contains regulatory sites for the chromatin modifying complexes Polycomb (Pc) and Trithorax (Trx) Group Proteins. These protein complexes are involved in long-term maintenance of Hox (and other) gene expression patterns. They do this through modification of histones and reorganization of chromatin structures. While Hox genes are initiated by a cascade of signaling transcription factors, these signals are generally transient and do not last through later stages of development (Ingham and Martinez Arias 1992). However, Hox genes must maintain their expression throughout development to obtain normal development (Struhl and Akam 1985) (Lewis 1963; Garcia-Bellido et al. 1976). The Pc and Trx group complexes read these activity states and maintain them by making chromatin modifications to designate the domain as either open (on) or closed (off) (Franke et al. 1992). The Pc Group Complex is associated with repressed domains whereas the Trx Group Complex is associated with open domains (Paro 1990; Kennison 1993; Simon 1995; Pirotta 1997). These activities allow cells to maintain their homeotic identities throughout the rest of development.

Pc and Trx Groups recognize their target domains through binding of maintenance elements called Polycomb Response Elements (PREs) or Trithorax Response Elements (TREs) (Muller and Bienz 1991; Simon et al. 1993; Chan et al. 1994; Fritsch et al. 1999; Busturia et al. 2001). The *bxd* locus contains a well-characterized PRE within its first intron. It is important

for maintenance of *Ubx* repression in appropriate tissues, like the wing disc. When mutated, *Ubx* activity is de-repressed in wing discs, leading to wing-to-haltere phenotypes (Sipos et al. 2007).

Additionally, this same *bxd* PRE is overlapped by three transcribed TREs (Fig. 4). It was shown that these transcripts are important for Trithorax Group Recruitment to the *Ubx* locus in cells in which *Ubx* is active to continue its expression in later development (Sanchez-Elsner et al. 2006). However, this model remains controversial and the definitive function of the *bxd* TREs is still not known.

The Complex Nature of the *bxd* Locus and its History

Although scientists have been studying the *bxd* locus for almost a hundred years, it has not been enough time to decipher all aspects of this extremely complex region. Its intertwining of numerous *cis* regulatory elements (both at the DNA and chromatin level of regulation) and its multitude of diversely expressed transcripts makes the separate characterization of these moieties very intractable. The majority of the research on *bxd* was performed during the 1980's and 90's when the Bithorax Complex was enjoying its heyday. Back then, the phenotype and regulatory studies relied mainly on mutant lines containing relatively large deletions, insertions and rearrangements within the locus. These methods did not allow for the fine level of functional and sequencing analysis that we enjoy today. Additionally, these genomic rearrangements obviously affected two different levels of *bxd* locus activities: both the transcripts and the *cis*-regulatory elements contained in the DNA sequence. It was impossible to completely decipher the individual effects of these separate elements. Indeed, Hogness et al. laments this very fact in his paper studying both *bxd* functions and expression (Hogness et al. 1985). Also, both the *in situ* expression and transcript sequencing techniques employed back then are not nearly as sensitive or precise as fluorescent *in situs* or fragment cloning through PCR, which are the routine methods used today to visualize expression or do enhancer stains. That being said, a majority of this previous work focused on the *cis* regulatory elements of the *pbx* and *bxd* enhancers and the Polycomb-regulating element of the PRE located within the *bxd* domain. Most of these studies concluded that the *cis* domains were the most important functional elements and that if the transcripts were contributing any significant regulatory information, it paled in comparison to the effects of the DNA domains (Hogness et al. 1985). With no obvious inroads into investigating the functions of the *bxd* transcripts, further research on this matter was eventually abandoned until recently when new techniques and mindsets for RNA research have fueled a resurgence in *bxd* transcript research.

Transcribed *bxd* TREs recruit the Trithorax Group to Promote *Ubx* Expression

In 2006, the Sauer group published a paper showing that three transcribed Trithorax Response Elements (TREs) within the 5' intron of *bxd* functioned to promote recruitment of the Trithorax Group of chromatin modifying proteins to the locus (Sanchez-Elsner et al. 2006). They claimed that interaction between the TRE transcripts and the Trithorax Group was instrumental in maintaining *Ubx* expression in appropriate cells throughout development. Transcription of the TREs was necessary to recruit the Trithorax Group as opposed to the Polycomb Group since the TREs overlap with a Polycomb Response Element (PRE) as well (Fig. 4). This is not a novel concept since many TREs are transcribed that overlap PRE sites (Schmitt et al. 2005). They claim that these TREs are expressed in cells in which *Ubx* is active

(including the halteres), that Ash1 (a Trithorax Group protein) interacts with the TREs *in vivo*, and that misexpression of the TREs leads to ectopic *Ubx* expression (Sanchez-Elsner et al. 2006). While this is an updated theory that describes functional transcription within the *bxd* domain, it does not discuss any novel concepts involving the *bxd* transcripts themselves.

bxd transcripts may repress *Ubx* activity through promoter interference

An opposing viewpoint on the function of the *bxd* locus was published by the Mazo group (Petruk et al. 2006). In their model, they claim that *bxd* transcription exists solely as a mechanism for repression of *Ubx* through the action of promoter interference. They suggest that while the *bxd* transcripts themselves are not functional, transcription of the locus causes the transcription machinery to run through the *Ubx* promoter, preventing recruitment of PolII to the promoter to properly transcribe the *Ubx* transcripts, effectively shutting off the gene. They support this theory through double *in situs* of *Ubx* and *bxd* that show that the two genes are expressed in mutually exclusive patterns and that these two genes are never on in the same cell at the same time. Additionally, they perform RT-PCR on the intergenic region between *bxd* and *Ubx* and show transcription within this region in cells in which *Ubx* is inactive. Also, they utilize a *Ubx* mutant, pbx^2, that has a deletion within the 5' *bxd* promoters (Irvine et al. 1991). This mutant is missing a large chunk of the *pbx* enhancer which regulates *Ubx* expression in parasegment 6. Within this mutant, however, *bxd* transcription from the 5' *bxd* promoters is abolished, leading to upregulation of *Ubx* transcription in the early embryo. Further study indicated that the Trithorax Group binds to this region to promote elongation of *bxd* and subsequent repression of *Ubx*.

bxd Function in *trans*

Neither of these theories fully addresses whether *bxd* transcripts have any sort of trans-acting function as was originally theorized (Hogness et al. 1985). Additionally, both of these theories are highly contentious and not fully accepted by the *Drosophila* research community.

The *Ubx* region of the Bithorax Complex is not yet fully understood or characterized. This is apparent from the recent novel discoveries of *Ubx* regulation through non-coding RNAs in the *iab-4* region. Additional mysteries exist as to the detailed methods of its transcriptional regulation through *bxd* and other non-coding RNAs. In this thesis, I discuss the efforts that I have made to describe the role of *bxd* and other transcripts in *Ubx* regulation.

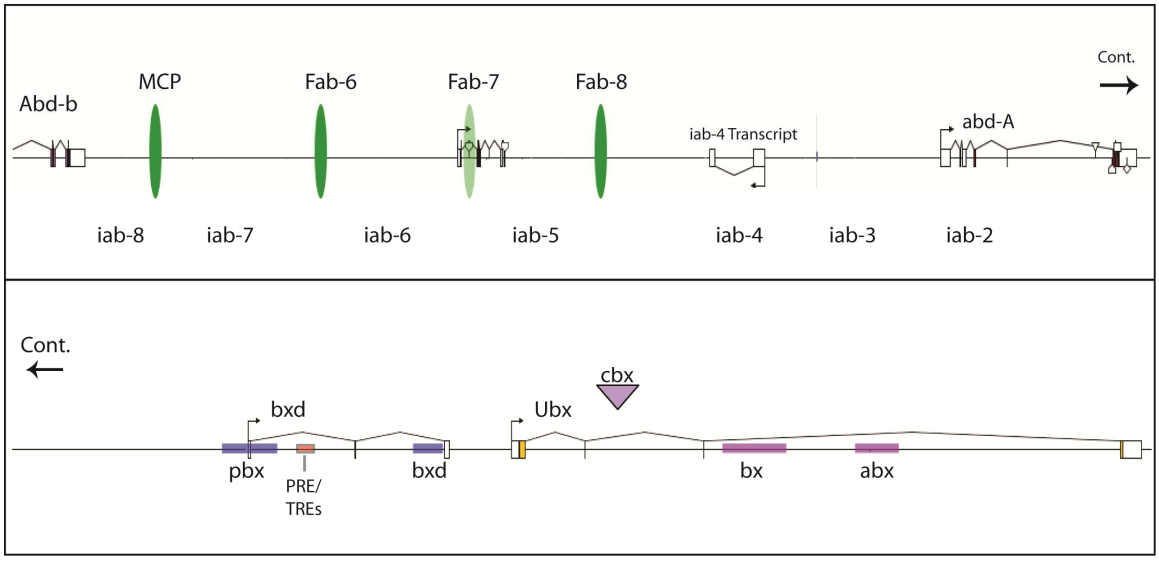


Figure 1 - Bithorax Complex Map This map includes the majority of the Bithorax Complex genes and regulatory elements., including the iab domains. Insulator elements are green ovals. *Ubx* enhancers for the *pbx*/*bxd* domain are blue. *Ubx* enhancers for the *abx*/*bx* domain are purple. *Cbx* is an insertion mutation of the *pbx* enhancer into the *abx*/*bx* domain. PRE/TRE element is in red.

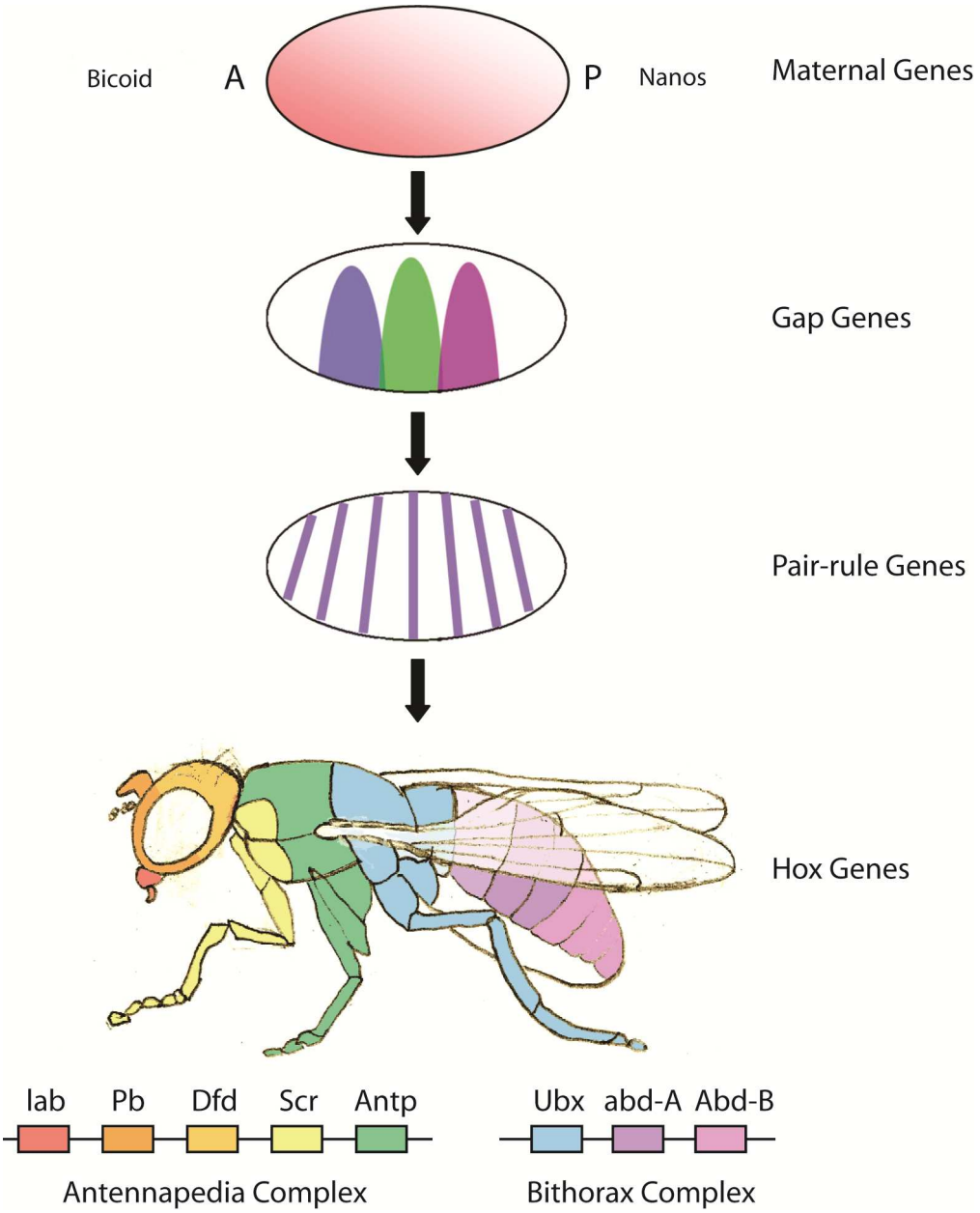


Figure 2 - Stages of Developmental Genes This figure shows the stages of early developmental genes in the *Drosophila* embryo. Maternally-deposited transcripts are transcribed to produce opposing gradients of protein. The Gap Genes (*giant*, *Kruppel*, *knirps*) are transcribed in broad domains. They interact to activate the Pair-rule genes which break the embryo into segments. The Hox genes give unique identities to these segments. Hox genes are separated into two complexes: the Antennapedia Complex and the Bithorax Complex. Genes are colored corresponding to the segments they pattern in the adult fly.

Chapter II:

Characterization of the Non-coding RNA *bithoraxoid*

Introduction

This chapter describes my undertakings of the characterization of the non-coding RNA *bithoraxoid* (*bxd*), including its expression patterns, interactions and possible functions. Located within the *Drosophila* Bithorax Complex, it has remained an enigma since its discovery nearly a hundred years ago by Calvin Bridges (Bridges and Morgan 1923). Expressed, and regulated like a Hox gene, its non-coding status has befuddled Hox researchers for decades. While many efforts have been made to determine its function and general characteristics, it still remains a locus of unanswered questions. In this chapter, I review the known attributes and current theories surrounding this locus and supplement it with my own unique studies of its expression and developmental effects.

bxd Expression Patterns in the Early Embryo

bxd expression begins in the early embryo about 2.5 hours after fertilization. It appears in a broad stripe of expression ranging from parasegment 6 to the posterior of the embryo (Fig. 3). It precedes *Ubx* expression, as shown in Fig. 3A, which is stained for *bxd* in red and *Ubx* in green. At this early stage, only a strong pattern of *bxd* expression is seen. Soon after, as shown in Fig. 3B, a faint band of *Ubx* expression begins to form at the anterior border of the *bxd* domain in parasegment 6. Additionally, the *bxd* pattern begins to narrow posteriorly as seen in Fig. 3D. As the embryo develops, *Ubx* expression becomes more prominent in parasegment 6 and begins to turn on in the more posterior parasegments in a banding pattern (Fig. 3E). At this point, the broad *bxd* stripe also begins to resolve into a more banded pattern (Fig. 3F). As gastrulation begins, *Ubx* and *bxd* resolve into banded patterns of stripes, each of which shows strongest expression in alternating parasegments (Fig. 3G). However, they maintain overlapping expression in cells on the borders of the parasegment boundaries (Fig. 3G and higher magnification in Appendix I). This overlap is conserved throughout the rest of embryogenesis as observed in the following pictures. As germ band elongation proceeds, *bxd* and *Ubx* stripes of expression become more distinct (Fig. 3I and 3J). Additionally, *bxd* appears to gain expression faintly anteriorly to and overlapping with *Ubx* expression in parasegment 6 (Fig. 3K and 3L). This overlap is seen more clearly in a ventral view of the same stage (Fig. 3M and 3N).

These images and descriptions of the *bxd* and *Ubx* expression patterns are mainly in line with previous reports about their activity in early embryogenesis (Hogness et al. 1985). However, in later stages, clear visualization of expression becomes difficult in such structured embryos and larvae. Previous studies used Northerns to analyze the transcripts present in later stages. Hogness et al. claimed that the early transcripts (all but one of the characterized transcripts) are not expressed in later stages and that only one very short RNA appeared at all later in development (1985). This was puzzling, since they had no function to assign to something so transiently expressed. However, other reports have suggested that *bxd* is expressed at later stages, including in the wing disc. Indeed, qPCR analysis of total RNA from embryo, larval stages, and wing discs agree with these later ideas and show that they do retain expression throughout development, albeit at different levels for different transcripts (Fig. 7). This suggests

a continuing function, or at least, activity of *bxd* transcription throughout development and not merely in the early embryo.

shmiR System Allows for Targeting of *bxd* transcripts independent of enhancers

One of the best tools that *Drosophila* researchers have at their disposal is the GAL4-UAS system of gene expression that was first described by Fischer et al. in 1988 (Fischer et al. 1988). GAL4 is a protein from *Saccharomyces* that acts as a transcriptional activator for genes that contain its DNA recognition sequence, called UAS, at their promoters (Giniger et al. 1985; Ma and Ptashne 1987c; Ma and Ptashne 1987a; Ma and Ptashne 1987b). When introduced into flies, the GAL4 gene is able to specifically express any UAS-labeled gene within the spatial and temporal region of GAL4 expression (Brand and Perrimon 1993) Since then, dozens of enhancer-trap GAL4 lines have been made that drive GAL4 in various tissues and times throughout development. In order for researchers to express their own transgene, it needs to be cloned into a UAS cassette and inserted into the genome. Since the harnessing of known transposable elements within the *Drosophila* genome, termed P-elements, it has become relatively easy to insert specific DNA sequences into the genome by attaching P-element sites to the construct (Engels 1996). Thus, *Drosophila* has become one of the easiest systems to study the effects of over- or mis-expression of any particular construct.

While this system is very powerful, it does not allow researchers to perform knock-down assays for particular genes since it is only useful for overexpression of genes. *Drosophila* is notoriously difficult to perform targeted gene knock-outs in, since its transformation rate for homologous recombination is painfully low (Maggert et al. 2008). P-elements can be used to disrupt genes, but since they generally insert randomly into the genome, this is not particularly useful for looking at specific genes. Additionally, gene insertions may lead to partially-functional products that may prevent full analysis of resultant phenotypes. One method of getting around this problem was to use dsRNA injection into *Drosophila* embryos. Once inside, these RNAs are processed into 21 nucleotide long hairpins that silence gene expression through the siRNA pathway (Elbashir et al. 2001). Though effective, this system has limited uses, however, since it can only be used in injectable tissues, it can not be easily targeted, and the amount of time the silencing is effective is unpredictable and uncontrollable. Additionally, these dsRNAs must be on the order of ~300bps which can produce many different siRNA sequences, leading to potentially abundant off-target effects (Jackson and Linsley 2004).

To overcome these problems, Haley et al. developed a customizable microRNA UAS expression construct termed "shmiRs" (Haley et al. 2008). This construct contains sequence that produces an ~80 bp "pre-miRNA" stem loop that is recognized by cellular factors, spliced, and preferentially loaded into the RISC complex for degradation of the targeted transcript. This system allows for control of the exact sequence of an expressed siRNA, thereby diminishing nonspecific, off-target phenotypes. Haley et al. showed greater than 95% knockdown of target mRNAs when using this system (Haley et al. 2008). Additionally, this system allows for greater control over expression timing and spatial deployment since it utilizes the UAS construct for its expression vector backbone. This allows the researcher to knock down expression of a single mRNA transcript at a particular time and place simply by crossing the shmiR-containing fly lines with any of the GAL4 enhancer trap lines that already exist. Not only does this make knock downs more approachable for *Drosophila* researchers, but it also targets these mRNAs in *trans*, without disrupting the gene's DNA sequence that may contain other *cis*-regulatory elements.

It was with this system that I decided to experiment with knocking down the *bxd* non-coding transcripts to test whether disrupting their *trans* functions would produce effects independent of the known *cis*-regulatory elements within their locus. I designed two sets of shmiRs that would be inserted into the genome in pairs. The first set called shmiR-1/2 targets the largest transcript of the *bxd* locus (*bxd-D*) (Fig. 4). A second set targets three other early transcripts (*bxd-A*, *B*, and *C*) (Fig. 4) and was called shmiR-3/4. Upon crossing these lines to a GAL4 expression line called MS1096 which drives GAL4 expression in wing discs, halteres, and the 3rd leg discs, a strong wing phenotype emerged (Fig. 5). The wings appear shrunken and improperly patterned. They are too disfigured to determine if wing veins appear but there are wing-like hairs around the edge suggesting that they retain some form of wing identity. When performing the same cross with the *bxd*-3/4 shmiR, the phenotypes are similar but more severe (Fig. 5). Since MS1096 is an X-linked driver, males show more severe phenotypes due to the fact that in males, the X chromosome is upregulated due to the mechanisms of dosage compensation. The fact that the shmiR-3/4 females look very similar to the shmiR-1/2 males suggests that they are producing different intensities of the same phenotype with these different shmiR lines. This might make sense, since the shmiR-3/4 shmiR is targeting more transcripts, and also more intensely expressed transcripts (see Fig. 4), than shmiR-1/2. shmiR-3/4 males have the most severe phenotype, causing the wings to look like tight balls of tissue more reminiscent of halteres than wings. To show that this was not a GAL4 line effect, these crosses were done in another X-linked wing driver, A9, with the same results (Fig. 5).

These results suggest that the *bxd* transcripts are in fact important regulators of wing development within the wing disc. The fact that all of the *bxd* transcripts that were tested are expressed in the wing disc (Fig. 7) backs up this idea. In order to perform further controls, and to establish that these phenotypes are not artifacts of shmiR expression, the GAL4 drivers were crossed to a non-specific shmiR (NS-shmiR) that has no significant binding to any transcript within the *Drosophila* genome. These shmiR expressions show no visible phenotype when crossed to either GAL4 line (Fig. 5 and Appendix III). To show that these phenotypes were specific to *bxd* knockdown, two additional shmiR pair lines targeting other positions within the *bxd* transcripts were crossed to MS1096 and revealed that the phenotypes were similar to the original *bxd* shmiR lines (Fig. 6).

Earlier knockdown reveals stronger homeotic phenotypes

In order to assess earlier *bxd* transcript activity in the embryo as opposed to later larval structures such as the wing disc, a different driver called KREG-GAL4 which drives early expression in a Kruppel pattern in addition to later expression was used to drive expression in the early embryo. When crossed to *bxd*-3/4 shmiRs, an even more severe wing phenotype appeared (Fig. 9C). These wings appear to have normal-looking wing hinges with very haltere-looking structures emerging from them (Fig. 9C). When magnified, the wings have a small ball shape reminiscent of a haltere but still have long wing-like hairs that suggest some wing patterning is taking place (Fig. 9D). Very severe curling of the 3rd legs was evident, which was also observed in the MS1096 crosses (Fig. 9B). This curling is evidence of *Ubx* perturbation in the leg disc. Additionally, the flies exhibited a phenotype of twisted genitalia (Fig. 9C and 9D). Normally pointing directly anterior, the male genitals in these flies have genitals that are rotated on their axis and sticking out from the abdomen (Fig. 9E and 9F). This is a phenotype often seen in embryos that have been perturbed for *Abd-B* expression, which is necessary for genital disc

patterning (Foronda et al. 2006). These results show that knocking down these *bxd* transcripts earlier is producing stronger and more varied homeotic phenotypes. This suggests that *bxd* may be having an effect on other genes located within the Bithorax Complex. This is particularly poignant since expressing the shmiRs ubiquitously in later larval stages produces phenotypes of the wing and abdomen, which are controlled by genes in the Bithorax Complex. *bxd* disruptions do not seem to perturb expression of other homeotic genes located in the other Hox complex, the Antennapedia Complex, which patterns the anterior thorax and head structures. This indicates that *bxd* transcripts may have a function related to genes limited to the physical location within Bithorax Complex.

bxd shmiRs create adequate knockdown in wing discs

In order to prove that the phenotypes are due to *bxd* transcript knock down, it is necessary to prove that the shmiRs are adequately targeting the transcripts for degradation and significantly lowering transcript levels in the appropriate tissues. To do this, qPCR analysis was performed on 3rd instar climbing larvae expressing shmiR-3/4 with the KREG driver. The NS-shmiR crossed to the KREG driver was the control. As shown, at least 50% knock down in expression is seen across the various transcripts (Fig. 8). It is interesting to note that the more intense knockdown is seen in the transcripts that have both shmiRs targeting them (as in *bxd-B/C*) than just one (*bxd-A*).

