10-28-2021 11:30 AM

Screens for Phenotypic Non-Specificity of Transcription Factor Function in Drosophila melanogaster

Sheng Cheng, *The University of Western Ontario*

Supervisor: Percival-Smith, Anthony, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biology

Abstract

Phenotypic non-specificity is a phenomenon in which the phenotypes associated with the expression of a given Transcription Factor (TF) are induced or rescued by multiple distinct TFs. Importantly, this phenomenon is observed with TFs from different protein families that recognize distinct DNA binding sites. To further analyze this phenomenon in *Drosophila melanogaster*, experiments were initiated for the purpose of integrating non-resident TFs into target *TF* loci via recombinase mediated cassette exchange (subsequent to the introduction of *attP* sites at the *TF* loci by CRISPR mediated homology directed repair). Contrary to expectations, no homologous recombinants were identified during the initial CRISPR mediated attempts at gene editing. However, three w^+ non-homologous recombinants were identified: two when targeting *bcd* (*Bcd 4* and *Bcd 39*) and one when targeting *Scr* (*Scr-D1*). *Bcd 4* and *Bcd 39* were the result of transposition of the w^{67c23} allele into the first intron of the *osp* gene; whereas *Scr-D1* was the result of the insertion of the *mini-white* gene from the *Scr* repair template into the genome (with hallmarks of transposition). These non-homologous recombination events suggest that DSBs activate transposable element mobilization. In an alternate approach for studying phenotypic non-specificity, the *UAS-GAL4* system was used to express non-resident TFs and assess the functional complementation of loss-of-function alleles at several *TF* loci. The rescue of six *TF* loci (*lab*, *Dfd*, *Scr*, *Ubx*, *dsx* and *fru*) was determined using at least 12 non-resident TFs. Five out of the six *TF* loci were rescued by non-resident TFs: *lab* was rescued by expression of DSX^M; *Scr* was rescued by expression of FOXO; *Ubx* was rescued by expression of ANTP and EY; *dsx* phenotypes were rescued to different extents by the expression of a majority of the non-resident TFs; and *fru* was rescued by expression of DISCO. In all cases,

the rescue was non-uniform across the pleiotropic phenotypes that depend upon the expression of the resident TF. This suggests that the phenomenon of phenotypic non-specificity is differentially pleiotropic.

Keywords

Transcription Factor, CRISPR, Transposable element, Phenotypic non-specificity, Differential pleiotropy, Limited specificity model

Summary for lay audience

Transcription is the process of copying the DNA sequence of a gene into RNA. “Transcription factors” (TFs) are a class of important proteins that regulate this process by binding to specific DNA sequences adjacent to the gene, thereby turning genes “on” or “off”. Traditionally, each transcription factor is thought to have its own distinct preference with respect to the DNA sequences it binds. Therefore, the function of transcription factors is specific (i.e., that a transcription factor can only regulate a certain number of genes). In my research, I observed multiple occasions of “phenotypic non-specificity” of transcription factors. The results of my research shows that specific transcription factors have the potential to regulate many more distinct genes than expected, and that the function of a transcription factor can be replaced or substituted by another transcription factor. My research indicates that current paradigms of transcription factor function and TF- DNA interaction are not comprehensive and that further studies in this area are needed. Furthermore, I discovered that DNA. damage (DNA double strand breaks) caused by the genetic tool, CRISPR, may destabilize the genome of the organism being manipulated and potentially create unexpected mutations. This discovery should be taken into consideration with regards to the future implementation of CRISPR, especially with respect to clinical trials of CRISPR mediated therapies.

Co-Authorship statement

Chapter 2. Characterization of CRISPR induced non-homologous recombination

I performed all the experimental procedures and genetic analysis with a few exceptions: The microinjections were performed by Dr. Anthony Percival-Smith and me together. Dr. Anthony Percival-Smith isolated the *Bcd 39* and *Bcd 4* transformants and mapped the insertions to chromosome 2. Dr. Anthony Percival-Smith finished the last step of the 5'RACE assay and analyzed the DNA sequence data. Dr. Anthony Percival-Smith made intellectual contributions to experimental design and assisted in writing the manuscript of this chapter.

Chapter 3: The differential pleiotropy of phenotypic non-specificity in *Drosophila melanogaster*

I performed all the experiments mentioned in the "Material and Methods" and "Results" of this chapter with a few exceptions: Dr. Anthony Percival-Smith did the statistical analysis for the data I collected. The SEM pictures of *Ubx* and *dsx* experiments were taken by Dr. Anthony Percival-Smith. Dr. Anthony Percival-Smith established and maintained most of the stocks used for the experiments. The videos of the courtship behavior of *fru* experiments were taken by Dr. Anthony Percival-Smith and I determined the courtship indexes. The panels C2, C3, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14 in Figure 1; A, B, C and D in Figure 3; P2 in Figure 7; J3 and j3 in Figure 8 were taken by Dr. Anthony Percival-Smith. Dr. Anthony Percival-Smith performed the cross to make *rnGAL4-UASfoxo*, *rnGAL4-UASScr* lines. Dr. Anthony Percival-Smith made intellectual contributions to experiment design and assisted in writing the manuscript of this chapter.

Acknowledgments

I want to take this opportunity to express my gratitude to my supervisor, Dr. Anthony Percival-Smith and my advisors, Dr. Greg Kelly and Dr. Jim Karagiannis for their continuous guidance and support during my PhD program.

I offer my special thanks to the members of the Percival-Smith laboratory Anirban Banerjee, Alaa Briek, Suhayr Solangi, Clair Newman, Kristie Ng, Sharon Law, Katherine Galang and Janet Tu for being supportive.

A special thanks to Xi Zhang from Dr. David Smith's lab for guiding me with Geneious and genetic analysis.

I want to thank my best friends in UWO Anirban Banerjee and Mohamed Gatie for being supportive and for the insightful and fun conversation we had.

I dedicate this thesis to my parents Huaming Cheng and Hongfei Lu, my wife Jing Cui and my daughter Amelia who made uncountable sacrifices for the sake of my education. I would never have succeeded without your support.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

abx	*anterobithorax* (regulatory region of *Ubx* gene)
A1	first abdominal segment
abd-A	*abdominal-A* (gene)
Abd-B	*Abdominal-B* (gene)
bx	*bithorax* (regulatory region of *Ubx* gene)
BX-C	Bithorax Complex
ANT-C	Antennapedia Complex
Antp	*Antennapedia* (gene)
ANTP	Antennapedia (protein)
ao	antenna or antennal organ
ap	*apterous* (gene)
AP	Apterous (protein)
bab1	*bric a brac 1* (gene)
BAB1	Bric a brac 1 (protein)
bcd	*bicoid* (gene)
BCD	Bicoid (protein)
br	*broad* (gene)
BR	Broad (protein)
br.Z1	*broad* Z1 isoform (DNA construct)
BR.Z1	Broad Z1 isoform (protein)
br.Z2	*broad* Z2 isoform (DNA construct)
BR.Z2	Broad Z2 isoform (protein)

bZIP	basic leucine zipper
Cas9	CRISPR-associated protein 9 (protein)
ci	cirri
CI	courtship index
CRE	*Cis*-regulatory elements
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CS	Campaniform sensilla
CTD	C-terminal domain
DA	dorsal arms
DB	dorsal bridge
DBD	DNA binding domains
Df	deficiency
Dfd	*Deformed* (gene)
DFD	Deformed (protein)
disco	*disconnected* (gene)
DISCO	Disconnected (protein)
DSB	double strand breaks
dsx	*doublesex* (gene)
DSX	Doublesex (protein)
dsx^F	*doublesex* female isoform (DNA construct)
DSX^F	Doublesex female isoform (protein)
dsx^M	*doublesex* male isoform (DNA construct)

DSX^M	Doublesex male isoform (protein)
EMS	ethyl methanesulfonate
en	*engrailed* (gene)
ey	*eyeless* (gene)
EXD	Extradenticle (protein)
EY	Eyeless (protein)
FB	Foldback
foxo	*forkhead box, sub-group O* (gene)
FOXO	Forkhead box, sub-group O (protein)
fru	*fruitless* (gene)
FRU	Fruitless (protein)
fru^M	*fruitless* male isoform (DNA construct)
FRU^M	Fruitless male isoform (protein)
Fst	*Frost* (gene)
HD	homeodomain
HDR	homologous directed repair
HER	Hermaphrodite (protein)
HLH	helix-loop-helix
Hox	*Homeotic selector* (genes)
HOX	Homeotic selector (protein)
IE	inserted element
IFM	indirect large flight muscle
IX	Intersex (protein)

Kr	*Krüppel* (gene)
lab	*labial* (gene)
LAB	Labial (protein)
Lb	Labial (head structure)
MC	Maxillary cirri
Md	Mandibular
MH / mh	mouth hook
msl-2	*male-specific lethal 2* (gene)
MSL-2	Male-specific lethal 2 (protein)
MT	medium tooth
Mx	Maxillary
MXO / mxo	maxillary sense organ
NHEJ	non-homologous end joining
osp	*outspread* (gene)
PAM	protospacer adjacent motif
pb	*proboscepedia* (gene)
PB	Proboscepedia (protein)
pbx	*postbithorax* (regulatory region of *Ubx* gene)
RMCE	Recombinase-Mediated Cassette Exchange
RS	repetitive sequence
run	*runt* (gene)
Scr	*Sex combs reduced* (gene)
SCR	Sex combs reduced (protein)

SEM	scanning electron microscopy
sqz	*squeeze* (gene)
SQZ	Squeeze (protein)
SRY	sex reversal locus
Sxl	*Sex-lethal* (gene)
SXL	Sex-lethal (protein)
T1	first thoracic segment
T2	second thoracic segment
T3	third thoracic segment
TF	Transcription Factor
tra	*transformer* (gene)
tracrRNA	*trans*-activating CRISPR RNA
TALEN	transcription activator-like effector nuclease
Tb	tubby
TE	Transposable element
TIR	terminal inverted repeat
TRA	Transformer (protein)
UAS	Upstream activation sequence
Ubx	*Ultrabithorax* (gene)
UBX	Ultrabithorax (protein)
VA	ventral arms
w	*white* (gene)
WT	wild type

XSE	X-encoded signal element
y	*yellow* (gene)
ZF	Zinc-finger
ZFN	Zinc-finger nuclease

Chapter 1. Introduction

1.1 Transcription factors and the regulation of development: A crucial questions in developmental biology relates to how a complex organism develops from a single-celled fertilized egg. Cell fate determination is vital for the initial genetically identical, totipotent cells to form the specialized cells required for constructing a complex organism. The diversity of specialized cells is achieved through the response of undetermined cells to external and internal information, which determines cell fate resulting in differential gene expression. The regulation of the rate of transcription initiation is a major mechanism that controls gene expression and is mediated by transcription factors. Transcription factors (TFs) are proteins that bind to *cis*-regulatory elements (enhancers and silencers) and regulate the rate of transcription initiation (Latchman, 1993). *Cis*-regulatory elements (CREs) are non-coding DNA sequences composed of binding sites for transcription factors (Ong & Corces, 2011; Wittkopp & Kalay, 2012). Enhancers are typical examples of CREs (Ong & Corces, 2011). Enhancers interact with promoters to regulate gene expression (Heintzman & Ren, 2009).

TFs control and regulate the expression of genes such that the correct set of genes are expressed in the correct cell and at the correct time during development. One common characteristic of TFs is that they possess a DNA binding domain, which binds to the transcription factor DNA-binding sites in target genes (Mitchell & Tjian, 1989; Ptashne & Gann, 1997). TFs are organized into protein families based on the amino acid sequence of their DNA binding domains (DBDs) (Jin *et al.*, 2014; Matys *et al.*, 2006; Ptashne & Gann, 1997; Wingender *et al.*, 2015). In eukaryotes, TFs with homeodomain (HD), C2H2-Zinc

finger (ZF), basic leucine zipper (bZIP) or basic helix-loop-helix family (HLH) DBDs are members of large protein families (Lambert *et al.*, 2018). TFs often form dimers (homodimers or heterodimers) and specifically bind to a short stretch of nucleotides (typically 6-12 nucleotides) (Gurdon, 2016; Vinson *et al.*, 2011).

1.2 Advantages of Drosophila as a model organism: *Drosophila melanogaster*, is an attractive model organism for research in the fields of molecular biology, developmental biology, genetics, and neuroscience. *D. melanogaster* was first used as a research organism by William E. Castle in 1901 at Harvard University. However, it was not used as a model organism for genetic studies until 1909, when Thomas Hunt Morgan from Columbia University found a fly with a white-eye mutation, which he subsequently characterized uncovering the chromosomal basis of inheritance (Morgan, 1910). During his 25-year-research career at Columbia, Morgan and his lab members made some of the most influential discoveries in Genetics using *Drosophila* as a genetic model system. These discoveries include the first genetic map by Sturtevant in 1913, the discovery of genetically inheritable homeotic mutants by Bridges in 1915 (Bridges, 1915), and the creation of balancer chromosomes by Muller in 1918 (Muller, 1918). Morgan won the Nobel prize in Physiology or Medicine in 1933 for his contributions to the establishment of the chromosome theory of inheritance. Since these discoveries and subsequent decades of continued research, *Drosophila melanogaster* has become a sophisticated genetic model organism.

The powerful genetic tools available for *D. melanogaster* allow investigators to elucidate

the basis of complex traits and gene-gene and gene-environment interactions. *Drosophila* takes only ten to fourteen days at room temperature from the time the egg is laid until an adult fly eclose from a pupal case. This short generation time greatly increases the rate of experimental analysis. Also, females are very fecund laying an average of 700-1000 eggs externally (Bownes *et al.*, 1989) which facilitates sample collection and experimental manipulation (Ashburner *et al.*, 2005). *Drosophila* embryonic development occurs after external oviposition allowing observation and genetic dissection of development (Ashburner *et al.*, 2005). Lastly, *Drosophila* is relatively inexpensive to maintain in the lab and is easy to work with in large numbers. Since the functions of many important genes are well conserved across evolution, information gained from the study of genetic pathways in *Drosophila* can be applied to other organisms that cannot be so easily manipulated in the laboratory (Ashburner *et al.*, 2005).

In addition to having a sophisticated genetic system, *Drosophila* is also a sophisticated genomic and developmental system. With more than 100 years of research, the life history, physiology, behavior, and life cycle of *Drosophila* are well characterized. The sophisticated analysis of the life cycle makes *Drosophila* an ideal organism to study development. In addition, the recent detailed analysis of the cell biology of neural connections (connectomes) makes *Drosophila* a sophisticated model organism for neurobiology. The genome of *Drosophila* is relatively small and is composed of 4 chromosomes (around 180 Mb in total) that carry 15,504 genes: the sex chromosomes (X and Y; **Figure 1**) and three sets of autosomes (chromosome 2, 3, and 4; **Figure 1**). Of particular value are the genomic tools available in *D. melanogaster*. The genome sequence

and annotation are of the highest quality of any sequenced organism (Adams *et al.*, 2000; Shah *et al.*, 2019). The genomic DNA sequence data of *Drosophila* is combined with knowledge derived from *Drosophila* genetics, biochemistry and physiology in a publicly accessible database, called FlyBase (dos Santos *et al.*, 2015). Also, many mutant and transgenic fly lines are available from stock centers. In 2007, the genomes of an additional 12 *Drosophila* species were also sequenced, making *Drosophila* a great model for the study of evolution (Drosophila 12 Genomes *et al.*, 2007). The availability of sequence data from these 12 species, which were chosen based on their evolutionary distance from *D. melanogaster*, has and will facilitate discovery of conserved motifs, the identification of new genes, and will assist in further annotation of the *D. melanogaster* genome.

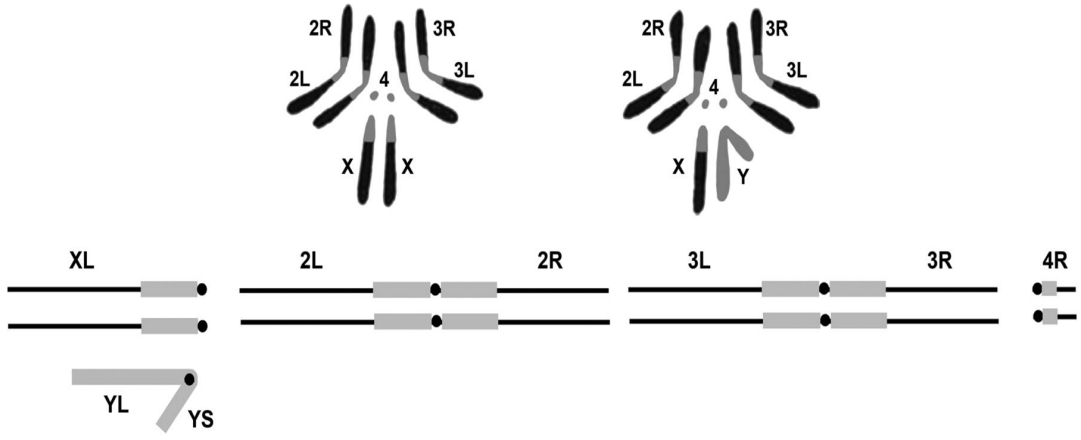


Figure 1. Mitotic chromosomes of *Drosophila* indicating euchromatic Regions, heterochromatic regions, and centromeres. The top two figures show the mitotic chromosomes of *Drosophila*, female on the left and male on the right. The bottom is a diagram of the structure of the chromosomes. Arms of the autosomes are designated 2L, 2R, 3L, 3R, and 4R. Arms of the sex chromosomes are designated XL, YL and YS. Grey color represents heterochromatin and black is euchromatin. This figure is adapted from Kaufman 2017 with permission.

GAL4-UAS: The GAL4-UAS system is a commonly used genetic tool in *Drosophila* to drive expression of a gene of interest (Brand & Perrimon, 1993). It is such a powerful and versatile genetic tool that it is often referred to as the "Swiss army knife" of *Drosophila* genetics (Duffy, 2002). The GAL4 protein is a yeast transcription factor that has no endogenous targets within the *Drosophila* genome. The upstream activation sequence (*UAS*) is an enhancer, which is specifically bound by the TF, GAL4. When the GAL4 protein binds to the *UAS* sequence, the gene fused to the *UAS* sequence is expressed. The UAS-GAL4 is a binary approach, in that the *UAS* sequence (fused to the specific gene of interest) is kept in one fly line and GAL4 (fused to a promoter and tissue-specific enhancer) is kept in another. Only when these two lines are crossed is the gene of interest expressed in cells expressing GAL4 of the subsequent progeny. The advantage of the binary approach is that the functions of different target genes can be analyzed when expressed at distinct times and in distinct cells using the array of tissue-specific GAL4 driver lines available (**Figure 2**).

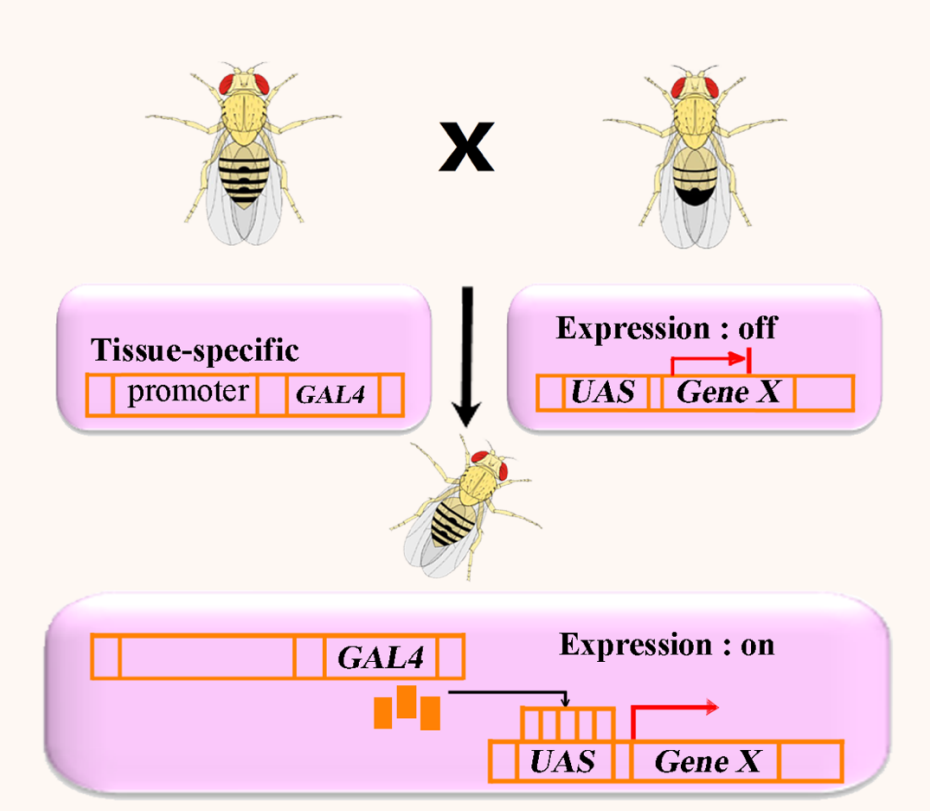


Figure 2. **The bipartite design of the *Drosophila* UAS-GAL4 system.** The gene of interest (*Gene X*) is fused to the UAS sequence and kept in one fly line. GAL4 is fused to a tissue-specific promoter and is kept in another. In the progeny of a cross between the two lines, the GAL4 protein will bind to the *UAS* sequence and activate the gene of interest in the specific tissue.

1.3 Genetic regulation of *Drosophila* embryogenesis: One common characteristic of the body plan of bilaterians is repeated, metameric units (segments) (Carroll *et al.*, 2004). In *Drosophila*, the body is segmented into 15 units: three head, three thoracic and nine abdominal segments (Martinez-Arias & Lawrence, 1985). The process of the segmentation and structure of the body plan is controlled by a regulatory hierarchy of five sets of genes: *maternal effect* genes, *gap* genes, *pair rule* genes, *segment polarity* genes, and *homeotic* genes (**Figure 3**) (Carroll *et al.*, 2004). Most of the genes of this hierarchy encode TFs (Carroll *et al.*, 2004).

Maternal effects genes are the first-class of genes to act in the hierarchy. During oogenesis, RNA transcripts of *maternal effect* genes are transported to the egg and translated after fertilization to organize the coordinates of the developing embryo (Johnston & Nüsslein-Volhard, 1992). For example, the mRNA of the maternal effect gene, *bicoid* (*bcd*), is localized in the cytoplasm at the future anterior pole of the egg and is translated after fertilization (Carroll *et al.*, 2004; Johnston & Nüsslein-Volhard, 1992). The Bicoid (BCD) protein forms a concentration gradient that determines the anterior to posterior (A-P) axis of the embryo (Carroll *et al.*, 2004; Johnston & Nüsslein-Volhard, 1992). A mother homozygous for a *bicoid* loss-of-function allele produces larval progeny in which the head and thoracic segments are missing (Driever & Nüsslein-Volhard, 1988). Maternal effect proteins regulate the expression of the second set of segmentation genes, the *gap* genes.

Gap genes are amongst the first zygotically expressed genes and are transcribed in spatially restricted expression domains along the A-P axis of the embryo and include: *huckebein*,

tailless, *giant*, *hunchback*, *Krüppel*, and *knirps* (Johnston & Nüsslein-Volhard, 1992). Loss-of-function alleles in *gap* genes result in the loss of multiple, contiguous segments. The Gap proteins regulate the expression of the third class of segmentation genes, the *pair rule* genes.

Pair rule genes are expressed in a pattern of seven one segment wide stripes (Rivera-Pomar & Jãckle, 1996). In embryos homozygous for *pair rule* loss-of-function alleles, every other segment is deleted (Nüsslein-Volhard *et al.*, 1984; Wakimoto & Kaufman, 1981). The Pair rule proteins regulate the expression of *segment polarity* genes, which establish the anterior – posterior polarity within a segment. After the body is segmented by the first four sets of genes, the expression of the fifth set, *Hox* genes, determine the distinct morphologies of the segments; the segmental identity (Capovilla *et al.*, 1994; Heffer *et al.*, 2010; Pearson *et al.*, 2005).

The term "homeosis" was used by William Bateson in 1894 to describe the phenomenon whereby one body part or organ of an organism is transformed into the likeness of another body part or organ (Bateson *et al.*, 1894). *Homeotic* (*Hox*) *selector* genes were identified by mutations that resulted in heritable homeotic transformations (Lewis, 1978; Nüsslein-Volhard & Wieschaus, 1980; Nüsslein-Volhard *et al.*, 1984). All HOX proteins contain a 60-amino-acid DNA-binding domain, the homeodomain (HD) (Levine & Hoey, 1988; McGinnis *et al.*, 1990). The sequence of HD is highly conserved among animal phyla (McGinnis & Krumlauf, 1992). *Hox* genes determine the unique segmental identity of individual body segments. *Hox* genes are important developmental regulatory genes that

are expressed in spatially restricted domains along the A-P axis of the embryo (Ingham & Arias, 1992). The phenotype of loss-of-function and gain-of-function mutations in *Hox* genes are homeotic transformations where one segment is transformed into the likeness of another (Bridges, 1915; Kaufman *et al.*, 1980; Kaufman *et al.*, 1990; Levine & Hoey, 1988; Lewis, 1978). The order of *Hox* genes along the chromosome corresponds to the head to tail order of the body segment in which they are expressed (Harding *et al.*, 1985; Lewis, 1978). This phenomenon is called "collinearity" (Carroll *et al.*, 2004; Lewis, 1985). *Hox* genes were first discovered in *D. melanogaster* and their functions in body patterning has been extensively studied both in *Drosophila* and many other species.

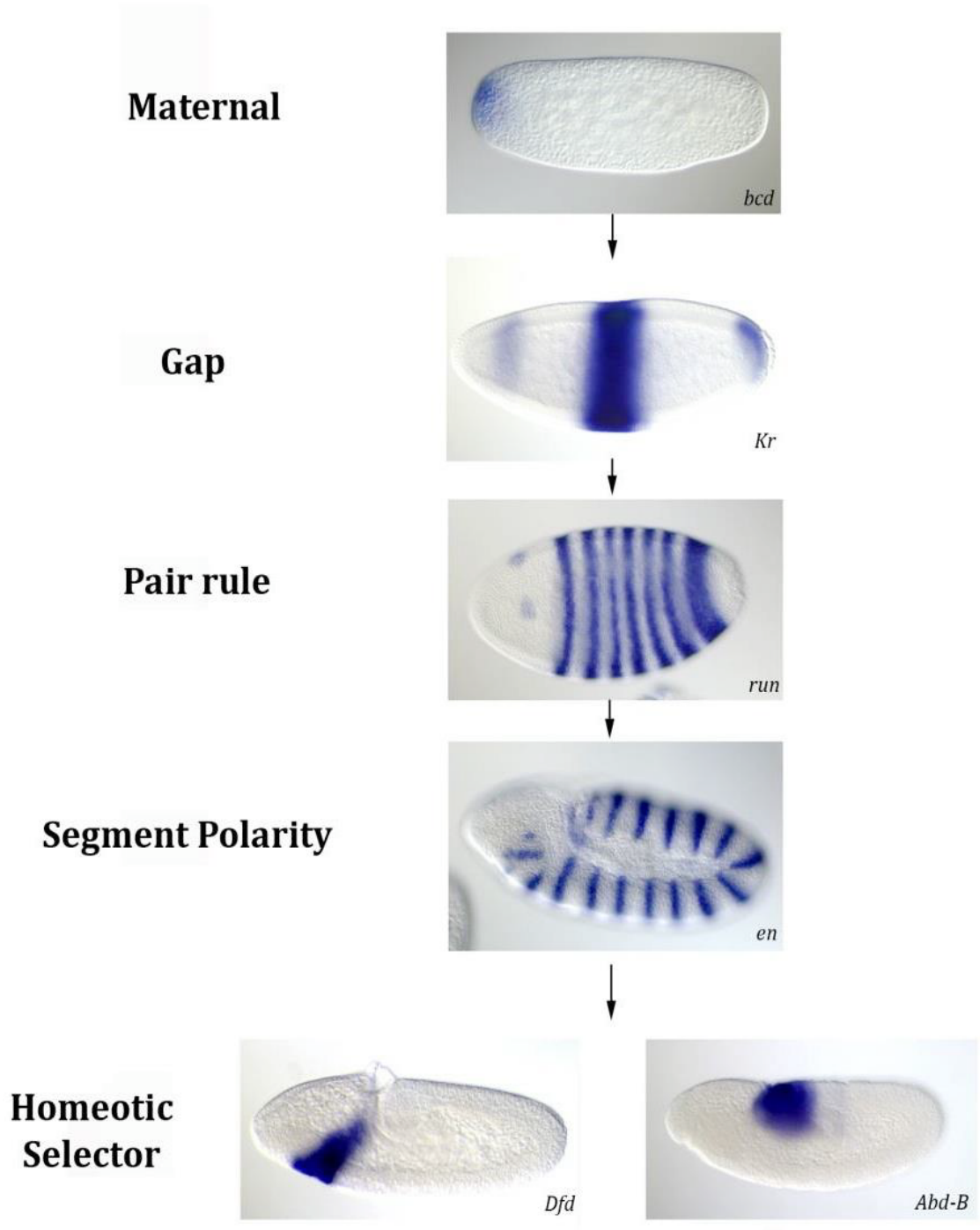

Maternal
bcd
Gap
Kr
Pair rule
run
Segment Polarity
en
Homeotic Selector
Dfd
Abd-B

Figure 3. Segmentation genes are expressed in a hierarchy that regulates the pattern of development along the A-P axis of a developing *Drosophila* embryo. These are *in situ* hybridizations to mRNA of representative genes from each set of segmentation gene in the segmental hierarchy. The protein products of genes expressed earlier in the hierarchy regulate the expression of genes further down in the hierarchy to segment the developing embryo and determine segmental identities. The expression patterns of the maternal coordinate gene, *bicoid* (*bcd*); gap gene, *Krüppel* (*Kr*); pair rule gene, *runt* (*run*); segmental polarity gene, *engrailed* (*en*), and homeotic selector genes, *Deformed* (*Dfd*) and *Abdominal-B* (*Abd-B*) are shown (Tomancak *et al.*, 2002; Tomancak *et al.*, 2007). This figure has been adapted from Sivanantharajah, 2013 with permission.