Ubx Upregulation in *bxd*-shmiR Wing Discs

Ubx is not normally expressed in wing discs, with the exception of the thin peripodial membrane that surrounds the disc, which is very different in appearance to the wing disc proper. Aberrant expression of *Ubx* in the wing disc leads to homeotic transformations of wings-to-halteres. This is evident in the *Contrabithorax* (*Cbx*) mutant (Fig. 10A and 10B) which was recovered by Ed Lewis in 1949. It has small wings whose bubble-like shape is reminiscent of haltere structures. Sequencing of the *Ubx* locus within this fly revealed that the mutation was caused by two DNA lesions (Bender et al. 1983). A 17 kb piece of DNA had been deleted from the *bxd*/*pbx* (Fig. 2) regulatory region and reinserted into the *abx*/*bx* domain (located within the *Ubx* 3' intron; see Fig. 2). Normally, *pbx* mutants show a transformation of posterior haltere into posterior wing, suggesting that the *pbx* enhancer is important for maintaining high levels of *Ubx* expression within the posterior haltere disc (Lewis 1963; Garcia-Bellido 1975; Garcia-Bellido et al. 1976). However, when this same enhancer was inserted near the *abx*/*bx* region, the phenotype was a strong transformation of posterior wing to haltere. Peifer et al. explained this phenomenon by suggesting that these enhancer regions must be within two separate domains of regulatory expression (Peifer 1987). When the *pbx* insert had been moved to the *abx*/*bx* domain of expression, it caused strong upregulation of *Ubx* improperly within parasegment 5, which patterns the posterior wing, causing its transformation into a haltere.

The fact that the *bxd* shmiR flies so strongly resemble the *Cbx* fly phenotype is a good indicator that the *bxd* phenotype could be due to upregulation of *Ubx* activity (Fig. 10A and 10B). Indeed, the *Cbx* flies even displayed the same curling of the 3rd leg (Fig. 10A arrow). If the *bxd* transcripts were actually important for repressing *Ubx* expression, it would be expected that *Ubx* protein would be visible in the wing discs of these *bxd*-shmiR lines. Using the cross with the strongest wing phenotype, shmiR-3/4 x KREG, wing discs were dissected and stained

for *Ubx* protein using the *Ubx* antibody from Kelsh et al (Kelsh et al. 1994). NS-shmiR x KREG discs were also dissected and stained in parallel as a control. The result showed no *Ubx* stain in the NS-shmiR wing discs, as expected, but showed intense staining of *Ubx* protein in the halteres proving that the staining actually did work (Fig. 10C). The shmiR-3/4 discs, however, showed a significant level of *Ubx* protein within the pouch of the wing disc (Fig. 10D). This would explain the significant homeotic-looking wing-to-haltere transformation observed in the *bxd* knockdowns and gives credence to the model that *bxd* transcripts have a role in repressing improper *Ubx* activity.

Heat shock time series suggests Polycomb Interaction

Although we now had a good indication as to overall *bxd* function, it was necessary to identify a mechanism as to how *bxd* is regulating these homeotic genes. To do this, an expression time series was set up using heat shock GAL4 drivers. These are GAL4 drivers that are under the regulatory control of the promoter of the heatshock *Drosophila* gene, *hsp70*. When placed within a hot environment (37^{o}C), heat shock genes activate and transcribe heat shock proteins. They act as chaperones that help keep sensitive proteins in the proper conformation in the increased heat (Arya et al. 2007). In this *hsp70*-GAL4 driver, this same response activates GAL4 ubiquitously at high levels. Therefore, timing of expression of the shmiR can be controlled with great precision.

To test the effects of *bxd* knockdown at different times throughout development, a time series of embryo laying was set up. Appropriate parents were placed together in a vial. They were allowed to lay eggs in that vial for 24 hours. Then, they were flipped into a new vial and allowed to lay for 24 hours again. In the meantime, the first vial is allowed to age without new embryos being laid into it. This produces a series of vials containing staged progeny. This was performed for seven days, meaning that the final product was seven different vials of genetically consistent progeny that were 1-7 days old. At that time, all of the vials were placed into a 37^{o}C water bath for 1 hour. Afterwards, they were removed, placed at room temperature and allowed to develop to adulthood. Emerged adults were then removed and counted.

In this experiment, a shmiR-1/2 line was used that had the transgene inserted on the X chromosome. A cross was set up (Fig. 11) in which only females of the next generation would inherit both the *hsp70*-GAL4 construct and the shmiR-1/2 expressing construct. Therefore, only females would be subjected to both the *bxd* shmiR effects and the *hsp70*-GAL4 effects, while the males would only be subjected to the hs-GAL4 effects. A NS-shmiR cross control was set up at the same time as a comparison. When the adults emerged, living males and females were counted for each of the vials and recorded. Ratios of female to males for the *bxd*-shmiR cross were then compared to the ratios of female to males for the NS-shmiR cross to judge differences in survivorship. The NS-shmiR results show the baseline ratio that would be expected with nonsense shmiR expression (Fig. 11). The graph shows that there was a significant effect on the *bxd* females in the first 3-4 days of development since almost no live females emerged from these vials. The fact that many males emerged suggests that this is not just due to the general experimental conditions of heatshocking and GAL4 expression. This indicates that knocking down *bxd* transcripts during the first 3-4 days of development has a significant lethality associated with it.

bxd may be key player for Polycomb Recruitment

The lethality timeline of the *hsp70*-GAL4 test suggested that *bxd* transcripts might be involved in proper establishment of the Polycomb group for repression of *Ubx* in appropriate cells, such as those found in the wing disc. While various transcription factors are responsible for the initiation of the expression patterns for the Hox genes, eventually these factors disappear in development. However, Hox gene expression is necessary until late in development (Lewis 1963; Garcia-Bellido 1975; Garcia-Bellido et al. 1976). The factors that maintain the appropriate expression patterns are the Pc and Trx groups as discussed above. In a *Pc* mutant, Hox genes will turn on in their appropriate expression patterns initially, but at about 14-16 hrs, these patterns begin to fade and eventually *Abd-B* becomes ubiquitously expressed throughout the embryo, repressing all of the other Hox genes (Pirotta 1997). Therefore, Pc proteins are not needed for initiation but are necessary for maintaining these proper expression patterns.

The fact that *hsp70*-GAL4 x shmiR-1/2 embryos show lethality in the first few days, but not after, suggests that they may be necessary for Pc establishment in the Bithorax Complex. To test this idea, a complex cross was set up between shmiR-1/2 and MS1096 flies and a line containing a PRE-*Ubx* construct (Muller and Kassis 2006). In this construct, the PRE from within the *bxd* intron (Fig. 12A) is placed upstream of a minimal *Ubx* promoter that is driving lacZ. When Pc is present, it binds the PRE and shuts off lacZ expression (Fig. 12B). This results in no staining, which is what is seen in a wing disc carrying this construct (Fig. 12D). A second construct contains the same *Ubx*-lacZ fragment with lambda DNA replacing the *bxd*-PRE (Fig. 12C). In this scenario, Pc will not bind the unrecognized sequence and lacZ is not active (Fig. 12E). This reporter acts as a sensor for proper Pc Group functioning in the wing disc.

To test whether knocking down *bxd* transcripts interferes with proper Pc regulation in the wing disc, crosses were set up to put all three constructs (shmiR-1/2, MS1096, PRE-lacZ) into the same flies and see whether or not lacZ expression would appear in the wing discs. Upon examination of these mutant wing discs, lacZ expression was visible in the wing discs (Fig. 12F). This suggests that somehow *bxd* transcripts are playing an important role in either recruiting or tethering the Pc complex to the *bxd*-PRE. When these transcripts are degraded, Pc is disrupted, *Ubx* is relieved of its repression, and spurious *Ubx* expression occurs. Phenotypes of wing-to-haltere transformations are common in Pc mutants. A study by Sipos et al. showed that when the *bxd*-PRE was deleted, posterior wing segments transformed into haltere segments (Sipos et al. 2007). These data suggest that the *bxd* shmiR phenotypes are due to an upregulation of *Ubx* in the wing disc caused by lack of Pc binding.

*E(*z) Genetic Mutants Antagonize *bxd*-shmiR Phenotype

To further investigate the role that Polycomb proteins might be playing in *bxd* activity, a series of crosses with genetic mutants and the *bxd*-shmiR were employed. In this scheme, a mutant fly line containing a P-element insertion into the gene *Enhancer of Zeste (E(z))*, a known component of the Polycomb Group 2 complex, was combined with the *bxd*-shmiR/MS1096-GAL4 expression pair (Fig. 13). *E(z)* contains a SET domain which means it is capable of binding single-stranded nucleic acids and is responsible for methylating histone H3 on lysines 27 and 9 to condense chromatin structure. It has been shown to be an active component at the *bxd*-PRE in *Drosophila* for Ubx repression (Kahn et al. 2006). A complex series of crosses was performed to produce a pool of progeny that would have various pairings of the three

components: *E(*z)^{G4251} mutant, *bxd*-shmiR, and MS1096-GAL4. Analyzing these progeny would determine whether or not a decrease in *E(*z) activity would worsen the *bxd*-shmiR phenotype, which would suggest that *E(*z) was interacting genetically with *bxd in vivo*.

The results of three crosses were analyzed. The first control cross included combining the *E(*z)^{G4251} mutant, MS1096-GAL4 and NS-shmiR. This is to see the effects of the *E(*z) mutant interacting with expression of any shmiR in the MS1096-GAL4 to determine whether any background phenotypes exist. As shown in Fig. 13, however, all progeny appeared completely normal, indicating that the *E(*z)^{G4251} mutant and the NS-shmiR have no phenotypes on their own. A second control cross involved combining the 3rd chromosome balancer in the original *E(*z) line with the *bxd*-shmiR and MS1096-GAL4 lines. This cross will set a baseline of expected phenotypes and ratios of phenotypes produced by the *bxd*-shmiR that will be compared against the test cross. This cross produced a spectrum of 3 distinguishable phenotypes in expected ratios: normal, a mild wing phenotype, and a moderate wing phenotype. It is expected that normal flies would result from the cross not only because the cross should produce a small percentage of genetically wildtype flies but also because the *E(*z)^{G4251} mutation does not produce a phenotype on its own as evidenced from the NS-shmiR cross. The mild and moderate wing phenotypes are due to the *bxd*-shmiR effects. However, these effects are much more subtle in the females than the previously described males since they do not get the boost of GAL4 expression from an upregulated X chromosome. Therefore, they display a milder range of phenotypes, from normal-looking to a moderate amount of wing disturbance (Fig. 13). The fact that fewer than half of the females from this cross show any phenotype suggests that half of the females carrying the *bxd*-shmiR had such subtle phenotypes that they were indistinguishable from normal females.

The test cross produced some similar-looking flies as the previous cross, however, they had four distinguishable classes of phenotypes instead of three, including normal, mild, moderate and severe. As seen in the figure, a new class of severe wing phenotype appeared in this cross and no other, suggesting that the *E(*z)^{G4251} mutation is in fact interacting with the *bxd* knockdown to produce an even stronger effect. Additionally, a higher percentage of progeny were also in the moderate and mild classes of phenotypes than the *bxd*-shmiR/MS1096/Balancer cross suggesting that this interaction, while producing a new severe category, also increased the penetrance of the *bxd*-shmiR phenotype in the rest of the flies.

These results suggest that *E(*z), or the entire PRC2 complex, is involved in moderating *bxd* function *in vivo*. This idea would be in line with previous data from humans involving the non-coding transcript HOTAIR. HOTAIR was discovered by the Rinn group and is a non-coding RNA within the *HOXC* cluster in the human genome (Rinn et al. 2007). In foreskin cells, it was shown that HOTAIR is normally active and that *HOXD* genes were repressed. When this transcript was targeted with siRNAs, repression of *HOXD* genes was alleviated. Further studies indicated that this was due to a lessening of PRC2 activity within this locus. Additionally, through RNA Immunoprecipitation experiments, it was shown that two PRC2 proteins that are known to have RNA-binding abilities, *E(*z) and *Su(z)12* (another PRC2 component necessary for H2K27 methylation), were bound to the HOTAIR transcript in normal foreskin cells, suggesting that this interaction was necessary for either recruitment or stabilization of the PRC2 complex within the *HOXD* locus. A more detailed mechanism for this activity is not currently known. This is not the only case of non-coding RNAs interacting with Pc complexes. As previously discussed, *Xist* also functions similarly. These data suggest that the *bxd* transcripts may be functioning in a homologous method as the HOTAIR transcript to maintain repression of the *Ubx*

gene in differentiated cells. This is a reasonable model since *bxd* and HOTAIR have striking similarities such as both being non-coding transcripts within Hox gene complexes. Further studies of the interaction of the PRC2 complex proteins with *bxd* transcripts would give more details about this activity, and RNA IPs with various Pc proteins are currently being done.

Overexpression of Certain *bxd* Transcripts Causes Phenotypes

One of the main arguments used to show that *bxd* does not function in *trans* is that overexpression of its cDNA does not show a phenotype (Petruk et al. 2006; Petruk et al. 2008). In order to test this idea, I generated two different constructs to overexpress *bxd* transcripts. Since the *bxd* locus is so large, the logical method of expressing the largest of the transcripts is through cDNA cloning. Using a cDNA library generated from PolyA (+) RNA from 2-4 hr embryos, the cDNA for the *bxd-D* transcript was cloned and inserted into an expression construct. No phenotype was noticeable when driven using the MS1096- or KREG-GAL4 drivers (Fig. 14). An additional construct was made to drive expression of the *bxd-B* and *bxd-C* transcripts by amplifying the genomic region containing these two transcripts and cloning it into an expression vector. Overexpressing this construct in wing discs using the MS1096-GAL4 driver resulted in strong wing phenotype that is unique from the phenotype observed from *bxd*-shmiRs. In this phenotype, the wings are clearly wings that show creases throughout the wing and seem to be improperly unfolded. This is a clear indication that the *bxd* transcripts can function in *trans* and their misexpression can result in misregulation of downstream wing patterning genes.

Discussion

bithoraxoid has remained a puzzling mystery since it was first discovered nearly 100 years ago. Here I attempted to present evidence that suggests that it has a role in Pc regulation of the Hox gene *Ultrabithorax*. Until the invention of the *Drosophila* shmiR system, the study of the *bxd* transcripts separate from the *cis* functions of their locus was nearly impossible. The specificity and inducible nature of their targeting allows for clean and sufficient knockdown of the individual transcripts for study while unaffecting the *cis* regulatory functions. Use of this system has revealed astounding wing phenotypes that arguably resemble homeotic wing-to-haltere transformations. Additionally, earlier expression of shmiRs using the KREG-GAL4 driver indicates that *bxd* may play an early role in regulation for the entire Bithorax Complex. Driving expression of the shmiRs outside of the posterior parasegments (with ubiquitous drivers; data not shown) produces no phenotypes in the anterior regions, such as the eye. The fact that the *bxd* shmiRs produce no phenotypes in the areas of the fly in which *bxd* is not normally expressed indicates that expression of the *bxd* shmiRs does not lead to artifactual phenotypes and that the wing phenotypes are genuinely due to knockdown of *bxd* transcripts.

Indeed, substantial controls were performed to be certain that the *bxd* phenotypes are in fact genuine. Control shmiR, NS-shmiR, which drives expression of a nonsense shmiR that targets no transcribed sequence within the *Drosophila* genome shows no phenotypes in any of the GAL4 backgrounds tested (Appendix III). Additionally, the *bxd* shmiR pairs 1/2 and 3/4 were broken apart and driven individually with the GAL4 drivers to ensure that they each reproduced the *bxd* knockdown phenotype, at least to a reduced extent (see Appendix II). Multiple additional shmiR pairs were made to *bxd* transcripts utilizing different binding sites

(Fig. 6) and showed similar phenotypes to the original shmiRs. These controls show beyond a reasonable doubt that the *bxd* knockdown phenotype is not an artifact but is consistently reproduced using shmiRs directed towards *bxd* transcripts.

When expressed in wing discs, the *bxd* shmiRs cause a wing phenotype reminiscent of a homeotic wing-to-haltere transformation. This suggests that *Ubx* may be inappropriately upregulated in wing discs since ectopic *Ubx* expression in wing discs does cause a wing-to-haltere transformation (Bender et al. 1983). This idea is supported by the fact that Cbx^{1} mutants, which have a regulatory mutation that causes inappropriate *Ubx* expression in wing discs, have a wing and T3 leg phenotype strikingly similar to those observed in *bxd*-shmiR flies (Fig 10A and B). Additionally, *Ubx* antibody stains of shmiR-3/4-expressing wing discs show ectopic *Ubx* expression in the wing pouch compared to wing discs from phenotypically normal NS-shmiR expressing discs (Fig 10C and D). This result indisputably suggests that normally *bxd* transcripts are somehow involved in repressing *Ubx* expression in inappropriate cells. This is in line with previous descriptions of *pbx* mutant embryos. These lines contain a deletion of the *pbx* regulatory domain which overlies the promoter for the *bxd*-A and *bxd-D* transcripts. In early embryos, these mutants show no expression of these two transcripts. Additionally, *Ubx* expression is inappropriately upregulated in the parasegments in which these *bxd* transcripts are missing (Petruk et al. 2006). Similar effects are seen in bxd^{113} mutants in which other areas of the *bxd* transcripts are disrupted (Hogness et al. 1985). However, these mutations do not lead to wing phenotypes. Instead, they cause haltere-to-wing transformations indicating that *Ubx* expression has been reduced in parasegment 6. This is likely due to the fact that disrupting the *cis*-regulatory enhancers has a much stronger effect on *Ubx* regulation than the *bxd* transcripts. Indeed, the *pbx* and *bxd* are strong positive *Ubx* enhancers and without their regulation, overexpression of *Ubx* would be unlikely (Hogness et al. 1985).

Previous studies have shown that Polycomb is an important element for continued repression of *Ubx* throughout development in appropriate cells. When Pc function is disrupted, ectopic expression of *Ubx* can occur in wing discs, resulting in wing-to-haltere transformations (Schwartz and Pirrotta 2007; Sipos et al. 2007; Schwartz and Pirrotta 2008). The fact that expression of *bxd*-shmiRs in wing discs causes *Ubx* upregulation suggests that it is somehow disrupting Pc function since maintenance repression would be occurring at this late stage of development. The heatshock-shmiR developmental series showed that lethality was induced in embryos that received a pulse of ubiquitous shmiR expression within the first 3-4 days of development (Fig. 11). This period of time is coincident with Pc initiation of Hox genes, suggesting that the *bxd*-shmiRs are disrupting this regulation, resulting in death. The shmiR/*bxd*-PRE/*Ubx*-lacZ construct experiment proved that in the presence of *bxd*-shmiRs, the Pc group is not adequately binding to the *bxd*-PRE in wing discs, leading to lacZ expression. This indicates that without *bxd* transcripts, Pc repression at the *Ubx* locus is somehow disrupted. Genetic crosses involving the *E(z)* mutant, which is a component of the PRC2 complex, and the expression of shmiR-1/2 in the wing discs lends further support to the idea that *bxd* transcripts normally repress *Ubx* through interaction with Pc proteins. In the *E(z)* background, shmiR-1/2 targeting of *bxd* transcripts leads to a worsening of the normal wing phenotype that is only seen with this combination of genetic elements (Fig. 13). This suggests that a decrease in *E(z)* function is further antagonizing the derepression of *Ubx* caused by a decrease in *bxd* transcripts, suggesting that they normally work in the same pathway to maintain *Ubx* repression in wing discs. The fact that the *E(z)* mutant shows no wing phenotypes on its own (as seen in the NS-

shmiR crosses) further proves that the wing phenotype enhancement is due to a genetic interaction and not just a compounding of two separate wing phenotypes.

The results above led me to construct a model of how *bxd* transcripts might lead to repression of *Ubx* in *Drosophila* embryos. Significant precedence exists for the role of non-coding RNAs in the proper targeting or recruitment of Pc complexes to specific loci (Brown et al. 1992; Clemson et al. 1996; Penny et al. 1996; Lee and Lu 1999; Rinn et al. 2007; Zhao et al. 2008). Currently, there is no detailed model as to how this functions biochemically *in vivo*. However, it is my opinion that secondary structure within these non-coding RNAs may act as a binding platform for the protein complexes either *en total* or through binding of individual components that then assemble to form a complete active structure (Fig. 15). Binding of the RNA may retain the complexes at the targeted locus, holding the Pc complexes there to maintain the repression throughout development. Targeting these transcripts for degradation would release the proteins, allowing the locus to become derepressed, leading to upregulation of *Ubx* activity.

Previous studies have identified several regions of complementarity between the transcribed *bxd* locus and the *Ubx* locus (Hogness et al. 1985). Hogness et al. suggested that these regions could form base-pairing between the *Ubx* DNA sequence and the *bxd* RNA sequence *in vivo*, though no evidence currently suggests that. It is possible that through complementarity of these regions the *bxd* transcripts may direct targeting of the Pc complex to the *Ubx* locus, causing repression of the *Ubx* locus while maintaining activity of the *bxd* locus. The fact that overexpression of the cDNA of the *bxd-D* transcript showed no phenotype, while the overexpression of the unspliced *bxd-B* and *bxd-C* transcripts did, suggests that it is perhaps the unspliced transcripts that are the functional unit and that splicing inactivates them. The fact that the regions of homology shared with *Ubx* are located within the *bxd* introns supports this idea. Studies of the Cbx^1 mutation have shown that the *Ubx* and *bxd* loci seem to exist within separate chromatin domains, since movement of the *pbx* enhancer into the *abx*/*bx* region activates the normally silent enhancer in wing discs, causing a dominant wing-to-haltere transformation (Peifer 1987). This would be in line with the model since *bxd* and *Ubx* are in general expressed in opposing expression patterns. However, there must be additional regulatory forces determining whether or not *bxd* represses *Ubx* in each cell, since *bxd* and *Ubx* do show consistent overlap of expression in cells bordering their expression domains that lasts at least throughout early embryogenesis (Appendix I).

Recent studies have provided alternative theories as to the function of *bxd* transcripts on *Ubx* regulation (Petruk et al. 2006; Sanchez-Elsner et al. 2006; Petruk et al. 2008). The Sauer group claims that expression of the Trithorax Response Elements in the *bxd* locus leads to activation of the *Ubx* locus through recruitment of the Trithorax Group of chromatin-modifying proteins. While he is presenting a generally activating role of the *bxd* locus on *Ubx* activity, which is in contrast to mine, his model deals only with the activity of the TRE transcripts, not the *bxd* transcripts proper. Seeing as I have done no experiments dealing with the TRE transcripts, I have no evidence to dispute him, nor is his model contradictory to mine.

A second group has recently made claims about an alternative model for *bxd* function. The Mazo group claims that the transcripts in and of themselves are not functional, however, the transcription of the *bxd* locus results in promoter interference of the *Ubx* promoter, causing repression (Petruk et al. 2006; Petruk et al. 2008). This theory is not necessarily mutually exclusive of mine since it is possible the transcripts could be recruiting Pc while the transcriptional machinery emanating from the *bxd* locus occludes the *Ubx* promoter. However,

this group's data is at odds with not only many of my results but also historical results from previous studies of the *bxd* locus. For instance, their main source of evidence that the *bxd* locus works through promoter interference is that their *in situs* show that *bxd* and *Ubx* are expressed in mutually exclusive expression patterns. However, my *in situs* contradict their result, since high magnification of *bxd* and *Ubx* double *in situs* shows *bxd* and *Ubx* overlapping expression within the same nuclei that border the two expression domains (Appendix IA and IB). Indeed, expression is seen from both *bxd* and *Ubx* loci on the same chromosome since the red and green dots completely overlap each other. This is not possible according to their model of promoter interference. Further, this is not a transient phenomenon since overlapping expression persists in later stages of *bxd* and *Ubx* expression (Appendix IC and ID). They also point to the *pbx* mutant as an example of *bxd* transcript abolishment leading to *Ubx* upregulation. However, they do not address the fact that *pbx* mutants exhibit a haltere-to-wing phenotype, indicating a downregulation of *Ubx* expression, not an upregulation. Additionally, the *pbx* mutation does not disrupt transcription of the 3' *bxd* transcripts (*bxd-B* and *bxd-C*) which would theoretically still be present to occlude the *Ubx* promoter. To demonstrate that *bxd* acts through a *cis*-regulatory mechanism and not in *trans*, they inject dsRNA against *bxd* transcripts into early embryos to knock down *bxd* levels and claim that resultant *in situs* show no *bxd* expression and no *Ubx* upregulation. This is highly unlikely, however, since dsRNAs may target free transcripts but are incapable of preventing transcription of nascent transcripts which would still be visible on an *in situ* stain. Indeed, several groups have used intronic probes to show *de novo* expression on the chromosome (Ronshaugen and Levine 2004). Also, these dsRNAs result in no phenotype that is described. As described above, injection of dsRNA has several drawbacks including lack of targeted control. Since these dsRNAs resulted in no phenotype such as the wing-to-haltere transformations seen with my shmiRs, it is likely that the dsRNA dispersed throughout the embryo and was diluted or degraded prior to the initial translation of functional *Ubx* protein at ~6 hr after egg laying (Simon et al. 1990). Finally, they claim that *bxd* elongation is maintained through association with the Trithorax Group which is needed for maintaining active chromatin states of a locus. This is in direct contradiction to dozens of papers on *Ubx* regulation through Polycomb Group-mediated repression of *Ubx* (Schwartz and Pirrotta 2007; Sipos et al. 2007; Schwartz and Pirrotta 2008). These groups have shown that disruption of Pc binding within wing discs leads to misexpression of *Ubx* (Sipos et al. 2007). However, there are no papers describing such an effect in Trithorax disruption. Overall, their model may be potentially possible, however, their results directly contradict not only my own observations of the *bxd* locus but also many *bxd* researchers throughout history.

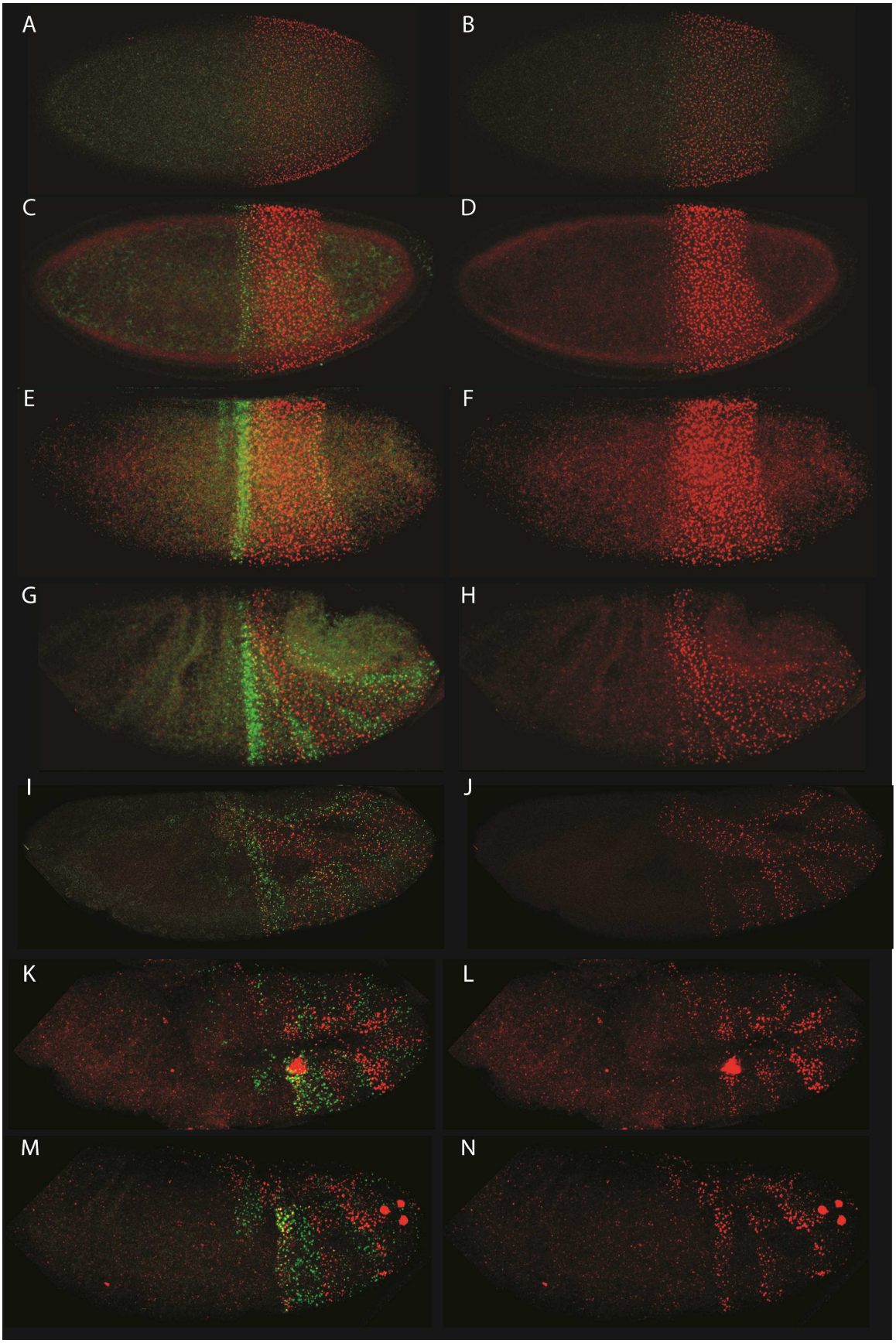

A
B
C
D
E
F
G
H
I
J
K
L
M
N

Figure 3 - *bxd* and *Ubx* expression through Early *Drosophila* Development Fluorescent *in situs* during 2-5 hrs of development. *bxd* is in red and *Ubx* is in green. A) *bxd* precedes *Ubx* expression in a broad domain at 2 hrs B) *Ubx* expression begins to come on in a weak stripe anterior to the *bxd* domain C), E), G), K), I) and M) show both *bxd* and *Ubx* expression in progressively older embryos D), F), H), J), L) and N) show only *bxd* expression in the same embryos.

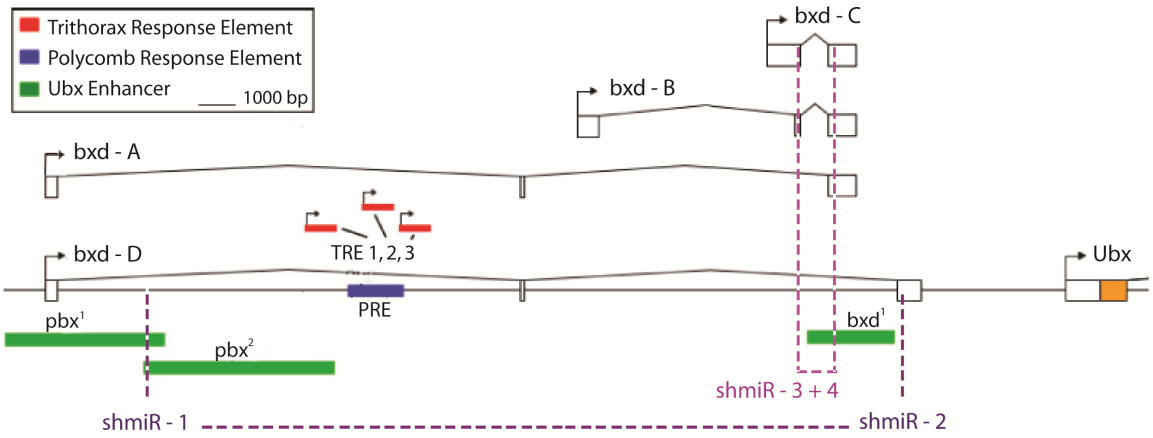


Figure 4 - shmiR Map of the *bxd* Locus The *bxd* locus contains several different transcripts of different spliceforms and originating from different promoters. It is ~8 kb upstream of *Ubx*. Within the *bxd* locus are several enhancers for *Ubx*, called *pbx* and *bxd*. Additionally, a PRE and overlapping transcribed TREs are within its first intron. Two pairs of shmiRs were made directed at *bxd* transcripts, shmiR-1/2and shmiR-3/4.

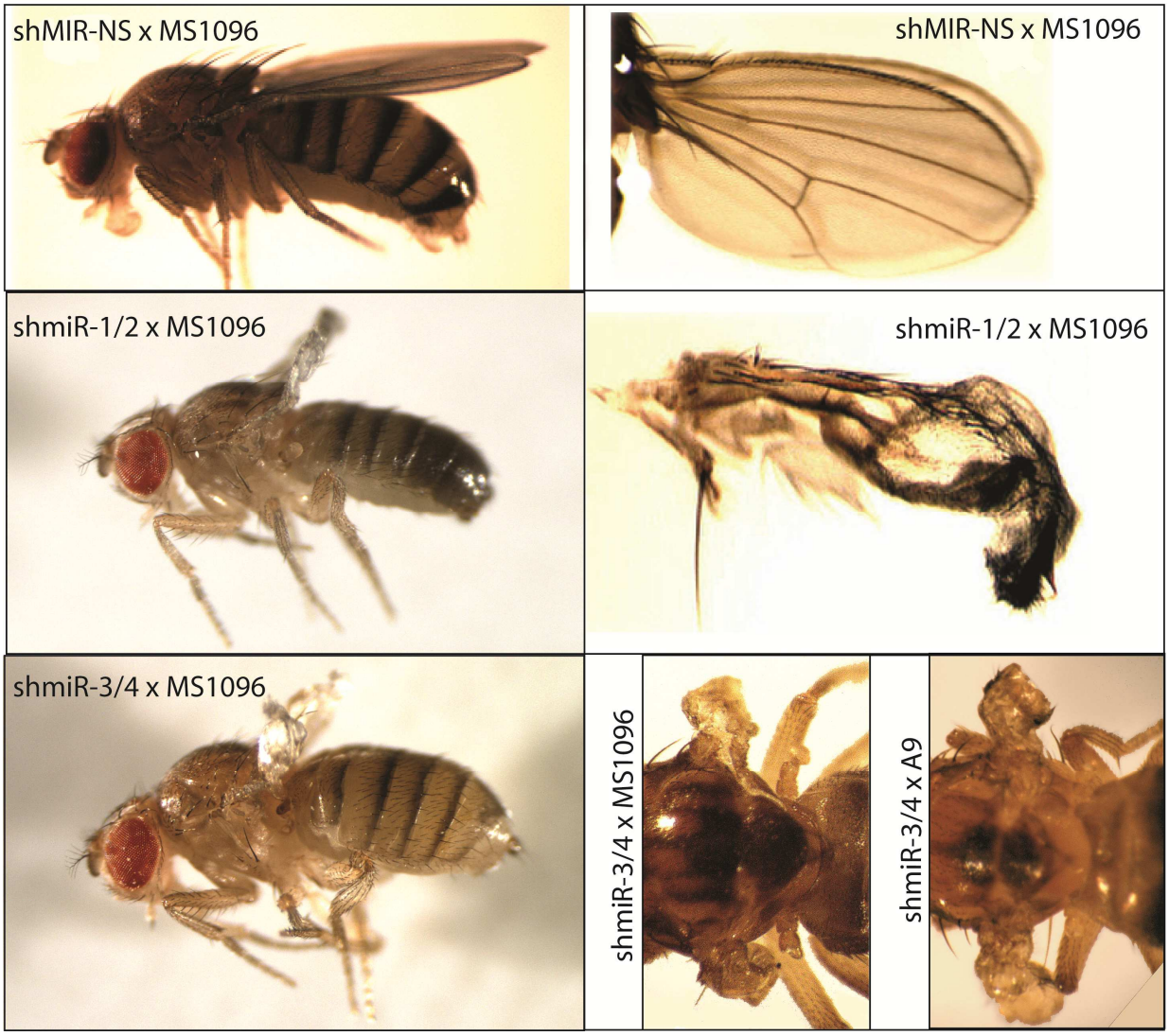


Figure 5 - Phenotypes of *bxd* knockdown in wing discs Transgenic flies containing the shmiR-1/2 construct were crossed to MS1096 – GAL4 driver flies, which drives GAL4 in the wing disc. Resultant male progeny show severe wing defects consisting of crumpled wings. Female progeny from a shmiR-3/4 x MS1096 – GAL4 cross looked similar. Males from this cross were more severe, with wings looking like crumpled balls. This is expected since the MS1096 – GAL4 driver is located on the X chromosome and due to dosage compensation this driver would be more highly expressed in males. Crosses made with a control shmiR construct, NS–shmiR, show no phenotype.

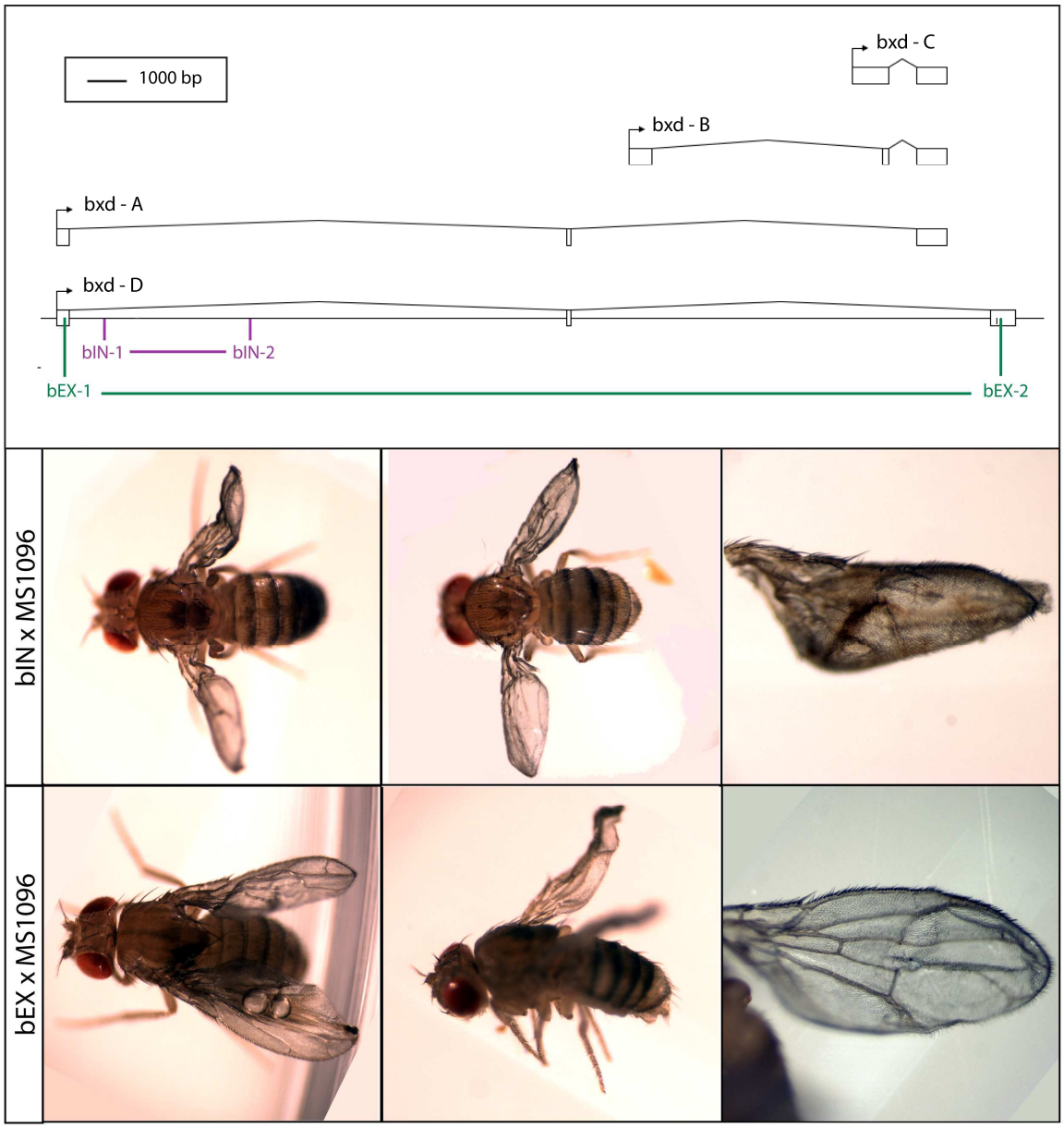


Figure 6 - Additional *bxd* shmiRs show a consistent phenotype Two additional shmiR construct pairs were made to target *bxd* transcripts at unique sites. When crossed to the MS1096 – GAL4 driver they show similar wing phenotypes consistent with the original *bxd* shmiR constructs.

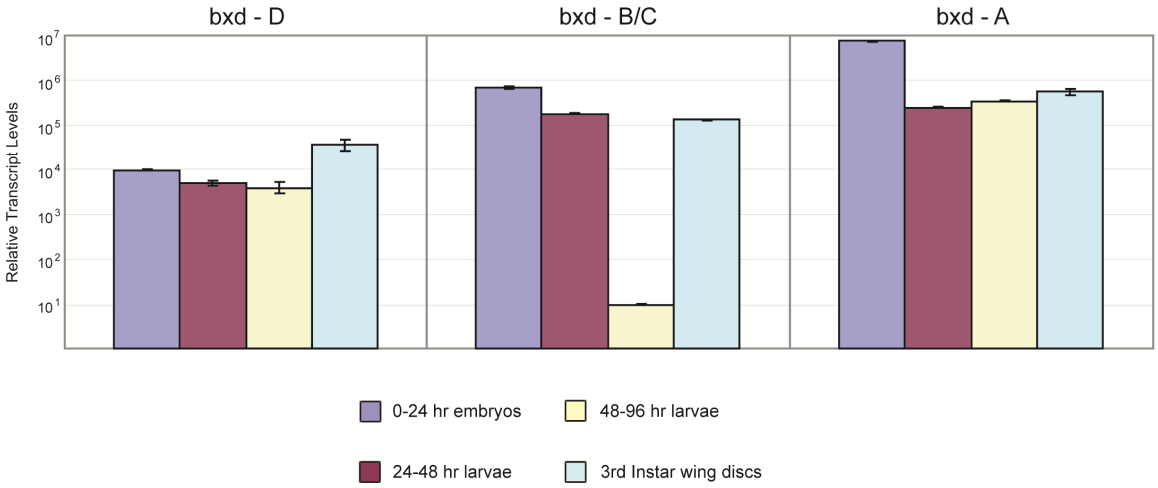


Figure 7 - *bxd* transcript levels vary throughout development qPCR analysis was done on samples from various stages of development. All *bxd* transcripts are continually expressed throughout all stages, including in the wing disc. The *bxd–D* transcript shows the lowest amount of expression, while the *bxd–B/C* and *bxd–A* transcripts show approximately equal levels but vary in their expression throughout different stages of development.

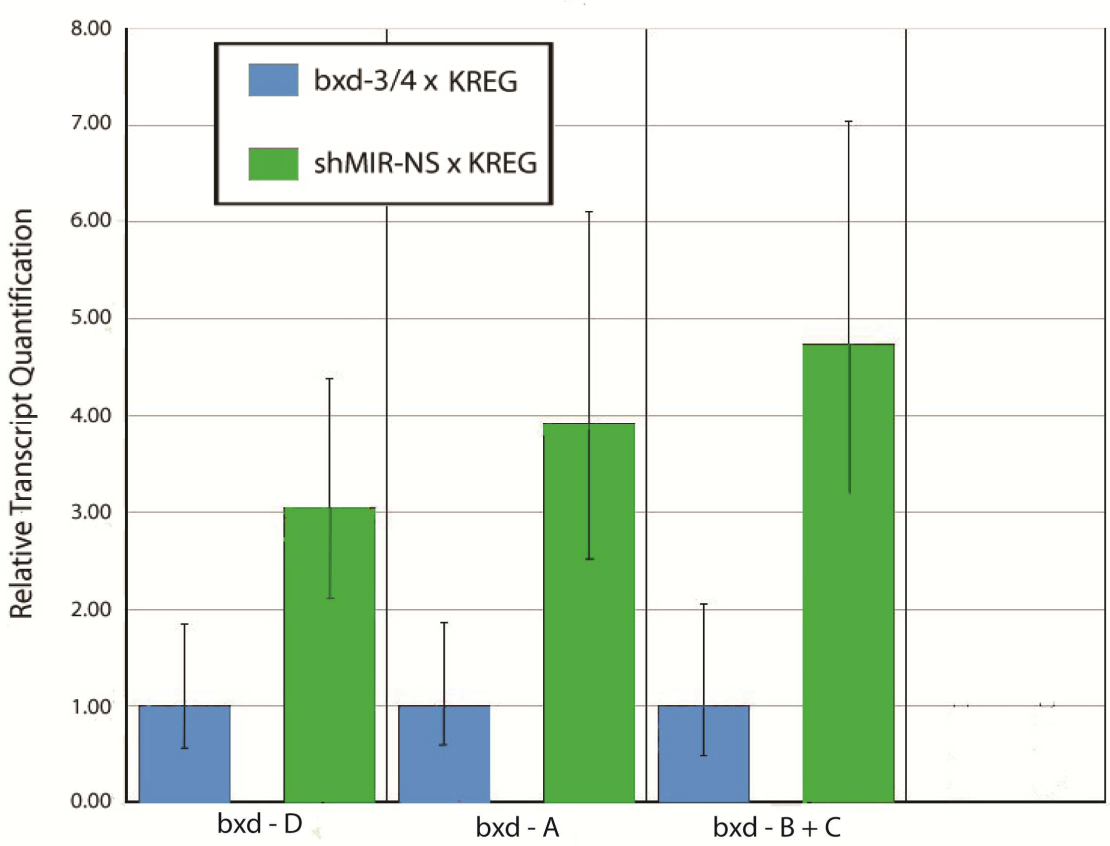


Figure 8 - qPCR shows knockdown of *bxd* transcripts in *bxd* – shmiR expressing larvae
qPCR was performed on third instar crawling larvae from shmiR- 3/4 x KREG – GAL4 crosses. Significant knockdown of transcript is seen compared to the control NS – shmiR x KREG – GAL4 cross.