1.4 *Drosophila Hox* genes: HOX proteins establish embryonic segment identities along the AP axis of bilaterian bodies (Capovilla *et al.*, 1994; Heffer *et al.*, 2010; Pearson *et al.*, 2005). In *D. melanogaster*, there are eight *Hox* genes: *labial* (*lab*), *proboscepedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) (**Figure 4**), which are located in two gene clusters on the right arm of the third chromosome. Five of the genes (*lab*, *pb*, *Dfd*, *Scr* and *Antp*) are found in the Antennapedia Complex (ANT-C) (Kaufman *et al.*, 1980), and the remaining three (*Ubx*, *abd-A* and *Abd-B*) are found in the Bithorax Complex (BX-C) (Lewis, 1978). The *D. melanogaster* embryo is divided into three head segments: Mandibular (Md), Maxillary (Mx) and Labial (Lb), three thoracic segments (T1-T3) and nine abdominal segments (A1-A9) (Carroll *et al.*, 2004). The *Hox* gene, *lab* (located at one end of the ANT-C), is expressed in the most anterior region of the embryo; whereas, the gene located at the other end of the BX-C, *Abd-B*, is expressed in the most posterior part of the embryo (abdominal segments 8 and 9) (**Figure 4)** (Carroll *et al.*, 2004; Lemons & McGinnis, 2006).

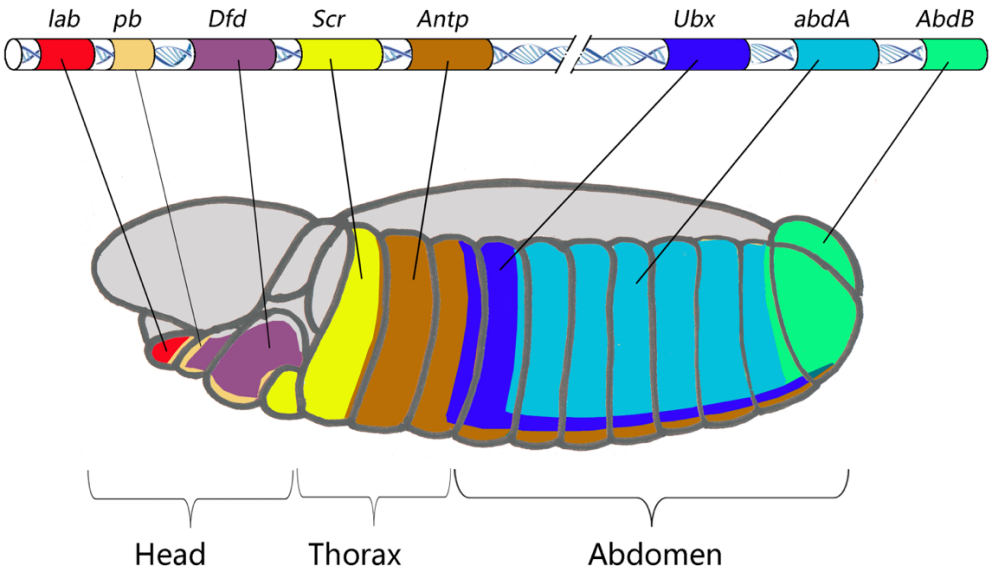


Figure 4. ***Homeotic selector* (*Hox*) gene expression in the Drosophila embryo**. Drosophila *Homeobox* (*Hox*) genes are located in two gene clusters and the expression of the genes along the A-P axis corresponds to their location along the fly chromosome. The expression of the *Hox* genes in the Antennapedia Complex, *labial* (*lab*; red), *proboscipedia* (*pb*; khaki), *Deformed* (*Dfd*; purple), *Sex combs reduced* (*Scr*; yellow) and *Antennapedia* (*Antp*; brown), and the Bithorax complex, *Ultrabithorax* (*Ubx*; blue), *abdominal-A* (*abd-A*; cyan) and *Abdominal-B* (*Abd-B*; green), are indicated on a diagram of an embryo after germ band retraction.

The following sections describe the phenotypes resulting from mutations in the genes that are central to the analysis presented in Chapter 3. The genes analyzed are *labial, Deformed, Sex combs reduced, Ultrabithorax, doublesex* and *fruitless*.

labial*:* The gene *lab* is the most anteriorly expressed member of the *Drosophila Hox* genes. It is expressed primarily in the mandibular segment and the intercalary region (an appendage-less segment between the antenna and mandible) of the head, and also in the midgut (Hughes & Kaufman, 2002). The *lab* gene was originally named "*labial*" because loss-of-function alleles disrupt development of the labial segment; however, the *lab* gene is not expressed in the labial segment (Carroll *et al.*, 2004; Hughes & Kaufman, 2002).

lab*^{14} / *lab*^{4}:*** *lab* null loss-of-function allele result in the failure of *Drosophila* embryo to undergo head involution, which is the internalization of the mouth and head structures that initially start to develop on the embryonic surface ectoderm. The two amorphic *labial* alleles used for experiments in this thesis are *lab*^{4} and *lab*^{14}. The *lab*^{4} allele, also known as *lab*^{f8}, is a homozygously lethal amorphic allele induced by ethyl methanesulfonate (EMS) mutagenesis. The cytology of this allele is normal; no inversion or large genome deletion occurred during mutagenesis (Merrill *et al.*, 1989). The *lab*^{14} allele, also known as *lab*^{vd1} allele, is another amorphic allele isolated after X-ray radiation. It is cytologically normal, but associated with a small deletion (< 2kb) in the *lab* gene (Diederich *et al.*, 1989).

The fruit fly larval cuticle provides a rich set of morphological characteristics to analyze the requirement of genes in the development of the body plan. The head structure of wild

type cuticle preparation is shown in **Figure 5**, with the mouth hooks, medium tooth, H-piece, ventral arms, dorsal arms, and dorsal bridge indicated. *Drosophila* larvae lacking LAB expression do not develop the H-piece, including the bridge and the lateral bar (Merrill *et al.*, 1989). Due to improper head involution, the two mouth hooks are widely separated (**Figure 5**) (Merrill *et al.*, 1989).

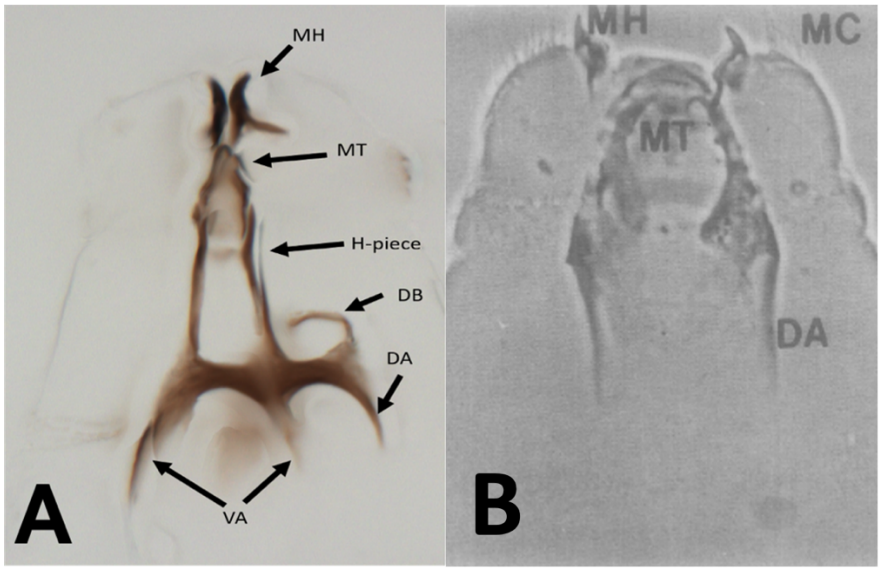


Figure 5. The larval cuticle of the head of WT (A) and a *lab* null mutant (B). The H-piece structure (Hb and Hl) is absent in the *lab* null mutant, and the MHs are widely separated when compared with the WT larval cuticle. Abbreviations: MH, mouth hook; MC, Maxillary cirri; MT, medium tooth; DA, dorsal arms; Hb, H-piece bridge; Hl, H-piece lateral bar; DB, dorsal bridge; DA, dorsal arms; VA, ventral arms. Panel B was adapted from Merrill et al., 1989 with permission.

Deformed: The *Dfd* gene is required for determining the segmental identity of the maxillary and mandibular segments in the larval head (Regulski *et al.*, 1987). Larvae homozygous or hemizygous for *Dfd* loss-of-function alleles develop with a failure of head involution and the loss of larval head structures. Adult loss-of-function phenotypes are deletion of parts of the head and transformations of head to thoracic identity (Brown *et al.*, 1999; Lohmann *et al.*, 2002; Mahaffey *et al.*, 2001; McGinnis *et al.*, 1990; Regulski *et al.*, 1987; Zeng *et al.*, 1994).

Dfd^{16} / Dfd^{12}: The two amorphic *Dfd* alleles used for studies in this thesis are Dfd^{16} and Dfd^{12}. The Dfd^{12} allele, also known as Dfd^{rR11}, is an amorphic allele resulting from a single nucleotide transversion (T to A mutation at 3R:6752954) as a result of EMS mutagenesis. This mutation is a nonsense allele that truncates DFD protein translation at amino acid 210 (Zeng *et al.*, 1994). The Dfd^{16} allele, also known as Dfd^{W21} or Dfd^{rW21}, is an EMS induced single nucleotide transition (G to A mutation at 3R:6793812) which is also a nonsense allele truncating translation at amino acid 346 of DFD (Zeng *et al.*, 1994).

Embryos hemizygous for Dfd^{16} / Dfd^{12} have displaced maxillary and mandibular segments due to ventral side supernumerary cell accumulation in both segments (Hueber *et al.*, 2007). Dfd^{16} / Dfd^{12} embryos lack the maxillary cirri primordium and the anterior boundary of the dorsal ridge between mandibular and maxillary segments (Lohmann *et al.*, 2002) and the Dfd^{16}/ Dfd^{12} embryos develop lacking the mouth hooks and cirri (**Figure 6**) (Brown *et al.*, 1999; Mahaffey *et al.*, 2001).

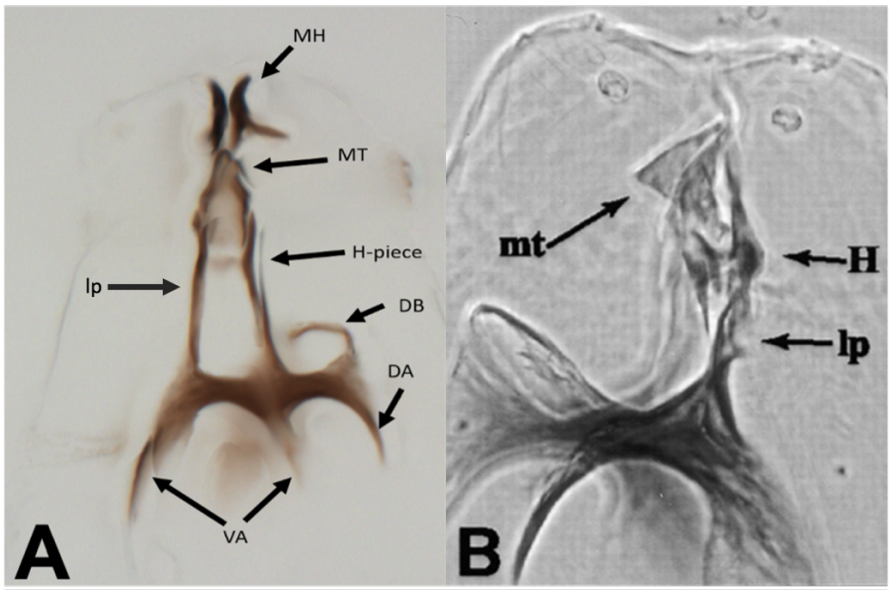


Figure 6. The larval cuticle of the head of WT (A) and *Dfd* null mutant (B). Relative to the WT (A), *Dfd*^{16} / *Dfd*^{12} larvae have no cirri and mouth hooks but the H-piece and lateral process form. Abbreviations: mh, mouth hook; ci, cirri; mt, medium tooth; H, H-piece bridge; lp, lateral process (H-piece lateral bar). Panel A is the same as in Figure 5 A. Panel B was adapted from Mahaffey et al., 2001 with permission.

Sex combs reduced: *Sex combs reduced* (*Scr*) is the fourth *Hox* gene in the *Drosophila* Antennapedia Complex. The SCR transcript is composed of three exons (exon 1, exon 2 and exon 3); exon 2 and exon 3 are coding (LeMotte *et al.*, 1989). In *Drosophila*, the SCR protein is expressed in the labial and T1 segments during embryogenesis. SCR is proposed to have two functions: SCR^{T1} required for prothoracic (e.g. T1 beards, salivary gland and sex comb) development, and SCR^{lab} required for labial derivatives (proboscis) development (Percival-Smith *et al.*, 2013). SCR^{T1} function is conserved, as ectopic expression of the murine HoxA5 protein (SCR homolog) is able to induce similar phenotypes as those caused by ectopic expression of SCR (Zhao *et al.*, 1993). However, SCR^{lab} activity is insect-specific and not conserved throughout bilaterians. Co-ectopic expression of murine HoxA5 with PB cannot induce ectopic proboscises while co-ectopic expression of *Drosophila* SCR and PB can (Percival-Smith *et al.*, 2013). Percival-Smith *et al.* (2013) proposed that during insect evolution expression of PB shifted posteriorly to the labial segment to assist in the switch from the bilaterian conserved SCR^{T1} function to the insect-specific SCR^{lab} function.

The SCR protein contains multiple motifs that are conserved at different taxonomic levels. The HD, octapeptide, YPWM motif and KMAS motif are universally conserved across all bilaterian SCR orthologs. The LASCY motif is conserved in protostome SCR orthologs. The SCKY, PQDL and NANGE motifs are conserved in arthropods SCR orthologs. The MVDYTQLQPQRL sequence (DYTQL motif) and the carboxy-terminal domain (CTD) are insect-specific. The YTPNL, DISPK and NEAGS are conserved in dipterans. Lastly, the NDPVT, QSLAS and VNVPM are *Drosophila*/genus-specific (Curtis *et al.*, 2001;

Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2015). The analysis of SCR function by examination of mutant alleles or by the expression of recombinant proteins uncovered differential pleiotropy which is the non-uniform behavior of alleles across different tissues (Carroll *et al.*, 2004; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2009, 2014, 2015). Differential pleiotropy suggests that SCR is composed of small independent protein motifs that alone make small, tissue-specific contributions to the overall activity of SCR (Sivanantharajah & Percival-Smith, 2009, 2014, 2015).

Scr^4 / Scr^2: The two amorphic *Scr* alleles used for my studies are Scr^4 and Scr^2. The Scr^2 allele is an amorphic allele caused by a single nucleotide transition (C to T mutation at 3R:6841841) that introduces a stop codon in the Scr^2 open-reading frame (Sivanantharajah & Percival-Smith, 2009). The Scr^4 allele, also referred to as Scr^{W21}, is another *Scr* amorphic allele caused by the single nucleotide transition (C to T mutation at 3R:6841790) that introduces a stop codon (Sivanantharajah & Percival-Smith, 2009). Both Scr^2 and Scr^4 encode truncated versions of the SCR proteins (**Figure 7**).

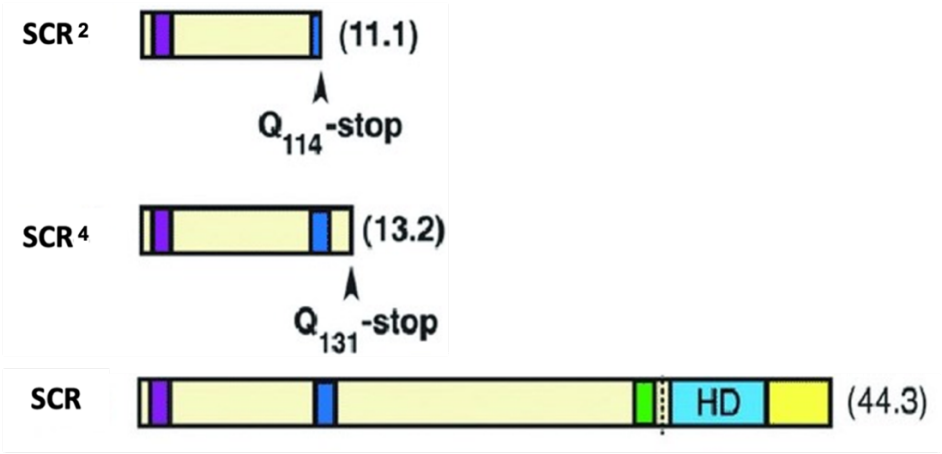


Figure 7. The structure of the SCR^2, SCR^4 and WT proteins (Sivanantharajah & Percival-Smith, 2009). Both *Scr^2* and *Scr^4* encode truncated forms of SCR proteins. The unit in brackets to the right of the proteins is kilodaltons (Sivanantharajah & Percival-Smith, 2009). The octapeptide motif (purple), DYTQL motif (blue), YPWM motif (green), HD (cyan) and CTD (yellow) are indicated in the figure. This figure has been adapted from Sivanantharajah and Percival-Smith, 2009 with permission.

SCR is required for the formation of the male sex combs. *Scr* is haplo-insufficient for determination of the number of sex combs that form; for example, Scr^{4}/+ heterozygous males have a reduction in the sex comb bristle number from the WT 10-12 to 6-7 (Bantignies *et al.*, 2011; Ragab *et al.*, 2006; Southworth & Kennison, 2002). Loss of SCR activity during embryogenesis results in failure of head involution and the larvae have a reduced number of T1 beard setae, duplication of the maxillary sense organ (mxo) (**Figure 8**), and a disrupted labial segment (labial derivatives, like salivary glands, are lost) (Mahaffey & Kaufman, 1987; Pederson *et al.*, 1996; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2009, 2014). The duplicated maxillary sense organ indicates a labial to maxillary segment transformation during embryogenesis (Pederson *et al.*, 1996). Adult viable hypomorphic *Scr* alleles result in adult flies with a decreased number of pseudotracheal rows and a decreased number of sex combs on the first legs (Bantignies *et al.*, 2011; Pattatucci *et al.*, 1991; Ragab *et al.*, 2006; Sivanantharajah & Percival-Smith, 2009; Southworth & Kennison, 2002). In clones of *Scr* null mutant cells in the proboscis primordia, the proboscis is transformed to a maxillary palp (Percival-Smith *et al.*, 1997).

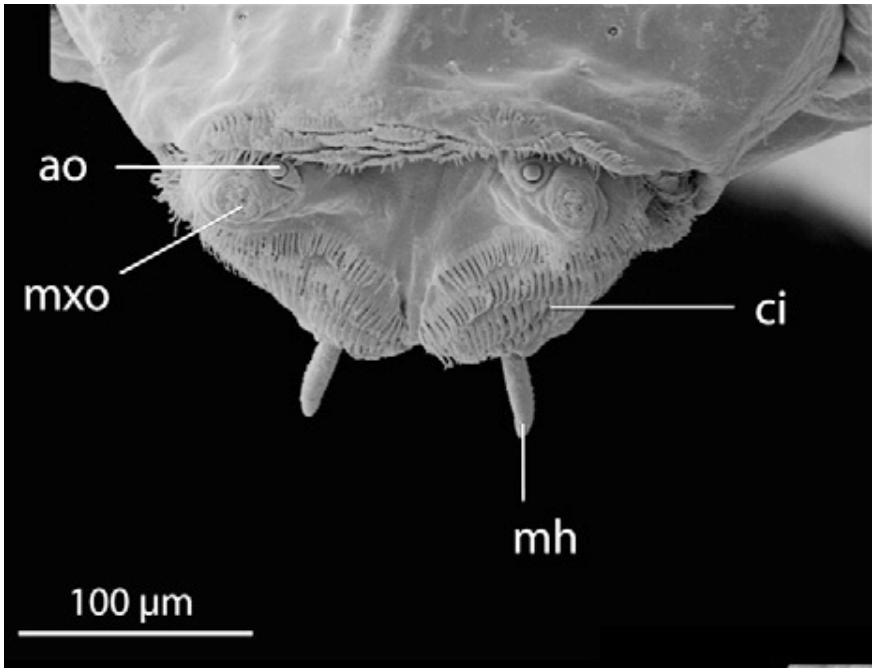


Figure 8. Scanning electron micrograph of the *Drosophila* larval head. Abbreviations: ao, antenna or antennal organ; ci, cirri; mh, mouth hooks; mxo, maxillary sense organ (Wipfler *et al.*, 2013). This figure has been adapted from Wipfler et al., 2013 with permission.

Ultrabithorax*:* The gene *Ubx* is the sixth gene of the *Drosophila* HOM-C gene clusters and the first in the BX-C. In *D. melanogaster*, at least six different isoforms of UBX protein exist (Carroll *et al.*, 2004; Lewis, 2004). *Ubx* is expressed in the third thoracic (T3) and first abdominal (A1) segments. The UBX protein is expressed throughout the haltere but not in the wing (Weatherbee *et al.*, 1998). In adult flies, the T2 segment houses the indirect large flight muscles (IFMs) and T3 segment houses the smaller haltere muscles (Rivlin *et al.*, 2001). Reduction of UBX function results in the transformation of haltere tissue into wing tissue (Kerridge & Morata, 1982; Morata & Garcia-Bellido, 1976; Morata & Kerridge, 1981; Rivlin *et al.*, 2001; Weatherbee *et al.*, 1998). Complete loss of function of UBX during metamorphosis leads to transformation of dorsal and ventral appendages of the third thoracic segment (T3), which includes the halteres and third pair of legs, into the counterparts on the second thoracic segment (T2), giving a four-winged (bithorax phenotype) fly (Lewis, 1978).

Campaniform sensilla (CS) are a class of insect mechanoreceptors, which receive proprioceptive and exteroceptive stimuli (Dinges *et al.*, 2021). In *Drosophila melanogaste*r, the CS are located on the wings, halteres, legs and thorax (Dinges *et al.*, 2021). Based on the morphology, CS in *Drosophila* are categorized into six groups (Cole & Palka, 1982). *Ubx* mutations transform CS on halteres to wing-like CS (Cole & Palka, 1982).

***$Ubx^{9.22}$ / $Ubx^{abx1,bx3,\ 61d,\ pbx1}$*:** The amorphic *Ubx* allele, $Ubx^{9.22}$, isolated after X-ray irradiation, deletes 1580 bps of DNA sequence that includes the exon 3, intron 3 and the first 48 codons of the homeodomain (Mastick *et al.*, 1995; Subramaniam *et al.*, 1994). The

Ubx$^{9.22}$ deletion affects all UBX isoforms (Subramaniam *et al.*, 1994). *Ubx*^{abx1,bx3,61d,pbx1} is a combination of a series of *Ubx* hypomorphic alleles: *anterobithorax*^{1} (*abx*^{1}), *bithorax*^{3} (*bx*^{3}), *postbithorax*^{1} (*pbx*^{1}) and *Ubx*^{61d}. The alleles *abx*, *bx* and *pbx* incorporate changes in regulatory regions of *Ubx* (Casanova *et al.*, 1985; Castelli-Gair & García-Bellido, 1990; Qian *et al.*, 1991; Slack & Bard, 1991). Flies which have single mutant (*abx*^{1} or *bx*^{3} or *pbx*^{1}), double mutants (*bx*^{3} and *pbx*^{1}), triple or quadruple mutants (*abx*^{1},*bx*^{3}, *pbx*^{1} or *abx*^{1},*bx*^{3},*Ubx*^{61d}, *pbx*^{1}) and hemizygous with a *Ubx* deficiency (*Df*(*3R*)*2P*) allele, have an increasing strength of the T3 segment to T2 segment homeotic transformations observed (**Figure 9**) (Rivlin *et al.*, 2001). *abx*^{1} transforms the T2 posterior compartment (T2p) to T1 posterior compartment (T1p) and the T3 anterior compartment (T3a) to T2 anterior compartment (T2a); *bx*^{3} transforms T3a to T2a; while *pbx*^{1} transforms T3p to T2p (Rivlin *et al.*, 2001). The double, triple and quadruple combinations of the alleles lead to more complete transformations of T3 segment into T2 segment and replace the halteres with a second pair of wings (Rivlin *et al.*, 2001).

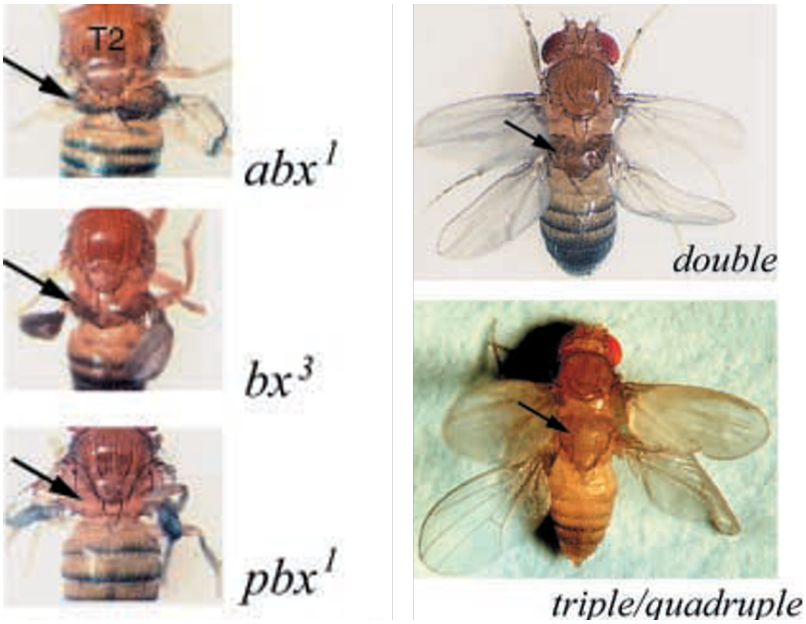


Figure 9. Adult phenotypes of *Ubx* hypomorphic alleles. The T3 to T2 transformations of abx^1, bx^3, pbx^1, double bx^3 and pbx^1 mutant, triple (abx^1,bx^3, pbx^1) or quadruple mutants (abx^1,bx^3,Ubx^{61d}, pbx^1) when they are hemizygous with *Ubx* deficiency (*Df*(*3R*)*2P*) allele (Rivlin *et al.*, 2001). The arrows indicate the transformed T3 segment. This figure is adapted from Rivlin et al., 2001 with permission.

1.5 Sex determination: Sex determination is the process by which the dimorphic sexual characteristics of an organism are determined. The evolution of the mechanisms of sex determination is hypothesized to be a reverse-order process (Wilkins, 1995). According to this theory, the upstream sex identity regulators (genes and mechanisms) in the sex determination pathway change very rapidly as new species are formed, and therefore, the upstream sex identity regulators have diverged between species. However, the downstream regulators which are directly responsible for determining sexual dimorphism, behavior and gamogenesis are conserved across species (Waterbury *et al.*, 1999).

The sex determination pathway of *Drosophila melanogaster* is well characterized. The key upstream sex determination regulator in *Drosophila melanogaster* is the protein product of the feminizing gene *Sex-lethal* (*Sxl*) (Cline, 1983; Salz *et al.*, 1987). In females, *Sxl* is ON orchestrating female morphological development. In males, *Sxl* expression is OFF and males develop (Cline, 1983; Salz *et al.*, 1987). However, the role of *Sxl* found in *D. melanogaster* is not conserved far beyond *Drosophila*; for example, *Sxl* does not determine sex in the closely related *Musca domestica* (house fly) (Meise *et al.*, 1998). *Musca-Sxl* is expressed in both males and females unlike the female specific expression observed in fruit flies (Meise *et al.*, 1998).

In contrast, the role of the downstream sexual morphology differentiation factor in *D. melanogaster*, *doublesex* (*dsx*), is conserved between divergent species consistent with the reverse-order theory. The *dsx* homolog, *mab-3*, is found in the nematode *Caenorhabditis elegans* and is required for masculinization (Raymond *et al.*, 1998). Both MAB-3 and DSX

are sexual differentiation factors in worms and flies and share a similar DNA binding domain (Raymond *et al.*, 1998). This indicates that the sexual differentiation function of DSX is conserved across different species. Therefore, studying DSX function in *D. melanogaster* helps in understanding the evolutionary history and phylogeny of the sex determination pathways.

Although both mammals and *Drosophila* have heterogametic sex chromosomes (XX females and XY males), the mechanisms of sexual determination are very different in mammals and *Drosophila* (Mauch & Schoenwolf, 2001). In mammals the presence or absence of the Y chromosome which carries the sex reversal locus (SRY) is the major determinant of sex; whereas, in *Drosophila* the presence or absence of the Y chromosome is unimportant for sex determination (Mauch & Schoenwolf, 2001). Prior to 2007, *Drosophila* sex determination was thought to be determined by the ratio of X chromosome (X) to autosomes (A) (Bridges, 1921, 1925). Generally, there are either one or two copies of the X chromosomes and two sets of autosomes (Bridges, 1921, 1925). If the X/autosome ratio is 1X:2A, the individual is male; whereas, when the ratio is 2X:2A the individual is female (Bridges, 1921, 1925; Mauch & Schoenwolf, 2001). The 2X:2A ratio activates the feminizing gene *Sxl* during the first two hours after fertilization (Cline, 1983; Salz *et al.*, 1987; Salz *et al.*, 1989). However, the experimental results of Erickson and Quintero (2007) has challenged this model and provided supporting evidence for an alternative idea that sex determination depends on the cumulative dosage of X-encoded signal element (XSE) proteins. Relative to the X/autosome ratio model, the alternative proposes that the cumulative dosage of XSE proteins in the cell before the cellularization stage determines

the sexual fate of the embryo during embryogenesis (Erickson & Quintero, 2007). The presence of 2 X chromosomes (2X:2A genome) results in sufficient expression of XSE proteins to initiate SXL expression before cellularization promoting female development; but a single X chromosome (1X:2A) does not produce an adequate dosage of XSE proteins before cellularization to activate SXL expression leading to male development (Erickson & Quintero, 2007).

***Drosophila* sex-determination pathway:** In male *Drosophila*, the dosage compensation mechanism increases the expression of X-linked genes by twofold to ensure male and female flies have the same dose of X-linked gene products (Grimaud & Becker, 2009). The presence of the SXL protein in the female embryo inhibits the dosage compensation mechanism by preventing the translation of the gene *male-specific lethal 2* (*msl-2*) (Lucchesi & Kuroda, 2015). SXL also activates the female-specific splicing of the RNA transcript of the downstream gene *transformer* (*tra*) for TRA expression and female somatic sex determination. In males that lack SXL expression, *tra* pre-mRNA is spliced into a mature transcript that prematurely terminates translation such that no active TRA product is expressed (Keyes *et al.*, 1992; Lucchesi & Kuroda, 2015; Robinett *et al.*, 2010). TRA interacts with the protein TRA-2 (splicing factor) that is expressed in both males and females and forms the TRA-TR2 protein complex (Ruiz & Sánchez, 2010). This TRA-TR2 complex functions as an RNA splicing factor that regulates the expression of the genes *doublesex* (*dsx*) and *fruitless* (*fru*) (Robinett *et al.*, 2010; Ruiz & Sánchez, 2010) (**Figure 10**).

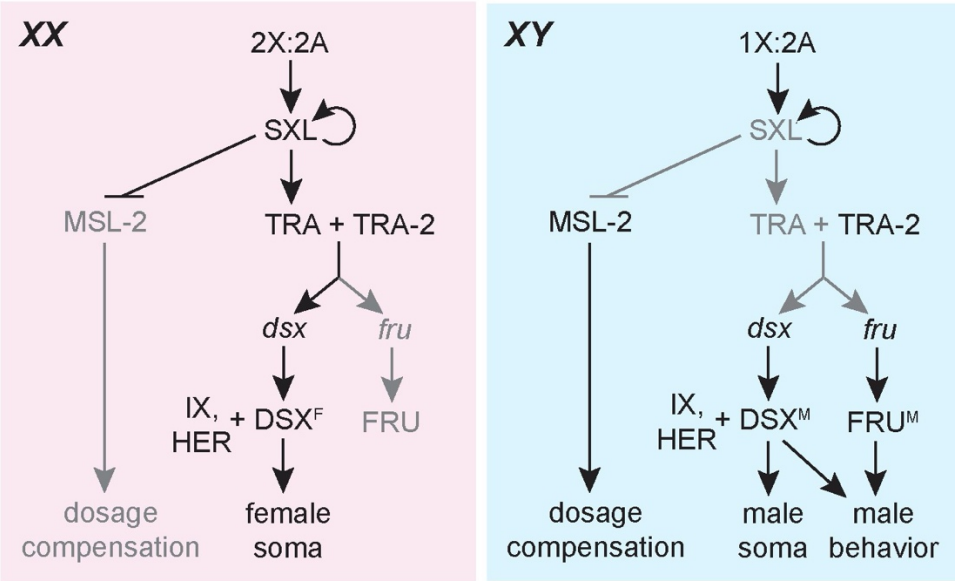


Figure 10. ***Drosophila*** **sex-determination pathway**. In males, SXL is not expressed and no TRA is expressed. In females, SXL protein inhibits the male specific dosage compensation mechanism by preventing the translation of MSL-2. SXL also activates the female-specific splicing of the transcript of the downstream gene *transformer* (*tra*). Without SXL (males), *tra* pre-mRNA is spliced to produce a transcript with a premature stop codon. In females, the protein product of *tra*, TRA, interacts with the universally expressed protein TRA-2 (splicing factor) and forms the TRA::TRA-2 protein complex (Ruiz & Sánchez, 2010). The complex functions as an RNA splicing factor that regulates the expression of the genes *doublesex* (*dsx*) and *fruitless* (*fru*) (Robinett *et al.*, 2010; Ruiz & Sánchez, 2010). IX and HER, encoded by genes *intersex* and *hermaphrodite* respectively, are two co-factors required for DSX^F to function (Robinett *et al.*, 2010). This figure has been adapted from Robinett et al., 2010 with permission.

doublesex: The *doublesex* (*dsx*) gene encodes a transcription factor required for both male and female sex determination of *Drosophila*. The *dsx* gene is transcribed in both males and females, but sex-dependent alternate splicing of exon 4 and exon 5 of *dsx* pre-mRNA produces transcripts that encode distinct DSX protein isomers (Baker *et al.*, 1987; Burtis & Baker, 1989). In females, when TRA protein is present, the *dsx* pre-mRNA is spliced to form the female-specific mature mRNA DSX^F which when subsequently translated produces the female-specific protein DSX^F (Burtis & Baker, 1989; Ryner & Baker, 1991). When the TRA protein is not present, *DSX* pre-mRNA will be spliced to form the male-specific transcript DSX^M, which when translated produces the protein DSX^M. DSX^M and DSX^F are identical for the first 397 amino acids, including the DNA binding domain (Cho & Wensink, 1997). The difference between these two DSX isoforms are the carboxyl terminal ends: DSX^M has 152-amino acid C-terminal addition and DSX^F has a distinct 30-amino acid C-terminal addition (Cho & Wensink, 1997; Yang *et al.*, 2008).

The DSX proteins are transcription factors that determine all aspects of male and female somatic sex determination (Robinett *et al.*, 2010). DSX^F activates female-specific genes promoting development of female somatic sexual characteristics and prevents the development of male somatic sexual characteristics (Ryner & Baker, 1991). Conversely, DSX^M promotes male development by activating male-specific genes and preventing female development (Ryner & Baker, 1991; Salz *et al.*, 1989). When both are absent both male and female genitalia develop.

When both DSX^F and DSX^M are expressed together, DSX^F and DSX^M compete with each other in regulating target genes (Cho & Wensink, 1997; Waterbury *et al.*, 1999). Ectopic expression of DSX^F in males (*P*[*dsxF 26B*];+/+) feminize the male genitalia; the male genitals are rotated (Waterbury *et al.*, 1999). This feminization is enhanced when one copy of the endogenous *dsx* is removed; in *P*[*dsxF 26B*];dsx^1/+, the frequency of genital rotation increases and a partial vagina forms (Waterbury *et al.*, 1999). Furthermore, male sex comb formation is affected (Demir & Dickson, 2005; Waterbury *et al.*, 1999). Expression of DSX^F in a *dsx* null mutant background (*P*[*dsxF 26B*];dsx^1/ *Df dsx*), results in flies that are transformed into pseudo-females, with the male genitalia suppressed and leaving the vagina and unformed sex combs (Waterbury *et al.*, 1999) (**Figure 11**).

In addition to morphological changes, ectopic expression of DSX^F affects fly sexual behaviors: males ectopically expressing DSX^F are courted by other males (Waterbury *et al.*, 1999). This is likely due to the expression of DSX^F inducing female pheromone production (Waterbury *et al.*, 1999). Furthermore, when DSX^F is ectopically expressed in males with *dsx* null background (*P*[*dsxF 26B*];dsx^1/ *Df dsx*); these "pseudo-females" are not only courted by, but are also copulated by wild type males (Waterbury *et al.*, 1999). Similarly, expression of DSX^M also induce masculinization in females (Rideout *et al.*, 2010).

Genetic studies have discovered four *dsx* dominant mutant alleles (dsx^{dom}) that are distinct from *dsx* loss-of-function recessive mutant alleles (Baker & Ridge, 1980; Duncan & Kaufman, 1975; Nagoshi & Baker, 1990; Nöthiger *et al.*, 1987). The dsx^{dom} alleles have

two effects: first, they do not express dsx^F function, and second, they constitutively express dsx^M (Baker & Ridge, 1980; Nagoshi & Baker, 1990). The phenotype of the dsx^{dom} alleles further suggests that the DSX^M and DSX^F TFs inhibit one another (Nagoshi & Baker, 1990). Chromosomally female (XX), dsx^{dom} hemizygous ($dsx^{dom}/Df\ dsx$) flies have male somatic phenotypes (Nagoshi & Baker, 1990). The fact that dsx^{dom} transforms heterozygous females ($dsx^{dom}/+$) to a sex neutral phenotype suggests that DSX^M inhibits DSX^F (Nagoshi & Baker, 1990).

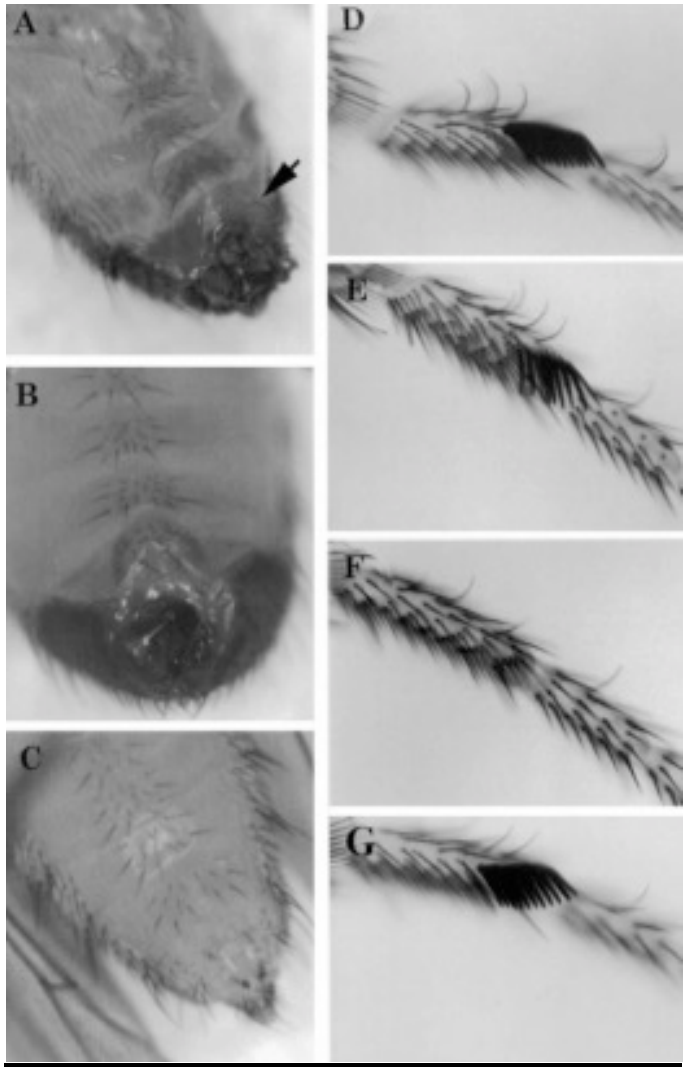


Figure 11. **Effects of DSX^{F} ectopic expression.** DSX^{F} expressed in males feminizes the male genital and sex combs. The feminization is enhanced when the endogenous *dsx* gene is removed. A, D: *P*[*dsxF 26B*];+/+; B, E: *P*[*dsxF 26B*];*dsx*^{1}/+; C, F: *P*[*dsxF 26B*];*dsx*^{1}/ *Df dsx*. G is WT male. This figure has been adapted from Waterbury et al., 1999 with permission.

dsx ^{1}/ dsx^{Gal4}: In this thesis two *dsx* alleles are used. The *dsx^{1}* allele is an amorphic allele of the *dsx* gene (Ota *et al.*, 1981). The *dsx^{Gal4}* allele (Robinett *et al.*, 2010) is a targeted insertion of GAL4 sequence into *dsx* disrupting the *dsx* gene and creating a *dsx* amorphic allele (Rideout *et al.*, 2010; Robinett *et al.*, 2010).

fruitless: The gene *fruitless* (*fru*) is a part of the sex determination regulatory hierarchy. The FRU^{M} protein is necessary and sufficient for male courtship behavior (Demir & Dickson, 2005). The *fru* gene is a complex gene with multiple promoters (P1-4) that express primary transcripts that undergo extensive alternative splicing (Anand *et al.*, 2001; Ryner *et al.*, 1996). The mature transcripts encode several distinct transcription factors (Anand *et al.*, 2001; Ryner *et al.*, 1996; Stockinger *et al.*, 2005). The transcript initiated from the P1 promoter undergoes sex-specific alternative splicing regulated by the TRA::TRA2 complex (Anand *et al.*, 2001; Robinett *et al.*, 2010; Ryner *et al.*, 1996; Stockinger *et al.*, 2005). In females, the TRA-TRA2 complex splices the *fru* P1 pre-mRNA to produce a mature mRNA that encodes a non-functional product (Ryner *et al.*, 1996). In males TRA is not expressed and the *fru* P1 pre-mRNA is spliced to give mature mRNAs that express multiple isoforms of the FRU^{M} protein (Anand *et al.*, 2001; Robinett *et al.*, 2010; Stockinger *et al.*, 2005). The FRU^{M} proteins are expressed in 2% of in the male central nervous system (CNS) neurons and are necessary and sufficient for the male sexual behavior (mating and courtship) (Demir & Dickson, 2005).

***fru$^{4-40}$* and *fru^{Gal4}* alleles**: In my thesis I use the *fru$^{4-40}$* allele and two *fru^{Gal4}* alleles. *fru$^{4-40}$* is a *fru* deficiency allele derived from the imprecise excision of a P- element from the allele

fru^{4} (Anand *et al.*, 2001). The P-element insertion of the *fru*^{4} allele is between the 5' exon of the P2 and P3 promoters in the *fru* locus (Anand *et al.*, 2001). The deletion caused by the imprecise excision of the P-element extends distally from the P-element insertion site for at least 70 kb removing DNA sequences upstream of the P3 promoter, including the P1 and P2 promoter. The P1 and P2 transcripts are not expressed from the *fru*$^{4\text{-}40}$ allele (Anand *et al.*, 2001). As the male P1 transcript is spliced to produce FRU^{M}, this deletion influences male courtship behaviors. The male flies heterozygous for the *fru*$^{4\text{-}40}$ allele (*fru*$^{4\text{-}40}$ / *fru*$^{+}$) are fertile but the *fru*$^{4\text{-}40}$/ *fru*^{1} hemizygous male flies (*fru*^{1} is another *fru* allele affecting expression of the P1 transcript) are sterile (Anand *et al.*, 2001). Furthermore, *fru*$^{4\text{-}40}$/ *fru*^{1} do not court females but will court males (Fan *et al.*, 2013). The two *fru*^{Gal4} alleles used in thesis are *fru*^{Gal4A} (Stockinger *et al.*, 2005) and *fru*^{Gal4B} (Kimura *et al.*, 2005). The *fru*^{Gal4A} was created by inserting the Gal4 sequence into the sex-specifically spliced exon, exon S, so that the transcript driven by the P1 promoter encodes GAL4 rather than FRU^{M}. The *fru*^{Gal4B} is an insertion of a P-element carrying *Gal4* in intron 2 of the *fru* locus. The *fru*^{Gal4B} is a loss-of-function-allele of *fru* with no detectable expression of male-specific *fru* transcript in the *fru*^{Gal4B} homozygous individual (Kimura *et al.*, 2005).

2.1 Specificity of transcription factor function: TFs mediate the transcription of the genes through interactions with DNA. A defining feature of TFs is that they usually contain at least one DNA-binding domain that recognizes a specific sequence (Mitchell & Tjian, 1989; Ptashne & Gann, 1997). The study of gene regulation by TFs dates back to the 1960s (Gilbert & Müller-Hill, 1966; Jacob & Monod, 1961; Ptashne, 1967a, 1967b). There has been a long lasting question ever since regarding how a TF locates a specific DNA

sequence in a complex genome (Kribelbauer *et al.*, 2019). Initially, it seemed that the DNA sequence preference of any TF could be predicted by a simple model (Pabo & Sauer, 1984). As more TF structures and DNA sequences were analyzed, it became obvious that TFs use a more complex system of mechanisms to recognize specific DNA sequences (Garvie & Wolberger, 2001; Luscombe *et al.*, 2001) and that a simple model for the prediction of the DNA sequence recognized may not exist (Pabo & Sauer, 1992; Slattery *et al.*, 2014).

The majority of the early models of TF structures and their DNA sequence preference came from the study of prokaryotes (Kribelbauer *et al.*, 2019). However, comparing DNA recognition of prokaryotic TFs with those of eukaryotic TFs reveals a major difference (Smith & Matthews, 2016). The DNA sequences bound by prokaryotic TFs are long enough for prokaryotic TFs to find specific binding sites in the genome (Wunderlich & Mirny, 2009). However, eukaryotic TFs recognize shorter DNA sequences 6-8 bp long, such that the DNA sequences do not contain enough information for eukaryotic TFs to find specific sites in the genome (Berger *et al.*, 2008; Wunderlich & Mirny, 2009). A mechanism proposed to alleviate this problem is that eukaryotic TF proteins interact cooperatively to increase the size and information of sequence recognized; for example, HOX proteins bind to the cofactor, Extradenticle (EXD) (Ryoo *et al.*, 1999). However, even though the size of the sequence recognized increases (relative to the size the genome), the sequence recognized is still very small such that spurious binding to multiple binding sites in the genome is still high. Although there are many ideas regarding how eukaryotic TFs may work to bring about the expression of specific genes, the central problem of how eukaryotic TFs work remains unanswered.

2.2 Phenotypic non-specificity of transcription factor function: Phenotypic non-specificity is a phenomenon where the phenotype(s) associated with the expression of a particular TF is induced or rescued by multiple distinct TFs. Phenotypic non-specificity of TF function is observed within and between TF families (Banreti *et al.*, 2014; Greig & Akam, 1995; Hirth *et al.*, 2001; Lelli *et al.*, 2011; Percival-Smith, 2017; Percival-Smith & Laing Bondy, 1999; Percival-Smith *et al.*, 2013; Percival-Smith *et al.*, 2005). For example, both HD containing TFs and non-HD TFs when ectopically expressed induce the eyeless, wingless and ectopic first thoracic beard phenotypes (Percival-Smith, 2017). Furthermore, the reduced maxillary palp phenotype caused by *pb*-null alleles is partially rescued by expression of a non-PB-homologous protein, DSX^{M}, which does not contain the HD domain (Percival-Smith, 2017). These observations cannot be explained by traditional models of TF function which emphasize the functional specificity of TF function. These observations led to the proposal of a model of limited specificity of transcription factor function. The novel "limited specificity model" helps to explain the phenomena of phenotypic non-specificity.

2.3 Model of limited specificity: The model of limited specificity was proposed as an explanation for the phenomenon of phenotypic non-specificity of TF function (Percival-Smith, 2018) and can be contrasted with the following alternatives: (1) complete non-specificity model — DNA binding domains of TFs have no preference to the DNA sequence and the cooperative interactions between TFs is not specific; (2) complete specificity model — the DNA sequence recognition is restricted to very specific sequences due to TF cooperative interactions being restricted to a small number of proteins. The

model of limited specificity proposes that the specificity of DNA sequence recognition and cooperative interactions between TFs is limited and this level of specificity is not sufficient to target the expression of a certain set of genes required for a particular phenotype (Percival-Smith, 2018).

The model of limited specificity explains phenotypic non-specificity of TF function. Limited specificity proposes that a TF regulates a set of genes much larger than the sub-set required for a certain phenotype. For example, TFa may be required for the expression of 200 genes but the expression of only a small group of these genes may be required for a certain phenotype. When other TFs are expressed in place of TFa they regulate sets of hundreds of genes, and every so often a TFb is able to regulate a set of genes that includes the small group of genes required for the phenotype resulting in rescue of the phenotype. For example, when DSX^M rescues the growth of maxillary palp development in the absence of PB, DSX^M is proposed to regulate the subset of PB regulated genes required for maxillary palp growth.

2.4 Pleiotropy: A pleiotropic gene encodes a product that is required for more than one phenotypic trait (Carroll *et al.*, 2004). Therefore, mutant alleles in pleiotropic genes affect more than one trait. For example, PB requirement for mouthparts development is pleiotropic being required for both maxillary palp growth/differentiation and for proboscis development in the suppression of tarsus determination and promotion of proboscis development. In uniform-pleiotropy, all hypomorphic loss-of-function alleles affect the different phenotypes to a similar degree such that the order of severity of the alleles for the

different phenotypes is the same for all phenotypes (Sivanantharajah & Percival-Smith, 2009). In differential pleiotropy, hypomorphic loss-of-function alleles have a differential effect on the phenotypes, such that the order of severity of the phenotypes is distinct between the phenotypes. The observation of differential pleiotropy in a locus encoding a TF suggests that the TF functional elements are dispersed as small protein elements throughout the protein and each of these elements make small, tissue specific contributions to overall TF function (Hittinger *et al.*, 2005; Joshi *et al.*, 2007; Merabet *et al.*, 2011; Percival-Smith *et al.*, 2013; Prince *et al.*, 2008; Sivanantharajah & Percival-Smith, 2009, 2014, 2015; Tour *et al.*, 2005).

2.5 Functional conservation: The term functional conservation of TF function means that the TF and its orthologs from another species have similar function. In the experiments designed to test TF functional conservation, the phenotype which depends on the expression of a specific TF in a species is assessed with the expression of the TF's orthologs from another species (Halder *et al.*, 1995; Hunter & Kenyon, 1995; Lutz *et al.*, 1996; Malicki *et al.*, 1990; Percival-Smith & Laing Bondy, 1999; Zhao *et al.*, 1993). This analysis of the functional conservation of TF function is based on an implicit presumption in the experimental design that TF function is specific for the regulation of the specific sets of genes required for the phenotype and that if the ortholog is functionally conserved it would regulate the same set of genes. The assumption of TF functional specificity is the foundation of the assessment of functional conservation. However, phenotypic non-specificity of TF function undermines the interpretation of these experiments by showing that non-orthologous and non-paralogous TFs induce or rescue the phenotype. For example:

the murine ortholog of PB HOXA2 partially rescues the *pb*-null phenotype in *Drosophila* suggesting functional conservation of HOXA2 and PB function. However, the partial rescue by DSX^{M} undermines the interpretation of functional conservation of PB and HOXA2 function. Therefore, claims of functional conservation of TF function during evolution require reconsideration because the underlying presumption in the experimental design may have no foundation.

3. Genome editing tools: The terms "genome editing or genomic engineering" or "gene editing" refer to genetic engineering methodologies that modify or replace DNA sequences in an organism's genome (Esvelt & Wang, 2013). Currently, the most commonly used genome editing techniques are (1) clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9), (2) transcription activator-like effector nucleases (TALENs), (3) zinc-finger nucleases (ZFNs), and (4) homing endonucleases or meganucleases (Gaj *et al.*, 2016).

Clustered Regularly Interspaced Palindromic Repeat (CRISPR): The technique, Clustered Regularly Interspaced Palindromic Repeat (CRISPR), is a well-known genome-editing tool used to achieve efficient and targeted genetic modification in *Drosophila* and other model and non-model organisms (Cho *et al.*, 2013; Cong *et al.*, 2013; DiCarlo *et al.*, 2013; Friedland *et al.*, 2013; Gratz *et al.*, 2013; Hwang *et al.*, 2013; Wang *et al.*, 2013). CRISPR arrays were first discovered in the genome of *Escherichia coli* in 1987 and later these arrays were found to participate in the prokaryotic adaptive immunity system which is used to suppress infection by foreign genetic elements (Barrangou *et al.*, 2007; Bhaya *et*

al., 2011; Ishino *et al.*, 1987). In a CRISPR II system, when the foreign genetic element invades a bacterial cell, CRISPR RNA (crRNA) that is complementary to the invading nucleic acid, and a constitutively expressed *trans*-activating CRISPR RNA (tracrRNA) direct the CRISPR-associated nuclease (Cas) to introduce a site specific double-strand-break (DSB) in the exogenous invading DNA (Bhaya *et al.*, 2011; Gaj *et al.*, 2013). The type II CRISPR system has been adapted to create a highly efficient genome editing tool for generating a site-specific DSB —– CRISPR/Cas9 system (Gratz *et al.*, 2013; Wang *et al.*, 2013).