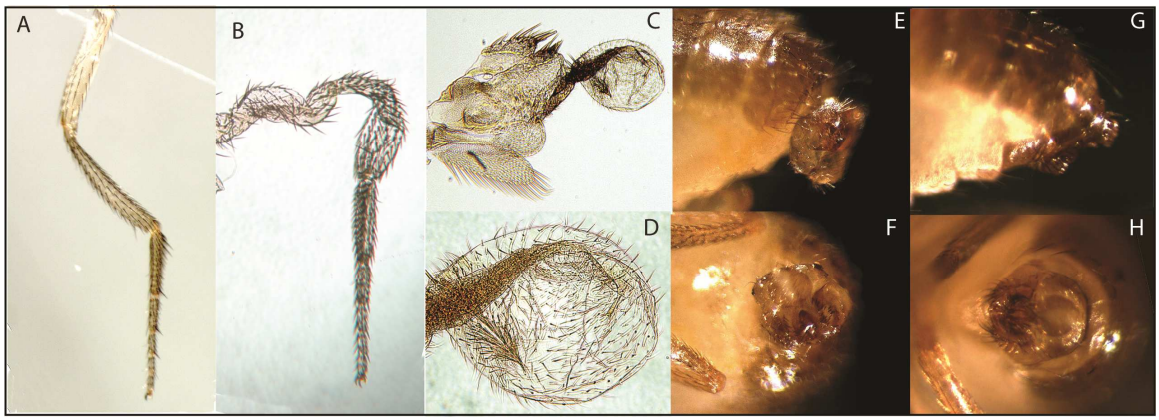


Figure 9 - shmiR – 3/4 shows more severe phenotypes with KREG – GAL4 driver A cross of the shmiR-3/4 line with the KREG –GAL4 line which drives expression of GAL4 in a Kruppel pattern shows more severe phenotypes than with a wing disc driver. A) NS–shmiR x KREG- GAL4 3rd leg shows normal phenotype B) shmiR-3/4 x KREG 3rd leg shows significant curling of the appendage C) wing from a shmiR-3/4 x KREG showing wing-to-haltere transformation D) close-up of wing from C) shows wing-like hairs on the wing E) side view of male genitalia from shmiR-3/4 cross shows abnormal protrusion compared to G) from a NS – shmiR cross F) ventral view of male shmiR-3/4 genitalia exhibiting twisting on its axis relative to H) normal genitalia from NS–shmiR cross

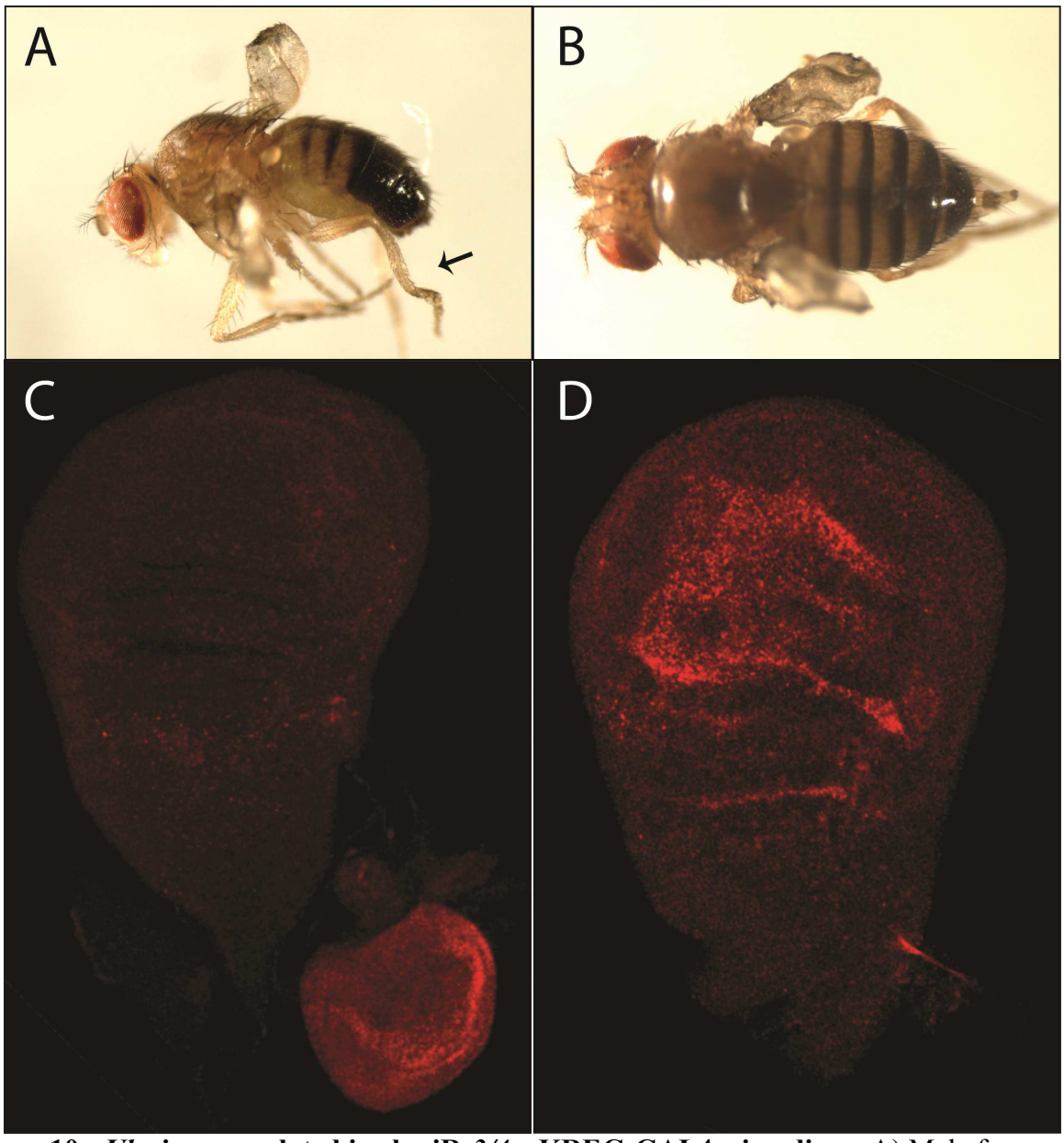


Figure 10 - *Ubx* is upregulated in shmiR-*3/4* x KREG-GAL4 wing discs A) Male from mutant line of $Ubx^{Cbx\text{-}1}$ flies that have inappropriate expression of *Ubx* in the wing discs. Arrow points to curling of the 3rd leg B) Female $Ubx^{Cbx\text{-}1}$ fly C) *Ubx* antibody stain on wing disc and haltere disc from NS-shmiR x KREG-GAL4 3rd instar larvae. *Ubx* protein stain is seen in the haltere disc but not in the wing disc D) *Ubx* protein stain shows *Ubx* expression in wing disc from shmiR-3/4 x KREG-GAL4 3rd instar larvae

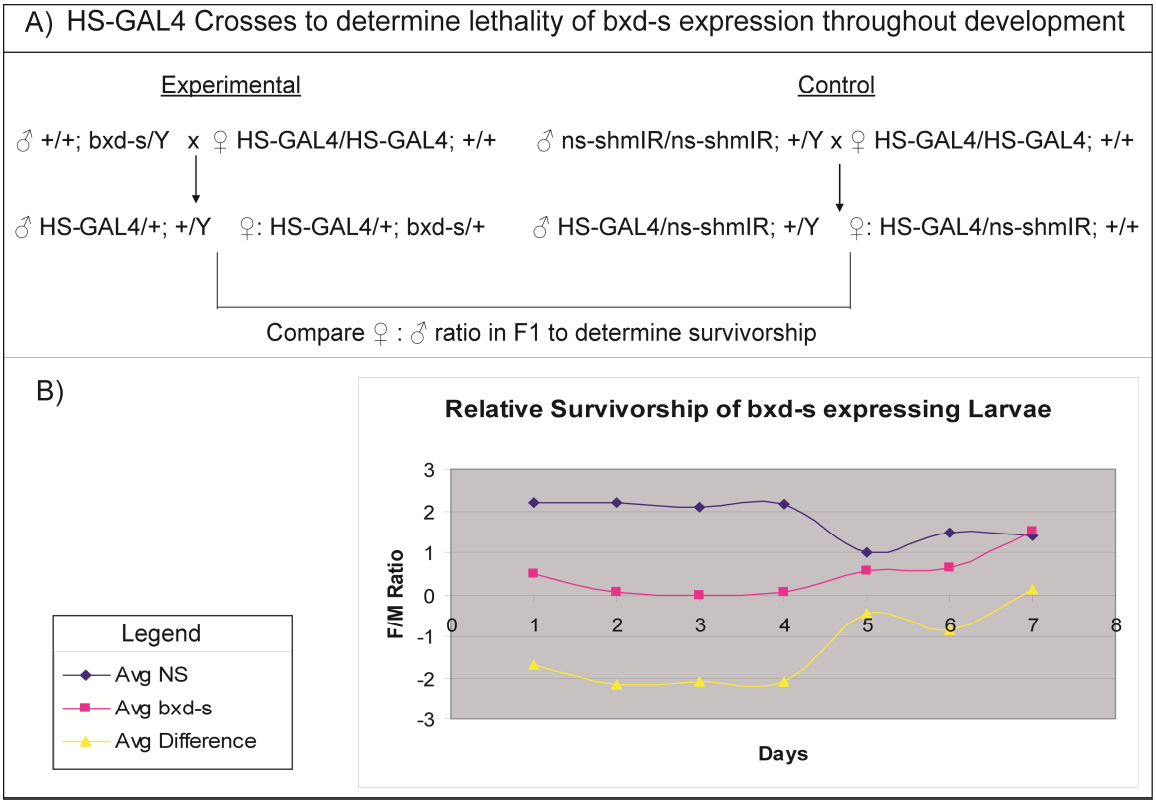


Figure 11 - *bxd* shmiR flies show lethality when induced ubiquitously during early development A) Cross scheme showing crosses performed using shmiR-1/2 and hsp70-GAL4 flies to test lethality throughout development B) Relative survivorship of flies containing both shmiR-1/2 and hsp70-GAL4 construct compared to NS-shmiR x hsp70-GAL4 flies shows lethality of *bxd* knockdown during first four days of development

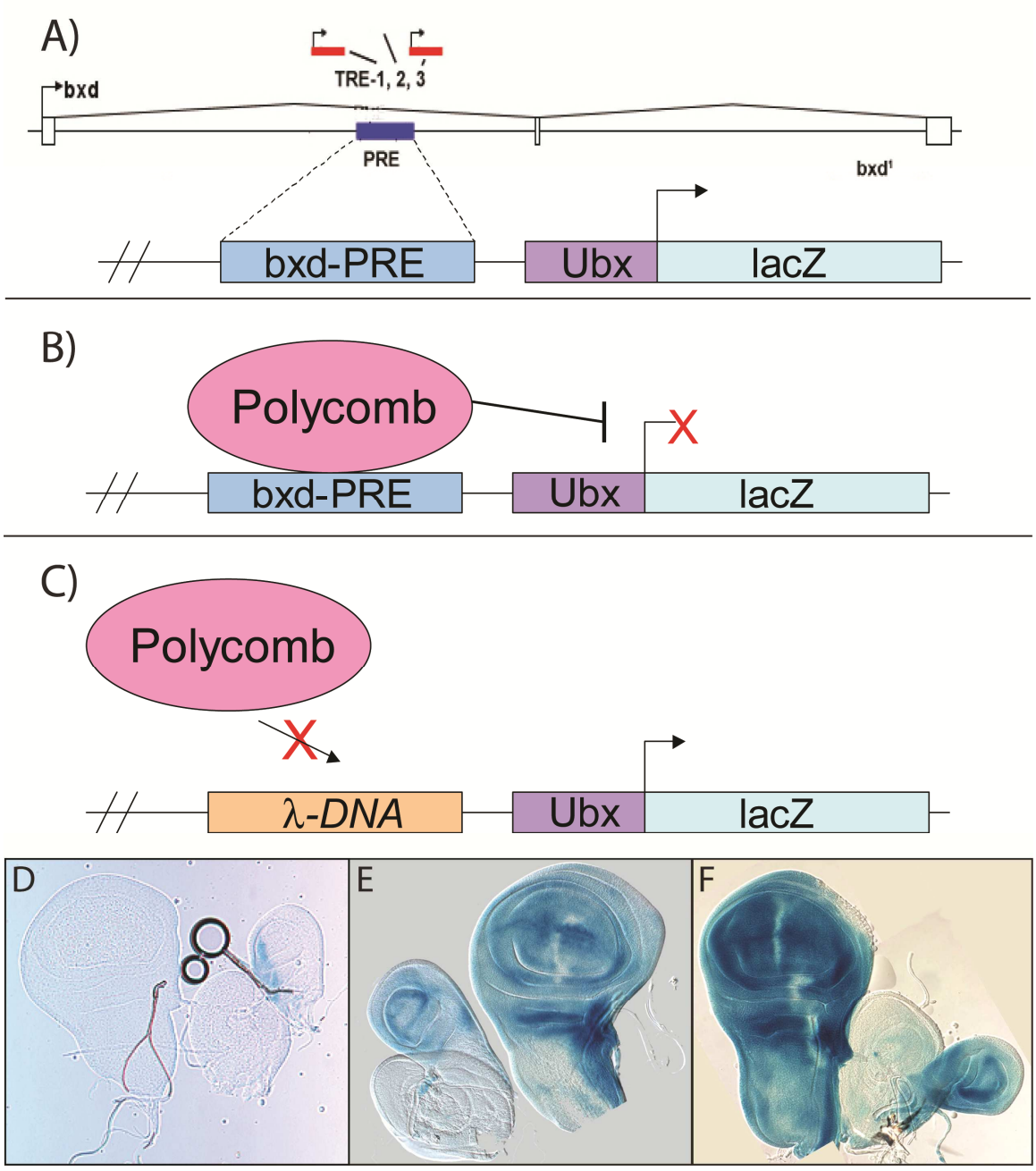
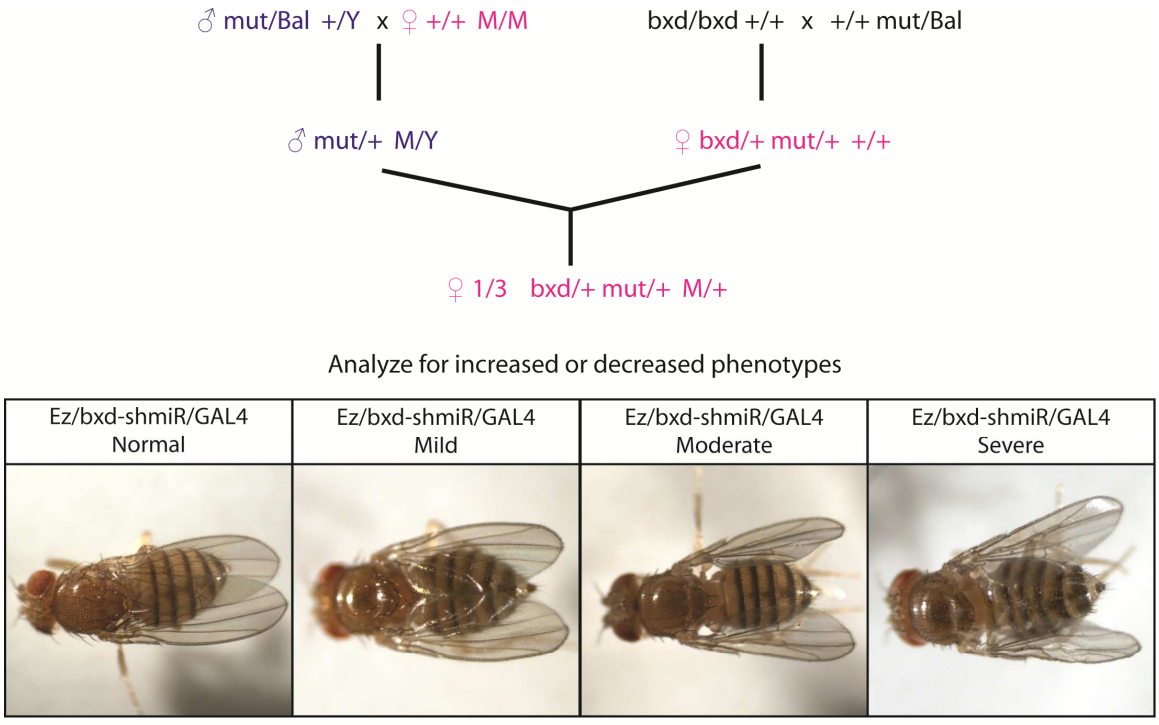


Figure 12 - *bxd* knockdown shows interruption of Pc binding at *bxd*-PRE This experiment tests for a role for *bxd* transcripts in normal Pc binding at the *bxd*-PRE A) A construct containing the *bxd*-PRE inserted upstream of a minimal *Ubx* promoter driving lacZ B) Flies containing the construct in A) will have Pc binding the PRE in wing discs, shutting off lacZ expression, producing no lacZ staining as seen in D) C) A control construct in which the *bxd*-PRE is replaced with random sequence will not have Pc binding, allowing lacZ expression in the wing discs as seen in E) F) Flies containing the *bxd*-PRE-*Ubx*-lacZ construct in a shmiR-1/2 x MS1096-GAL4 background show lacZ staining in the wing discs similar to E)



Ez/bxd-shmiR/GAL4 Phenotypic Ratios		Bal/bxd-shmiR/GAL4 Phenotypic Ratios		Ez/NS-shmiR/GAL4 Phenotypic Ratios	
Phenotype	%	Phenotype	%	Phenotype	%
Severe	20	Severe	0	Severe	0
Moderate	17	Moderate	5	Moderate	0
Mild	13	Mild	25	Mild	0
Normal	50	Normal	70	Normal	93
n =	150	n =	86	n =	93

Figure 13 - *E(z)* Genetic Mutants compound phenotype of *bxd*-shmiRs Crosses were done using shmiR-1/2, MS1096-GAL4, and an *E(z)* mutant containing a P-element insertion within the *E(z)* reading frame (Bellen 2009). The resultant F2 progeny would have females with a mixture of genotypes, including 1/3 of which would have all three constructs. Females within this group showed a progressive range of wing phenotypes resembling the phenotype normally seen with shmiR-1/2 x MS1096-GAL4 crosses. These phenotypes are displayed in the control cross with the *bxd*-shmiR, GAL4 and the Balancer chromosome from the Ez^{27149} line that does not contain the *E(z)* mutation. *E(z)/bxd*-shmiR/GAL4 flies show an additional phenotypic category in which the wing mutation is significantly more severe. The control cross with NS-shmiR have progeny that are all phenotypically normal.

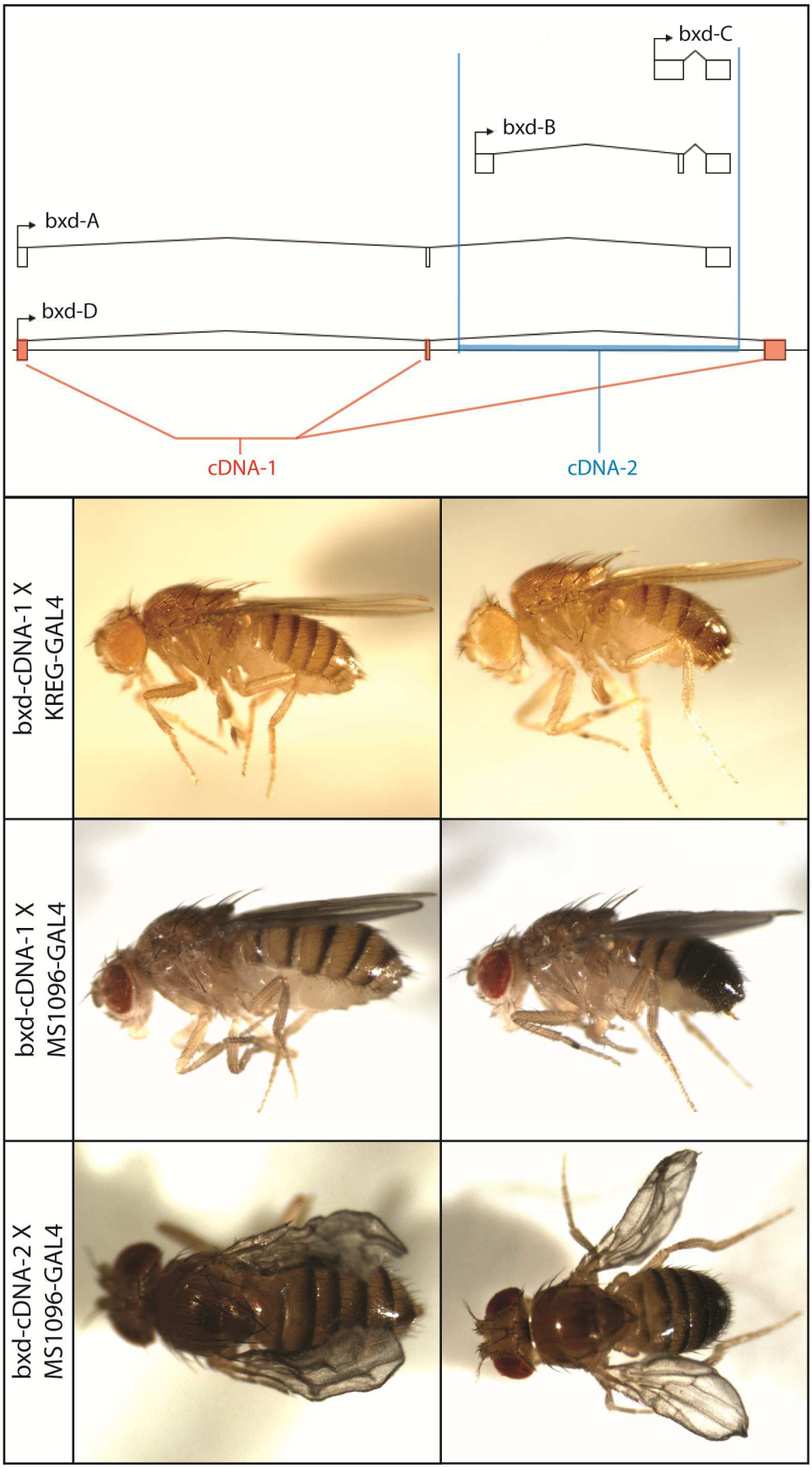

bxd-C
bxd-B
bxd-A
bxd-D
cDNA-1
cDNA-2
bxd-cDNA-1 X
KREG-GAL4
bxd-cDNA-1 X
MS1096-GAL4
bxd-cDNA-2 X
MS1096-GAL4

Figure 14 - Overexpression of *bxd* transcripts shows phenotypes in certain cases Two different *bxd* constructs were made for overexpression. *bxd*-cDNA-1 contained only the exons of *bxd-D* transcript, meaning that it was an actual cDNA. This construct showed no phenotypes when overexpressed in the wing and halteres using MS1096-GAL4 or early in the embryo using KREG-GAL4. *bxd*-cDNA-2 contained the entire genomic region surrounding *bxd-B* and *bxd-C* transcripts. When expressed using MS109-GAL4 it showed a significant wing phenotype, however, it is significantly different compared to the phenotypes seen with *bxd*-shmiR expression.

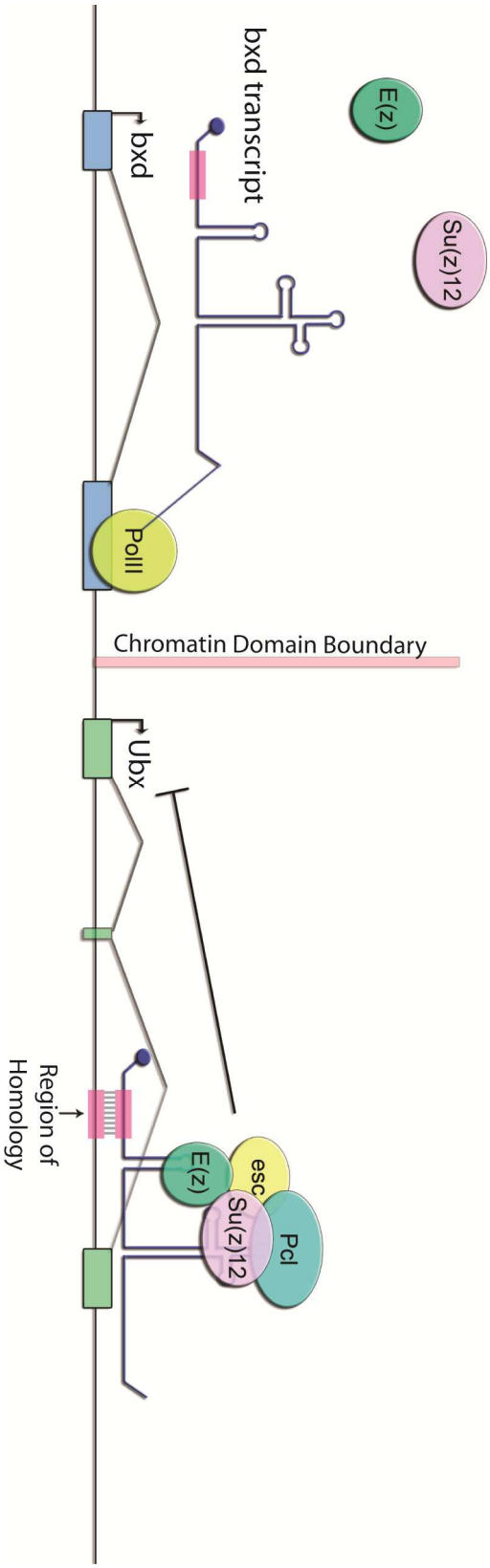

E(z)
Su(z)12
bxd transcript
bxd
PolII
Chromatin Domain Boundary
Ubx
Region of Homology
E(z)
esc
Su(z)12
Pcl

Figure 15 – Model for *bxd-B*ased repression of *Ubx* involving Pc Proteins In this model, *bxd* and *Ubx* are present within two separate chromatin domains as shown with the pink theoretical chromatin boundary element. Upon transcription, it is supposed that the *bxd* transcript(s) fold into a particular secondary structure which acts as a binding and assembly site for E(z) and Su(z)12 proteins. Pcl and esc proteins are recruited to assemble with the E(z) and Su(z)12 proteins, forming an active complex. The *bxd* transcript is targeted to the *Ubx* gene and associates with the region of homology that they both share. The Pc complex then acts to shut off the domain through chromatin condensation.

Chapter III:

Novel Non-coding RNAs in the Bithorax Complex

Introduction

The Bithorax Complex exists as a massive 320 kb of highly important sequence in the *Drosophila* Genome. Despite its enormity, the complex contains only three confirmed protein-coding genes: *Abdominal-B (Abd-B), abdominal-A (abd-A)* and *Ultrabithorax* (*Ubx*). It has long been thought that the majority of this non-coding sequence contained a complex web of enhancers and other *cis*-regulatory information. While many of these regulatory units have in fact been described, other mysterious features of this complex, such as the significant amount of non-coding transcription that exists, are only being appreciated in this modern era of RNA research. Many distinct and stable transcripts have been discovered over the years, including *bxd*, *iab-4*, and *Fab-7* (Bae et al. 2002). However, until the advent of whole genome tiling arrays it was impossible to gauge the full extent of the promiscuous transcription. Biemar et al was the first to produce a comprehensive study of the transcribed areas of the whole *Drosophila* genome during early embryonic development (Biemar et al. 2005; Biemar et al. 2006). Its revelation, that the Bithorax Complex is riddled with uncharacterized transcription in both intergenic and intronic regions, suggests that many more important transcripts exist than were previously thought.

In this chapter, I discuss my exploits to identify and characterize novel non-coding RNAs within the Bithorax Complex utilizing the whole genome tiling array described in Biemar et al. Using additional tools such as the Vista Genome Browser (http://pipeline.lbl.gov/cgi-bin/gateway2) for sequence conservation analysis, I identified a previously uncharacterized non-coding transcript emerging from the 3' intron of *Ubx*.

Discovery of a Novel Transcript

By combing through the Bithorax Complex of the tiling array data and highlighting peaks of expression, I was able to identify new potential transcripts. Candidate peaks were then analyzed for sequence conservation between other related *Drosophilid* species, such as *Drosophila pseudobscura*. Peaks displaying good conservation and expression potential were cloned and converted into RNA probes. Using these probes, I performed RNA FISH on 2-4 hr yw *Drosophila* embryos and looked for novel expression patterns. One ~500 bp region located within the 3' intron of *Ubx*, dubbed *750*, produced a unique expression pattern consisting of seven stripes (Fig. 17) appearing at ~2 hrs of embryonic development. The posterior four stripes of this transcript overlap with *Ubx* expression, which would be expected since it resides within a *Ubx* intron (Fig. 16). The anterior stripes, however, never appear as part of any known *Ubx* transcript since *Ubx* is not expressed anterior to parasegment 5. *750* expression continues into later stages but eventually weakens in expression and blends into normal *Ubx* staining (Fig. 17).

Several other early embryonic genes display a stripe pattern similar to the *750* pattern, particularly the pair-rule gene *ftz*. To determine whether these were overlapping expression patterns, I performed a triple *in situ* with *750* and *Ubx* probes, and Ftz antibody (Fig. 18). While *Ubx* and *750* have limited overlapping expression, *750* and Ftz contain no overlapping stripes and their expression patterns are shifted by several nuclei. Additionally, there is no significant

sequence similarity that exists between the *750* region and any of the pair rule genes, nor any other location in the genome.

Effects in Genetic Mutants

To further understand the nature of *750* expression, *750 in situs* were performed in mutant genetic backgrounds and analyzed for responses to possible regulatory inputs. In a Kr^2 mutant, one with decreased *Kr* expression, *750* exhibits a significant decrease in intensity in the third stripe, within the *Kr* expression domain indicating that *Kr* functions upstream of *750* to promote expression within this segment (Fig. 19). Additionally, in a ftz^{11} background in which *ftz* expression is decreased, *750* stripes are squished together, especially the posterior four that overlap *Ubx* expression. However, the stripes are not abolished or decreased in intensity as would be expected if these in fact were Ftz stripes. Finally, the *750* stripes also respond to a Ubx^{abx} mutant which would not be expected if these stripes were due to other pair-rule genes since this class of genes is upstream of Hox gene effects. In this particular mutant, *Ubx* is downregulated so that heterozygotes show haltere-to-wing defects when combined with other *Ubx* mutations. The *Ubx* pattern in the 2-4 hr embryo responds with a decreased intensity in the posterior stripes and a more compressed expression pattern. *750* mirrors this change and is also missing an anterior stripe, suggesting that the *abx* enhancer is somehow responsible for its expression pattern.

Characterization of the *750* Domain of Expression

Additional probing around the *750* region revealed that just a few kb up- or downstream of the region yielded expression patterns matching that of *Ubx* (Fig. 20). To get a better idea of the limits of the *750* transcript, 5'and 3' RACE (Rapid Amplification of Cdna Ends) was performed. The results suggested a ~3 kb region surrounding the original *750* probe was the extent of the transcript. A region encompassing this area (Construct 1, Fig. 20) was amplified and cloned into a UAS expression vector. It was randomly inserted into the *Drosophila* genome and crossed to various GAL4 driver lines to drive expression in the *Drosophila* embryo. However, no phenotype was ever produced suggesting that misexpression or overexpression in trans of this transcript does not cause any sort of genotypic misregulation or that the RACE analysis did not reveal the entirety of the transcript and that additional exons may exist elsewhere in the region.

Ubx 4.7 kb Non-coding Transcript Connection

One potential explanation for the existence of the *750* intronic expression pattern is that it is actually part of a larger non-coding transcript emanating from the *Ubx* promoter. This *4.7* kb non-coding transcript contains the 5' exon and an additional unique exon within the middle of the large 3' intron (Fig. 21). This exon is known to overlap the *abx* enhancer and is likely to overlap the *750* region as well.

This *4.7* kb non-coding *Ubx* transcript was first described by Akam et al in 1985 during the early days of *Ubx* expression characterization using *in situs* (Akam and Martinez-Arias 1985; Akam et al. 1985). It is described as a non-coding, spliced, non-polyadenylated transcript that remains localized to the nucleus. It does not share the 3' exon containing the homeobox. It is

expressed strongly in the early embryo between 3-6 hrs, and it exhibits a typical *Ubx* pattern with its highest intensity in parasegment 6. However, based on early reports, it is expected to lack expression in later stages of embryonic development. No function was ever discovered or seriously suggested, and its further characterization was ignored in favor of the other protein-coding transcripts.

In order to investigate the possible connection between my *750* transcript and *4.7*, I made my own probe to the purported extended region of the 5' exon that *4.7* encompasses (Fig. 21). Upon visualization with another *Ubx* probe, it appears that *4.7* expression and *Ubx* intronic expression completely overlap in both spatial and time characteristics (Fig. 21). Due to the lack of anterior stripes, it would suggest that the *750* stripes are not part of *4.7* expression and are in fact a unique entity. However, closer inspection to the ventral anterior region of the embryo suggests a hint of expression reminiscent of *750* stripes that is not shared by the intronic *Ubx* probe. It is possible that additional *in situs* or Northern analyses would reveal a link between these two transcripts.

To investigate this further, I made two sets of shmiR constructs against both the 5' exon of *4.7* and the *750* transcript location (Fig. 21). One construct (shmiR-5/6) contains both constructs and should target both exons of the *4.7* transcript. The other construct (shmiR-6) contains only the second shmiR that resides within the *750* region. Because these shmiRs, while in the intronic region, do share homology with pre-spliced protein-coding *Ubx* transcripts, it is possible that they may disrupt levels of *Ubx* protein in addition to the non-coding transcripts. Because of this reason, I decided to test these within the wing disc, since it is known that *Ubx* is not expressed there.

These knockdowns were performed by crossing the UAS-shmiRs with two wing disc GAL4 lines: MS1096 and A9. Both drivers are located on the X chromosome and therefore show increased expression in males compared to females due to upregulated expression of most male X-linked genes due to dosage compensation. When expressed in the wing discs, shmiR-5/6 produced flies with curled held-out wings in both an MS1096 and A9 background. Surprisingly, shmiR 2 also produced this phenotype only with the MS1096 driver, suggesting that this driver may have a stronger or more targeted GAL4 expression than A9. This weakened duplication of phenotype suggests that the two shmiRs are in fact targeting the same transcript and that both are needed for stronger targeting action. However, this does not necessarily indicate that *750* and *4.7* are one in the same transcript but that, at the very least, the *750* region is in fact located within the *4.7* 3' exon.

Additionally, these data suggest that *4.7* expression is not limited merely to early embryonic expression but does exist within the wing disc and possibly additional locations as well. The fact that no perturbation is seen when the shmiRs are expressed using an earlier driver, KREG-GAL4, suggests that the early expression is not *4.7*'s main time of action. To determine whether or not the shmiR 2 is actually affecting *750* expression, it would be useful to express the shmiR during early embryogenesis and perform an *in situ* to see whether or not *750* expression is diminished in this environment. However, this would be a difficult operation since a hazard of the GAL4 system is that there are no current GAL4 drivers that create strong early expression in the very early embryo. Additionally, shmiR expression and processing and target degradation would require an additional 1-3 hours after initial GAL4 expression, indicating that it would likely not be possible to visualize perturbations of the *750* expression pattern even with an adequate GAL4 driver.

750 Enhancer Activity Mirrors Expression Pattern

One peculiar hallmark of the Bithorax Complex is the preponderance of transcribed enhancers. The fact that they have unique and restricted expression patterns along the A/P axis like Hox genes suggests that they are more than mere spurious transcripts. There are many theories about the significance of this non-coding transcription, including opening up the chromatin for easier access to the promoter or recruiting factors sensitive to the transcription, however, for most of these moieties there is no known function. To delve more deeply into the potential role of the *750* region, I decided to test to see if this expressed fragment might also be a transcribed enhancer.

To do this, I cloned a 2.5 kb fragment surrounding the *750* probe into the E2G-lacZ reporter construct and inserted it into the *Drosophila* genome. I then performed an *in situ* for lacZ expression on early embryos. Amazingly, this fragment drives expression in a pattern remarkably similar to the *750 in situ* pattern in the early embryo (Fig. 22 and 23). In stage 4 expression, the fragment drives seven stripes of expression reminiscent of the *750 in situ* in spatiotemporal characteristics. In fact, even the middle stripe within parasegment 6 that would overlap normal *Ubx* expression is the brightest, just as it is in the *750* transcript and *Ubx* expression proper. As the enhancer progresses through the next stages of development, it becomes more like the typical *Ubx* expression pattern, just as *750* does.

While it is not surprising that this *750* fragment has enhancer activity since it likely overlaps the *abx* enhancer, it is quite amazing that it drives expression in a strikingly similar pattern to its expressed transcript, much like other known transcribed enhancers of the Bithorax Complex (Bae et al. 2002).. At the very least, this proves that *750* is not an artifact due to nonspecific probe interactions with a pair-rule gene. An interesting comparison to the hemizygote phenotype of *abx* mutants described by the Bender group states that the flies "hold their wings straight up, suggesting an effect on the flight muscles" (Peifer and Bender 1986). This description is reminiscent of the *4.7*/*750* shmiR phenotypes in which the flies have a distinct "held out" phenotype in which the wings stick up into the air and are not capable of being held into a normal horizontal position. If these phenotypes do in fact match, it would suggest that knocking down the transcripts that emanate from the *abx* region either somehow interferes with its enhancer ability or that the original *abx* mutation phenotypes ascribed to interference of enhancer function were in actuality due to the interruption of a then unknown transcript.

Previous studies of the pattern of the *abx* enhancer suggest that it displays a more traditional *Ubx* pattern with stripes of staining with an anterior limit in parasegment 5 (Simon et al. 1990). However, certain early staining patterns from the Bender group suggest a possible hint of striped expression anterior to this boundary. It is possible that these constructs did actually show a *750*-like pattern but early *in situ* techniques were not sensitive enough to fully characterize them. New stains with these historical lines might reveal further congruence with the *750* region.

Discussion

Presented here was evidence for a previously unknown mysterious transcribed region with enhancer activity within the large 3' exon of the Hox gene *Ubx*. Riddled with non-coding inter- and intragenic transcription, the Bithorax Complex likely contains many undiscovered regulatory mechanisms. Considering the level of importance of the Hox genes in overall

developmental control, their tight and accurate regulation must be given high priority in terms of energy and resource expenditure of the embryo. With this in mind, the Bithorax Complex most likely employs a diverse set of regulatory mechanisms to not only fine-tune expression patterns and levels but also to serve as fail-safes for the main regulatory players.

The bulk of study of the Bithorax Complex was performed in an earlier period of *Drosophila* and biological research. Molecular techniques were generally crude and lacking the utmost sensitivity needed to detect and analyze much of the unusual characteristics of this locus. Additionally, the mindset of the period was one in which RNA and non-coding transcripts were overlooked in favor of more tractable and "important" protein-coding genes and *cis*-regulatory mechanisms. With the advancement of molecular techniques and a resurgence in the diverse field of functional RNAs, it is now time to take another look at the Bithorax Complex to identify its secrets that were originally missed.

Indeed, in the past few years alone, several groups have identified functions for known non-coding transcripts in the Bithorax Complex. The *iab* enhancers of the *Abd-B* gene were shown to express non-coding transcripts in the same segment as their known phenotypic action (Bae et al. 2002). Additionally, Aravin et al. characterized two miRNAs located in the *iab-4* region, which they referred to as mir-iab-4-5p and mir-iab-4-3p, denoting the orientation of each hairpin (Aravin et al. 2003). These miRNAs are derived from the primary transcript, pri-mir-*iab-4*, which was previously referred to as *iab-4* (Cumberledge et al. 1990). A 2005 paper from the Levine lab that I contributed to established that *Ubx* is a target of miR-iab4-5p and that it directly affects *Ubx* expression *in vivo* (Ronshaugen et al. 2005). In fact, ectopic expression of this miRNA attenuates endogenous *Ubx* protein accumulation and induces a weak classical homeotic mutant phenotype: the transformation of halteres into wings. Additionally, other groups more recently identified a second miRNA produced from antisense transcription of the *iab-4* locus, called miR-iab-8 (Bender 2008; Tyler et al. 2008). Ectopic expression of this miR causes repression of both *Ubx* and *abd-A* and results in stronger homeotic transformations than those caused by mir-iab-5p (Stark et al. 2008; Tyler et al. 2008). microRNAs of this nature in the Bithorax Complex are suspected to play an endogenous role of tightening borders of Hox gene expression, regulating levels and functioning as a fail-safe for aberrant regulation of primary regulatory mechanisms (Ronshaugen et al. 2005). Indeed, this latter model has the most evidence so far, since it has been shown that deletion of individual microRNAs does not usually produce a phenotype but that deletion of microRNA clusters does (Miska et al. 2007). The Horvitz group suggests that this is due to functional redundancy in many microRNAs and that it is very likely that targets have multiple microRNAs (2007).

These recent exploits indicate that there is much more room for discovery of additional regulatory mechanisms for non-coding transcripts in the Bithorax Complex. My data contribute to this concept and suggest that the *750* transcript within the *Ubx* intron is a novel functional RNA involved in homeotic regulation. While many groups have studied the enhancer activity associated with this *Ubx* intron, until recently, methods to easily untangle primary *Ubx* transcription from other RNAs in intronic regions did not exist. While using a whole genome tiling array allowed me to identify this peak of transcription, current methods using whole genome Solexa sequencing techniques would yield even further information about the limits of the transcript and further potential exonic pieces connected to it. Despite these limitations, the pool of evidence that I have collected characterizing this mysterious *750* fragment, combined with information gleaned from previous papers, leads me to believe that I have identified non-coding functional transcription of the historical *Ubx* enhancer, *anterobithorax* or *abx*.

anterobithorax, or *abx*, was first discovered by Ed Lewis along with several other homeotic mutations dubbed *bx*, *pbx*, c*bx* and *Ubx* (Lewis 1978). He believed them each to be a transacting substance that worked together in an enzymatic series that correctly governed organ patterning of individual parasegments within the abdomen. However, later work later revealed these to be not individual genes, but separate *cis*-acting regulatory regions that worked together to correctly pattern the one trans-acting substance, *Ubx* (Peifer and Bender 1986). More in depth investigation of their individual functions using mutants showed that the *abx*/*bx* mutations define a regulatory region that programs the spatial distribution of *Ubx* protein in parasegment 5 (White and Wilcox 1985). *abx* mutants show a general patterning defect in parasegment 5 identity, usually involving partial haltere-to-wing transformations (Peifer and Bender 1986)). In order to visualize the enhancer activity, Simon et al. dissected the *abx* locus into various fragments and tested their activity using a *Ubx*-lacZ construct (Simon et al. 1990). They found that a majority of the fragments showed patterns similar to normal *Ubx in situ* patterns, with striped staining in the posterior half of the embryo beginning with parasegment 5. Because of its prominent staining in parasegments 5 and 6 and the previously known mutation data, they suggested that *abx* was important for establishing the anterior boundary of *Ubx* expression within these two parasegments.

However, an oddity associated with a few of their fragments suggested that there were aspects of the *abx* region that were contradictory to their model. Certain fragments located within a particular region of the *abx* locus produced expression patterns that mimicked *Ubx* expression except that they drove expression of two fainter stripes anterior to parasegment 5. Further dissection of these fragments produced similar results. The hypothesis of this phenomenon is discussed in the paper, but no definitive explanation is reached and this artifact is forgotten throughout the rest of *abx* history.

These forgotten patterns, though faint, clearly resemble both the transcript and expression patterns that I discovered within the *750* region. Also, based on the crude maps of the *abx* mutants (Simon et al. 1990) the *750* transcript easily falls within the annotated *abx* mutations. Additionally, the fragment that I cloned surrounding the *750* transcript to test the expression pattern is included within some of the fragments Simon et al. found to drive the *abx* anterior stripes. If *750* is indeed a piece of the transcribed *abx* enhancer, this would be in line with the general trend of transcription patterns of enhancers within the Bithorax Complex reproducing their enhancer expression patterns (Bae et al. 2002). While there are many documented cases of transcribed enhancers in this locus, there is currently evidence for only a few of these transcripts having a known function (Ronshaugen et al. 2005). The *750* shmiR data presented here suggests that targeting this transcript for degradation with microRNAs produces a wing phenotype (Fig. 21), suggesting that this transcript has a function distinct from the *cis* regulation of its sequence. If this is the case, it would be at odds with most known *abx* phenotypes, which usually involve haltere-to-wing transformations (Saari and Bienz 1987). However, Peifer et al. did describe (not pictured) an *abx* mutant with wing effects similar to the phenotypes I observed with my shmiR lines (Peifer and Bender 1986). It is also possible that previous *abx* mutants were disrupting actions at other points or locations in development that I have not targeted with my shmiR, such as very early expression in the 2-6 hr embryo during which *abx* is known to be active (Simon et al. 1990). It is unknown as to why the *abx* enhancer might give expression patterns in regions of the embryo in which *Ubx* is not and should not be expressed. However, it is likely explained by the fact that several *cis*-regulatory elements of *Ubx* exist (*bx*, *pbx*, c*bx*…) and that further

interactions of these enhancers might contain repressor sites that when combined with the *abx* regulatory effects would repress inappropriate anterior *Ubx* expression.

RACE analysis of the *750* locus suggested that the limits of the transcript were within a ~3kb boundary, but misexpression of this segment produced no obvious phenotypic results (Fig. 20). This could be explained through several lines of reasoning. First of all, it is possible that to be effective, this *750* transcript might need to be transcribed within the endogenous locus in order to effectively open chromatin or recruit other regulatory factors. When expressed in a foreign locus, these effects might be drowned out by other nearby regulatory mechanisms or may just not cause any misregulation in the area in which the element landed. Secondly, it could be assumed that the 3kb construct did not represent the entirety of the transcript and that additional components may be necessary for an effect. The fact that targeting the transcript for knockdown produced a phenotype while overexpression did not supports this idea. To fully express a transcript one needs to identify its entirety while targeting it for degradation requires only a small piece. The RACE analysis may have identified a portion of the transcript, but it may have missed additional exonic regions located further from the *750* locus. Indeed, the discovery of the *4.7*kb non-coding *Ubx* transcript, an exon of which overlaps the *abx* region (Fig. 21), suggests that this may be the case. Further analyses of these transcripts through Northerns and Solexa sequencing of the early embryo are needed to disseminate this possibility.

At this time, there is no definitive model for the function of these non-coding transcripts, including *750*, associated with enhancers in the Bithorax Complex. The only ones to have an assigned function have contained microRNAs (Ronshaugen et al. 2005). However, thorough scans of the region through computational hairpin prediction analysis and small RNA Solexa sequencing libraries have identified no further likely microRNAs within the complex (personal communication with Ben Haley and David Hendrix of the Levine Lab). This would suggest that these non-coding transcripts might have a diversified set of functions, some of which have been proposed earlier in this discussion and some of which may have yet to be identified. It will take a revisit to the Bithorax Complex with modern methods and thinking to tackle these mysterious elements.