In the CRISPR/Cas9 genome editing system, the RNA-guided CRISPR-associated nuclease, Cas9, isolated from bacterial *Streptococcus pyogenes* (Gratz *et al.*, 2013) and a genetically engineered chimeric RNA (chiRNA) (a fusion of crRNA and tracrRNA also referred to as guide RNA or gRNA) containing the complementary sequence to the target site of the host genome interacts with the Cas9 protein to direct a specific DSB in the host genome (Boucherat *et al.*, 2013; Shen *et al.*, 2004) (**Figure 12**). Target-site recognition relies on Watson-Crick base pairing between the spacer of chiRNA and one strand of the target DNA (protospacer), which is immediately followed by a "NGG" tri-nucleotide protospacer adjacent motif (PAM) recognized by Cas9 (Bhaya *et al.*, 2011). Following recognition Cas9 initiates a DSB that will be repaired either by non-homologous end joining (NHEJ) or homologous directed repair (HDR) (Gratz *et al.*, 2013). If a homologous DNA template used to repair the DNA contains genetic modifications, the modifications will be incorporated into the genome through HDR, thereby editing the genome.

The CRISPR/Cas9 system has proven to be an invaluable research tool and has greatly accelerated the generation of new insights into the function and regulation of biological systems (Adli, 2018; Zhang *et al.*, 2015). However, there are still limitations of this genome editing technique. First, the "off-target" effects of CRISPR are a major concern. It has been reported that the frequency of unintended DNA modifications at untargeted genomic sites created by CRISPR/Cas9 is greater than 50% of the intended on-target modifications (Cho *et al.*, 2014; Corrigan-Curay *et al.*, 2015; Fu *et al.*, 2013; Hsu *et al.*, 2013; Mali *et al.*, 2013; Zhang *et al.*, 2015) questioning the precision of CRISPR/Cas9 editing. Moreover, the ability to trigger HDR after the DSB is a factor influencing successful gene editing. If the DSB is at an active euchromatic gene region, the epigenetic marker, H3K9me3, will recruit the homologous recombination enzymes to the damage site and promote HDR, whereas NHEJ will most likely be triggered if the DSB is in a silenced gene region (Aymard *et al.*, 2014)

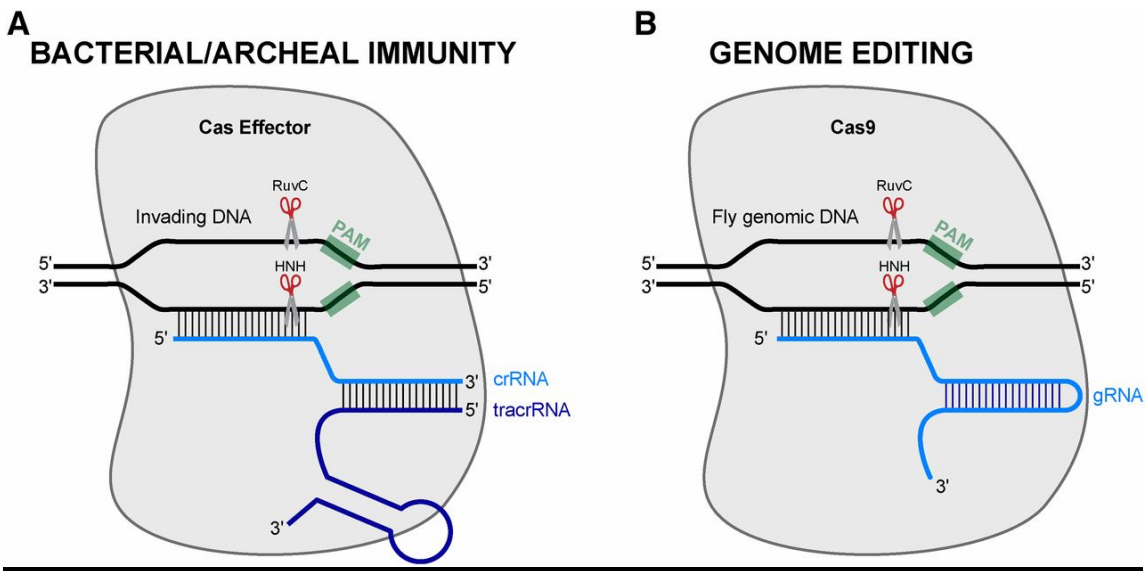


Figure 12. CRISPR/Cas9 system. **(A)** crRNA and tracrRNA direct the CRISPR-associated nuclease (Cas) to introduce site specific DSBs in the exogenous invading DNA **(B)** The CRISPR/Cas9 system only requires an engineered chiRNA or gRNA (Bier *et al.*, 2018) in order to recognize a specific sequence and induce a DSB. RuvC (Recombination UV C) and HNH (Histidine-Asparagine-Histidine) are the endonuclease domains of Cas9 protein. This figure is adapted from Bier et al, 2018 with permission.

***Recombinase-Mediated Cassette Exchange* (*RMCE*)** RMCE is the exchange of specific DNA segments between two DNA molecules. RMCE allows the incorporation of any DNA sequence at a single position in the genome. The recombinase, ΦC31 integrase, catalyzes the specific and unidirectional exchange of DNA cassettes between *att* site-specific recombination sites (*attP* site and *attB* site), and is used for RMCE mediated introduction of DNA into the genome (Groth *et al.*, 2004). Two 39 base pair sequences, the *attP* sites, are present on the *Drosophila* chromosome and serve as a landing site. The ΦC31 integrase facilitates the precise integration of DNA flanked by two *attB* sites carried on a plasmid

into the genome at the landing site through site-specific recombination between *attP* and *attB* sites (Bischof *et al.*, 2007) (**Figure 13**). After recombination, the *attP* and *attB* sites are converted to *attR* and *attL* sites (Bateman *et al.*, 2006; Bateman & Wu, 2008; Groth *et al.*, 2004) and because the ΦC31 integrase alone cannot recognize these *attL* and *attR* sites, the exchanged DNA is stably inherited.

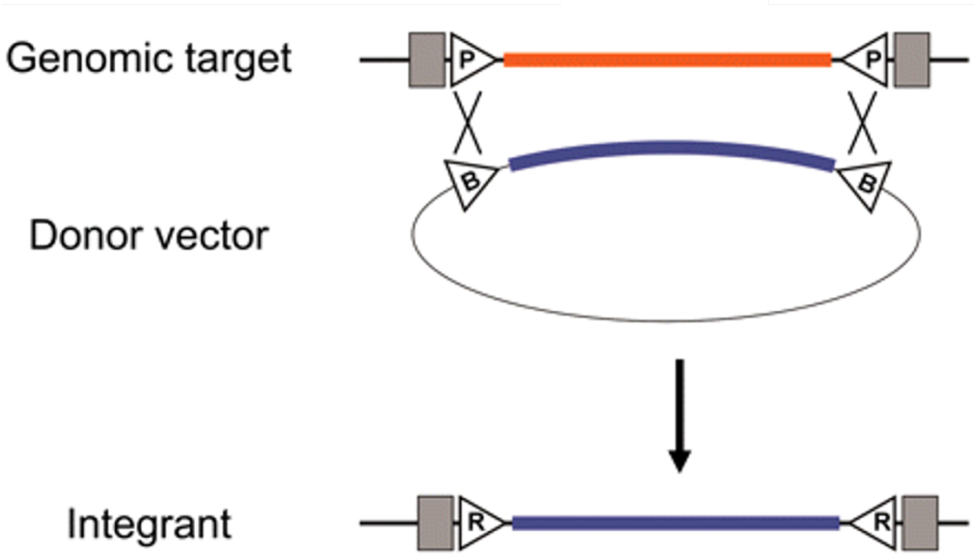


Figure 13. RMCE system. The ΦC31 integrase facilitates the integration of vector DNA (blue) flanked by two *attB* sites into the genome through the recombination between *attP* and *attB* sites. The genome sequence (orange) is replaced with the vector DNA after the recombination. This figure is adapted from Bateman & Wu, 2008 with permission.

yellow: The *yellow* gene (*y*) is located on the tip of the X chromosome (1B1). The y^1 allele is a *yellow* gene amorphic allele caused by an A to C transversion at the start codon (ATG) of the *yellow* open reading frame (Geyer *et al.*, 1990). The visible phenotype of y^- individuals is the yellow pigmentation of the adult cuticle and larval setae and mouthparts (Biessmann, 1985) (**Figure 14**).

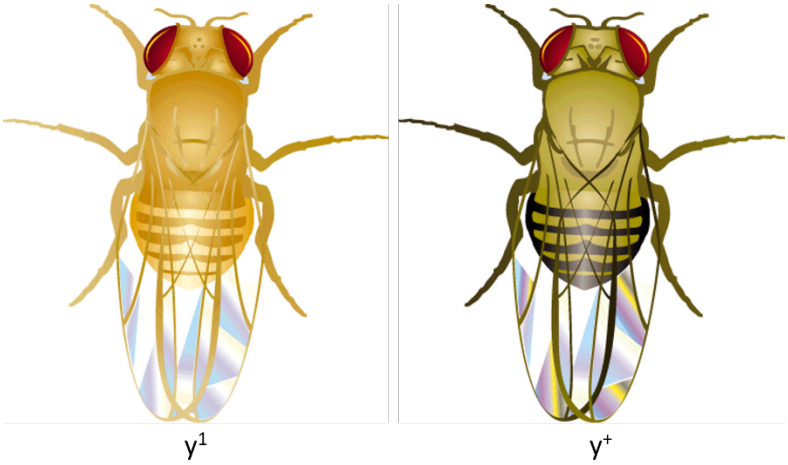


Figure 14. The pigmentation difference between y^1 and y^+ flies. This figure was obtained from https://annex.exploratorium.edu/exhibits/mutant_flies/mutant_flies.html

white: The *white* gene in *Drosophila* was first identified by Thomas Morgan in 1910 (Morgan, 1910), and it codes for an ATP cassette transporter which transports guanine and tryptophan (the red and brown eye pigment precursors) into the developing eye tissue during pupation (Mackenzie *et al.*, 1999). Loss-of-function alleles in *white* cause the *Drosophila* eye pigmentation to change from red to white (**Figure 15**). The amorphic w^{67c23} allele was derived from the allele, *white-crimson* (w^c), which itself is a partial revertant of the *white-ivory* (w^i) allele (Collins & Rubin, 1982, 1984). The w^i allele results from a 2.96 kb tandem duplication of *white* sequence (from intron 1 to exon 3) in the *white* locus

(Bhadra *et al.*, 1997; Sabl & Birchler, 1993; Suárez *et al.*, 1996). The w^c allele results from the insertion of a 10kb FB element into the w^i duplication and reverts to w^i or generates novel *w* mutant alleles at a high frequency (Collins & Rubin, 1982, 1984). The amorphic w^{67c23} allele is a deletion derivative of the w^c allele in which several hundred kb of DNA upstream of the *white* gene is deleted (including the first exon, start codon and promoter region) (Moschetti *et al.*, 2004). In addition, a transposon-like element, NOF, flanked by two FB transposable elements is introduced between the deletion breakpoints such that the FB-NOF-FB element is directly upstream of the exon 2 sequence of the *white* gene (Moschetti *et al.*, 2004).

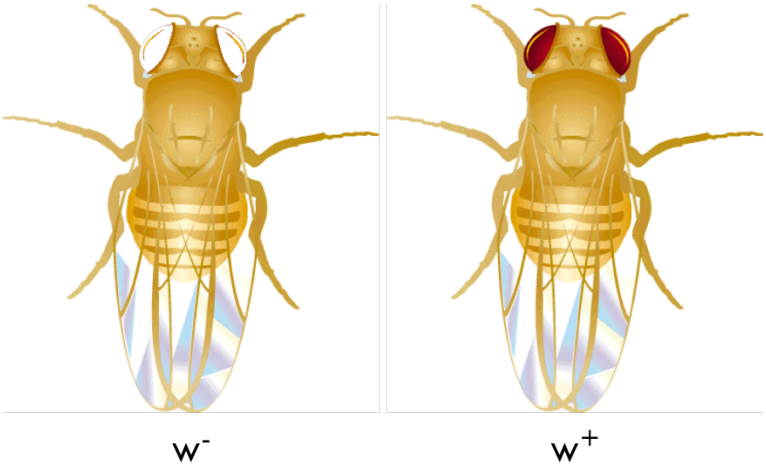


Figure 15. The pigmentation difference between w^- and w^+ flies. This figure was obtained from https://annex.exploratorium.edu/exhibits/mutant_flies/mutant_flies.html

***Transposable elements*:** Transposable elements (TEs) are DNA sequences that can move their positions in the genome. TE genetic elements contribute significantly to genetic variation in all living organisms (Capy, 1998). The structure and biochemistry of

transposition categorize TE into two groups (Craig *et al.*, 2002; Finnegan, 1992). Class I elements, or retrotransposons, move via an RNA intermediate and reverse-transcription and usually possess long terminal repeats. In contrast, Class II elements transpose from one position to another via a DNA intermediate and have inverted repeats at their ends. Foldback (FB) elements are a group of poorly described TEs that have not been assigned to a particular class (Feschotte & Pritham, 2007). Their mechanism of transposition is unknown. The inverted terminal repeats of FB sequences contain different numbers of short direct repeats. This characteristic of FB elements allows the formation of extensive secondary DNA structures which cause DNA modifications like deletions, duplications and other chromosomal rearrangements at high frequency (Kaminker *et al.*, 2002; Potter *et al.*, 1980)

FB mediated w^{67c23} allele migration: The w^{67c23} allele is an amorphic allele. Flies with the w^{67c23} allele exhibit a w^- phenotype. However, flies with the w^{67c23} allele are able to revert to w^+ via migration/transposition of the w^{67c23} DNA to a new location. This migration was observed after screening for w^+ flies upon injection of the *pBari1_47D*w^+ plasmid into the w^{67c23} fly embryos. In three w^+ fly lines, the endogenous w^{67c23} allele flanked by FB elements was found to have transposed into introns located in three different genes (**Figure 16**) (Moschetti *et al.*, 2004). The transposed w^{67c23} allele utilized the promoter of the gene creating a chimeric mRNA that included *white* and in addition the first AUG of the gene was used to initiate translation to create an active chimeric protein containing W protein sequence (Moschetti *et al.*, 2004).

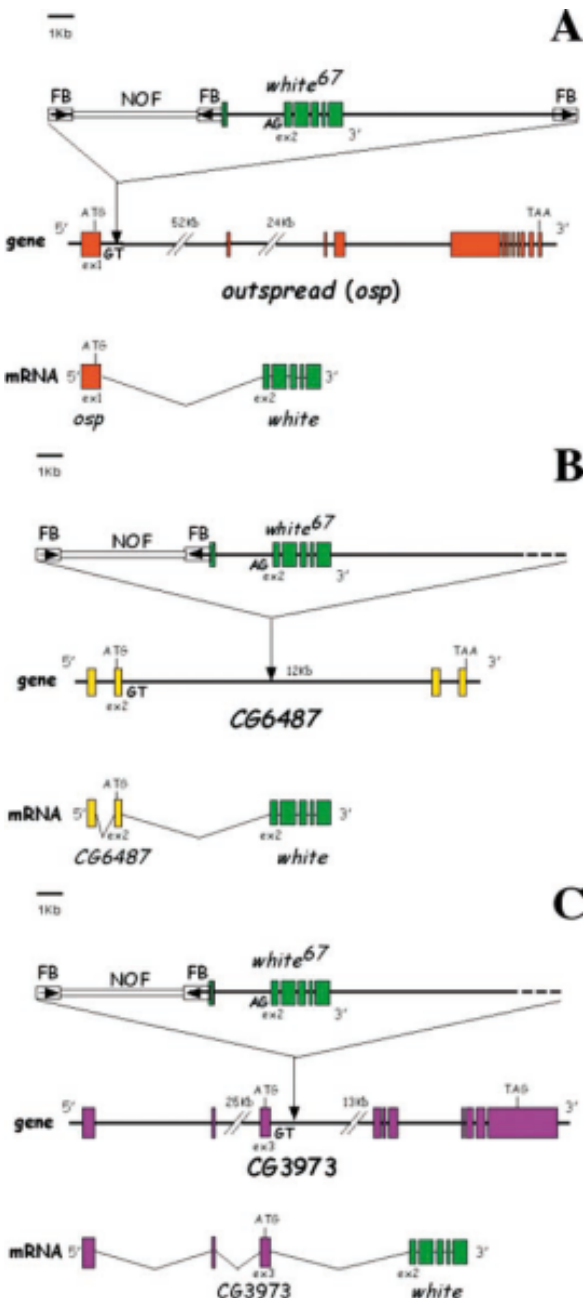


Figure 16. Three examples of FB mediated w^{67c23} transposition. After injection of *pBari1_47Dw*$^{+}$ plasmid, the w^{67c23} allele was observed to migrate to different locations in the genome. Three different gene promoters drive the expression of the w^{67c23} allele, each creating a chimeric *white* mRNA that expresses a White protein. This figure has been adapted from Moschetti et al. 2004 with permission.

4. Objectives: The original objective of this thesis was to study both the conservation of Sex combs reduced transcription factor functions and phenotypic non-specificity of transcription factor function. The original experimental design was to integrate resident and non-resident TFs through RMCE into a *Scr* locus that had been edited using CRISPR and HDR. However, genome editing with CRISPR identified three w^+ transformants that were not the result of homologous recombination. Thus, the final objectives of this thesis were revised to the following:

1. To characterize the CRISPR induced non-homologous recombination transformants (Chapter 2).
2. To study transcription factor functional non-specificity using functional complementation through the application of the *UAS-GAL4* expression system (Chapter3).

4.1 Chapter 2: The specific objective of this chapter was to characterize the origins of the three w^+ transformants identified using CRISPR. I hypothesized that these three transformants were created by non-homologous recombination events induced by CRISPR in the *Drosophila melanogaster* genome. My goal was to identify the position of integration of *white* DNA and determine the potential mechanism of the non-homologous recombination events.

4.2 Chapter 3: The specific objective of this chapter was to study TF functional non-specificity using functional complementation with the (*UAS-GAL4*) system. I hypothesized that phenotypic non-specificity of TF function would be commonly observed due to the

limited specificity of transcription factor function. To test this hypothesis. I screened 12 non-resident TFs for rescue of six target *TF* loci using the *UAS-GAL4* system for phenotypic non-specificity.

REFERENCE

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., & Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science, 287*(5461), 2185-2195.

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications, 9*(1), 1911.

Anand, A., Villella, A., Ryner, L. C., Carlo, T., Goodwin, S. F., Song, H.-J., Gailey, D. A., Morales, A., Hall, J. C., Baker, B. S., & Taylor, B. J. (2001). Molecular Genetic Dissection of the Sex-Specific and Vital Functions of the *Drosophila melanogaster* Sex Determination Gene *fruitless*. *Genetics, 158*(4), 1569-1595.

Ashburner, M., Golic, K. G., & Hawley, R. S. (2005). *Drosophila : a laboratory handbook*. Cold Spring Harbor Laboratory Press.

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., & Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature Structural & Molecular Biology, 21*(4), 366-374.

Baker, B. S., Nagoshi, R. N., & Burtis, K. C. (1987). Molecular genetic aspects of sex determination in Drosophila. *BioEssays, 6*(2), 66-70.

Baker, B. S., & Ridge, K. A. (1980). Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. *Genetics, 94*(2), 383-423.

Banreti, A., Hudry, B., Sass, M., Saurin, A. J., & Graba, Y. (2014). Hox proteins mediate developmental and environmental control of autophagy. *Dev Cell, 28*(1), 56-69.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., & Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell, 144*(2), 214-226.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science, 315*(5819), 1709.

Bateman, J. R., Lee, A. M., & Wu, C. t. (2006). Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. *Genetics, 173*(2), 769-777.

Bateman, J. R., & Wu, C. t. (2008). A Simple Polymerase Chain Reaction-Based Method for the Construction of Recombinase-Mediated Cassette Exchange Donor Vectors. *Genetics, 180*(3), 1763.

Bateson, W., Pierpont, F. H., & Clay, C. J. (1894). *Materials for the study of variation treated with especial regard to discontinuity in the origin of species*. Macmillan and co.

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S. A., Morris, Q. D., Bulyk, M. L., & Hughes, T. R. (2008). Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell, 133*(7), 1266-1276.

Bhadra, U., Pal-Bhadra, M., & Birchler, J. A. (1997). A Sex-Influenced Modifier in Drosophila That Affects a Broad Spectrum of Target Loci Including the Histone Repeats. *Genetics, 146*(3), 903.

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics, 45*(1), 273-297.

Bier, E., Harrison, M. M., O'Connor-Giles, K. M., & Wildonger, J. (2018). Advances in Engineering the Fly Genome with the CRISPR-Cas System. *Genetics, 208*(1), 1.

Biessmann, H. (1985). Molecular analysis of the yellow gene (y) region of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America, 82*(21), 7369-7373.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., & Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proceedings of the National Academy of Sciences of the United States of America, 104*(9), 3312-3317.

Bokor, P., & DiNardo, S. (1996). The roles of hedgehog, wingless and lines in patterning the dorsal epidermis in Drosophila. *Development, 122*(4), 1083.

Boucherat, O., Montaron, S., Bérubé-Simard, F.-A., Aubin, J., Philippidou, P., Wellik, D. M., Dasen, J. S., & Jeannotte, L. (2013). Partial functional redundancy between Hoxa5 and Hoxb5 paralog genes during lung morphogenesis. *American Journal of Physiology - Lung Cellular and Molecular Physiology, 304*(12), L817-L830.

Bownes, M., Abrahamssen, N., Gilman, C., Howden, R., Martinez, A., & Slee, R. (1989). Drosophila: A Laboratory Handbook. By Michael Ashburner. New York: Cold Spring Harbor Laboratory. 1989.

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

Bridges, C. B. (1915). A linkage variation in Drosophila. *Journal of Experimental Zoology, 19*(1), 1-21.

Bridges, C. B. (1921). TRIPLOID INTERSEXES IN DROSOPHILA MELANOGASTER. *Science, 54*(1394), 252.

Bridges, C. B. (1925). Sex in Relation to Chromosomes and Genes. *The American Naturalist, 59*(661), 127-137.

Brown, S., Holtzman, S., Kaufman, T., & Denell, R. (1999). Characterization of the TriboliumDeformed ortholog and its ability to directly regulate Deformed target

genes in the rescue of a DrosophilaDeformed null mutant. *Development Genes and Evolution, 209*(7), 389-398.

Burtis, K. C., & Baker, B. S. (1989). Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell, 56*(6), 997-1010.

Capovilla, M., Brandt, M., & Botas, J. (1994). Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. *Cell, 76*(3), 461-475.

Capy, P. (1998). *Dynamics and evolution of transposable elements*. Springer.

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2004). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Wiley.

Casanova, J., Sánchez-Herrero, E., & Morata, G. (1985). Prothoracic transformation and functional structure of the *Ultrabithorax* gene of Drosophila. *Cell, 42*(2), 663-669.

Castelli-Gair, J. E., & García-Bellido, A. (1990). Interactions of Polycomb and trithorax with cis regulatory regions of Ultrabithorax during the development of Drosophila melanogaster. *The EMBO journal, 9*(13), 4267-4275.

Castelli-Gair, J. E., Micol, J. L., & García-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.

Cho, S., & Wensink, P. C. (1997). DNA Binding by the Male and Female Doublesex Proteins of Drosophila melanogaster. *Journal of Biological Chemistry, 272*(6), 3185-3189.

Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology, 31*(3), 230-232.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research, 24*(1), 132-141.

Cline, T. W. (1983). The interaction between daughterless and sex-lethal in triploids: A lethal sex-transforming maternal effect linking sex determination and dosage compensation in Drosophila melanogaster. *Developmental Biology, 95*(2), 260-274.

Cole, E., & Palka, J. (1982). The pattern of campaniform sensilla on the wing and haltere of Drosophila melanogaster and several of its homeotic mutants. *Journal of embryology and experimental morphology, 71*, 41-61.

Collins, M., & Rubin, G. M. (1982). Structure of the drosophila mutable allele, *white-crimson*, and its *white-ivory* and wild-type derivatives. *Cell, 30*(1), 71-79.

Collins, M., & Rubin, G. M. (1984). Structure of chromosomal rearrangements induced by the FB transposable element in Drosophila. *Nature, 308*(5957), 323-327.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819.

Corrigan-Curay, J., O'Reilly, M., Kohn, D. B., Cannon, P. M., Bao, G., Bushman, F. D., Carroll, D., Cathomen, T., Joung, J. K., Roth, D., Sadelain, M., Scharenberg, A. M., von Kalle, C., Zhang, F., Jambou, R., Rosenthal, E., Hassani, M., Singh, A., & Porteus, M. H. (2015). Genome Editing Technologies: Defining a Path to Clinic: Genomic Editing: Establishing Preclinical Toxicology Standards, Bethesda, Maryland 10 June 2014. *Molecular Therapy, 23*(5), 796-806.

Craig, N. L., Craigie, R., Gellert, M., & Lambowitz, A. M. (2002). *Mobile DNA II*. American Society of Microbiology.

Curtis, C. D., Brisson, J. A., DeCamillis, M. A., Shippy, T. D., Brown, S. J., & Denell, R. E. (2001). Molecular characterization of Cephalothorax, the Tribolium ortholog of Sex combs reduced. *genesis, 30*(1), 12-20.

Demir, E., & Dickson, B. J. (2005). *fruitless* Splicing Specifies Male Courtship Behavior in *Drosophila*. *Cell, 121*(5), 785-794.

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research, 41*(7), 4336-4343.

Diederich, R. J., Merrill, V. K., Pultz, M. A., & Kaufman, T. C. (1989). Isolation, structure, and expression of labial, a homeotic gene of the Antennapedia Complex involved in Drosophila head development. *Genes & Development, 3*(3), 399-414.

Dinges, G. F., Chockley, A. S., Bockemühl, T., Ito, K., Blanke, A., & Büschges, A. (2021). Location and arrangement of campaniform sensilla in Drosophila melanogaster. *Journal of Comparative Neurology, 529*(4), 905-925.

dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., Emmert, D. B., Gelbart, W. M., & FlyBase, C. (2015). FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome

assembly and large-scale migration of genome annotations. *Nucleic acids research, 43*(Database issue), D690-D697.

Driever, W., & Nüsslein-Volhard, C. (1988). The *bicoid* protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell, 54*(1), 95-104.