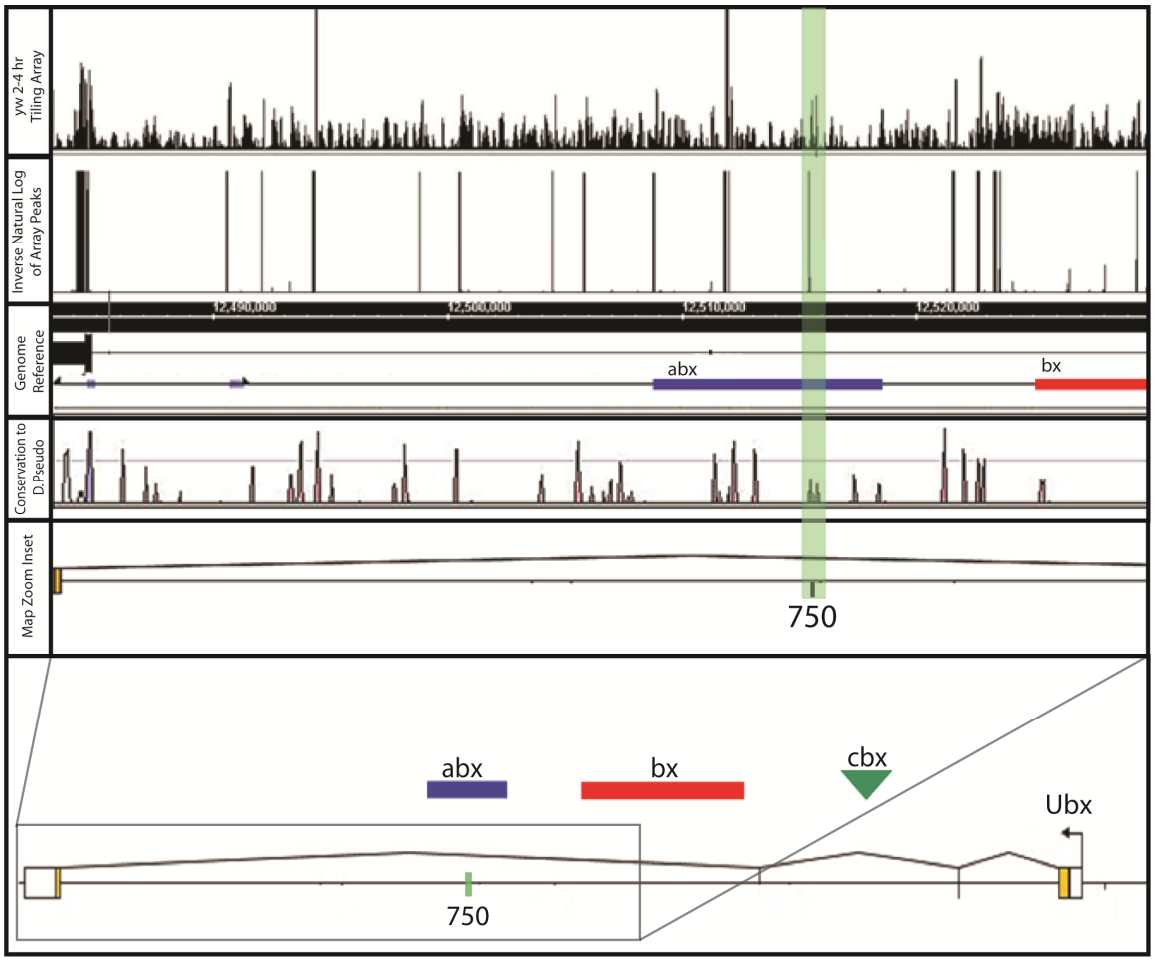


Figure 16 – Map, microarray and conservation of the *750* transcript Through use of the *Drosophila melanogaster* whole genome tiling array data using yw 2-4 hr RNA, a region of potential transcription was identified (Biemar et al. 2005; Biemar et al. 2006). Sequence conservation to *D. Pseudobscura* indicated relatively high conservation of a transcript named *750*. As shown in the map, it is thought to overlay the region for the known *Ubx* enhancer, *abx*.

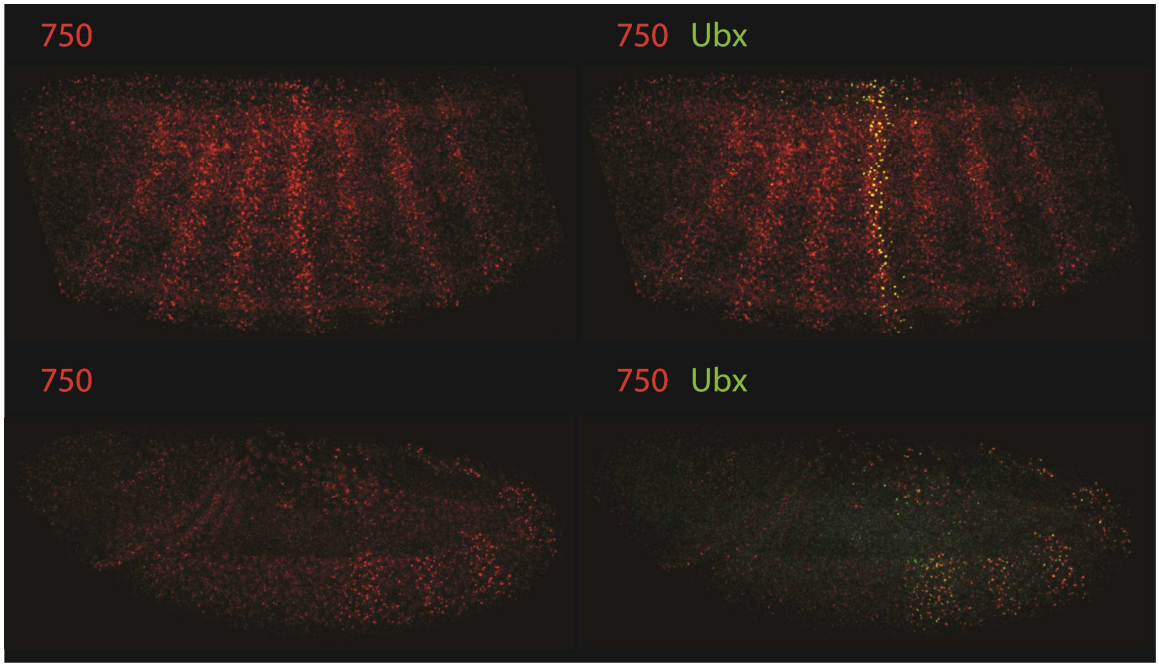


Figure 17 – *750* transcript shows *Ubx*-like pattern with additional anterior stripes An RNA probe was made to the region identified as *750*. *in situs* on 2-5 hr *D. melanogaster* embryos were done in concert with a probe to the 5' intron of *Ubx*. *750* shows overlapping posterior expression with *Ubx* but also shows three additional stripes anterior to *Ubx* expression.

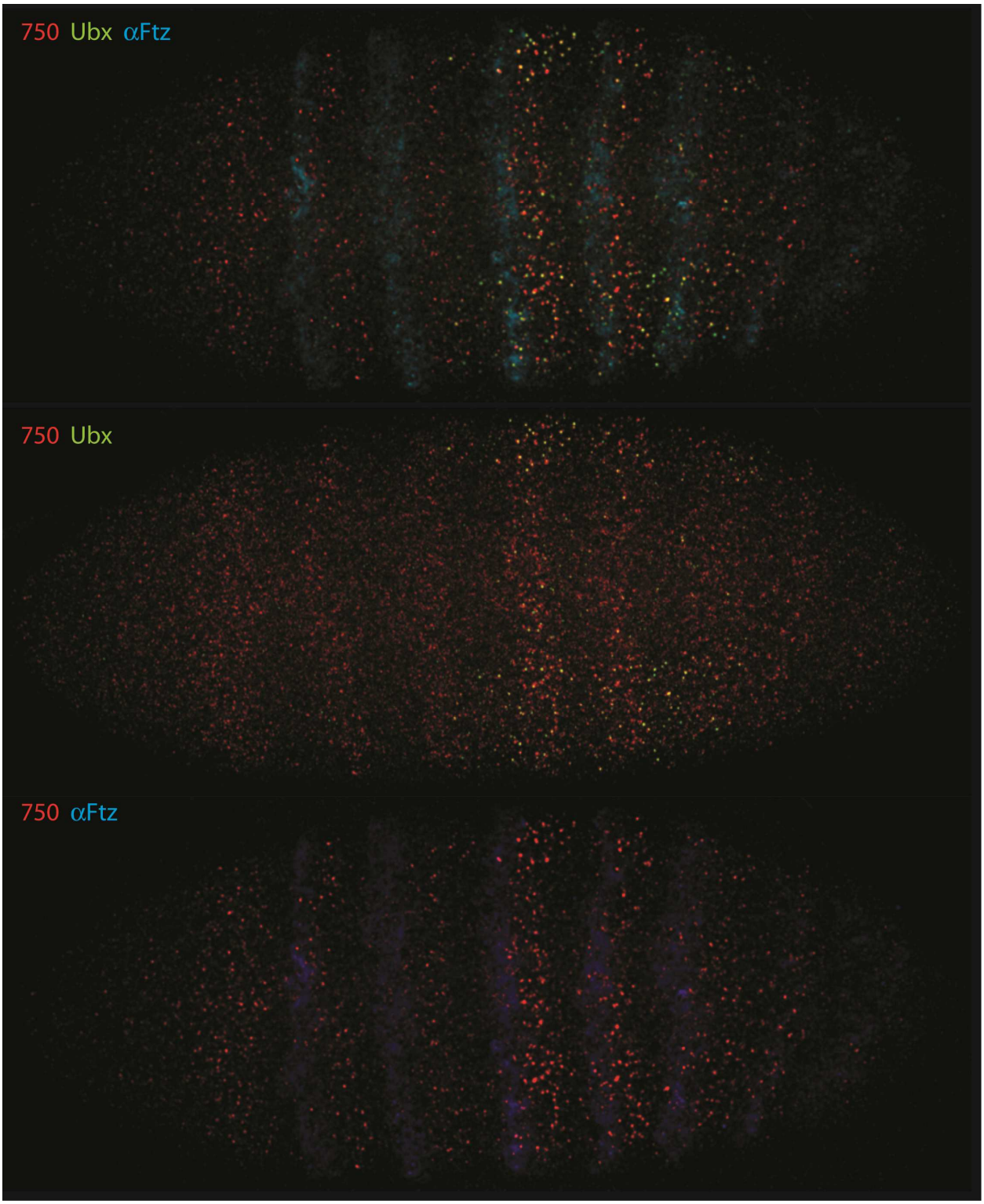


Figure 18 – *750* expression does not match that of Ftz staining An *in situ* with *750* (red) and *Ubx* (green) was performed in addition to an antibody stain for Ftz (blue) on 2-4 hr embryos. *750* and *Ubx* show overlapping expression in the posterior but Ftz is out of register with the other two probes.

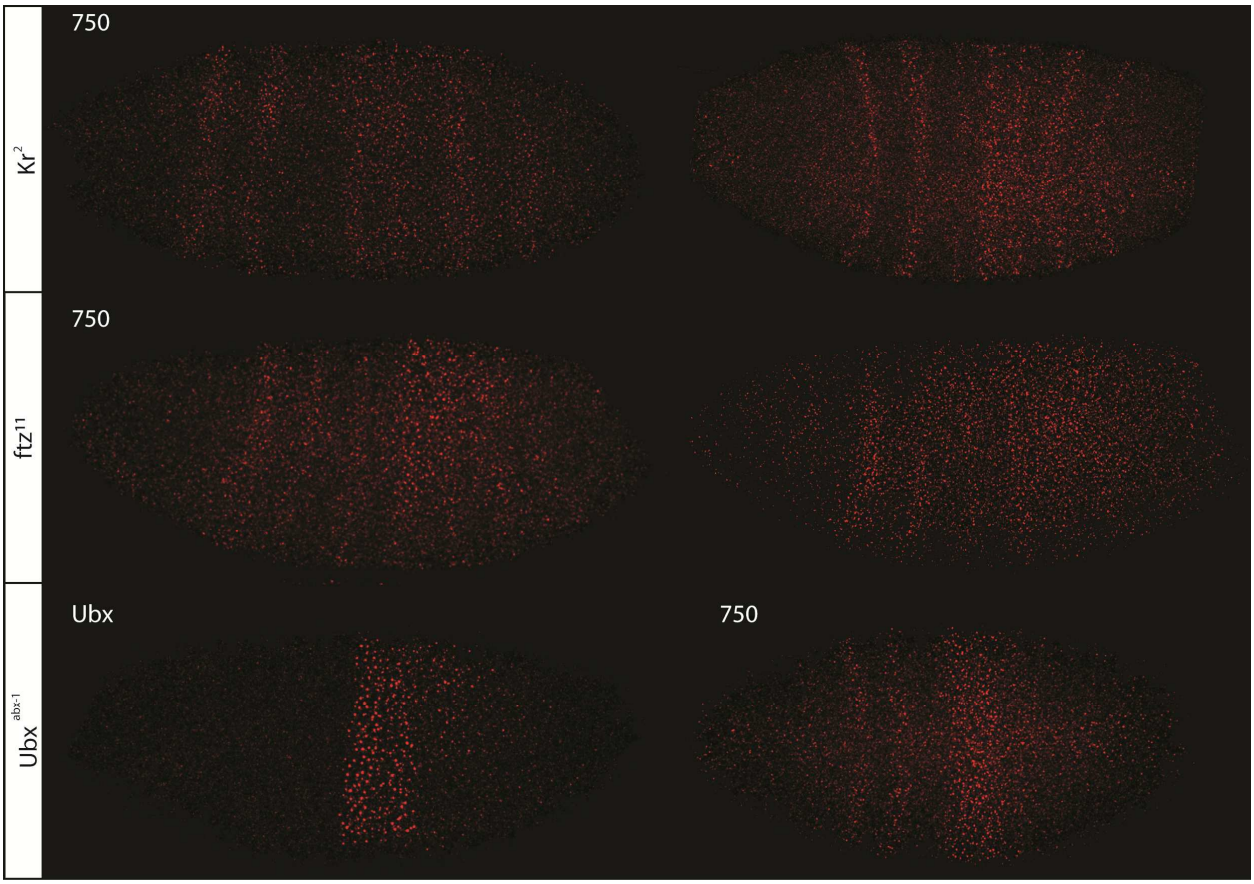


Figure 19 – i*n situ*s for *750* on genetic mutants *in situs* for *750* were performed on 2-4 hr embryos from three separate genetic backgrounds. *Kr*^{2} mutants show a loss or reduction of the third anterior stripes. *ftz*^{11} mutants show a compaction of stripes in the posterior. *Ubx*$^{abx\text{-}1}$ embryos show an upregulation of *Ubx* expression in parasegments 5 and 6 but decrease in expression posterior to this. *750* expression mimics this pattern with the loss or compaction of the third anterior stripe.

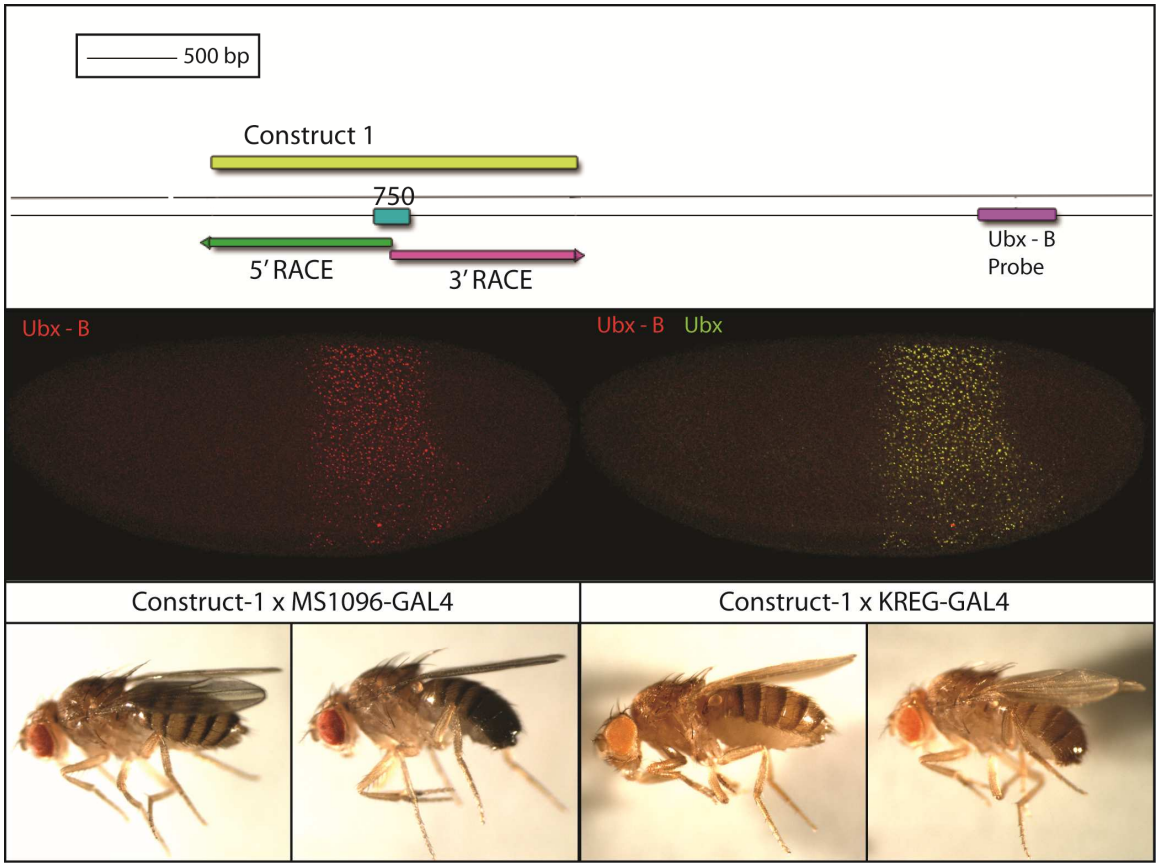


Figure 20 – *750* RACE shows potential exonic domains 5' and 3' RACE were performed on 2-4 hr embryos for the extent of *750* expression. Extension on either side suggests boundaries of the transcript. Construct-1 encompasses this region and was inserted into a pUAST expression construct. Overexpression of this construct using MS1096 and KREG-GAL4 lines produced no obvious phenotypes. A second probe to *Ubx* 3' intron several kb downstream of the *750* region shows only normal *Ubx* expression.

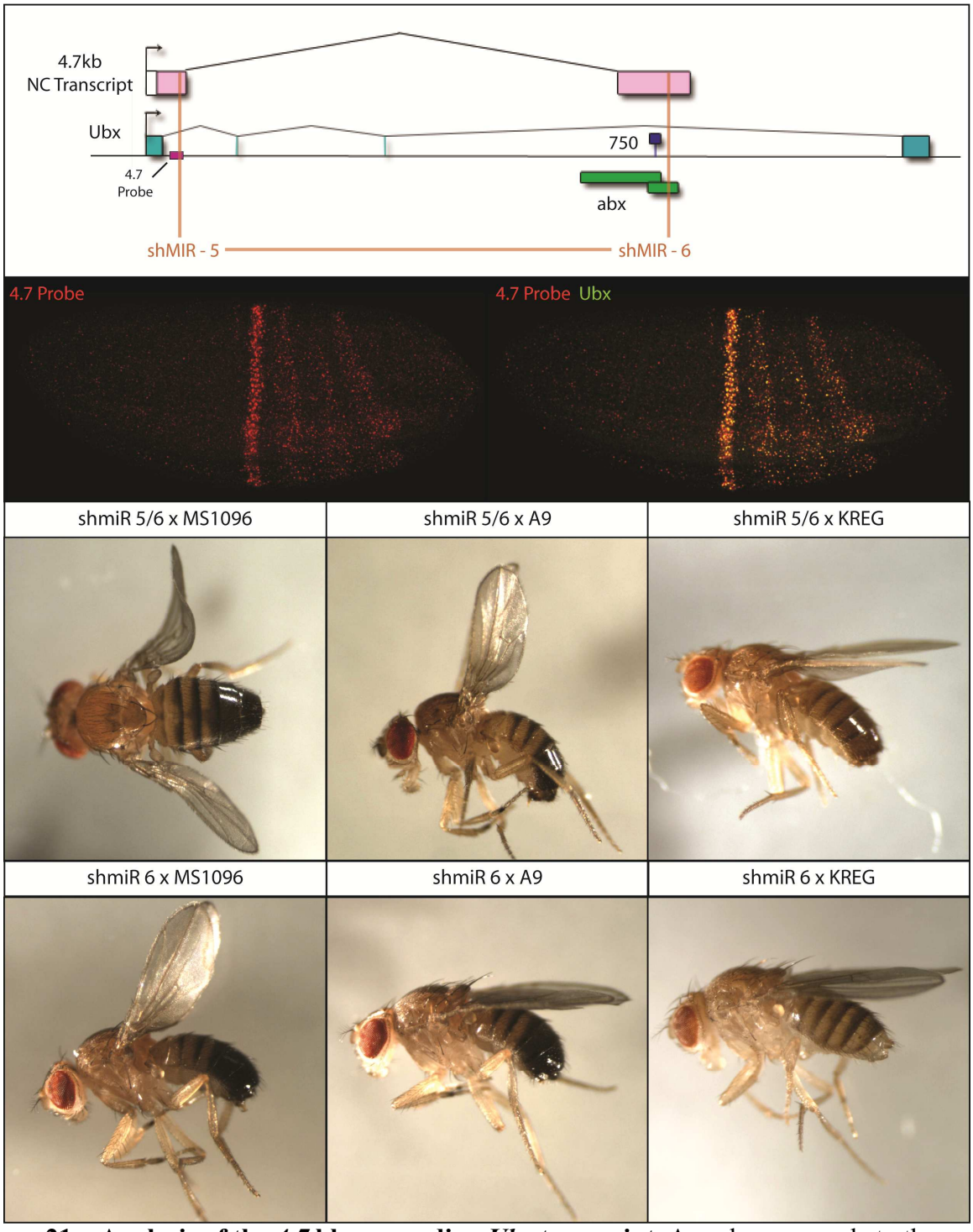


Figure 21 – Analysis of the *4.7* kb non-coding *Ubx* transcript A probe was made to the purported *4.7* kb non-coding transcript of the *Ubx* locus, whose 3' exon is thought to overlay the *750* region. Expression pattern shows mainly normal *Ubx* expression although a hint of anterior stripes is visible. Two shmiR constructs were made, one containing both shmiR-5 and -6 and one containing only shmiR-6. Expression of shmiR-5/6 in wing discs using MS1096 and A9-GAL4 shows a held-out wings phenotype that is not visible when crossed to KREG-GAL4. Expression of shmiR-6 shows this phenotype only when crossed to MS1096-GAL4.

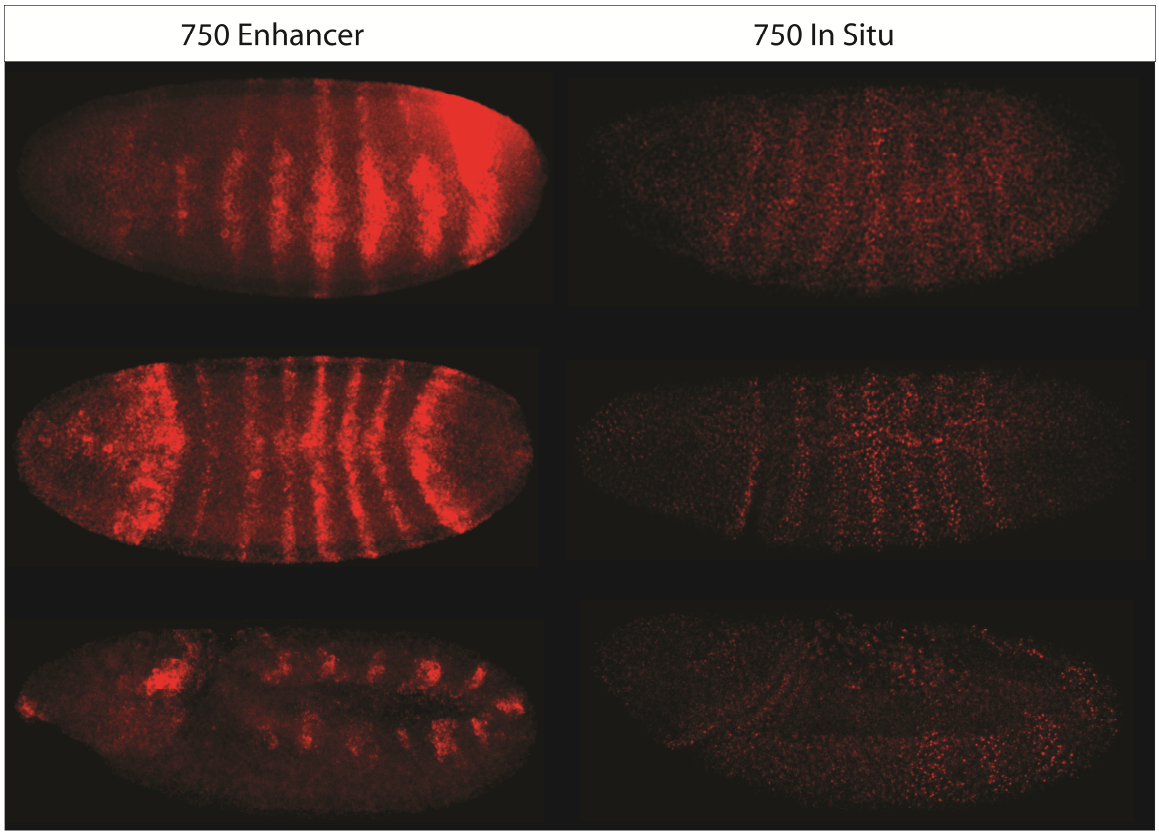


Figure 22 – *750* enhancer stain matches *750 in situ* expression A ~2 kb piece surrounding the *750* region was inserted into an enhancer reporter construct driving lacZ. An *in situ* done for lacZ expression in these 2-4 hr embryos shows a pattern very similar to that of the expression pattern for *750 in situ*s

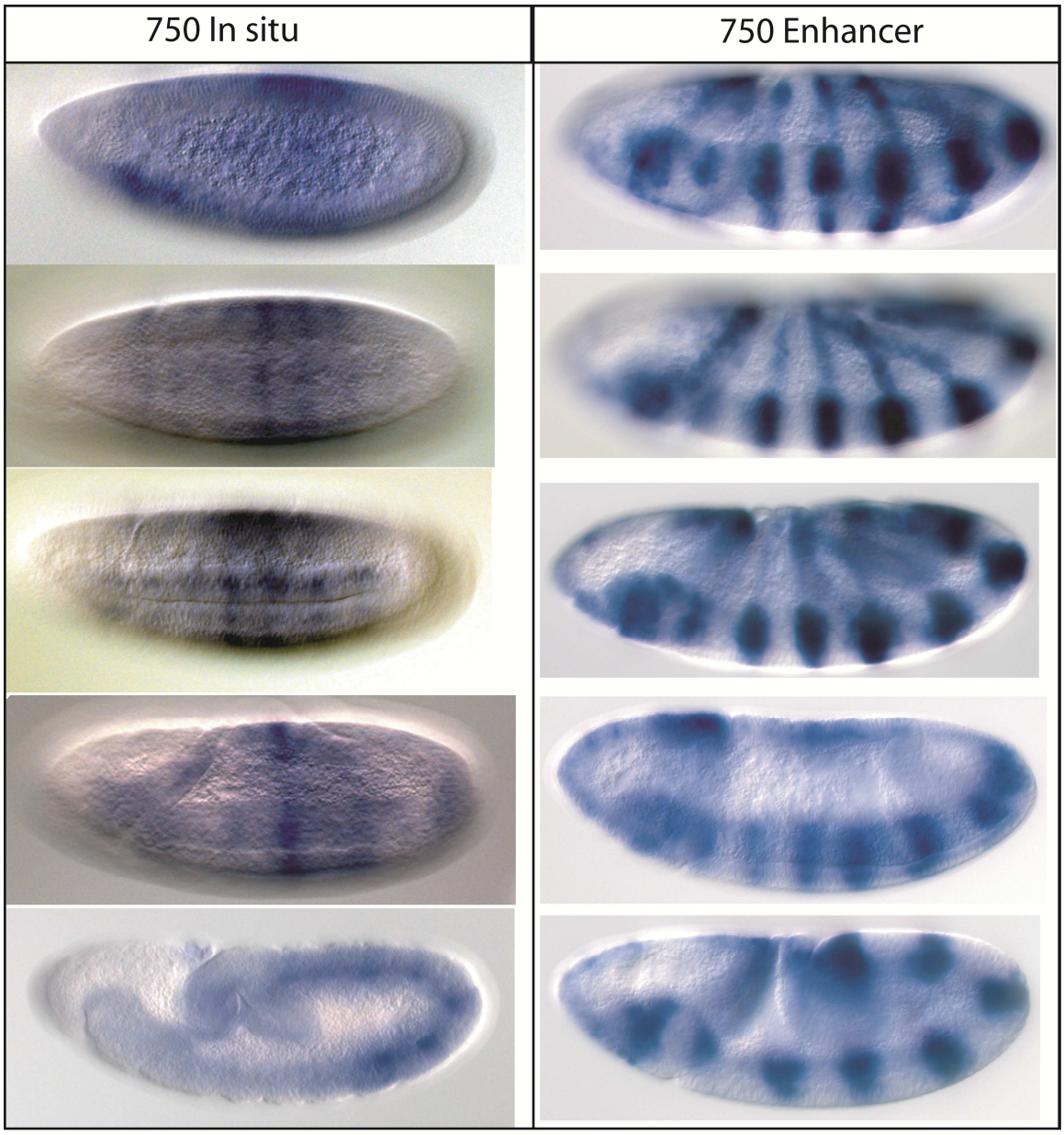


Figure 23 – Colorimetric *750* lacZ stain Additional colorimetric stains were performed on the embryos described in Fig. 22 above. Again, *750 in situ* expression and *750* enhancer expression show a similar pattern.

Chapter IV:

Concluding Remarks

It has become clear throughout the history of biological research that the only thing one can reliably expect in this field is the unexpected. Indeed, there were many important discoveries throughout the years, such as the structure of DNA, that many people thought heralded the end of the complex mysteries of life, and that further biological research would merely be "cleaning up the details." However, these displays of human hubris were usually soon proceeded with new and unexpected discoveries that illustrated just how little we yet knew about life on this planet. The discovery of the immense role of non-coding RNAs in the regulation of genes was one of those moments.

Although the sequencing of the human genome gave biologists new reasons to look beyond proteins as the source of organismal complexity, it has been only in the past few years that RNA has gained the respect it deserves for its regulatory abilities. Since then, small RNAs have become highly characterized throughout many different organisms, opening the door for researchers to delve into the vast unknown of long non-coding RNAs. Despite their protracted history, much less is known about these RNAs, which exist in many different forms and locations throughout the genome. Even throughout the heyday of the Bithorax Complex, non-coding RNAs were not given serious consideration in terms of function. Based on detailed genetic data, Ed Lewis predicted the existence of nine trans-acting units within the Bithorax Complex (1978). Once the sequence of the complex revealed the existence of only three protein-coding genes, however, it was widely regarded that the missing six elements were instead cis-regulatory modules. Despite the fact that non-coding RNAs were known to exist within the complex, there was no widespread historical precedent to indicate that these transcripts could match the impact of the protein-coding Hox genes and eventually the mysterious RNAs were pushed aside.

With the advent of new RNA-based technologies and a departure from the old protein-centric view of molecular biology, it is time to revisit the Bithorax Complex and peel apart its thick mesh of intercalary transcribed regions in order to discover the novel regulatory mechanisms that it likely holds. Hox genes are potent modulators of development and it is necessary to maintain a complex network of mechanisms to keep the Hox genes under tight regulatory control. Therefore, it is likely that the Bithorax Complex contains additional non-coding transcripts whose discovery would advance our understanding of regulatory mechanisms in gene regulation and possibly human disease. Indeed, the human HoxC non-coding regulatory transcript HOTAIR was recently found to be systematically dysregulated during breast cancer progression, leading to changes in chromatin states that encouraged metastasis (Gupta et al.). The data that I presented in this thesis regarding *bithoraxoid* and *750* suggest that additional discoveries are yet to be made. It is likely that our further exploration of long non-coding RNAs in the Bithorax Complex and throughout the genome may answer questions about gene regulation that we never even knew we had.

Chapter V:

Materials and Methods

Fly Stocks

The *Drosophila melanogaster* strain used for p-element mediated transgenesis was yw^{67}, as described previously (Stathopoulos and Levine 2002). All experiments involving p-elements were performed with a minimum of three independent lines to be certain of authenticity of phenotypes. PhiC mediated transgenesis was carried out using 86Fb as described by (Bischof et al. 2007). The *bxd*-PRE/*Ubx*/lacZ and NOPRE/*Ubx*/lacZ fly lines were a gift from J. Muller (Muller and Kassis 2006). All other stocks, including GAL4 lines and genetic mutants, were obtained from Bloomington.

Fly Crosses

pUAST over-expression constructs were balanced using the 612 double balancer (*yw; Sp/CyO; PrDr/TM3 Sb Ser*) and balanced or homozygous males were crossed to virgin females from the appropriate GAL4 line. Depending on the cross, crosses and subsequent development were carried out at 18°C, room temperature, 25°C, or 29°C.

Cloning and injection of transgenic constructs

DNA fragments were amplified from genomic DNA purified from yw^{67} flies. Fragments were amplified using the primers listed in Appendix IV. The *750*E enhancer fragment was cloned into the AttB equivalent of the reporter line nE2g (Eve minimal promoter, *lacZ* reporter) (Hare et al. 2008). The full length *bxd-D* cDNA was amplified from cDNA generated using a BD Marathon RACE kit from 2-4 hr yw^{67} embryos. All overexpression lines were cloned into pUAST (Brand and Perrimon 1993) or the AttB equivalent (Bischof et al. 2007). Constructs were introduced into the *Drosophila* germ line as previously described (Rubin and Spradling 1982; Groth et al. 2004).

qPCR

RNA was extracted from embryos or larvae by bleaching off the chorion and crushed using an automated pestle. Total RNA was isolated using the Trizol© reagent and method (Invitrogen). PolyA+ RNA was isolated using Poly-d(T) Dynabeads (Invitrogen). RT-PCR was performed using the SuperScript VILO kit (Invitrogen). qPCR amplification was done using the Roche CyberGreen Master Mix (Roche) on a Applied Biosystems RT 7000 machine. qPCR primers used are listed in Appendix IV.

Embryo fixation and staining

Drosophila embryo fixation and *in situ* hybridizations were carried out according to methods described previously (Kosman et al. 2004; Zinzen et al. 2006). All imaginal discs were isolated from 3rd instar climbing larvae in PBS and fixed and stained as previously described

(Ronshaugen et al. 2005). Ubx antibody was a gift from Akam (Akam et al. 1985). For RNA detection, embryos were hybridized with dioxygenin or dinitrophenyl (Perkin Elmer) labeled probes and visualized colorimetrically (Jiang et al. 1991) or fluorescently (Kosman et al. 2004) together with antibody to Ftz protein or nuclear lamin (Developmental Studies Hybridoma Bank). Probes were generated with the primers listed in Appendix IV and in vitro transcription using Sp6 or T7 reagents (Promega). Visualization of fluorescent *in situ*s was done using secondary antibodies conjugated to AlexaFluor Dyes (Molecular Probes) on a Leica SP2 confocal microscope. Image processing was carried out either in ImageJ or Imaris Software. Final color balance and contrast was adjusted in Photoshop.

Fly Pictures

Photos of flies were produced by gassing the flies on fly gas pads for ~10-20 mins. Photos were then taken on a brightfield dissecting scope using ScopePro imaging software.

shmiRs

shmiR sequences used are listed in Appendix IV. Insert sequences were generated using the Hairpin Design Tool listed in (Haley et al. 2008) and were ordered as oligos from IDT (www.idtdna.com). Inserts were cloned into the pHB and pNE3 vectors using the methods described in (Haley et al. 2008).

References

(2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011): 931-945.

Akam, M.E. and Martinez-Arias, A. (1985). The distribution of Ultrabithorax transcripts in Drosophila embryos. *Embo J* **4**(7): 1689-1700.

Akam, M.E., Martinez-Arias, A., Weinzierl, R., and Wilde, C.D. (1985). Function and expression of ultrabithorax in the Drosophila embryo. *Cold Spring Harb Symp Quant Biol* **50**: 195-200.

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

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. *Dev Cell* **5**(2): 337-350.

Arya, R., Mallik, M., and Lakhotia, S.C. (2007). Heat shock genes - integrating cell survival and death. *J Biosci* **32**(3): 595-610.

Bachellerie, J.P., Cavaille, J., and Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* **84**(8): 775-790.

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. *Proc Natl Acad Sci U S A* **99**(26): 16847-16852.

Bartel, B. (2005). MicroRNAs directing siRNA biogenesis. *Nat Struct Mol Biol* **12**(7): 569-571.

Beachy, P.A., Helfand, S.L., and Hogness, D.S. (1985). Segmental distribution of bithorax complex proteins during Drosophila development. *Nature* **313**(6003): 545-551.

Bellen, H. 2009. In (ed. Bloomington).

Bender, W. (2008). MicroRNAs in the Drosophila bithorax complex. *Genes Dev* **22**(1): 14-19.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B., and Hogness, D.S. (1983). Molecular Genetics of the Bithorax Complex in Drosophila melanogaster. *Science* **221**(4605): 23-29.

Berezikov, E. and Plasterk, R.H. (2005). Camels and zebrafish, viruses and cancer: a microRNA update. *Hum Mol Genet* **14 Spec No. 2**: R183-190.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M., and Snyder, M. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**(5705): 2242-2246.

Biemar, F., Nix, D.A., Piel, J., Peterson, B., Ronshaugen, M., Sementchenko, V., Bell, I., Manak, J.R., and Levine, M.S. (2006). Comprehensive identification of Drosophila dorsal-ventral patterning genes using a whole-genome tiling array. *Proc Natl Acad Sci U S A* **103**(34): 12763-12768.

Biemar, F., Zinzen, R., Ronshaugen, M., Sementchenko, V., Manak, J.R., and Levine, M.S. (2005). Spatial regulation of microRNA gene expression in the Drosophila embryo. *Proc Natl Acad Sci U S A* **102**(44): 15907-15911.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**(9): 3312-3317.

Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W.P., Weber, G., Lee, K., Fraioli, R.E., Cho, S.H., Yung, R., Asch, E., Ohno-Machado, L., Wong, W.H.,

and Cepko, C.L. (2004). Genomic analysis of mouse retinal development. *PLoS Biol* **2**(9): E247.

Brand, A.H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**(2): 401-415.

Bridges, C.B. and Morgan, T.H. (1923). The Third-Chromosome Group of Mutant Characters of Drosophila melanogaster. *The Carnegie Institution, Washington, DC*.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**(3): 527-542.

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**(11): 2163-2173.

Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E., Iorio, M.V., Visone, R., Sever, N.I., Fabbri, M., Iuliano, R., Palumbo, T., Pichiorri, F., Roldo, C., Garzon, R., Sevignani, C., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M., and Croce, C.M. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* **353**(17): 1793-1801.

Carninci, P. Kasukawa, T. Katayama, S. Gough, J. Frith, M.C. Maeda, N. Oyama, R. Ravasi, T. Lenhard, B. Wells, C. Kodzius, R. Shimokawa, K. Bajic, V.B. Brenner, S.E. Batalov, S. Forrest, A.R. Zavolan, M. Davis, M.J. Wilming, L.G. Aidinis, V. Allen, J.E. Ambesi-Impiombato, A. Apweiler, R. Aturaliya, R.N. Bailey, T.L. Bansal, M. Baxter, L. Beisel, K.W. Bersano, T. Bono, H. Chalk, A.M. Chiu, K.P. Choudhary, V. Christoffels, A. Clutterbuck, D.R. Crowe, M.L. Dalla, E. Dalrymple, B.P. de Bono, B. Della Gatta, G. di Bernardo, D. Down, T. Engstrom, P. Fagiolini, M. Faulkner, G. Fletcher, C.F. Fukushima, T. Furuno, M. Futaki, S. Gariboldi, M. Georgii-Hemming, P. Gingeras, T.R. Gojobori, T. Green, R.E. Gustincich, S. Harbers, M. Hayashi, Y. Hensch, T.K. Hirokawa, N. Hill, D. Huminiecki, L. Iacono, M. Ikeo, K. Iwama, A. Ishikawa, T. Jakt, M. Kanapin, A. Katoh, M. Kawasawa, Y. Kelso, J. Kitamura, H. Kitano, H. Kollias, G. Krishnan, S.P. Kruger, A. Kummerfeld, S.K. Kurochkin, I.V. Lareau, L.F. Lazarevic, D. Lipovich, L. Liu, J. Liuni, S. McWilliam, S. Madan Babu, M. Madera, M. Marchionni, L. Matsuda, H. Matsuzawa, S. Miki, H. Mignone, F. Miyake, S. Morris, K. Mottagui-Tabar, S. Mulder, N. Nakano, N. Nakauchi, H. Ng, P. Nilsson, R. Nishiguchi, S. Nishikawa, S. Nori, F. Ohara, O. Okazaki, Y. Orlando, V. Pang, K.C. Pavan, W.J. Pavesi, G. Pesole, G. Petrovsky, N. Piazza, S. Reed, J. Reid, J.F. Ring, B.Z. Ringwald, M. Rost, B. Ruan, Y. Salzberg, S.L. Sandelin, A. Schneider, C. Schonbach, C. Sekiguchi, K. Semple, C.A. Seno, S. Sessa, L. Sheng, Y. Shibata, Y. Shimada, H. Shimada, K. Silva, D. Sinclair, B. Sperling, S. Stupka, E. Sugiura, K. Sultana, R. Takenaka, Y. Taki, K. Tammoja, K. Tan, S.L. Tang, S. Taylor, M.S. Tegner, J. Teichmann, S.A. Ueda, H.R. van Nimwegen, E. Verardo, R. Wei, C.L. Yagi, K. Yamanishi, H. Zabarovsky, E. Zhu, S. Zimmer, A. Hide, W. Bult, C. Grimmond, S.M. Teasdale, R.D. Liu, E.T. Brusic, V. Quackenbush, J. Wahlestedt, C. Mattick, J.S. Hume, D.A. Kai, C. Sasaki, D. Tomaru, Y. Fukuda, S. Kanamori-Katayama, M. Suzuki, M. Aoki, J. Arakawa, T. Iida, J. Imamura, K. Itoh, M. Kato, T. Kawaji, H.

Kawagashira, N. Kawashima, T. Kojima, M. Kondo, S. Konno, H. Nakano, K. Ninomiya, N. Nishio, T. Okada, M. Plessy, C. Shibata, K. Shiraki, T. Suzuki, S. Tagami, M. Waki, K. Watahiki, A. Okamura-Oho, Y. Suzuki, H. Kawai, J. and Hayashizaki, Y. (2005). The transcriptional landscape of the mammalian genome. *Science* **309**(5740): 1559-1563.

Casares, F. and Sanchez-Herrero, E. (1995). Regulation of the infraabdominal regions of the bithorax complex of Drosophila by gap genes. *Development* **121**(6): 1855-1866.

Castelli-Gair, J.E., Micol, J.L., and Garcia-Bellido, A. (1990). Transvection in the Drosophila Ultrabithorax gene: a Cbx1 mutant allele induces ectopic expression of a normal allele in trans. *Genetics* **126**(1): 177-184.

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A* **97**(26): 14311-14316.

Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A., and Bachellerie, J.P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a non-coding RNA gene in rats. *J Biol Chem* **276**(28): 26374-26383.

Celniker, S.E., Sharma, S., Keelan, D.J., and Lewis, E.B. (1990). The molecular genetics of the bithorax complex of Drosophila: cis-regulation in the Abdominal-B domain. *Embo J* **9**(13): 4277-4286.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *Embo J* **13**(11): 2553-2564.

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z., and Rowley, J.D. (2004). Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* **32**(16): 4812-4820.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S., and Gingeras, T.R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**(5725): 1149-1154.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132**(3): 259-275.

Croft, L.J., Lercher, M.J., Gagen, M.J. and Mattick, J.S. . (2003). Is prokaryotic comoplexity limited by accelerated growth in regulatory overhead? *Genome Biology Preprint Depository*.

Cumberledge, S., Zaratzian, A., and Sakonju, S. (1990). Characterization of two RNAs transcribed from the cis-regulatory region of the abd-A domain within the Drosophila bithorax complex. *Proc Natl Acad Sci U S A* **87**(9): 3259-3263.

Dahary, D., Elroy-Stein, O., and Sorek, R. (2005). Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res* **15**(3): 364-368.

Dai, Z.H. and Chen, L.B. [The impact of microRNAs on the evolution of metazoan complexity]. *Yi Chuan* **32**(2): 105-114.

DiNardo, S., Heemskerk, J., Dougan, S., and O'Farrell, P.H. (1994). The making of a maggot: patterning the Drosophila embryonic epidermis. *Curr Opin Genet Dev* **4**(4): 529-534.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C., Sunkin, S.M., Crowe, M.L., Grimmond, S.M., Perkins, A.C., and Mattick, J.S. (2008a). Long non-coding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* **18**(9): 1433-1445.

Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008b). Differentiating protein-coding and non-coding RNA: challenges and ambiguities. *PLoS Comput Biol* **4**(11): e1000176.

Du, T. and Zamore, P.D. (2005). microPrimer: the biogenesis and function of microRNA. *Development* **132**(21): 4645-4652.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat Cell Biol* **10**(9): 1106-1113.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**(2): 188-200.

Engels, W.R. 1996. P-elements in Drosophila. in *Transposable Elements* (ed. H.S.a.A. Gierl), pp. 103-123, Berlin.

Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in Drosophila. *Nature* **332**(6167): 853-856.

Foronda, D., Estrada, B., de Navas, L., and Sanchez-Herrero, E. (2006). Requirement of Abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule. *Development* **133**(1): 117-127.

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.W., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. *Embo J* **11**(8): 2941-2950.

Frith, M.C., Pheasant, M., and Mattick, J.S. (2005). The amazing complexity of the human transcriptome. *Eur J Hum Genet* **13**(8): 894-897.

Fritsch, C., Brown, J.L., Kassis, J.A., and Muller, J. (1999). The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development* **126**(17): 3905-3913.

Galloni, M., Gyurkovics, H., Schedl, P., and Karch, F. (1993). The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. *Embo J* **12**(3): 1087-1097.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. *Ciba Found Symp* **0**(29): 161-182.

Garcia-Bellido, A., Ripoll, P., and Morata, G. (1976). Developmental compartmentalization in the dorsal mesothoracic disc of Drosophila. *Dev Biol* **48**(1): 132-147.

Giniger, E., Varnum, S.M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**(4): 767-774.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**(4): 1775-1782.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R.B., van de

Vijver, M.J., Sukumar, S., and Chang, H.Y. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**(7291): 1071-1076.

Haley, B., Hendrix, D., Trang, V., and Levine, M. (2008). A simplified miRNA-based gene silencing method for Drosophila melanogaster. *Dev Biol* **321**(2): 482-490.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008). Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. *PLoS Genet* **4**(6): e1000106.

Ho, M.C., Schiller, B.J., Goetz, S.E., and Drewell, R.A. (2009). Non-genic transcription at the Drosophila bithorax complex functional activity of the dark matter of the genome. *Int J Dev Biol* **53**(4): 459-468.

Hoch, M. and Jackle, H. (1993). Transcriptional regulation and spatial patterning in Drosophila. *Curr Opin Genet Dev* **3**(4): 566-573.

Hogness, D.S., Lipshitz, H.D., Beachy, P.A., Peattie, D.A., Saint, R.B., Goldschmidt-Clermont, M., Harte, P.J., Gavis, E.R., and Helfand, S.L. (1985). Regulation and products of the Ubx domain of the bithorax complex. *Cold Spring Harb Symp Quant Biol* **50**: 181-194.

Huttenhofer, A., Schattner, P., and Polacek, N. (2005). Non-coding RNAs: hope or hype? *Trends Genet* **21**(5): 289-297.

Ingham, P.W. (1988). The molecular genetics of embryonic pattern formation in Drosophila. *Nature* **335**(6185): 25-34.

Ingham, P.W. and Martinez Arias, A. (1992). Boundaries and fields in early embryos. *Cell* **68**(2): 221-235.

Irish, V.F., Martinez-Arias, A., and Akam, M. (1989). Spatial regulation of the Antennapedia and Ultrabithorax homeotic genes during Drosophila early development. *Embo J* **8**(5): 1527-1537.