Drosophila 12 Genomes, C., Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., Akashi, H., Anderson, W. W., Aquadro, C. F., Ardell, D. H., Arguello, R., Artieri, C. G., Barbash, D. A., Barker, D., Barsanti, P., Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S. A., Bradley, R. K., Brand, A. D., Brent, M. R., Brooks, A. N., Brown, R. H., Butlin, R. K., Caggese, C., Calvi, B. R., Bernardo de Carvalho, A., Caspi, A., Castrezana, S., Celniker, S. E., Chang, J. L., Chapple, C., Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S. W., Comeron, J. M., Costello, J. C., Coyne, J. A., Daub, J., David, R. G., Delcher, A. L., Delehaunty, K., Do, C. B., Ebling, H., Edwards, K., Eickbush, T., Evans, J. D., Filipski, A., Findeiß, S., Freyhult, E., Fulton, L., Fulton, R., Garcia, A. C., Gardiner, A., Garfield, D. A., Garvin, B. E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J., Good, R., Gotea, V., Gravely, B., Greenberg, A. J., Griffiths-Jones, S., Gross, S., Guigo, R., Gustafson, E. A., Haerty, W., Hahn, M. W., Halligan, D. L., Halpern, A. L., Halter, G. M., Han, M. V., Heger, A., Hillier, L., Hinrichs, A. S., Holmes, I., Hoskins, R. A., Hubisz, M. J., Hultmark, D., Huntley, M. A., Jaffe, D. B., Jagadeeshan, S., Jeck, W. R., Johnson, J., Jones, C. D., Jordan, W. C., Karpen, G. H., Kataoka, E., Keightley, P. D., Kheradpour, P., Kirkness, E. F., Koerich, L. B., Kristiansen, K., Kudrna, D., Kulathinal, R. J., Kumar, S., Kwok, R., Lander, E., Langley, C. H., Lapoint, R., Lazzaro, B. P., Lee, S. J., Levesque, L., Li, R., Lin, C. F., Lin, M. F., Lindblad-Toh, K., Llopart, A., Long, M., Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C. A., Makalowski, W., Marzo, M., Matsuda, M., Matzkin, L., McAllister, B., McBride, C. S., McKernan, B., McKernan, K., Mendez-Lago, M., Minx, P., Mollenhauer, M. U., Montooth, K., Mount, S. M., Mu, X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M. A., O'Grady, P., Pachter, L., Papaceit, M., Parisi, M. J., Parisi, M., Parts, L., Pedersen, J. S., Pesole, G., Phillippy, A. M., Ponting, C. P., Pop, M., Porcelli, D., Powell, J. R., Prohaska, S., Pruitt, K., Puig, M., Quesneville, H., Ram, K. R., Rand, D., Rasmussen, M. D., Reed, L. K., Reenan, R., Reily, A., Remington, K. A., Rieger, T. T., Ritchie, M. G., Robin, C., Rogers, Y. H., Rohde, C., Rozas, J., Rubenfield, M. J., Ruiz, A., Russo, S., Salzberg, S. L., Sanchez-Gracia, A., Saranga, D. J., Sato, H., Schaeffer, S. W., Schatz, M. C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R. S., Sirot, L., Sirota, M., Sisneros, N. B., Smith, C. D., Smith, T. F., Spieth, J., Stage, D. E., Stark, A., Stephan, W., Strausberg, R. L., Strempel, S., Sturgill, D., Sutton, G., Sutton, G. G., Tao, W., Teichmann, S., Tobari, Y. N., Tomimura, Y., Tsolas, J. M., Valente, V. L., Venter, E., Venter, J. C., Vicario, S., Vieira, F. G., Vilella, A. J., Villasante, A., Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R. K., Wing, R. A., Wolfner, M. F., Wong, A., Wong, G. K., Wu, C. I.,

Wu, G., Yamamoto, D., Yang, H. P., Yang, S. P., Yorke, J. A., Yoshida, K., Zdobnov, E., Zhang, P., Zhang, Y., Zimin, A. V., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Acer, S. C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I., Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M., Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray, S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P., Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher, S., Foley, C. D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C. R., Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K., Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M. D., Hughes, L., Hurhula, B., Husby, M. E., Kamat, A., Kanga, B., Kashin, S., Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F., LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y., Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C., McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O., Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L., Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C., Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O. L., Patti, C., Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise, C., Rogov, P., Ross, K., Ryan, E., Settipalli, S., Shea, T., Sherpa, N., Shi, L., Shih, D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Jaffe, D. B., Alvarez, P., Brockman, W., Butler, J., Chin, C., Gnerre, S., Grabherr, M., Kleber, M., Mauceli, E., & MacCallum, I. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature, 450*(7167), 203-218.

Duffy, J. B. (2002). GAL4 system in drosophila: A fly geneticist's swiss army knife. *genesis, 34*(1‐2), 1-15.

Duncan, I. W., & Kaufman, T. C. (1975). Cytogenic analysis of chromosome 3 in Drosophila melanogaster: mapping of the proximal portion of the right arm. *Genetics, 80*(4), 733-752.

Erickson, J. W., & Quintero, J. J. (2007). Indirect Effects of Ploidy Suggest X Chromosome Dose, Not the X:A Ratio, Signals Sex in Drosophila. *PLOS Biology, 5*(12), e332.

Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Molecular systems biology, 9*, 641-641.

Fan, P., Manoli, Devanand S., Ahmed, Osama M., Chen, Y., Agarwal, N., Kwong, S., Cai, Allen G., Neitz, J., Renslo, A., Baker, Bruce S., & Shah, Nirao M. (2013). Genetic and Neural Mechanisms that Inhibit *Drosophila* from Mating with Other Species. *Cell, 154*(1), 89-102.

Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics, 41*(1), 331-368.

Finnegan, D. J. (1992). Transposable elements. *Current Opinion in Genetics & Development, 2*(6), 861-867.

Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiácovo, M. P., Church, G. M., & Calarco, J. A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nature Methods, 10*(8), 741-743.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology, 31*(9), 822-826.

Gaj, T., Gersbach, C. A., & Barbas, C. F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology, 31*(7), 397-405.

Gaj, T., Sirk, S. J., Shui, S.-L., & Liu, J. (2016). Genome-Editing Technologies: Principles and Applications. *Cold Spring Harbor perspectives in biology, 8*(12), a023754.

Garvie, C. W., & Wolberger, C. (2001). Recognition of specific DNA sequences. *Mol Cell, 8*(5), 937-946.

Geyer, P. K., Green, M. M., & Corces, V. G. (1990). Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *The EMBO journal, 9*(7), 2247-2256.

Gilbert, W., & Müller-Hill, B. (1966). Isolation of the lac repressor. *Proc Natl Acad Sci U S A, 56*(6), 1891-1898.

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J., & O'Connor-Giles, K. M. (2013). Engineering of Drosophilawith the CRISPR RNA-Guided Cas9 Nuclease. *Genetics, 194*(4), 1029.

Greig, S., & Akam, M. (1995). The role of homeotic genes in the specification of the *Drosophila* gonad. *Current Biology, 5*(9), 1057-1062.

Grimaud, C., & Becker, P. B. (2009). The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. *Genes & Development, 23*(21), 2490-2495.

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage φC31. *Genetics, 166*(4), 1775-1782.

Gurdon, J. B. (2016). Chapter Twenty-Four - Cell Fate Determination by Transcription Factors. In P. M. Wassarman (Ed.), *Current topics in developmental biology* (Vol. 116, pp. 445-454). Academic Press.

Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science, 267*(5205), 1788.

Harding, K., Wedeen, C., McGinnis, W., & Levine, M. (1985). Spatially regulated expression of homeotic genes in Drosophila. *Science, 229*(4719), 1236-1242.

Heffer, A., Shultz, J. W., & Pick, L. (2010). Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution. *Proceedings of the National Academy of Sciences, 107*(42), 18040-18045.

Heintzman, N. D., & Ren, B. (2009). Finding distal regulatory elements in the human genome. *Current Opinion in Genetics & Development, 19*(6), 541-549.

Hirth, F., Loop, T., Egger, B., Miller, D. F., Kaufman, T. C., & Reichert, H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. *Development, 128*(23), 4781-4788.

Hittinger, C. T., Stern, D. L., & Carroll, S. B. (2005). Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development, 132*(23), 5261.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827-832.

Hueber, S. D., Bezdan, D., Henz, S. R., Blank, M., Wu, H., & Lohmann, I. (2007). Comparative analysis of Hox downstream genes in *Drosophila*. *Development, 134*(2), 381.

Hughes, C. L., & Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan1. *Evolution & Development, 4*(6), 459-499.

Hunter, C. P., & Kenyon, C. (1995). Specification of anteroposterior cell fates in Caenorhabditis elegans by Drosophila Hox proteins. *Nature, 377*(6546), 229-232.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., & Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology, 31*(3), 227-229.

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

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology, 169*(12), 5429-5433.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol, 3*, 318-356.

Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic acids research, 42*(Database issue), D1182-D1187.

Johnston, D. S., & Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell, 68*(2), 201-219.

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., & Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell, 131*(3), 530-543.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster : II. Zygotic loci on the third chromosome. *Wilehm Roux Arch Dev Biol, 193*(5), 283-295.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., & Celniker, S. E. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biology, 3*(12), research0084.0081.

Kaufman, T. C. (2017). A Short History and Description of *Drosophila melanogaster* Classical Genetics: Chromosome Aberrations, Forward Genetic Screens, and the Nature of Mutations. *Genetics, 206*(2), 665.

Kaufman, T. C., Lewis, R., & Wakimoto, B. (1980). Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics, 94*(1), 115-133.

Kaufman, T. C., Seeger, M. A., & Olsen, G. (1990). Molecular and Genetic Organization of The Antennapedia Gene Complex of Drosophila melanogaster. In T. R. F. Wright (Ed.), *Advances in Genetics* (Vol. 27, pp. 309-362). Academic Press.

Kerridge, S., & Morata, G. (1982). Developmental effects of some newly induced *Ultrabithorax* alleles of *Drosophila*. *Journal of embryology and experimental morphology, 68*(1), 211.

Keyes, L. N., Cline, T. W., & Schedl, P. (1992). The primary sex determination signal of Drosophila acts at the level of transcription. *Cell, 68*(5), 933-943.

Kimura, K.-I., Ote, M., Tazawa, T., & Yamamoto, D. (2005). Fruitless specifies sexually dimorphic neural circuitry in the Drosophila brain. *Nature, 438*(7065), 229-233.

Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J., & Mann, R. S. (2019). Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annual review of cell and developmental biology, 35*(1), 357-379.

LaJeunesse, D., & Shearn, A. (1995). Trans-regulation of thoracic homeotic selector genes of the Antennapedia and bithorax complexes by the trithorax group genes: absent, small, and homeotic discs 1 and 2. *Mechanisms of Development, 53*(1), 123-139.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018, 2018/02/08/). The Human Transcription Factors. *Cell, 172*(4), 650-665.

Latchman, D. S. (1993). Transcription factors: an overview. *International journal of experimental pathology, 74*(5), 417-422.

Lelli, K. M., Noro, B., & Mann, R. S. (2011). Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. *Proc Natl Acad Sci U S A, 108*(52), 21122-21127.

Lemons, D., & McGinnis, W. (2006). Genomic Evolution of Hox Gene Clusters. *Science, 313*(5795), 1918.

LeMotte, P. K., Kuroiwa, A., Fessler, L. I., & Gehring, W. J. (1989). The homeotic gene Sex Combs Reduced of Drosophila: gene structure and embryonic expression. *The EMBO Journal, 8*(1), 219-227.

Levine, M., & Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. *Cell, 55*(4), 537-540.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature, 276*(5688), 565-570.

Lewis, E. B. (1985). Regulation of the genes of the bithorax complex in Drosophila. *Cold Spring Harb Symp Quant Biol, 50*, 155-164.

Lewis, E. B. (2004). Genes and Developmental Pathways. In H. D. Lipshitz (Ed.), *Genes, Development and Cancer: The Life and Work of Edward B. Lewis* (pp. 175-203). Springer US.

Little, J. W., Byrd, C. A., & Brower, D. L. (1990). Effect of abx, bx and pbx mutations on expression of homeotic genes in Drosophila larvae. *Genetics, 124*(4), 899-908.

Lohmann, I., McGinnis, N., Bodmer, M., & McGinnis, W. (2002). The *Drosophila Hox* Gene *Deformed* Sculpts Head Morphology via Direct Regulation of the Apoptosis Activator *reaper*. *Cell, 110*(4), 457-466.

Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage Compensation in Drosophila. *Cold Spring Harbor Perspectives in Biology, 7*(5).

Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic acids research, 29*(13), 2860-2874.

Lutz, B., Lu, H. C., Eichele, G., Miller, D., & Kaufman, T. C. (1996). Rescue of Drosophila labial null mutant by the chicken ortholog Hoxb-1 demonstrates that the function of Hox genes is phylogenetically conserved. *Genes & Development, 10*(2), 176-184.

Mackenzie, S. M., Brooker, M. R., Gill, T. R., Cox, G. B., Howells, A. J., & Ewart, G. D. (1999). Mutations in the white gene of Drosophila melanogaster affecting ABC transporters that determine eye colouration. *Biochimica et Biophysica Acta (BBA) - Biomembranes, 1419*(2), 173-185.

Mahaffey, J. W., Griswold, C. M., & Cao, Q.-M. (2001). The Drosophila Genes *disconnected* and *disco*-related Are Redundant With Respect to Larval Head Development and Accumulation of mRNAs From Deformed Target Genes. *Genetics, 157*(1), 225.

Mahaffey, J. W., & Kaufman, T. C. (1987). Distribution of the Sex combs reduced gene products in Drosophila melanogaster. *Genetics, 117*(1), 51-60.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., & Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology, 31*(9), 833-838.

Malicki, J., Schughart, K., & McGinnis, W. (1990). Mouse *Hox-2.2* specifies thoracic segmental identity in Drosophila embryos and larvae. *Cell, 63*(5), 961-967.

Martinez-Arias, A., & Lawrence, P. A. (1985). Parasegments and compartments in the Drosophila embryo. *Nature, 313*(6004), 639-642.

Mastick, G. S., McKay, R., Oligino, T., Donovan, K., & López, A. J. (1995). Identification of target genes regulated by homeotic proteins in Drosophila melanogaster through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics, 139*(1), 349-363.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., & Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic acids research, 34*(Database issue), D108-D110.

Mauch, T., & Schoenwolf, G. (2001). Developmental Biology. Sixth Edition. By Scott F. Gilbert. *American Journal of Medical Genetics, 99*, 170-171.

McGinnis, N., Kuziora, M. A., & McGinnis, W. (1990). Human Hox-4.2 and Drosophila Deformed encode similar regulatory specificities in Drosophila embryos and larvae. *Cell, 63*(5), 969-976.

McGinnis, W., & Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell, 68*(2), 283-302.

Meise, M., Hilfiker-Kleiner, D., Dubendorfer, A., Brunner, C., Nothiger, R., & Bopp, D. (1998). Sex-lethal, the master sex-determining gene in Drosophila, is not sex-specifically regulated in Musca domestica. *Development, 125*(8), 1487.

Merabet, S., Litim-Mecheri, I., Karlsson, D., Dixit, R., Saadaoui, M., Monier, B., Brun, C., Thor, S., Vijayraghavan, K., Perrin, L., Pradel, J., & Graba, Y. (2011). Insights into Hox Protein Function from a Large Scale Combinatorial Analysis of Protein Domains. *PLOS Genetics, 7*(10), e1002302.

Merrill, V. K. L., Diederich, R. J., Turner, F. R., & Kaufman, T. C. (1989). A genetic and developmental analysis of mutations in labial, a gene necessary for proper head formation in Drosophila melanogaster. *Developmental Biology, 135*(2), 376-391.

Micol, J. L., Castelli-Gair, J. E., & García-Bellido, A. (1990). Genetic analysis of transvection effects involving cis-regulatory elements of the Drosophila Ultrabithorax gene. *Genetics, 126*(2), 365-373.

Mitchell, P. J., & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science, 245*(4916), 371.

Morata, G., & Garcia-Bellido, A. (1976). Developmental analysis of some mutants of the bithorax system of Drosophila. *Wilhelm Roux's archives of developmental biology, 179*(2), 125-143.

Morata, G., & Kerridge, S. (1981). Sequential functions of the bithorax complex of Drosophila. *Nature, 290*(5809), 778-781.

Morgan, T. H. (1910). SEX LIMITED INHERITANCE IN DROSOPHILA. *Science, 32*(812), 120.

Moschetti, R., Marsano, R. M., Barsanti, P., Caggese, C., & Caizzi, R. (2004). FB elements can promote exon shuffling: a promoter-less white allele can be reactivated by FB mediated transposition in Drosophila melanogaster. *Molecular Genetics and Genomics, 271*(4), 394-401.

Muller, H. J. (1918). Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal Factors. *Genetics, 3*(5), 422-499.

Müller, J., & Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. *The EMBO journal, 10*(11), 3147-3155.

Nagoshi, R. N., & Baker, B. S. (1990). Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes & Development, 4*(1), 89-97.

Nöthiger, R., Leuthold, M., Andersen, N., Gerschwiler, P., Grüter, A., Keller, W., Leist, C., Roost, M., & Schmid, H. (1987). Genetic and developmental analysis of the sex-determining gene ‘double sex’ (dsx) of Drosophila melanogaster. *Genetical Research, 50*(2), 113-123.

Nüsslein-Volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature, 287*, 795.

Nüsslein-Volhard, C., Wieschaus, E., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. *Wilhelm Roux's archives of developmental biology, 193*(5), 267-282.

Ong, C.-T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics, 12*(4), 283-293.

Ota, T., Fukunaga, A., Kawabe, M., & Oishi, K. (1981). Interactions between sex-transformation mutants of Drosophila melanogaster. I. Hemolymph vitellogenins and gonad morphology. *Genetics, 99*(3-4), 429-441.

Pabo, C. O., & Sauer, R. T. (1984). Protein-DNA recognition. *Annu Rev Biochem, 53*, 293-321. https://doi.org/10.1146/annurev.bi.53.070184.001453

Pabo, C. O., & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem, 61*, 1053-1095.

Pattatucci, A. M., Otteson, D. C., & Kaufman, T. C. (1991). A functional and structural analysis of the Sex combs reduced locus of Drosophila melanogaster. *Genetics, 129*(2), 423-441.

Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning [10.1038/nrg1726]. *Nat Rev Genet, 6*(12), 893-904.

Pederson, J. D., Kiehart, D. P., & Mahaffey, J. W. (1996). The Role of HOM-C Genes in Segmental Transformations: Reexamination of theDrosophila Sex combs reduced Embryonic Phenotype. *Developmental Biology, 180*(1), 131-142.

Percival-Smith, A. (2017). Non-specificity of Transcription Factor function in Drosophila melanogaster. *Development Genes and Evolution*.

Percival-Smith, A. (2018). Phenotypic Nonspecificity as the Result of Limited Specificity of Transcription Factor Function. *Genetics research international, 2018*, 7089109-7089109.

Percival-Smith, A., & Laing Bondy, J. A. (1999). Analysis of murine HOXA-2 activity in Drosophila melanogaster. *Developmental Genetics, 24*(3-4), 336-344.

Percival-Smith, A., Sivanantharajah, L., Pelling, J. H., & Teft, W. (2013). Developmental competence and the induction of ectopic proboscises in Drosophila melanogaster. *Development Genes and Evolution, 223*(6), 375-387.

Percival-Smith, A., Teft, W. A., & Barta, J. L. (2005). Tarsus determination in Drosophila melanogaster. *Genome, 48*(4), 712-721.

Percival-Smith, A., Weber, J., Gilfoyle, E., & Wilson, P. (1997). Genetic characterization of the role of the two HOX proteins, Proboscipedia and Sex Combs Reduced, in determination of adult antennal, tarsal, maxillary palp and proboscis identities in Drosophila melanogaster. *Development, 124*(24), 5049-5062.

Potter, S., Truett, M., Phillips, M., & Maher, A. (1980). Eucaryotic transposable genetic elements with inverted terminal repeats. *Cell, 20*(3), 639-647.

Prince, F., Katsuyama, T., Oshima, Y., Plaza, S., Resendez-Perez, D., Berry, M., Kurata, S., & Gehring, W. J. (2008). The YPWM motif links Antennapedia to the basal transcriptional machinery. *Development, 135*(9), 1669.

Ptashne, M. (1967a). ISOLATION OF THE lambda PHAGE REPRESSOR. *Proc Natl Acad Sci U S A, 57*(2), 306-313.

Ptashne, M. (1967b). Specific binding of the lambda phage repressor to lambda DNA. *Nature, 214*(5085), 232-234.

Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature, 386*(6625), 569-577.

Qian, S., Capovilla, M., & Pirrotta, V. (1991). The bx region enhancer, a distant cis-control element of the Drosophila Ubx gene and its regulation by hunchback and other segmentation genes. *The EMBO journal, 10*(6), 1415-1425.

Ragab, A., Thompson, E. C., & Travers, A. A. (2006). High Mobility Group Proteins HMGD and HMGZ Interact Genetically With the Brahma Chromatin Remodeling Complex in Drosophila. *Genetics, 172*(2), 1069.

Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J., & Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature, 391*(6668), 691-695.

Regulski, M., McGinnis, N., Chadwick, R., & McGinnis, W. (1987). Developmental and molecular analysis of Deformed; a homeotic gene controlling Drosophila head development. *The EMBO journal, 6*(3), 767-777.

Rideout, E. J., Dornan, A. J., Neville, M. C., Eadie, S., & Goodwin, S. F. (2010). Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster. *Nature Neuroscience, 13*(4), 458-466.

Rivera-Pomar, R., & Jãckle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. *Trends in Genetics, 12*(11), 478-483.

Rivlin, P. K., Gong, A., Schneiderman, A. M., & Booker, R. (2001). The role of Ultrabithorax in the patterning of adult thoracic muscles in Drosophila melanogaster. *Development Genes and Evolution, 211*(2), 55-66.

Robinett, C. C., Vaughan, A. G., Knapp, J.-M., & Baker, B. S. (2010). Sex and the Single Cell. II. There Is a Time and Place for Sex. *PLOS Biology, 8*(5), e1000365.

Ruiz, M.-F., & Sánchez, L. (2010). Effect of the gene transformer of Anastrepha on the somatic sexual development of Drosophila. *Int J Dev Biol, 54*(4), 627-633.

Ryner, L. C., & Baker, B. S. (1991). Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes & Development, 5*(11), 2071-2085.

Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J., & Wasserman, S. A. (1996). Control of Male Sexual Behavior and Sexual Orientation in Drosophila by the fruitless Gene. *Cell, 87*(6), 1079-1089.

Ryoo, H. D., Marty, T., Casares, F., Affolter, M., & Mann, R. S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development, 126*(22), 5137-5148.

Sabl, J. F., & Birchler, J. A. (1993). Dosage dependent modifiers of white alleles in Drosophila melanogaster. *Genet Res, 62*(1), 15-22.

Salz, H. K., Cline, T. W., & Schedl, P. (1987). Functional Changes Associated with Structural Alterations Induced by Mobilization of a *P* Element Inserted in the *Sex-lethal* Gene of Drosophila. *Genetics, 117*(2), 221.

Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W., & Schedl, P. (1989). The Drosophila female-specific sex-determination gene, Sex-lethal, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes & Development, 3*(5), 708-719.

Shah, K., Cao, W., & Ellison, C. E. (2019, Jun 5). Adenine Methylation in Drosophila Is Associated with the Tissue-Specific Expression of Developmental and Regulatory Genes. *G3 (Bethesda), 9*(6), 1893-1900.

Shen, C.-N., Burke, Z. D., & Tosh, D. (2004). Transdifferentiation, Metaplasia and Tissue Regeneration. *Organogenesis, 1*(2), 36-44.

Sivanantharajah, L., & Percival-Smith, A. (2009). Analysis of the Sequence and Phenotype of Drosophila Sex combs reduced Alleles Reveals Potential Functions of Conserved Protein Motifs of the Sex combs reduced Protein. *Genetics, 182*(1), 191-203.

Sivanantharajah, L., & Percival-Smith, A. (2014). Acquisition of a Leucine Zipper Motif as a Mechanism of Antimorphy for an Allele of the Drosophila Hox Gene Sex Combs Reduced. *G3: Genes|Genomes|Genetics, 4*(5), 829-838.

Sivanantharajah, L., & Percival-Smith, A. (2015). Differential pleiotropy and HOX functional organization. *Developmental Biology, 398*(1), 1-10.

Slack, J. M. W., & Bard, J. B. L. (1991). *From Egg to Embryo: Regional Specification in Early Development*. Cambridge University Press.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., & Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci, 39*(9), 381-399.

Smith, N. C., & Matthews, J. M. (2016). Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Current Opinion in Structural Biology, 38*, 68-74.

Southworth, J. W., & Kennison, J. A. (2002). Transvection and silencing of the Scr homeotic gene of Drosophila melanogaster. *Genetics, 161*(2), 733-746.

Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirián, L., & Dickson, B. J. (2005). Neural Circuitry that Governs *Drosophila* Male Courtship Behavior. *Cell, 121*(5), 795-807.

Suárez, S., Cabré, O., Velázquez, A., Creus, A., Marcos, R., & Xamena, N. (1996). Molecular study of the germinal reversions induced at the white-ivory locus in Drosophila melanogaster. *Mutagenesis, 11*(6), 559-563.

Subramaniam, V., Bomze, H. M., & López, A. J. (1994). Functional differences between Ultrabithorax protein isoforms in Drosophila melanogaster: evidence from elimination, substitution and ectopic expression of specific isoforms. *Genetics, 136*(3), 979-991.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biology, 3*(12), RESEARCH0088-RESEARCH0088.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis [journal article]. *Genome Biology, 8*(7), R145.

Tour, E., Hittinger, C. T., & McGinnis, W. (2005). Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax. *Development, 132*(23), 5271.

Vinson, C., Chatterjee, R., & Fitzgerald, P. (2011). Transcription factor binding sites and other features in human and Drosophila proximal promoters. *Sub-cellular biochemistry, 52*, 205-222.

Wakimoto, B. T., & Kaufman, T. C. (1981). Analysis of larval segmentation in lethal genotypes associated with the Antennapedia gene complex in Drosophila melanogaster. *Developmental Biology, 81*(1), 51-64.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell, 153*(4), 910-918.

Waterbury, J. A., Jackson, L. L., & Schedl, P. (1999). Analysis of the doublesex female protein in Drosophila melanogaster: role on sexual differentiation and behavior and dependence on intersex. *Genetics, 152*(4), 1653-1667.

Weatherbee, S. D., Halder, G., Kim, J., Hudson, A., & Carroll, S. (1998). Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. *Genes & Development, 12*(10), 1474-1482.

Weinzierl, R. A., Axton, J. M., Ghysen, A., & Akam, M. (1987). Ultrabithorax mutations in constant and variable regions of the protein coding sequence. *Genes & Development, 1*(4), 386-397.

Wilkins, A. S. (1995). Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *BioEssays, 17*(1), 71-77.

Wingender, E., Schoeps, T., Haubrock, M., & Dönitz, J. (2015). TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic acids research, 43*(Database issue), D97-D102.

Wipfler, B., Schneeberg, K., Löffler, A., Hünefeld, F., Meier, R., & Beutel, R. G. (2013). The skeletomuscular system of the larva of Drosophila melanogaster (Drosophilidae, Diptera) – A contribution to the morphology of a model organism. *Arthropod Structure & Development, 42*(1), 47-68.

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics, 13*(1), 59-69.

Wunderlich, Z., & Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics, 25*(10), 434-440.

Yang, Y., Zhang, W., Bayrer, J. R., & Weiss, M. A. (2008). Doublesex and the Regulation of Sexual Dimorphism in Drosophila melanogaster: structure, function, and mutagenesis of a female-specific domain. *Journal of Biological Chemistry, 283*(11), 7280-7292.

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N., & McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. *The EMBO journal, 13*(10), 2362-2377.

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids, 4*, e264.

Zhao, J. J., Lazzarini, R. A., & Pick, L. (1993). The mouse Hox-1.3 gene is functionally equivalent to the Drosophila Sex combs reduced gene. *Genes & Development, 7*(3), 343-354.

Chapter 2. Characterization of CRISPR induced non-homologous recombination

Introduction:

Genome editing is a form of genetic engineering that is used to alter genomic DNA sequences in a defined manner within living organisms (Esvelt & Wang, 2013). Importantly, the DNA modifications induced by genome editing are restricted to specific genomic loci (Esvelt & Wang, 2013). This is in contrast to many other transgenic methods (e.g. P element-mediated germline transformation in *Drosophila*) where transgene insertions are not targeted (Majumdar & Rio, 2015). While many effective genome editing systems have been developed (e.g. those based on transcription activator-like effector nucleases or zinc finger nucleases), the method commonly referred to as "CRISPR/Cas9" or sometimes simply "CRISPR" has become the most widely adopted due to its relative ease of use and high precision (Ran *et al.*, 2013).