Irvine, K.D., Helfand, S.L., and Hogness, D.S. (1991). The large upstream control region of the Drosophila homeotic gene Ultrabithorax. *Development* **111**(2): 407-424.

Jackson, A.L. and Linsley, P.S. (2004). Noise amidst the silence: off-target effects of siRNAs? *Trends Genet* **20**(11): 521-524.

Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early Drosophila embryos. *Genes Dev* **5**(10): 1881-1891.

Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J., and Vasicek, T.J. (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Res* **15**(7): 1007-1014.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. *J Biol Chem* **281**(39): 29064-29075.

Karch, F., Bender, W., and Weiffenbach, B. (1990). abdA expression in Drosophila embryos. *Genes Dev* **4**(9): 1573-1587.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., and

Wahlestedt, C. (2005). Antisense transcription in the mammalian transcriptome. *Science* **309**(5740): 1564-1566.
Kelsh, R., Weinzierl, R.O., White, R.A., and Akam, M. (1994). Homeotic gene expression in the locust Schistocerca: an antibody that detects conserved epitopes in Ultrabithorax and abdominal-A proteins. *Dev Genet* **15**(1): 19-31.
Kennison, J.A. (1993). Transcriptional activation of Drosophila homeotic genes from distant regulatory elements. *Trends Genet* **9**(3): 75-79.
Kishore, S. and Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**(5758): 230-232.
Kiyosawa, H., Mise, N., Iwase, S., Hayashizaki, Y., and Abe, K. (2005). Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. *Genome Res* **15**(4): 463-474.
Kornberg, T.B. and Tabata, T. (1993). Segmentation of the Drosophila embryo. *Curr Opin Genet Dev* **3**(4): 585-594.
Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. *Science* **305**(5685): 846.
Lee, J.T. and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99**(1): 47-57.
Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**(5): 843-854.
Levine, M. and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* **424**(6945): 147-151.
Lewis, E.B. (1963). Genes and developmental pathways. *Am Zool*(3): 33.
-. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**(5688): 565-570.
Little, J.W., Byrd, C.A., and Brower, D.L. (1990). Effect of abx, bx and pbx mutations on expression of homeotic genes in Drosophila larvae. *Genetics* **124**(4): 899-908.
Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., and Golub, T.R. (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**(7043): 834-838.
Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615-651.
Ma, J. and Ptashne, M. (1987a). The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. *Cell* **50**(1): 137-142.
-. (1987b). Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**(5): 847-853.
-. (1987c). A new class of yeast transcriptional activators. *Cell* **51**(1): 113-119.
Macias, A., Casanova, J., and Morata, G. (1990). Expression and regulation of the abd-A gene of Drosophila. *Development* **110**(4): 1197-1207.
Maden, B.E. and Hughes, J.M. (1997). Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma* **105**(7-8): 391-400.
Maggert, K.A., Gong, W.J., and Golic, K.G. (2008). Methods for homologous recombination in Drosophila. *Methods Mol Biol* **420**: 155-174.
Marques, A.C. and Ponting, C.P. (2009). Catalogues of mammalian long non-coding RNAs: modest conservation and incompleteness. *Genome Biol* **10**(11): R124.

Martin, C.H., Mayeda, C.A., Davis, C.A., Ericsson, C.L., Knafels, J.D., Mathog, D.R., Celniker, S.E., Lewis, E.B., and Palazzolo, M.J. (1995). Complete sequence of the bithorax complex of Drosophila. *Proc Natl Acad Sci U S A* **92**(18): 8398-8402.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* **2**(11): 986-991.

-. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**(10): 930-939.

Mattick, J.S. and Gagen, M.J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other non-coding RNAs in the development of complex organisms. *Mol Biol Evol* **18**(9): 1611-1630.

Mattick, J.S. and Makunin, I.V. (2005). Small regulatory RNAs in mammals. *Hum Mol Genet* **14 Spec No 1**: R121-132.

-. (2006). Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**: R17-29.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genet* **3**(12): e215.

Muller, J. and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. *Embo J* **10**(11): 3147-3155.

Muller, J. and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr Opin Genet Dev* **16**(5): 476-484.

Naguibneva, I., Ameyar-Zazoua, M., Polesskaya, A., Ait-Si-Ali, S., Groisman, R., Souidi, M., Cuvellier, S., and Harel-Bellan, A. (2006). The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nat Cell Biol* **8**(3): 278-284.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of non-coding RNAs: lack of conservation does not mean lack of function. *Trends Genet* **22**(1): 1-5.

Paro, R. (1990). Imprinting a determined state into the chromatin of Drosophila. *Trends Genet* **6**(12): 416-421.

Peifer, M. and Bender, W. (1986). The anterobithorax and bithorax mutations of the bithorax complex. *Embo J* **5**(9): 2293-2303.

Peifer, M., Karch, F., and Bender, W. (1987). The bithorax complex: Control of segmental identity. *Genes Dev*(1): 891-898.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* **379**(6561): 131-137.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd non-coding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* **127**(6): 1209-1221.

Petruk, S., Smith, S.T., Sedkov, Y., and Mazo, A. (2008). Association of trxG and PcG proteins with the bxd maintenance element depends on transcriptional activity. *Development* **135**(14): 2383-2390.

Pirotta, V. (1997). Chromatin-silencing mechanisms in Drosophila maintain patterns of gene expression. *Trens Genet*(13): 314-318.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**(5616): 131-135.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., Grimmond, S.M., Hume, D.A., Hayashizaki, Y., and Mattick, J.S. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* **16**(1): 11-19.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. *Cell* **129**(7): 1311-1323.

Rogelj, B. and Giese, K.P. (2004). Expression and function of brain specific small RNAs. *Rev Neurosci* **15**(3): 185-198.

Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E.C. (2005). The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev* **19**(24): 2947-2952.

Ronshaugen, M. and Levine, M. (2004). Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the Drosophila embryo. *Dev Cell* **7**(6): 925-932.

Rubin, G.M. and Spradling, A.C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**(4570): 348-353.

Saari, G. and Bienz, M. (1987). The structure of the Ultrabithorax promoter of Drosophila melanogaster. *Embo J* **6**(6): 1775-1779.

Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* **40**(6): 719-721.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Non-coding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. *Science* **311**(5764): 1118-1123.

Sanchez-Herrero, E. (1991). Control of the expression of the bithorax complex genes abdominal-A and abdominal-B by cis-regulatory regions in Drosophila embryos. *Development* **111**(2): 437-449.

Sanchez-Herrero, E. and Akam, M. (1989). Spatially ordered transcription of regulatory DNA in the bithorax complex of Drosophila. *Development* **107**(2): 321-329.

Sanchez-Herrero, E., Guerrero, I., Sampedro, J., and Gonzalez-Reyes, A. (1994). Developmental consequences of unrestricted expression of the abd-A gene of Drosophila. *Mech Dev* **46**(3): 153-167.

Sanchez-Herrero, E., Vernos, I., Marco, R., and Morata, G. (1985). Genetic organization of Drosophila bithorax complex. *Nature* **313**(5998): 108-113.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev* **19**(6): 697-708.

Schwabish, M.A. and Struhl, K. (2007). The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol Cell Biol* **27**(20): 6987-6995.

Schwartz, Y.B. and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* **8**(1): 9-22.

-. (2008). Polycomb complexes and epigenetic states. *Curr Opin Cell Biol* **20**(3): 266-273.

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during Drosophila development. *Curr Opin Cell Biol* **7**(3): 376-385.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev Biol* **158**(1): 131-144.

Simon, J., Peifer, M., Bender, W., and O'Connor, M. (1990). Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. *Embo J* **9**(12): 3945-3956.

Sipos, L., Kozma, G., Molnar, E., and Bender, W. (2007). In situ dissection of a Polycomb response element in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **104**(30): 12416-12421.

Stark, A., Bushati, N., Jan, C.H., Kheradpour, P., Hodges, E., Brennecke, J., Bartel, D.P., Cohen, S.M., and Kellis, M. (2008). A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands. *Genes Dev* **22**(1): 8-13.

Stathopoulos, A. and Levine, M. (2002). Linear signaling in the Toll-Dorsal pathway of Drosophila: activated Pelle kinase specifies all threshold outputs of gene expression while the bHLH protein Twist specifies a subset. *Development* **129**(14): 3411-3419.

Struhl, G. and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *Embo J* **4**(12): 3259-3264.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**(2): 103-105.

Tyler, D.M., Okamura, K., Chung, W.J., Hagen, J.W., Berezikov, E., Hannon, G.J., and Lai, E.C. (2008). Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. *Genes Dev* **22**(1): 26-36.

van Nimwegen, E. (2003). Scaling laws in the functional content of genomes. *Trends Genet* **19**(9): 479-484.

Walter, G. (1986). Origin of life: The RNA world. *Nature* **319**(618).

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., and Wong, G.K. (2004). Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature* **431**(7010): 1 p following 757; discussion following 757.

Werner, A. and Berdal, A. (2005). Natural antisense transcripts: sound or silence? *Physiol Genomics* **23**(2): 125-131.

White, R.A. and Wilcox, M. (1985). Distribution of Ultrabithorax proteins in Drosophila. *Embo J* **4**(8): 2035-2043.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. *Genes Dev* **5**(10): 1813-1824.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**(5): 855-862.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of non-coding RNAs finds a repressor of NFAT. *Science* **309**(5740): 1570-1573.

Yue, R., Kang, J., Zhao, C., Hu, W., Tang, Y., Liu, X., and Pei, G. (2009). Beta-arrestin1 regulates zebrafish hematopoiesis through binding to YY1 and relieving polycomb group repression. *Cell* **139**(3): 535-546.

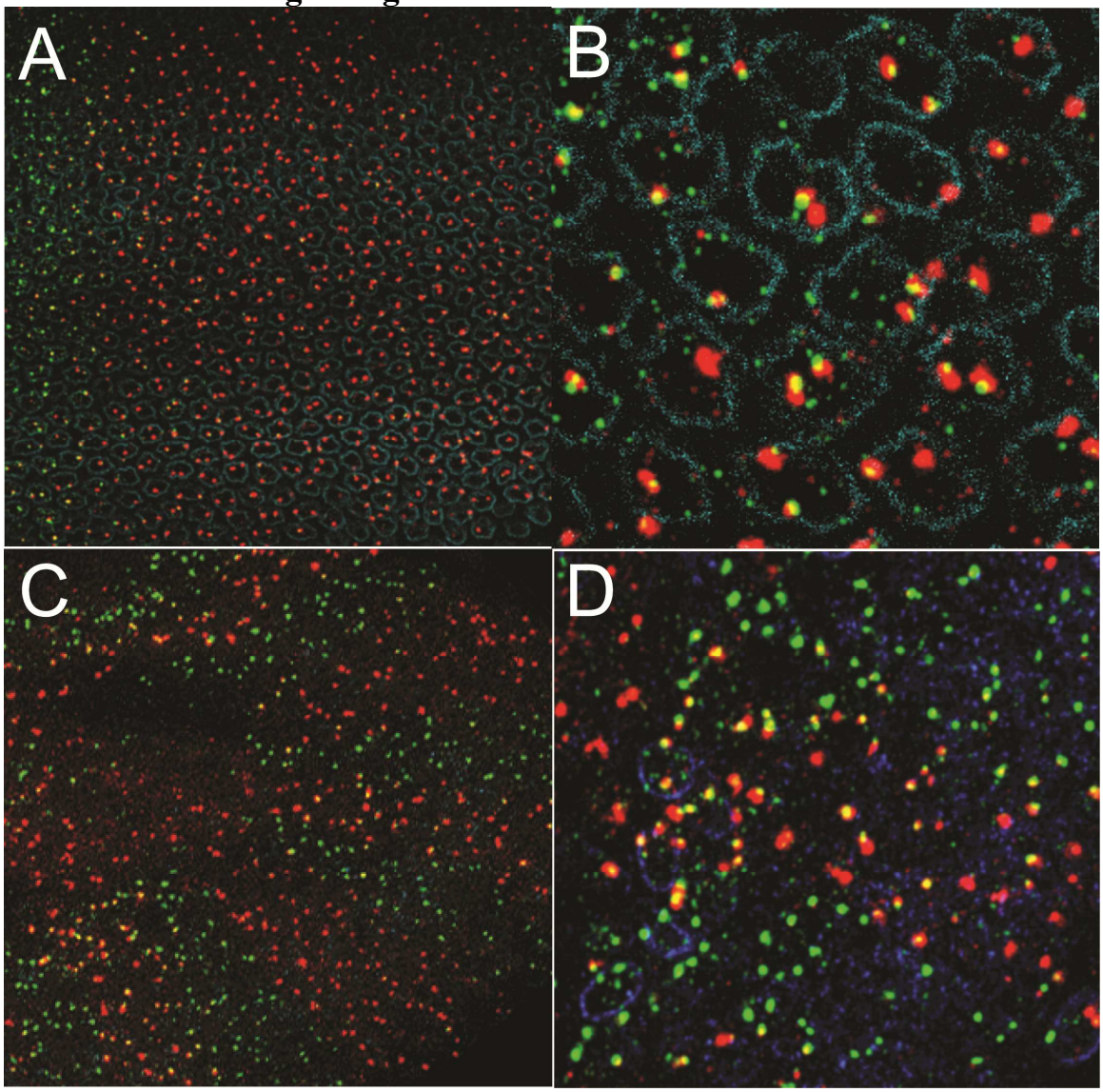
Zamore, P.D. and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. *Science* **309**(5740): 1519-1524.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**(5902): 750-756.

Zhou, J., Ashe, H., Burks, C., and Levine, M. (1999). Characterization of the transvection mediating region of the abdominal-B locus in Drosophila. *Development* **126**(14): 3057-3065.

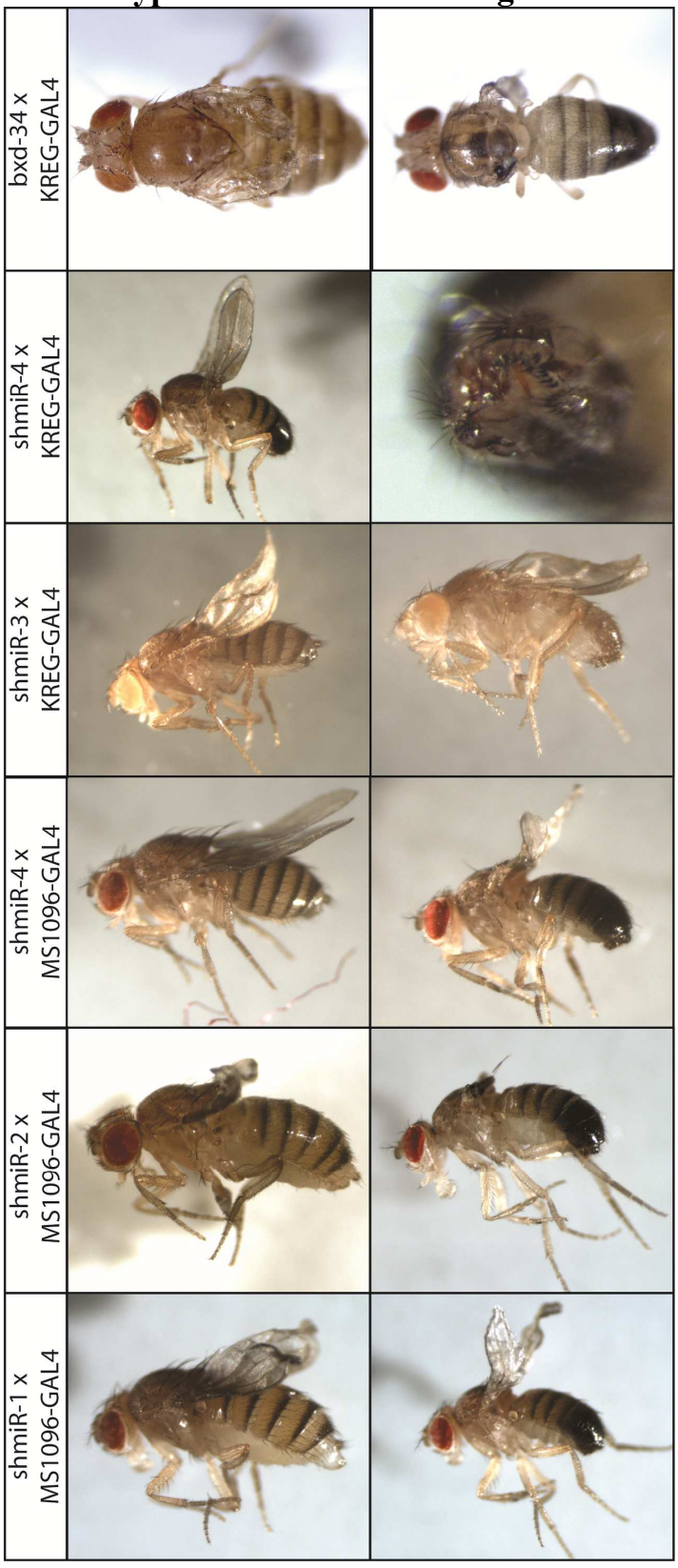
Zinzen, R.P., Cande, J., Ronshaugen, M., Papatsenko, D., and Levine, M. (2006). Evolution of the ventral midline in insect embryos. *Dev Cell* **11**(6): 895-902.

Appendix I:
High Magnification *bxd* and *Ubx in situ*s



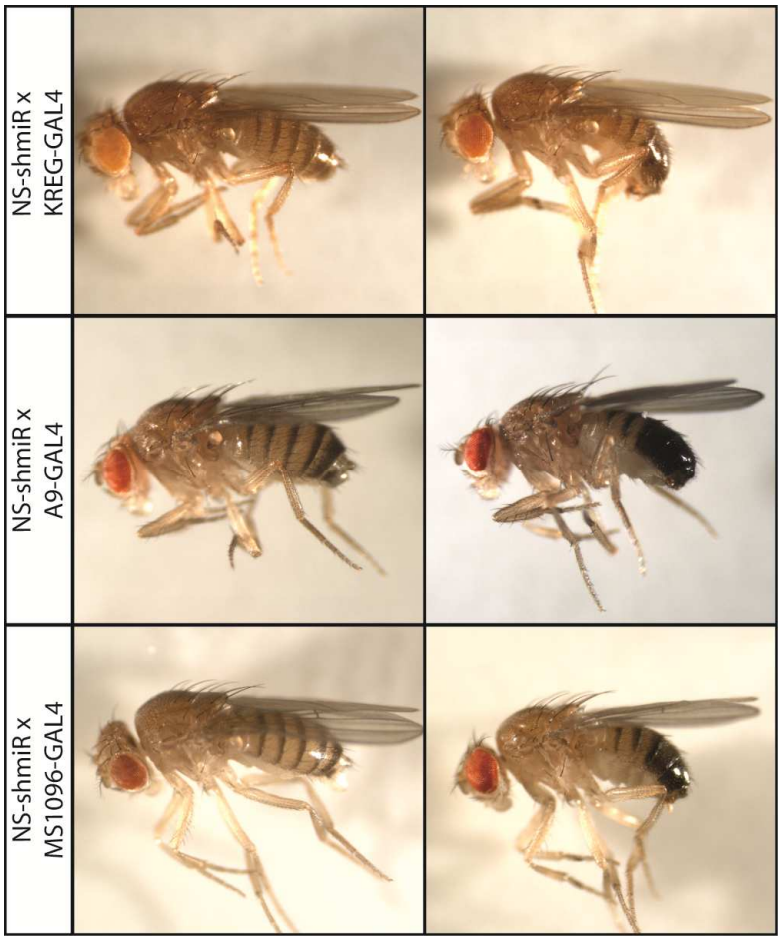
Appendix I – High magnification *in situ*s of *bxd* (red) and *Ubx* (green) probes co-stained with antibody to nuclear lamin (blue) A) Close-up of the posterior half of a ~2.5 hr embryo showing expression of *bxd* and *Ubx* enclosed in a circle of nuclear lamin stain B) High magnification zoom in of the *bxd*/*Ubx* boundary region in the same embryo. One to two dots are within each nucleus showing *de novo* expression from *bxd* and *Ubx* loci. Overlapping dots indicate expression from the same chromosome. C) Close-up of a ~3.5 hr embryo stained the same as listed above D) High magnification zoom in of the same embryo

Appendix II:
Additional Phenotypes for Double and Single shmiR Constructs



Appendix II – This figure shows additional photos of shmiR constructs described but not pictured in the text. The full image of the male and female phenotypes of shmiR-3/4 x KREG-GAL4 F1 flies. Additionally, shmiR constructs containing individual shmiRs are depicted to show that when driven individually, *bxd* shmiRs have similar phenotypes to those phenotypes caused when driven in pairs. All of the shmiRs pictured have typical wing phenotypes and shmiR-4 x KREG-GAL4 males have the twisted genitalia.

Appendix III:
NS-shmiR Crosses with GAL4 Lines



Appendix III – This figure contains images of crosses between the control shmiR NS-shmiR and all GAL4 lines used in this thesis. All progeny are phenotypically normal and show no artifactual phenotypes of non-specific shmiR expression.

Appendix IV:
Primers and Oligonucleotides

***bxd* Cloning**	
long-3s	AAAGCGGACCAGAGTTGAGA
long-3as	ACCCAAGTTCCGTGACGTAT
bxd-RA-cDNA-s	CGATCATCTCAAAAATATTCATTTC
bxd-RABC-cDNA-as	TTGGCCAGCATCTTGATGAG

***bxd* Probe Primers**	
bxd-5' ex-s	GTTGCGGATCGGCTTAAA
bxd-5' ex-as	TGCAGAAATATTACCGGAGAA
bxd-3' exon-s	AGCACGATTCCGGTCATAAT
bxd-3' exon-as	TCAAATATCCTTTCCGCAGTG
Ubx -s	gccaagaacaaagccaagag
Ubx--as	gccaaaatcacacattgcac
Ubx--B-s	TTCAAAAGCATCTCCGCATA
Ubx-B-as	TTCGCTGCAGCTGTTTTATG

***bxd* shmiRs**	
shmiR-1	TTTCTGACCCGTAATGAGCCC
shmiR-2	TTGTTTACGGCATTTGTCGAG
bIN-1	TTTCTACCTAGCTATGTGCCG
bIN-2	TTATTCGCAACCTCTGTCAGT
bEX-1	TTTGGCCTGGTCTCTTTCGCT
bEX-2	TTGGCAAGCACATCTTTAACA
shmiR-3	TATTCCACTAGTTAACTTGGT
shmiR-4	TTGTTCTCGTGAAATTGGGCG
NS-shmiR	TCTGCTTCTGCTCTAGAGTGG

750 Cloning	
750-E-s	CAAGAAGGAGGAGCTGATGG
750-E-as	AATGCAATGCGATGTCATGT
750-3s-EcoRI-P	CCGGAATTCCCCATAAATTGCCATTGTGA
750-3as-XhoI-P	GGCCTCGAGGTTCCACAGCACTCAATTCG

750 Probe Primers	
12515750-s	AGATGTAGGCCATGGTTTCG
12515750-as	GAGCAGCCAAATAGCCAAAG
4.7 Probe-s	GAAAGCAGGCAGAACAGACC
4.7 Probe-as	CGCAAGGCTCTCGAATTTAG
Ubx-B-s	TTCAAAAGCATCTCCGCATA
Ubx-B-as	TTCGCTGCAGCTGTTTTATG

750 shmiRs	
750-1	AAGGTGAGTCGCAGTTGCTGC
4.7-1	TTTGTTTTAAGCGTTTTCGCC
4.7-2	TAAATCGCTGAAATTGGCGGC

qPCR Primers	
bxd-M1-RT-s	GCAGACATTTTGCCACAAAC
bxd-M1-RT-as	AGAGTCTGGCCTCAAAACGA
bxd-RD-RT-s	GGGCTGCTATCCATGAAAAA
bxd-RD-RT-as	CCCCCTCTACTTTTCGGTTT
RP49-s	CGCACCAAGCACTTCATCC
RP49-as	AGCGGCGACGCACTCTGT