The acronym CRISPR refers to the clustered regularly interspaced short palindromic repeats of DNA that were initially observed in bacteria and which, together with the CRISPR associated Cas9 DNA endonuclease, form part of an antiviral defense system referred to as the CRISPR-Cas adaptive immune system (Bhaya *et al.*, 2011; Deveau *et al.*, 2010; Horvath & Barrangou, 2010; Makarova *et al.*, 2011). Through the study and manipulation of this bacterial antiviral defense system, CRISPR/Cas9 was developed into an efficient and programmable tool for genome editing. The system consists of the RNA-

guided CRISPR-associated nuclease, Cas9, and a chimeric RNA (chiRNA) which contains sequences complementary to the target DNA in the host genome (Jinek *et al.*, 2012). The RNA sequence information and PAM recognition of Cas9, guide the chiRNA-Cas9 riboprotein complex to the desired genomic sequence where Cas9 makes a double-stranded break (DSB) (Jinek *et al.*, 2012). The DSB can be repaired with either non-homologous end joining (NHEJ), or if a DNA repair template is provided, through homology directed repair (HDR) (Gratz *et al.*, 2013). In the absence of a repair template, NHEJ can result in point mutations or small deletions at the target locus (Aymard *et al.*, 2014). In contrast, if a repair template is provided, modifications that are present within the repair template construct can be incorporated into the genome via HDR (Gratz *et al.*, 2014).

The CRISPR/Cas9 system has proven to be an invaluable research tool and has greatly accelerated the generation of new insights into the function and regulation of biological systems. However, the high frequency (>50%) of unintended DNA modifications at untargeted genomic sites (i.e. "off-target" effects) is a major concern, especially for clinical and therapeutical applications (Cho *et al.*, 2014; Corrigan-Curay *et al.*, 2015; Fu *et al.*, 2013; Hsu *et al.*, 2013; Mali *et al.*, 2013; Zhang *et al.*, 2015). Moreover, the ability to trigger HDR after the DSB is an important factor influencing successful gene editing. If the DSB is at an active euchromatic gene region, the epigenetic marker, H3K9me3, will recruit the homologous recombination enzymes to the damage site and promote HDR, whereas NHEJ will most likely be triggered if the DSB is in a silenced gene region (Aymard *et al.*, 2014).

The original objective of the project was to the creation of genetically manipulable sites at gene loci encoding proteins required for determining the *Drosophila* body plan. These genes included the maternal effect gene *bicoid* (*bcd*) and the *H*ox gene, *Sex combs reduced* (*Scr*). To insert the genetically manipulable site (*attP—y*$^+$*/w*$^+$*—attP*) at a target locus, HDR between a repair template containing the *attP* flanked site and the endogenous locus was to be initiated by two genomic DSBs made by CRISPR Cas9s. This modification of specific loci was originally intended as the starting point of a broader study related to the phenotypic non-specificity of transcription factor function.

Three w^+ transformants were collected from the experiments: two targeting *bcd* and one targeting *Scr*. However, none of the three w^+ transformants were the result of homologous recombination. Analysis of these non-homologous recombination events revealed that two of the w^+ transformants were the result of the mobilization of the *white* gene on a transposon and that the third was the result of the insertion of the *mini-white* gene of the *Scr* repair template into the genome (with the important hallmarks of transposition). These results raise important concerns regarding the unintended consequences of CRISPR based genetic manipulations and their effects on the stability of the genome.

Materials and Methods:

***Drosophila melanogaster* stocks and media:** *Drosophila melanogaster* stocks were obtained from the Bloomington Drosophila Stock Center (Indiana University, Bloomington, Ind.) and maintained at room temperature (23-25 ºC) on corn meal media [1% (w/v) *Drosophila*-grade agar, 6% (w/v) sucrose, 10% (w/v) cornmeal, 1.5% (w/v) yeast

and 0.375% (w/v) 2-methyl hydroxybenzoate]. For collection of Drosophila embryos, female flies were allowed to lay eggs on apple juice plates [2.5% (w/v) Drosophila-grade agar, 6% sucrose, 50% apple juice and 0.3% (w/v) 2-methy hydroxybenzoate.

chiRNAs: Two different chiRNAs (a 3' and a 5' chiRNA) were designed to recognize and target the respective coding regions of the *Scr* and *bcd* genes. For example, the *Scr* 5' chiRNA targets a region 5' of *Scr* exon 2 and the *Scr* 3' chiRNA targets a region 3' of *Scr* exon 3. For efficient and specific target recognition, every chiRNA contained 18 - 20 nucleotides of sequence complementary to the respective genomic target (first nucleotide must be a guanine) (Jinek *et al.*, 2012). Cleavage by Cas9 also requires that the 3' end of the genomic target sequence contain di-guanines (NGG), known as the proto-spacer adjacent motif (PAM) (Jinek *et al.*, 2012). Two forms of chiRNA were created: a pU6-*Bbs*I vector based DNA plasmid form (Gratz *et al.*, 2013) and an *in vitro* transcribed RNA form (Bassett *et al.*, 2013).

***In vitro* transcribed chiRNAs:** Two oligonucleotides were used to generate the chiRNA template for *in vitro* transcription (**Figure 1**). The forward oligonucleotide (GAAATTAATACGACTCACTATAGGN_{18}GTTTTAGAGCTAGAAATAGC) contains an upstream sequence and a T7 promoter (underlined) that are required for *in vitro* transcription. This is followed by GGN_{18} sequence (N_{18} indicates the target specific sequence) and a portion of the chiRNA stem loops (Bassett *et al.*, 2013). The reverse oligonucleotide (AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATT TCT AGC TCT AAA AC) encodes the

entire chiRNA sequence that comes after the targeting sequence (**Figure 1**) (Bassett *et al.*, 2013). The overlapping forward and reverse oligonucleotides were used in a PCR reaction (in the absence of any other template) and the resulting product purified using a PCR purification kit (Qiagen, Hilden, Germany). *In vitro* transcription of these templates was then performed using the Megascript T7 Kit (Ambion, Austin, USA), with 300 ng of purified DNA template for four hours at 37°C. The reaction was extracted with phenol chloroform and the RNA product precipitated with ethanol (Bassett *et al.*, 2013). ChiRNAs were aliquoted in DEPC-treated water and stored at -80°C. The target specific sequences of the chiRNAs are listed in **Table 1**.

Table 1. *In vitro* transcribed chiRNA target specific sequences:

Target gene	5' ChiRNA	3' ChiRNA
Scr	**GGCAGCGGTGGAGGGGCGGG**	**GGTGCGCGAACTGCGACGGA**
bcd	**GGATGTTGGTGATGTGGGTG**	**GGGCGAAGGCTTGCCAAATT**
Fst	**GCCTTGGTGGCAGTGGCTTC**	

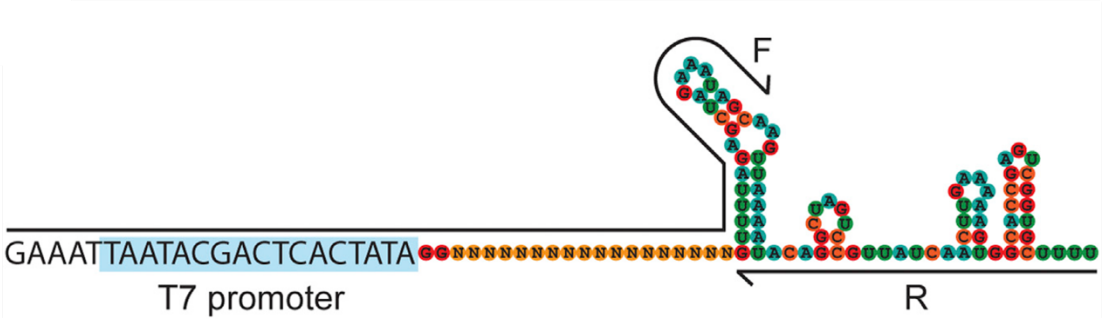


Figure 1. Schematic illustration of the oligonucleotides used to generate the chiRNA template for *in vitro* transcription. "F" indicates the forward oligonucleotide and "R" stands for the reverse oligonucleotide. The T7 promoter sequence is highlighted in blue. The N_{18} sequence (orange) indicates the target-specific sequence. This figure is adapted from Bassett et al., (2013) with permission.

Construction of pU6-chiRNAs: The target-specific sequences for both of the *Bcd* chiRNAs were synthesized as 5'-phosphorylated oligonucleotides, annealed, and ligated into the *Bbs*I restriction sites of pU6-*Bbs*I-chiRNA (Gratz *et al.*, 2013). The 5'*Bcd* chiRNA targets exon 1 of *Bcd* and the 3'*Bcd* chiRNA targets sequence 3' of *Bcd* exon 4.The *Scr* pU6-chiRNAs were made by Laura Garofalo prior my arrival to the lab (**Table 2**) (Garofalo, 2015) and the *Fst* pU6-chiRNAs was made by Anthony Percival-Smith and used as a positive control (**Table 2**) (Newman *et al.*, 2017).

Table 2. Phosphorylated oligonucleotides used in the cloning of pU6- chiRNAs:

	Oligonucleotides (5' to 3')	
chiRNAs	**Forward**	**Reverse**
***Bcd* 5'**	**CTTCGGATGTTGGTGATGTGGGTG**	**AAACCACCCACATCACCAACATCC**
***Bcd* 3'**	**CTTCGGGCGAAGGCTTGCCAAATT**	**AAACAATTTGGCAAGCCTTCGCCC**
***Scr* 5'**	**CTTCGATTTTTGAATTTATGGCAA**	**AAACTTGCCATAAATTCAAAAATC**
***Scr* 3'**	**CTTCGCGTGGCACTTTTCGGGTAC**	**AAACGTACCCGAAAAGTGCCACGC**
***Fst* 5'**	**CTTCGGCCTTGGTGGCAGTGGCTTC**	**AAACGAAGCCACTGCCACCAAGGC**

Donor template: The donor templates for targeted homologous recombination were cloned into the pFus_A plasmid using a Goldengate approach and consisted of the following arrangement of sequences: 5'homologous arm (*5'HA*) + *attP* + *marker* + *attP* + 3'homologous arm (*3'HA*). The eye color *white* (w^+) or the body color *yellow* (y^+) marker are flanked by inverted *attP* Φ31 recombination sites.

The *bcd* donor template contains the DNA sequences from *bcd* exon 1 and *bcd* exon 4 plus marker (y^+ or w^+) flanked by inverted *attP* sites. *bcd 5'HA* sequence (1434 bp) was PCR amplified from *y w D. melanogaster* genomic DNA with primers 5'HA-*Bcd*-*Bsa*I-F and 3'HA-*Bcd*-*attP*-*Bsa*I-R (**Table 3**). *bcd 3'HA*(1516 bp) was PCR amplified from *yw D. melanogaster* genomic DNA with primers *Bcd*-*Bsa*I-*attP*-F and *Bcd*-*Bsa*I-R (**Table 3**). The primers used to amplify homologous recombination arms added the *attP* recombination site sequence (39 bp) and a *Bsa*I restriction site to the 3' of *5'HA* and a *Bsa*I restriction site to 5' end of *5'HA*. Similarly, primers added a *Bsa*I restriction site and the *attP* recombination site sequence (39 bp) to the 5' of *3'HA* and a *Bsa*I restriction site to 3' end of *3'HA*. The y^+ gene was PCR amplified from MiMIC plasmid (Venken *et al.*, 2011) with primers *y*-*Bsa*I-F and *y*-*Bsa*I-R, which added *Bsa*I restriction sites to the 5' and 3' ends (**Table 3**). Similarly, the w^+ gene was PCR amplified with primers *w*-*Bsa*I-F and *w*-*Bsa*I-R from a plasmid (**Table 3**). The three DNA fragments, *5'HA* (with 3' *attP*), *3'HA* (with 5' *attP*), and the marker sequence, were digested with *Bsa*I generating unique 5' overhangs. *5'HA*, *3'HA*, and marker (y^+ or w^+) DNA fragments and *Bsa*I-digested and dephosphorylated pFus_A (Addgene, Watertown, USA) were ligated together in an ordered assembly reaction and transformed into DH5α cells. The *Scr* (y^+) donor template

was constructed by Laura Garofalo using the same overall design prior to my arrival in the lab (Garofalo, 2015). The w^+ fragment for the *Scr* (w^+) donor template was created by PCR amplifying the w^+ sequence with primers *w-SanDI*-F and *w-BamHI*-R (**Table 3**) thereby adding *SanDI* and *BamHI* restriction sites to the 5' and 3' ends, respectively. The *Scr* (y^+) donor template and the amplified w^+ sequence were digested with restriction enzymes *SanDI* and *BamHI*, DNA fragments were isolated and then ligated together to switch the y^+ marker for w^+ marker creating the *Scr* (w^+) donor template.

Table 3. Amplification oligonucleotides used in constructing *bcd* and *Scr* donor templates.

Amplicon	***Bcd* and *Scr* (w^+) donor template Amplification Oligonucleotides**	
	Forward	**Reverse**
***Bcd* 5' HA**	**5'HA-*Bcd-BsaI*-F: CAGCTAGGTCTCGCTATTTGGGCTTTCCCTATGCGAAC**	**5'HA-*Bcd-attP-BsaI*-R: CAGCTAGGTCTCCCATGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGTTTCCCCAAACACTCCGCC**
***white* (w^+)**	***w-BsaI*-F: CAGCTAGGTCTCCCATGGCATGCGGCCGCTCTAGATAAC**	***w-BsaI*-R: CAGCTAGGTCTCGGTCCCAAGATCCCCCGGATCCATAAC**
***yellow* (y^+)**	***y-BsaI*-F: CAGCTAGGTCTCCCATGCGACTATTAAATGATTATCGCC**	***y-BsaI*-R: CAGCTAGGTCTCGGTCCTCGACCTGCAGGTCAACGGATC**
***Bcd* 3' HA**	**3'HA-*Bcd-BsaI-attP*-F: CAGCTAGGTCTCGGGACCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCCTGGATGAAGAGGCGTGTTAGAG**	**3'HA-*Bcd-BsaI*-R: CAGCTAGGTCTCCCGCCCCATGTTAATGGGTCACTGTGCAC**
white* (w^+) of *Scr	***w-SanDI*-F: GACCCAGCACTATCATTGAACCCTAACACCGTTTGTAGCGTTACCTAGCG**	***w-BamHI*-R: CGGATCCGGTTATTGCGCCTTCACTGTATGCCATGGCCCTAATTTTAC**

Embryo genotypes and Cas9 sources: The following lines were used for *Drosophila* embryos injections: (1) the non-Cas9-expressing lines y^1 w^{67c23} (*yw*), (2) the transgenic Cas9-expressing line, *act-cas9* (y^1 *M*[*Act5c-cas9.P*] *ZH-2A* w^+; Bloomington stock center, 54590) where Cas9 is expressed from a constitutively expressed *actin5C* promoter, and (3) the transgenic Cas9-expressing line, *nos-cas9* (y^1 *M*[*nos-Cas9.P*] *ZH-2A* w^+; Bloomington stock center, 54591) where Cas9 is expressed from a germline-specific *nanos* promoter (Port *et al.*, 2014). When injecting *yw* embryos, the Cas9 source was either pHsp70-Cas9 DNA plasmids (Addgene, Massachusetts, USA) or Cas9 mRNA (Thermo Fisher, Massachusetts, USA).

Injection of embryos: *Drosophila melanogaster* embryos were collected on a yeasted apple juice plate every 30 min at room temperature. Embryos were dechorionated for 1 min with 3% sodium hypochlorite and washed with distilled water. Embryos were aligned on an apple juice/agar strip and then transferred onto double-sided tape on a microscope slide. Embryos were partially desiccated for 3-4 min and then covered in halocarbon oil. DNA was injected into the posterior end of the embryo using a glass needle (FHC Inc., Maine, USA) attached to a syringe filled with halocarbon oil and viewed on an inverted microscope (Wilovert, Wetzlar, Germany). All injections were performed at room temperature 30-45 min AEL, where a majority of embryos are at the syncytial blastoderm stage of development.

Injection medium: The various injection media used for CRISPR/Cas9-mediated HDR are indicated below. All media was prepared in 1X PBS to the indicated final concentrations.

1. **Cas9 plasmid / chiRNA plasmid**: 500 ng/µl pHsp70-Cas9, 250 ng/µl chiRNA (5' and 3'), 120 ng/µl donor template, 10% (w/v) glycerol.
2. **Cas9 mRNA / chiRNA plasmid**: 100 ng/µl Cas9 mRNA, 250 ng/µl chiRNA (5' and 3'), 500 ng/µl donor template, 10% (w/v) glycerol.
3. **Cas9 expressing embryos / *in vitro* transcribed chiRNA**: 500 ng/µl chiRNA (5' and 3'each), 500 ng/µl donor template, 10% (w/v) glycerol.
4. **Cas9 plasmid / *in vitro* transcribed chiRNA**: 500 ng/µl pHsp70-Cas9, 500 ng/µl chiRNA (5' and 3'either together or individually), 120 ng/µl donor template, 10% (w/v) glycerol.
5. **Cas9 mRNA / *in vitro* transcribed chiRNA**: 100 ng/µl Cas9 mRNA, 500 ng/µl chiRNA (both 5' and 3'), 300 ng/µl donor template, 10% (w/v) glycerol.

Screening for transformants: Hatched larvae (injection survivors) were collected and transferred onto corn meal media 72 hours after injection. In instances where the y^+ marker was being scored, adults of the G_0 generation were screened for the presence of y^+ patches on the cuticle. All adult G_0 flies were crossed to *yw* flies. The G_1 progeny of the fertile crosses were screened for wild type (y^+) body color or red (w^+) eye color.

Drosophila DNA extraction for Illumina sequencing: 2-4 flies were collected and frozen in liquid nitrogen and were then ground and dissolved in 400 µl $LiCl/CH_3COOK$ solution

(1 part 5M CH_3COOK and 2.5 parts 6M LiCl). The DNA was isolated from the lysate using DNeasy kit (Qiagen, Hilden, Germany) and purified with ethanol precipitation. The A_{260}/A_{280} and A_{260}/A_{230} ratios of the samples were greater than 2.0 and 1.8, respectively. Sequencing was performed by the London Regional Genomics Centre (London, ON) using the Illumina HiSeq2000 platform. The sequencing data was analysed using Geneious software.

5' RACE: Poly(A)+ mRNA was extracted from adult flies using the Sigma-Aldrich mRNA Isolation Kit (Sigma-Aldrich, St. Louis, USA). The Smarter RACE 5'/3' kit (Takara, Kusatsu, Japan) was used to perform RACE. Two nested primers based on sequence in the second and third exons of the *white* gene were used: *w*_outer (5'-GGA GCC GAT AAA GAG GTC AT-3') and *w*_inner (5' CCA GGC ATA GGT GAG GTT CT- 3'). Sequencing was performed by the London Regional Genomics Centre (London, ON).

PCR strategy for verifying translocation and insertion events: A PCR based strategy was utilized to verify the translocation and insertion events in the genome. The primers used are listed in **Table 4**. The PCR products were sequenced by the London Regional Genomics Center (London, ON).

PCR strategy used to determine the junction sequence of the *ScrD1* insertion: The following PCR strategy was used to determine the junction sequence of the *ScrD1* insertion into repetitive sequences. Primers were designed as illustrated in **Figure 2**. Forward primer

1 (F1) was from the repetitive sequence (RS) towards the inserted element (IE). A unique tag (5' CTAATACGACTCACTATAGGGC 3') was added to the end of the F1 primer. The reverse primer (R) was from IE toward RS. Forward primer 2 (F2) was designed to be complementary to the tag sequence. The two PCR reactions were performed: first, the standard PCR mixture with F1 and template DNA was ran for 10 to 15 cycles. After this step, the product, which contains the junction site and the tag sequence at the 5' end was amplified with the forward primer F2 and reverse primer R. The PCR products were sequenced by the London Regional Genomics Center. This strategy was used to verify the left junction sequence of the *Scr-D1* insertion event. The primers used to determine the *Scr-D1* left junction are listed in **Table 4**.

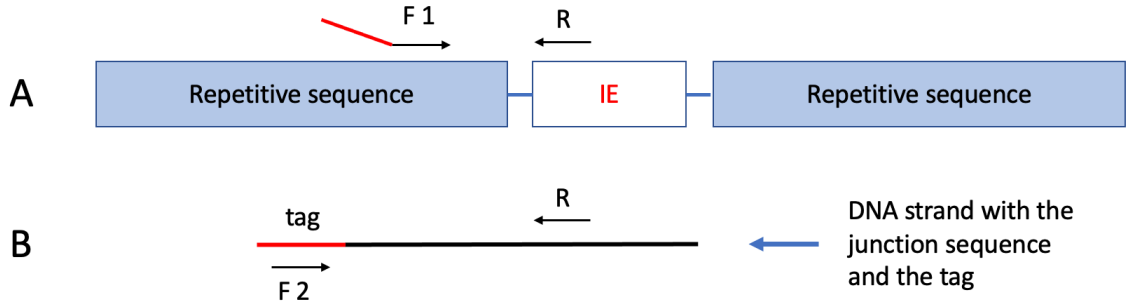


Figure 2. Schematic of primers used in the PCR strategy for insertion in repetitive sequence region. The inserted element (IE) is inserted in repetitive sequence. Panel A: Forward primer 1 (F1) has a unique tag added to the 5' end. The reverse primer (R) is designed from IE towards the repetitive sequence. Panel B: The DNA strand with junction sequence and the tag is amplified by primers F2 (complementary to tag) and reverse primer R.

Table 4. Oligonucleotides used in PCR for insertion verification.

Amplicon	Oligonucleotides used in PCR for insertion verification	
	Forward	Reverse
Bcd 4 left junction	5'-F: TATTGTGCCAGGCATAGGTG	4-5'-R: TCATACCCTTGAATAAGTTG
Bcd 4 right junction	4-3'-F: TCTGTGGGTAAGCTTTACTC	3'-R: AATCAGCGTTTGATTTACGC
Bcd 39 left junction	5'-F: TATTGTGCCAGGCATAGGTG	39-5'-R: CCTTTGAGGATAGTTAGTTC
Bcd 39 right junction	39-3'-F: TGTATACTTCTCTGACAAAC	3'-R: AATCAGCGTTTGATTTACGC
Scr-D1 left junction	D1-5'-F1: CTAATACGACTCACTATAGGGCTT GAGTCTATTAAATGGAC	D1-5'-R: TCAAAAAACAAACAAAAATAAG
Scr-D1 left junction	D1-5'-F2: GATTATGCTGAGTGATATCCCG	

Crossing scheme: The crossing scheme used to map the chromosomal location of *ScrD1* is shown in (**Figure 3)**.

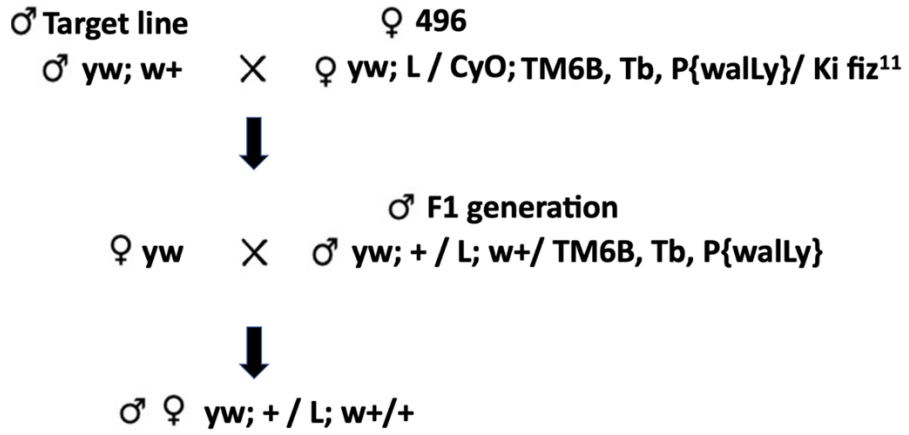


Figure 3. Crossing scheme used of map the chromosomal location of *ScrD1*

"*yw; w*$^{+}$" indicates the G1 transformant. The genomic background for transformant is yw and it has a *w*$^{+}$ insertion in the genome. The TM6B balancer is the Chromosome 3 balancer. The segregation of "*w*$^{+}$" from the TM6B balancer indicates that the insertion of "*w*$^{+}$" is on the 3rd Chromosome.

Results:

The introduction of a genetically manipulable site at two target gene loci using CRISPR/Cas9 mediated homologous recombination

As part of a broader study aimed at better understanding phenotypic non-specificity in *Drosophila*, we initially attempted to introduce a genetically manipulable site (*attP—marker* (y^+/w^+)*—attP*) into the maternal effect gene *bicoid* (*bcd*), or the *H*ox gene, *Sex combs reduced* (*Scr*) (**Figure 4**). A method based on CRISPR/Cas9 targeting (in conjunction with HDR) was chosen as the most practical strategy to replace the coding sequences of the respective target genes with the *attP—marker—attP* construct. As part of this strategy, the Cas9 endonuclease, together with the respective chiRNAs and donor templates, were injected into syncytial blastoderm embryos. Guided by the chiRNAs, Cas9 would then be expected to induce DSBs upstream of each of the PAM sequences in the coding regions of the respective target genes and induce HDR allowing the insertion of the genetically manipulable site into the desired loci. Furthermore, successfully transformed flies (with *yw; attP—*y^+/w^+*—attP* construct) that incorporated the cassette would be easily identifiable by the wild type body marker (y^+) or red eye color (w^+).

Four Cas9 sources were used for the microinjections: the *pHsp70-Cas9* DNA plasmid, Cas9 mRNA, and two transgenic Cas9-expressing lines (*act-cas9*, and *nos-cas9*). Furthermore, two sources of chiRNA were used for the microinjections: the *pU6-BbsI* vector based DNA plasmid form (Gratz *et al.*, 2013) and the *in vitro* transcribed RNA form (Bassett *et al.*, 2013). In addition, the *Frost* donor template was used as a positive control for the injections (Newman *et al.*, 2017).

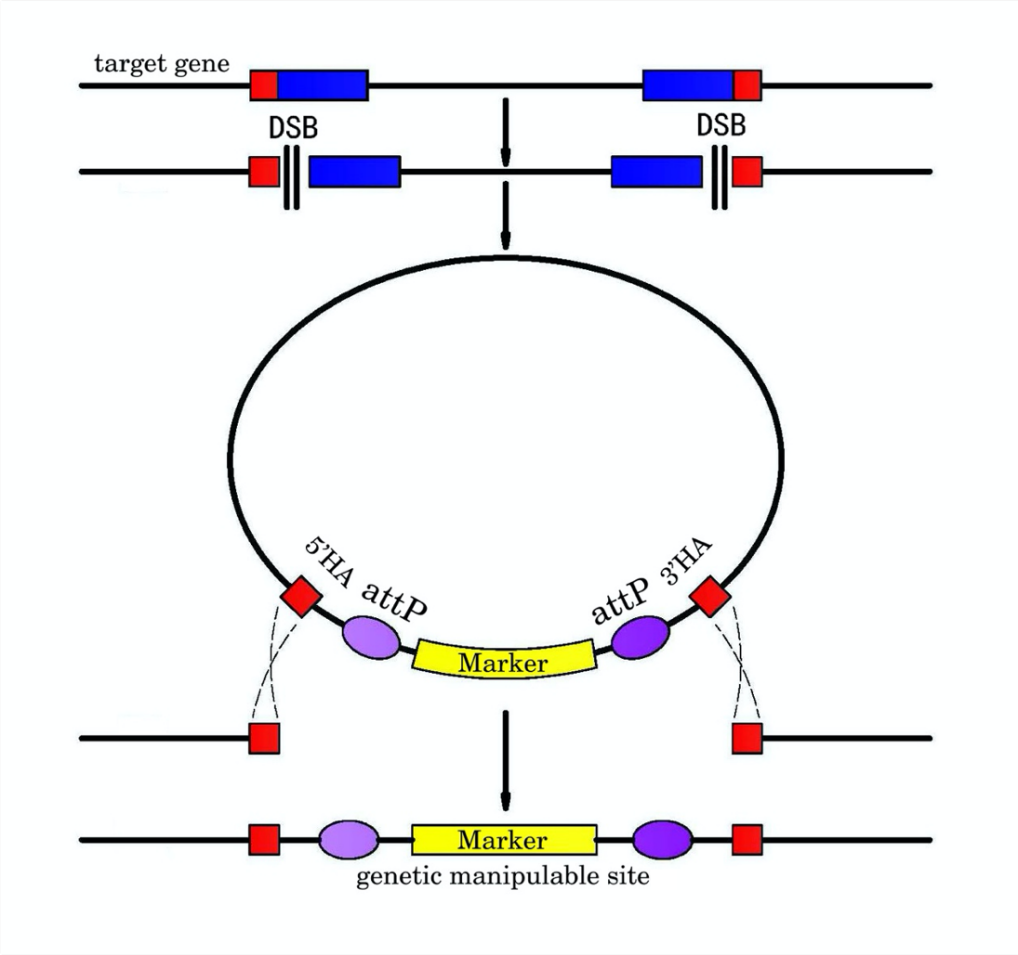


Figure 4. Schematic of CRISPR/Cas9-mediated HDR at the target gene locus. Red and blue rectangles indicate the untranslated regions and coding regions of the exons of the target gene respectively. Two chiRNAs (not shown) target the Cas9 nuclease (not shown) to the desired loci where they induce double-stranded breaks (DSBs) upstream of the PAM sequence (not shown). CRISPR/Cas9-mediated homologous recombination repair initiated by the DSBs occurs between the targeted locus and the donor template DNA, containing homologous arms of target gene (5'HA and 3'HA), such that the color marker (y^+ or w^+) and the *attP* sequences are inserted into the target locus. This results in the target gene coding region being replaced by two inverted *attP* recombination sites that flank a marker.

The results of the injections are summarized in **Table 5**. The *Frost* control yielded w^+ transformants (2 / 173 fertile crosses, Cas9 plasmid and chiRNA plasmid were used). For *bcd* and *Scr* experiments, some G_0 flies with injection of y^+ donor templates had y^+ clones of cells when the chiRNA source was RNA or plasmid encoded, indicating somatic transformation. No G_1 germ-line y^+ events were recovered. For injections with the w^+ donor templates, three independent w^+ G_1 transformants were collected from the injections using plasmid DNA as the Cas9 source and plasmid DNA derived chiRNAs. Among the three transformants, two were obtained during experiments targeting *bcd*, and one when targeting *Scr* (named *Bcd 4*, *Bcd 39* and *Scr-D1*) (**Table 5**).

The *Bcd* 4 and *Bcd* 39 events were not a result of homologous recombination since the w^+ trait of *Bcd 4* and *Bcd 39* segregated with the second chromosome, and the *bcd* locus is on the third chromosome. Therefore, the integration of w^+ sequences in *Bcd 4* and *Bcd 39* represent non-homologous recombination events. Furthermore, we reasoned that *Scr-D1* was also not a homologous recombination event based on the following logic: *Scr* null alleles are haplo-insufficient for the formation of sex combs, resulting in a reduction in the number of sex combs from 10-12 to only 5-6 (Bantignies *et al.*, 2011; Ragab *et al.*, 2006; Sivanantharajah & Percival-Smith, 2009; Southworth & Kennison, 2002). Integration of w^+ into the *Scr* locus, deleting the *Scr* coding region, thus should have reduced the number of male sex combs. However, *Scr-D1* males exhibited a normal sex comb bristle number (data not shown) indicating that *Scr-D1* did not represent a homologous recombination event. To characterize the nature of the three non-homologous recombination events, full genome sequencing experiments were initiated.

Table 5. Injection result summary.

The "Genotype" column indicates the fly line injected: either y^1w^{67c23} (*yw*) or transgenic Cas9-expressing lines. The "Gene" column shows the target gene name and the marker used in the donor template. The "Cas9 source" column indicates the Cas9 source: pHsp70-Cas9 DNA plasmid, Cas9 mRNA or expression from the Cas9 transgenic lines. The "chiRNA" column indicates the chiRNA source used, pU6-*Bbs*I vector-based DNA or *in vitro* transcribed RNA, and whether two (5' and 3') or one (5' or 3') chiRNA were used. The "crosses #" column indicates the number of crosses with injection survivors. The "fertile #" indicates the number of fertile crosses. "G_0 mosaic transformant" and "G_1 transformant" indicated the number of successful G_0 and G_1 transformants, respectively.

Fly Genotype	Gene	Cas9 source	ChiRNA	crosses #	fertile #	G0 mosaic transformant	G1 transformant
yw	*Bicoid* (y^+)	DNA plasmid	RNA (5'+3')	89	49	6	0
yw	*Bicoid* (y^+)	DNA plasmid	RNA 5'	38	20	1	0
yw	*Bicoid* (y^+)	DNA plasmid	RNA 3'	53	30	1	0
yw	*Bicoid* (y^+)	DNA plasmid	DNA (5'+3')	92	64	4	0
yw	*Bicoid* (w^+)	DNA plasmid	RNA (5'+3')	44	29	N/A	0
yw	*Bicoid* (w^+)	DNA plasmid	RNA 5'	26	21	N/A	0
yw	*Bicoid* (w^+)	DNA plasmid	RNA 3'	33	24	N/A	0
yw	*Bicoid* (w^+)	DNA plasmid	DNA (5'+3')	201	132	N/A	2
yw	*Scr* (y^+)	DNA plasmid	RNA (5'+3')	113	83	6	0
yw	*Scr* (y^+)	DNA plasmid	DNA (5'+3')	117	89	0	0
nos-cas9	*Scr* (y^+)	Transgenic	DNA (5'+3')	208	107	0	0
act-cas9	*Scr* (y^+)	Transgenic	DNA (5'+3')	44	21	0	0
yw	*Scr* (w^+)	DNA plasmid	DNA (5'+3')	287	215	N/A	1
yw	*Frost* (w^+)	DNA plasmid	RNA 5'	97	74	N/A	0
yw	*Frost* (w^+)	mRNA	RNA 5'	67	47	N/A	0
yw	*Frost* (w^+)	DNA plasmid	DNA 5'	264	173	N/A	2

The sequence of the w^{67c23} allele

The Illumina sequencing data obtained was first used to determine the sequence and detailed structure of the w^{67c23} allele. The w^{67c23} allele is a deletion derivative of an unstable allele, w^{c} (Collins & Rubin, 1982, 1984). In the w^{67c23} allele, about 130 kb (from X, 2795604 to 2924488) is deleted compared to the *Drosophila melanogaster* genome sequence (release r6.40) (Hoskins *et al.*, 2015). The deletion includes the promoter region, first exon and start codon of the *white* gene. The FB-NOF-FB element is in the deleted region next to a 364 bp long direct duplication of *w* sequence (from exon 2 to exon 3 and including intron 2) located upstream of intron 1 (**Figure 5**). Furthermore, a FB element about 9.2 kb downstream of *w* was found (**Figure 5**).

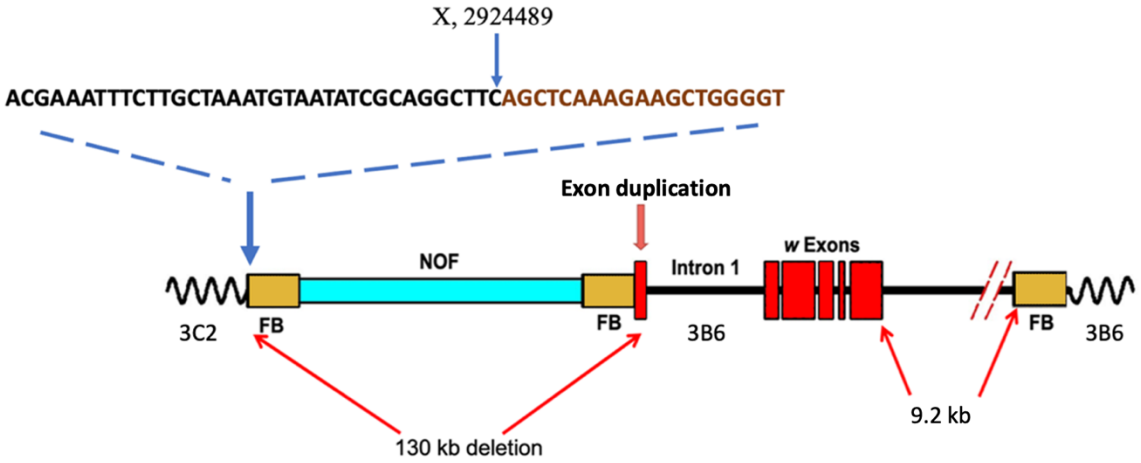


Figure 5. *w*^{67c23} allele structure. 130 kb of sequence is deleted (from X, 2795604 to 2924488) upstream of the *white* locus. The FB-NOF-FB element is inserted into the genome upstream of the duplicated *w* sequence (364 bp-long and same orientation as the *w* gene). Another FB element is located 9.2 kb downstream of the *white* gene. The fusion point of the FB element occurs in the 3C2 region of the genome. FB element sequence is indicated in brown. "X, 2924489" indicates the exact nucleotide position within the *D. melanogaster* genome (release r6.40) (Hoskins *et al.*, 2015). See Appendix 1 for the sequence.

***Bcd 4* and *Bcd 39* have no *mini-white* in the genome.**

The w^+ allele on the donor templates is "*mini-white*" and has most of intron 1 removed (Hazelrigg *et al.*, 1984; Levis *et al.*, 1985; Pirrotta *et al.*, 1985). The junction created by the removal of intron 1 is a unique characteristic of the *mini-white* gene that distinguishes it from the endogenous *white* locus. Aligning the sequence data to the reference genome revealed that the deletion junction of *mini-white* is present in the *Scr-D1* genome, but not in *Bcd 4* or *Bcd 39* genomes. Furthermore, no sequence from the *Bcd 4* and *Bcd 39* genomes aligned to the *white* promoter and exon 1 sequence. This suggested that the *white* insertion in *Scr-D1* contained *mini-white*, but that the w^+ phenotype observed in *Bcd 4* and *Bcd 39* came about through a different mechanism (one that allows *white* expression without its native promoter, exon 1, and a start codon).

Transposition of w^{67c23}

The w^{67c23} locus can be mobilized on a transposon (flanked by the NOF sequence and a FB element) and inserted into a gene in such a way that the gene's promoter and start codon are used to produce a chimeric transcript and protein with *white*, thereby resulting in a w^+ phenotype (Moschetti *et al.*, 2004). We hypothesized that such an event had occurred in the *Bcd 4* and *Bcd 39* genomes. To test this hypothesis 5' RACE was performed on mRNA extracted from *Bcd 39* and *Bcd 4* to isolate and identify the chimeric mRNAs. Surprisingly, the cDNAs of *Bcd 4* and *Bcd 39* chimeric mRNAs had identical sequences. The chimeric RNAs have the *osp* exon 1 spliced in frame to *w* exon 2, indicating that the w^{67c23} allele had migrated to the *outspread* (*osp*) locus (**Figure 6**). The short, duplication of the second to third *w* exons next to the NOF sequence are skipped during RNA splicing.

In the genomes of *Bcd 4* and *Bcd 39*, a 14-kb-long DNA fragment between the upstream FB element (5' end of NOF sequence upstream of w^{67c23} allele) and the downstream FB element (9.2 kb downstream of w^{67c23} locus) transposed into intron 1 of the *osp* locus. To characterize the exact insertion sites of the w^{67c23} allele in intron 1 of *osp*, PCR was used to obtain the 5' and 3' junction sequences of the insertion site for both *Bcd 4* and *Bcd 39*. The w^{67c23} allele is inserted at position (2L, 14642091) with a 9-bp tandem repeat (TAGTTTGTT) on both sides of the insertion in *Bcd 4* and at the position (2L, 14683661) with a 11-bp tandem repeat (CTGACAGTGTG) on both sides of the insertion in *Bcd 39* **(Figure 6).** See Appendices 2 and 3 for the sequence.

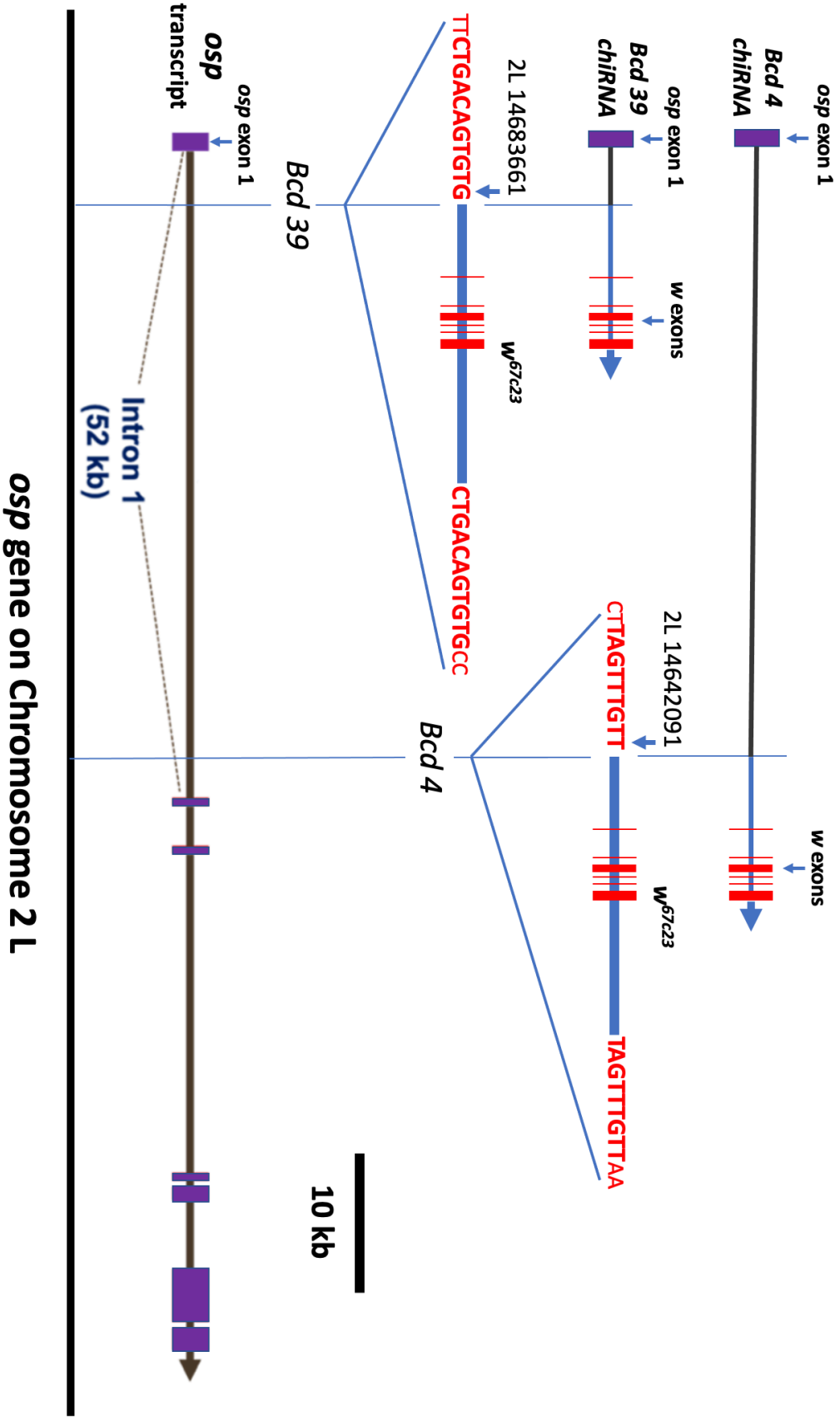

osp exon 1
Bcd 4 chiRNA
w exons
2L 14642091
CTTAGTTTGTT
w67c23
TAGTTTGTTAA
Bcd 4
osp exon 1
Bcd 39 chiRNA
w exons
2L 14683661
TTCTGACAGTGTG
w67c23
CTGACAGTGTGCC
Bcd 39
osp transcript
osp exon 1
Intron 1 (52 kb)
10 kb
osp gene on Chromosome 2 L

Figure 6. FB mediated w^{67c23} allele migration in *Bcd 4* and *Bcd 39* genome. The purple boxes on the *osp* transcript are *osp* exons. The w^{67c23} allele migrated and inserted into intron 1 of the *osp* gene. Unbolded red letters indicate the *osp* gene sequence. Bolded red letters indicate the tandem repeat (release r6.40) (Hoskins *et al.*, 2015). The insertion positions are indicated below the first tandem repeat. The promoter of *osp* drives the expression of the w^+ allele creating the chimeric mRNA (*osp* exon1 + *w* exons). The scale bar indicates a length of 10 kb.

Characterization of *Scr-D1*

Since the penetrance of the variegated eye phenotype of *Scr-D1* was low (< 2%), PCR was used to follow the *mini-white* allele. The *Scr-D1* males were crossed with *y w; L/CyO; Kiftz*^{11}*/TMB6* females and DNA extracted from F1 flies (**Figure 7 A**). The *mini-white* was present in both male and female offspring, and therefore, not X linked (**Figure 7 B**). Males from the F1 generation (with the marker *L* and the balancer *TM6B*) were crossed with *y w* female virgins and the F2 generation offspring collected (**Figure 7 C**). PCR was performed on DNA extracted from the F2 generation. *Mini-white* segregated from the balancer *TM6B* and not from marker *L*, indicating insertion on the third chromosome (**Figure 7 D**).

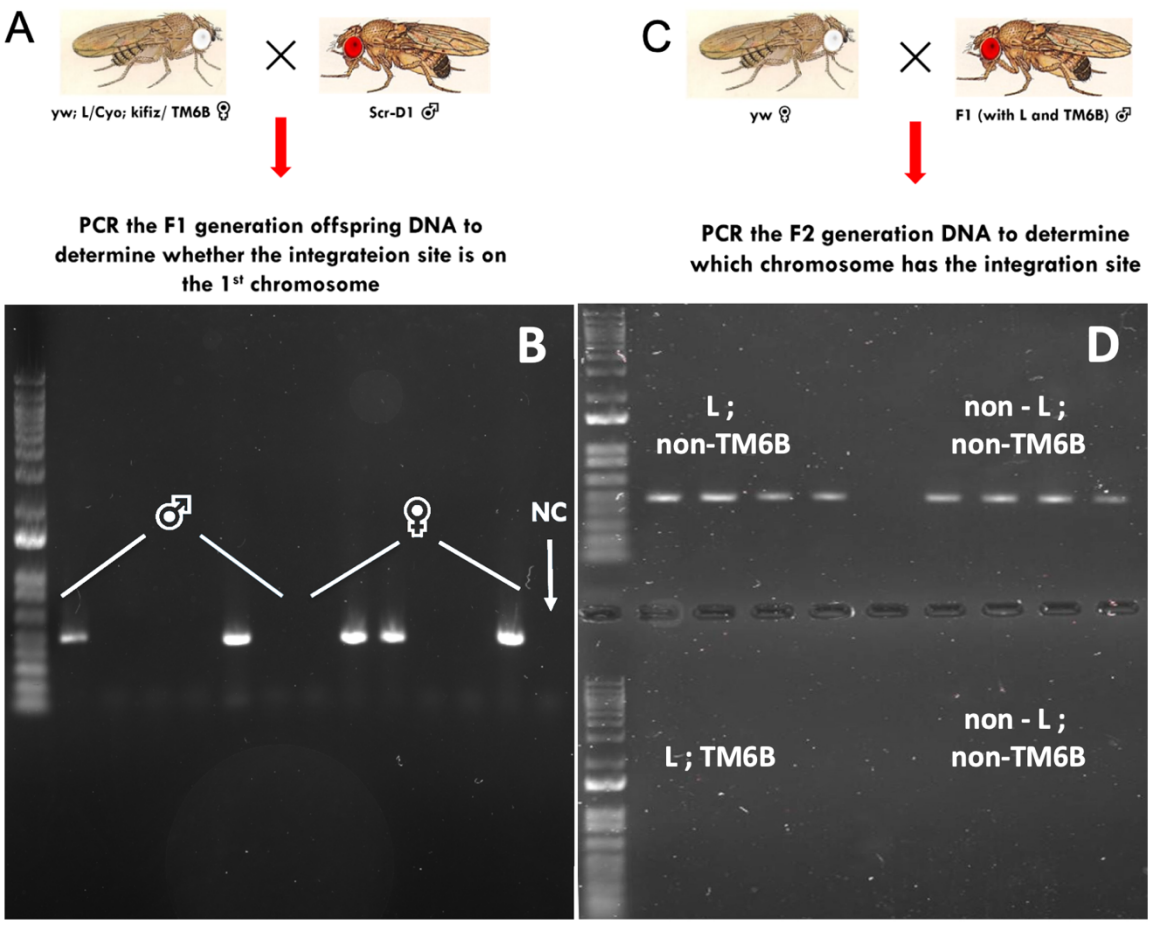


Figure 7. Chromosome segregation assay of *Scr-D1* mapping "w^+" to the 3rd chromosome. The *Scr-D1* males were crossed with *yw; L/CyO; Kiftz*^{11}*/TMB6* females and DNA extracted from F1 flies (Panel A). NC: negative control — no w^+ insertion in *yw* genome. The *mini-white* gene was present in both male and female offspring, and therefore, not X linked. Males from the F1 generation (with the marker *L* and the balancer *TM6B*) were crossed with *yw* female virgins and the F2 generation offspring collected (Panel C). PCR was performed on DNA extracted from the F2 generation (Panel D). *Mini-white* segregated from the balancer *TM6B* and not from marker *L*, indicating insertion on the third chromosome. The ladder used in B and D is 1 kb+ DNA ladder (Thermo Fisher, Massachusetts, USA).

Analysis of paired end reads of the Illumina sequencing data suggested that *ScrD1* was inserted into repetitive DNA (**Figure 8**). Characterization of the sequence of the DNA at the junction of the insertion indicated the insertion is in a *TE 17.6* transposable element on the 3rd chromosome at 81F (3R, 1280556) (release r6.40) (Hoskins *et al.*, 2015) close to pericentric heterochromatin, which explains the variegated eye color phenotype (**Figure 8 B**). Furthermore, the *mini-white* cassette and small portions of the *Scr* w^+ donor template were integrated into the genome, and the DNA inserted was flanked by 6 to 2 bp-long tandem repeats (AGGGTT to AG) at the integration site. The ends of the donor plasmid DNA inserted lack inverted repeats. See Appendices 4 and 5 for the sequence.

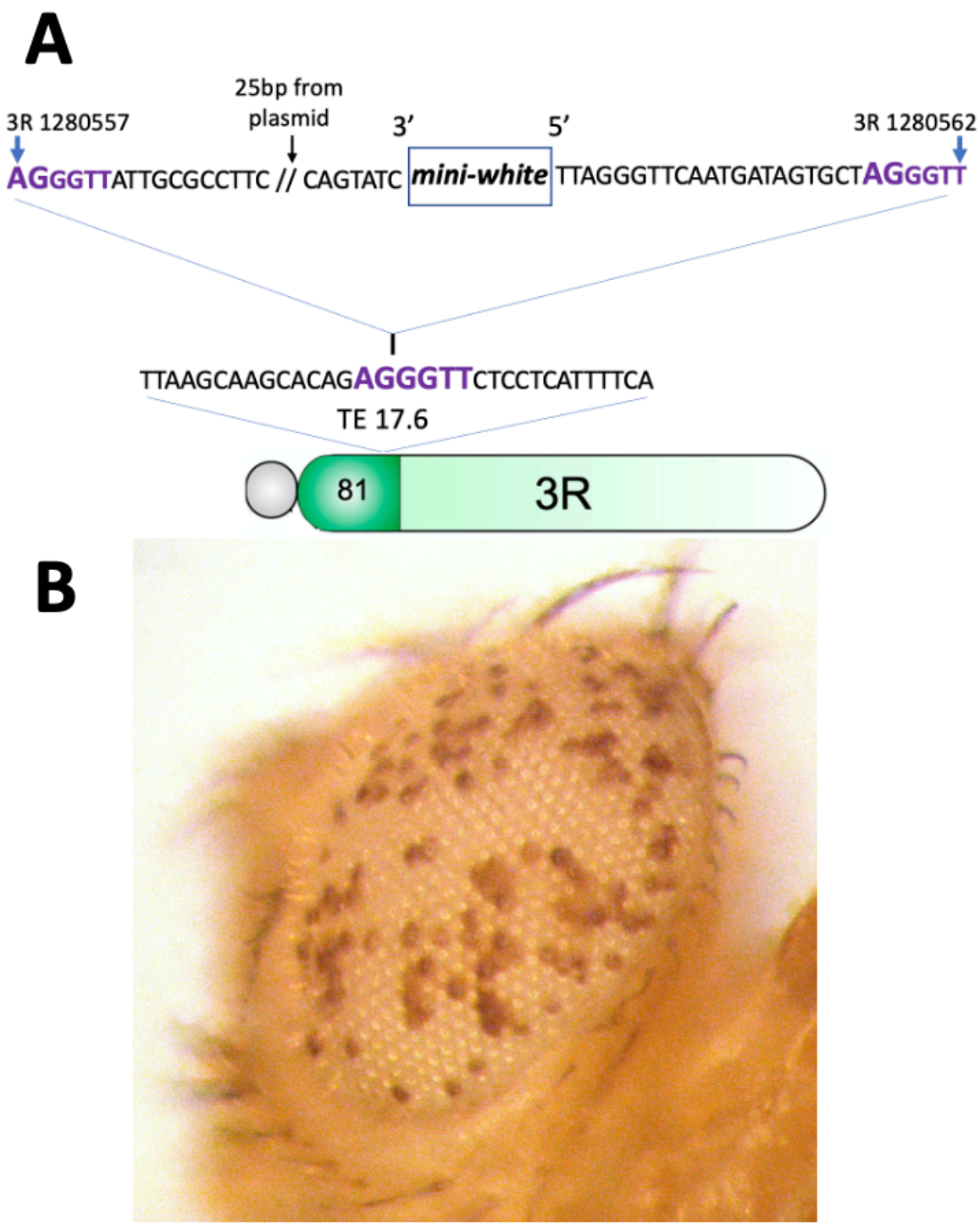

A
25bp from plasmid
3R 1280557
3'
5'
3R 1280562
AGGGTTATTGCGCCTTC // CAGTATC mini-white TTAGGGTTCAATGATAGTGCTAGGGTT
TTAAGCAAGCACAGAGGGTTCTCCTCATTTTCA
TE 17.6
81
3R
B

Figure 8. *Scr-D1* insertion site and *Scr-D1* phenotype. Panel A: Location of w^+ insertion in *Scr-D1* genome (3R, 81F, 1280556). The inserted w^+ fragment is flanked by a 6 - 2bp-long tandem repeats (**AG**GGTT) at the integration site. Portions of the *Scr* w^+ donor template (black letters) and *mini – white* cassette were integrated into the genome. See Appendices 4 and 5 for the sequence. The additional 25 bp sequence: ACT GTA TGC CAT GGC CCT AAT TTTA. Panel B: Variegated eye phenotype of *Scr -D1*.

Discussion:

Here I present the results of my investigation into the origin of three w^+ transformants (*Bcd 4*, *Bcd 39* and *Scr-D1*) collected during the course of experiments initially intended to result in the targeted integration of genetically manipulable sites via CRISPR. Unfortunately, none of the transformants were the result of homologous recombination. Two of the transformants (*Bcd 4* and *Bcd 39*) were the result of the mobilization of the w^{67c23} allele into the *osp* gene locus. The third transformant, *Scr-D1*, was the result of an insertion of *mini-white* into the genome (with hallmarks of transposition). In neither case (*bcd* locus or *Scr* locus) was the DSB repaired by the homologous template.

The common characteristic of the three non-homologous recombination events is transposition. These are trans events because the w^{67c23} allele and the *Scr* donor plasmid are not on the third chromosome where CRISPR induced the DSBs. This is in contrast to similar experiments where the *Fst* locus was successfully targeted using a similar approach (Newman *et al.*, 2017). For these reasons it is important to consider factors common to the targeting of *bcd* and *Scr*, and in contrast to the targeting of *Fst*, so that the observed non-homologous events might be better understood. One major difference is that *bcd* and *Scr* induce two DSBs 2.6kb and 16.8 kb apart, respectively, versus a single DSB for *Fst*. This raises the possibility that widely separated DSBs may increase the number of non-homologous events. A second commonality is that both *bcd* and *Scr* are in the Antennapedia complex and *Fst* is not. Within 50kb of either side of the DSBs of *Fst* there is one TE; whereas, for *bcd* and *Scr* there are three and five TEs, respectively.

The w^{67c23} allele

The y^1w^{67c23} line is commonly used in *Drosophila* research and has a yellow body color and white eyes. The w^{67c23} allele is an amorphic allele derived from the *white-crimson* (w^c) allele, which is a partial revertant of the *white-ivory* (w^i) allele (Collins & Rubin, 1982, 1984). The w^i allele is a 2.96 kb tandem duplication of *white* sequence (from intron 1 to exon 3) in the *white* locus (Bhadra *et al.*, 1997; Sabl & Birchler, 1993; Suárez *et al.*, 1996). The w^c allele is an insertion of a 10kb FB element into the w^i duplication that reverts the w^i phenotype and is unstable either reverting to w^i or generating *w* mutant alleles at a high frequency (Collins & Rubin, 1982, 1984).

From my characterization of the DNA sequence, the w^{67c23} allele is a deletion derivative of allele w^c in which 130kb of DNA upstream of the *white* gene is deleted, including the first exon, start codon and promoter region of the *white* gene (Moschetti *et al.*, 2004). In addition, a transposon-like element, NOF, flanked by two FB transposable elements is between the deletion breakpoint such that the FB-NOF-FB element is directly upstream of a 364 bp long of duplicated sequence of the *white* gene from exon 2 and exon 3. It is possible that the duplicated sequence next to the NOF FB element was part of the 2.96 kb tandem duplication of the initial w^i allele.

Transposable elements

Transposable elements (TEs) are DNA sequences that can move their positions in the genome. The mechanism of transposition is commonly used to categorize TEs: Class I elements or retrotransposons transpose using an RNA intermediate that is copied to DNA

using reverse-transcription, while Class II elements transpose using a DNA intermediate either by a replicative or "cut and paste" mechanism (Craig *et al.*, 2002; Finnegan, 1992). Foldback (FB) elements are a distinct group of poorly described TEs that are difficult to classify as the mechanism of transposition is unknown (Feschotte & Pritham, 2007). Furthermore, the transposases used have not been identified (Marzo *et al.*, 2008). Approximately 10% of FB elements are associated with a NOF sequence (4 kb) which codes for a 120-kDa protein of unknown function (Harden & Ashburner, 1990; Templeton & Potter, 1989). Their inverted terminal repeats contain different number of short repeats in direct orientation. This characteristic allows FB elements to easily form extensive secondary structures which can lead to DNA modifications like deletions, duplications and other chromosomal rearrangements at high frequency (Potter *et al.*, 1980). While the exact mechanism of FB mediated transposition is not fully understood, it is known that the FB-NOF element is a non-autonomous transposition element (Badal *et al.*, 2013). The protein coded by NOF sequence lacks any known transposase motifs and has structural similarity with hydrolases (Badal *et al.*, 2013).

In *Bcd 39*, the tandem repeat at the w^{67c23} allele insertion sites is 11-bp and in *Bcd 4*, the tandem repeat is 9-bp. As FB-NOF is a non-autonomous transposable element, and the tandem repeats at the respective insertion sites are of different sizes, it is likely that these two transposition events are mediated by two different transposases (Linheiro & Bergman, 2012). The tandem repeat reported by Moschetti *et al.* (2004) is 9-bp, which is the same as we observed in *Bcd 4*. In our experiments, no external transposons or transposases were introduced during injection. The w^{67c23} allele migrations must have been catalyzed by

endogenous transposases, adding to evidence that FB-NOF element is a non-autonomous transposable element.

w^{67c23} migration events

Migration of the w^{67c23} allele has been reported previously (Moschetti *et al.*, 2004). When plasmids carrying the *mariner*-like transposable element *Bari*, *pBari1_47Dw*$^{+}$, were injected into the w^{67c23} embryos, the w^{67c23} allele migrated to the introns of other genes. In each of the three characterized events, the w^{67c23} allele transposed into an intron of either *osp*, *CG6487*, or *Cg3973*. The promoters of the genes, *osp*, *CG6487*, and *Cg3973* drove transcription of w^{67c23} allele, creating chimeric *white* mRNA where the initial reading frame encoded by the gene is in frame with the *w* reading frame leading to expression of a chimeric protein that is functional for deposition of pigments.

Considering our data together with the work of Moschetti *et al.* (2004), a total of five w^{67c23} migration events have been characterized. In three cases, the insertion site is in the *osp* gene. The *osp* gene locus may thus represent a "hot-spot" for transposition. For the *bcd* experiments, two w^{67c23} migration events were observed among 132 fertile crosses which constitute about 13,200 gametes screened. It is likely, however, that other w^{67c23} migration events occurred but where the w^{+} phenotype simply could not be detected. Insertion in the reverse orientation would not express a chimeric transcript encoding the *white* open reading frame. Two thirds of the insertions in the correct orientation would not be in-frame. Therefore, only 1/6 insertion are expected to result in the w^{+} phenotype (12 out of 13,200 gametes). In addition, the insertions are required to be inserted in the introns of genes,

which constitutes about half the genome sequence (24 out of 13,200 gametes). Insertions into heterochromatic regions may never be identified, and furthermore, only insertions in genes that are expressed during eye pigmentation will express the w^+ phenotype, such that the rate of detection of these events could be in the range of 1/24 to 1/48 or even less. Therefore, the rate of transposition of w^{67c23} could be 48 to 96 out of 13,200 gametes. This rate is low and would not be expected to have much chance of turning up in the same fly with an independent HDR event. However, if all unmarked transposons in the *Drosophila* genome migrated at the same frequency as w^{67c23} then the rate is very high.

There are 5,373 terminal inverted repeat (TIR) elements in the genome; therefore, on average 18-36 transposon events are expected per gamete (Mérel *et al.*, 2020). If migration occurs for all 34,805 transposons at the same frequency as w^{67c23} then on average 126-253 transposon events are expected per gamete. The major problem is the random insertion of an unmarked TE close to the locus being modified by CRISPR and causing an independent phenotype that may be mistakenly ascribed to the CRISPR modification. The transformation procedure includes backcrosses which would reduce the number of transposition events recovered if not closely linked to the CRISPR modified locus. The total length of the genetic map is 284 mu, so the number of TE inserted 10 mu away from the targeted locus (closely linked) is 17.2 in the worst-case scenario. Therefore, it is reasonable to suggest that at least two independent CRISPR induced alleles are assessed for a phenotype given the potential for induction of major genome instability by CRISPR. This potential problem with DSB breaks inducing transposition needs to be quantitated for all transposons in the genome in the future. In addition, the *Scr-D1* has the *mini-white*

sequence inserted into the pericentric heterochromatin in Chromosome 3R. The inserted fragment is flanked by 6 to 2-bp-long tandem repeats (**AG**GGTT), which indicates that this event is catalyzed by a transposase. However, inverted repeats characteristic of TIR TEs where not found in the sequence transposed. The mobilization of w^{67c23} and the insertion of *mini-white* suggests a general activation of transposition by CRISPR. The activation of transposition by DSBs in humans would be a major concern for clinical applications of CRISPR as random insertion of DNA can lead to cancer (Anwar *et al.*, 2017).

The mobilization of TEs

Both the induction of DSBs and the introduction of the *Bari1* transposon by microinjection result in the mobilization of the w^{67c23} allele. Although without direct evidence, Moschetti et al, 2004 speculated that the w^{67c23} allele migrations observed were catalyzed by the protein, BARI1, encoded by the injected *Bari1* transposon. *Bari1* is a member of the *Tc1-mariner* superfamily which belong the Type II class of transposons, which transpose from one position to another in the genome using a DNA intermediate. During transposition of Type II transposons, the transposases bind to the terminal repeats of the transposons and induce double-strand DNA breaks (Craig *et al.*, 2002; Finnegan, 1992). The induction of DSBs with CRISPR and potentially with injection of *Bari1* suggests a common mechanism for the initiation of the mobilization of the w^{67c23} allele and the *mini-white* gene.

If the induction of DSBs initiates the activation of TE in the genome, then how might this occur? We speculate that the formation of DSBs inhibits the Piwi-interacting small interfering RNAs (piRNAs) and small interfering RNAs (siRNAs) systems that repress TE

expression and transposition (Chung *et al.*, 2008; Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Lau *et al.*, 2009; Roy *et al.*, 2020). The piRNAs and siRNAs bind to TE transcripts through sequence complementarity removing TE transcripts (Chung *et al.*, 2008; Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Lau *et al.*, 2009; Roy *et al.*, 2020). Furthermore, piRNAs and siRNAs will guide RNA-protein complexes to euchromatic TEs recruiting DNA and histone methyltransferases to the TEs and resulting in an enrichment of repressive epigenetic markers (like H3K9me2) in regions with high TE densities (Choi & Lee, 2020). We speculate that DSBs counteract this gene silencing epigenetic mechanism.

One potential mechanism may involve the *Drosophila* H2A variant (H2Av). H2Av is the functional and structural chimeric of two eukaryotic conserved H2A variants: H2AX and H2AZ (Baldi & Becker, 2013). Studies have shown that H2AZ is involved in gene transcription regulation and heterochromatin silencing (Billon & Côté, 2012), whereas H2AX is responsible for the repair of DNA damage, especially the phosphorylation of the C-terminal end of H2AX is crucial of mediating the machinery of DNA damage repair (Scully & Xie, 2013). H2Av is also phosphorylated at sites of DSBs (Joyce *et al.*, 2011; Lake *et al.*, 2013; Madigan *et al.*, 2002). The DSBs initiated via CRISPR in our experiments and via the TE *Bari1* in Moschetti's study might cause the phosphorylation of H2Av and override TE repressing mechanisms (inactivating demethylases that remove H3K9me2 markers in the TE regions) resulting in increased TE mobilization. Within 50kb of either side of the DSBs of *bcd* and *Scr* there are three and five TEs respectively. Alternatively, the DSBs may affect different epigenetic mechanisms that results in fewer

methyltransferases being recruited to the TE region, increasing TE activity and thus resulting in the observed increase in transposition (Lee & Karpen, 2017).

Significance and implications

The DSBs induced by the CRISPR/Cas9 system initiate the activation of TEs in the genome. The transposition of the w^{67c32} allele is catalyzed by different transposases and provides evidence supporting the proposal that the FB-NOF element is non-autonomous (Badal *et al.*, 2013). The potential of the induction of transposition at all TE integration sites raises more concerns regarding the unintended consequences of CRISPR/Cas9 based genetic manipulations by identifying the additional potential problem of DSBs inducing transposition. The unintended activation of TE transposition may result in the phenotype of mutant alleles in genes closely linked to the target gene being ascribed to the target gene. Therefore, the assessment of the phenotype of at least two independent CRISPR alleles is suggested. In addition, if this occurs in humans then it may affect the ability to safely modify human genetic conditions with CRISPR. Finally, these w^{67c23} allele migration events are examples of exon shuffling and transduction transposition, which refers to the phenomenon where exons from different genes are transposed into new genomic contexts and under control of new promoters (Gilbert, 1978; Moran *et al.*, 1999; Moschetti *et al.*, 2004). There is the potential that DSB initiated transposition may be associated with exon migration that may facilitate the generation of novel genetic functions during the evolution. The activation of TE with DSBs may provide an experimental paradigm to study the mechanism of activation of TEs and gene evolution.

Limitations of CRISPR experiment

The CRISPR injections in this project were not successful. No on-target transformants were collected from *Scr* and *bcd* experiments. Several limitations of CRISPR experiment design need to be addressed. First of all, the efficiency of the designed chiRNAs needs to be verified through algorithmically designed software (Naeem *et al.*, 2020). The specificity of Cas9 is very high in bacteria genome. However, due to the genome size and complexity, the off-target effects in eukaryotic genome are much higher in eukaryotic genome than bacteria (Pattanayak *et al.*, 2013). In order to increase the specificity and efficiency of CRISPR technique, researchers have developed algorithm-based computational tools to help design the chiRNA, such as: CasOT, Cas-OFFinder, Digenome-seq, SITE-seq, GUIDE-seq and etc. (Naeem *et al.*, 2020). The chiRNAs used in this project, especially the chiRNAs targeting *Scr*, were designed manually to target the region to be deleted and next to the deletion boundary, and did not use these algorithms. Potential off-target sites in the *Drosophila* genome or even in the donor template sequences were not detected. Second, the DNA sequence of the target locus in the *yw* line genome needs to be sequenced. The chiRNAs were designed based on the reference genome sequence of *Drosophila* on Flybase. There might be polymorphisms between the reference genome and the genome of *yw* line. The target efficiency of CRISPR system is determined through 20 nucleotide sequences of chiRNA and the PAM sequence (Fu *et al.*, 2013). The potential polymorphisms may disrupt the chiRNA design as more than three mismatches between target sequences and 20 nucleotides of chiRNA can result in off-target effects and four mismatches in distal end of PAM may also induce off-target effects (Fu *et al.*, 2013; Singh *et al.*, 2016). Third, two DSBs are used instead of one DSB. Comparing to the *Fst*

experiment (Newman *et al.*, 2017), the positive control, one major difference is that *bcd* and *Scr* induce two DSBs 2.6kb and 16.8 kb apart, respectively, versus a single DSB for *Fst*. It is possible that widely separated DSBs may increase the number of non-homologous events and reduce the efficiency of the CRISPR design. Fourth, using y^1w^{67c23} as the target line should be avoided. As the DSB will induce the w^{67c23} allele migration inside the genome, using y^1w^{67c23} as the target line for CRISPR injection should be avoided. Other *w* amorphic allele like w^{1118} may be a better option.

REFERENCE

Anwar, S. L., Wulaningsih, W., & Lehmann, U. (2017). Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *International journal of molecular sciences, 18*(5), 974.

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., & Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature Structural & Molecular Biology, 21*(4), 366-374.

Badal, M., Xamena, N., & Cabré, O. (2013). FB-NOF is a non-autonomous transposable element, expressed in Drosophila melanogaster and present only in the melanogaster group. *Gene, 526*(2), 459-463.

Baldi, S., & Becker, P. B. (2013, 2013/08/01). The variant histone H2A.V of Drosophila—three roles, two guises. *Chromosoma, 122*(4), 245-258.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., & Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell, 144*(2), 214-226.

Bassett, A. R., Tibbit, C., Ponting, C. P., & Liu, J.-L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Reports, 4*(1), 220-228.

Bhadra, U., Pal-Bhadra, M., & Birchler, J. A. (1997). A Sex-Influenced Modifier in Drosophila That Affects a Broad Spectrum of Target Loci Including the Histone Repeats. *Genetics, 146*(3), 903.

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics, 45*(1), 273-297.

Billon, P., & Côté, J. (2012, 2012/03/01/). Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1819*(3), 290-302.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research, 24*(1), 132-141.

Choi, J. Y., & Lee, Y. C. G. (2020). Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLOS Genetics, 16*(7), e1008872.

Chung, W. J., Okamura, K., Martin, R., & Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol, 18*(11), 795-802.

Collins, M., & Rubin, G. M. (1982). Structure of the drosophila mutable allele, *white-crimson*, and its *white-ivory* and wild-type derivatives. *Cell, 30*(1), 71-79.

Collins, M., & Rubin, G. M. (1984). Structure of chromosomal rearrangements induced by the FB transposable element in Drosophila. *Nature, 308*(5957), 323-327.

Corrigan-Curay, J., O'Reilly, M., Kohn, D. B., Cannon, P. M., Bao, G., Bushman, F. D., Carroll, D., Cathomen, T., Joung, J. K., Roth, D., Sadelain, M., Scharenberg, A. M., von Kalle, C., Zhang, F., Jambou, R., Rosenthal, E., Hassani, M., Singh, A., & Porteus, M. H. (2015). Genome Editing Technologies: Defining a Path to Clinic: Genomic Editing: Establishing Preclinical Toxicology Standards, Bethesda, Maryland 10 June 2014. *Molecular Therapy, 23*(5), 796-806.

Craig, N. L., Craigie, R., Gellert, M., & Lambowitz, A. M. (2002). *Mobile DNA II*. American Society of Microbiology.

Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., & Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. *Nature, 453*(7196), 798-802.

Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas System and Its Role in Phage-Bacteria Interactions. *Annual Review of Microbiology, 64*(1), 475-493.

Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Molecular systems biology, 9*, 641-641.

Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics, 41*(1), 331-368.

Finnegan, D. J. (1992). Transposable elements. *Current Opinion in Genetics & Development, 2*(6), 861-867.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology, 31*(9), 822-826.

Garofalo, L. E. (2015). Determination of the phylogenetic range of Sex combs reduced activity in Drosophila melanogaster. *Electronic Thesis and Dissertation Repository., 3067*.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., & Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science, 320*(5879), 1077-1081.

Gilbert, W. (1978). Why genes in pieces? *Nature, 271*(5645), 501-501.

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J., & O'Connor-Giles, K. M. (2013). Engineering of Drosophilawith the CRISPR RNA-Guided Cas9 Nuclease. *Genetics, 194*(4), 1029.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*. *Genetics, 196*(4), 961.

Harden, N., & Ashburner, M. (1990). Characterization of the FB-NOF transposable element of Drosophila melanogaster. *Genetics, 126*(2), 387-400.

Hazelrigg, T., Levis, R., & Rubin, G. M. (1984). Transformation of *white* locus DNA in Drosophila: Dosage compensation, *zeste* interaction, and position effects. *Cell, 36*(2), 469-481.

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science, 327*(5962), 167.

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., Pfeiffer, B. D., George, R. A., Svirskas, R., Krzywinski, M., Schein, J., Accardo, M. C., Damia, E., Messina, G., Méndez-Lago, M., de Pablos, B., Demakova, O. V., Andreyeva, E. N., Boldyreva, L. V., Marra, M., Carvalho, A. B., Dimitri, P., Villasante, A., Zhimulev, I. F., Rubin, G. M., Karpen, G. H., & Celniker, S. E. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome research, 25*(3), 445-458.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827-832.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science, 337*(6096), 816.

Joyce, E. F., Pedersen, M., Tiong, S., White-Brown, S. K., Paul, A., Campbell, S. D., & McKim, K. S. (2011). Drosophila ATM and ATR have distinct activities in the

regulation of meiotic DNA damage and repair. *Journal of Cell Biology, 195*(3), 359-367.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C., & Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature, 453*(7196), 793-797.

Lake, C. M., Korda Holsclaw, J., Bellendir, S. P., Sekelsky, J., & Hawley, R. S. (2013). The Development of a Monoclonal Antibody Recognizing the Drosophila melanogaster Phosphorylated Histone H2A Variant (γ-H2AV). *G3 Genes|Genomes|Genetics, 3*(9), 1539-1543.

Lau, N. C., Robine, N., Martin, R., Chung, W. J., Niki, Y., Berezikov, E., & Lai, E. C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res, 19*(10), 1776-1785.

Lee, Y. C. G., & Karpen, G. H. (2017). Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife, 6*, e25762.

Levis, R., Hazelrigg, T., & Rubin, G. M. (1985). Separable cis-acting control elements for expression of the white gene of Drosophila. *The EMBO journal, 4*(13A), 3489-3499.

Linheiro, R. S., & Bergman, C. M. (2012). Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. *PloS one, 7*(2), e30008-e30008.

Madigan, J. P., Chotkowski, H. L., & Glaser, R. L. (2002). DNA double‐strand break‐induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation‐induced apoptosis. *Nucleic acids research, 30*(17), 3698-3705.

Majumdar, S., & Rio, D. C. (2015). P Transposable Elements in Drosophila and other Eukaryotic Organisms. *Microbiology spectrum, 3*(2), MDNA3-2014.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology, 9*(6), 467-477.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., & Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology, 31*(9), 833-838.

Marzo, M., Puig, M., & Ruiz, A. (2008). The *Foldback*-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. *Proceedings of the National Academy of Sciences, 105*(8), 2957.

Mérel, V., Boulesteix, M., Fablet, M., & Vieira, C. (2020). Transposable elements in Drosophila. *Mobile DNA, 11*(1), 23.

Moran, J. V., DeBerardinis, R. J., & Kazazian, H. H. (1999). Exon Shuffling by L1 Retrotransposition. *Science, 283*(5407), 1530.

Moschetti, R., Marsano, R. M., Barsanti, P., Caggese, C., & Caizzi, R. (2004). FB elements can promote exon shuffling: a promoter-less white allele can be reactivated by FB mediated transposition in Drosophila melanogaster. *Molecular Genetics and Genomics, 271*(4), 394-401.

Naeem, M., Majeed, S., Hoque, M. Z., & Ahmad, I. (2020). Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells, 9*(7), 1608. https://doi.org/10.3390/cells9071608

Newman, C. E., Toxopeus, J., Udaka, H., Ahn, S., Martynowicz, D. M., Graether, S. P., Sinclair, B. J., & Percival-Smith, A. (2017). CRISPR-induced null alleles show that Frost protects Drosophila melanogaster reproduction after cold exposure. *J Exp Biol, 220*(Pt 18), 3344-3354.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology, 31*(9), 839-843. https://doi.org/10.1038/nbt.2673

Pirrotta, V., Steller, H., & Bozzetti, M. P. (1985). Multiple upstream regulatory elements control the expression of the Drosophila white gene. *The EMBO journal, 4*(13A), 3501-3508.

Port, F., Chen, H.-M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proceedings of the National Academy of Sciences, 111*(29), E2967.

Potter, S., Truett, M., Phillips, M., & Maher, A. (1980). Eucaryotic transposable genetic elements with inverted terminal repeats. *Cell, 20*(3), 639-647.

Ragab, A., Thompson, E. C., & Travers, A. A. (2006). High Mobility Group Proteins HMGD and HMGZ Interact Genetically With the Brahma Chromatin Remodeling Complex in Drosophila. *Genetics, 172*(2), 1069.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols, 8*(11), 2281-2308.

Roy, M., Viginier, B., Saint-Michel, É., Arnaud, F., Ratinier, M., & Fablet, M. (2020). Viral infection impacts transposable element transcript amounts in *Drosophila*. *Proceedings of the National Academy of Sciences, 117*(22), 12249.

Sabl, J. F., & Birchler, J. A. (1993). Dosage dependent modifiers of white alleles in Drosophila melanogaster. *Genet Res, 62*(1), 15-22.

Scully, R., & Xie, A. (2013, 2013/10/01/). Double strand break repair functions of histone H2AX. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 750*(1), 5-14.

Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., & Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature Communications, 7*, 12778-12778.

Sivanantharajah, L., & Percival-Smith, A. (2009). Analysis of the Sequence and Phenotype of Drosophila Sex combs reduced Alleles Reveals Potential Functions of Conserved Protein Motifs of the Sex combs reduced Protein. *Genetics, 182*(1), 191-203.

Southworth, J. W., & Kennison, J. A. (2002). Transvection and silencing of the Scr homeotic gene of Drosophila melanogaster. *Genetics, 161*(2), 733-746.

Suárez, S., Cabré, O., Velázquez, A., Creus, A., Marcos, R., & Xamena, N. (1996). Molecular study of the germinal reversions induced at the white-ivory locus in Drosophila melanogaster. *Mutagenesis, 11*(6), 559-563.

Templeton, N. S., & Potter, S. S. (1989). Complete foldback transposable elements encode a novel protein found in Drosophila melanogaster. *The EMBO journal, 8*(6), 1887-1894.

Venken, K. J. T., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A., & Bellen, H. J. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nature Methods, 8*(9), 737-743.

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids, 4*, e264.

Chapter 3. The differential pleiotropy of phenotypic non-specificity in *Drosophila melanogaster*

Introduction

The regulation of the rate of transcription initiation is a major mechanism controlling gene expression. The rate is mediated by transcription factors (TFs) that bind to specific DNA binding sites in the regulatory sequences of a gene. The set of TFs expressed in a cell are responsible for the transcription of a unique set of genes that dictate the phenotype of a cell or group of cells. A common view of TF function is that the unique expression pattern observed in a cell is dependent on the binding of TFs to specific *cis*-acting elements in the regulatory sequences of genes. This is achieved through DNA binding domains that recognize a specific DNA sequence, together with specific cooperative protein-protein interactions such that a restricted set of genes are regulated to bring about a phenotype. For example, the yeast mating type TF, α1, in a cooperative interaction with the MCM1 protein, activates the expression of both the "α"pheromone and "α" pheromone receptor required for the α mating type phenotype (Elble & Tye, 1991). Based on traditional models, there is little expectation that the function of TF α1 can be substituted with by another TF that recognizes a distinct DNA binding sequence. Thus, the observation of phenotypic non-specificity, where multiple distinct TFs are capable of inducing or rescuing the same

phenotype, is surprising and hard to reconcile with a very specific model of TF function (Percival-Smith, 2017, 2018).

Phenotypic non-specificity is observed within and between TF families (Banreti *et al.*, 2014; Greig & Akam, 1995; Hirth *et al.*, 2001; Lelli *et al.*, 2011; Percival-Smith, 2017; Percival-Smith & Laing Bondy, 1999; Percival-Smith *et al.*, 2013; Percival-Smith *et al.*, 2005). However, most of the phenotypes assessed for phenotypic non-specificity using TFs from different families were the result of the ectopic expression of the TF; the only example of functional complementation being the rescue of PB-dependent growth of the maxillary palp by DSX^{M} (Percival-Smith, 2017). Examples of phenotypic non-specificity are not limited to *Drosophila*. For example, the three OSK (Oct3/4, Sox2, Klf4) TFs induce pluripotency with very low efficiency that is increased by co-expression of either Myc or Glis1 (Kulcenty *et al.*, 2015). In addition, single cell transcriptomics has uncovered phenotypic convergence where distinct sets of TFs regulate the same phenotype in the optic lobe of *Drosophila*; a situation easily likened to phenotypic non-specificity where multiple TFs induce or rescue the same phenotype (Konstantinides *et al.*, 2018).

Changes in TF function during evolution are largely attributed to changes in *cis*-regulatory sequences that alter TF expression or expression of TF target genes (Carroll *et al.*, 2004; Ludwig, 2002; Simpson, 2002; Stern, 2000; Tautz, 2000; Wray *et al.*, 2003). Because of the autonomous nature of *cis* regulatory sequences, mutations that affect *cis* regulatory elements have limited effects. In contrast, due to the constraint and pleiotropy of TFs, protein coding mutations are thought to be deleterious and subject to strong purifying

selection. However, these ideas are based on uniform pleiotropy, where each coding mutation affects every trait or function for which a given protein is required. Differential pleiotropy (the non-uniform effect of mutations) and functional redundancy are two mechanisms that reduce mutational pleiotropy. The examples of differential pleiotropy observed with genes encoding HD-containing proteins suggests that the transcriptional functions of HD-containing TFs are dispersed as small protein elements throughout the protein, and that each of these elements make small, tissue-specific contributions to overall TF function (Hittinger *et al.*, 2005; Joshi *et al.*, 2007; Merabet *et al.*, 2011; Percival-Smith *et al.*, 2013; Prince *et al.*, 2008; Sivanantharajah & Percival-Smith, 2009, 2014, 2015; Tour *et al.*, 2005). Phenotypic non-specificity suggests functional redundancy between TFs. It is therefore appropriate to determine whether the rescue exhibits uniform or differential pleiotropy in examples of functional complementation.

A model of limited specificity of TF function explains phenotypic non-specificity (Percival-Smith, 2018). The model predicts phenotypic non-specificity to be a widespread phenomenon and thus easily observable in nature. In this chapter, I screened for phenotypic non-specificity using functional complementation of six TF loci with at least 12 non-resident TFs. I found that phenotypic non-specificity was frequently observed and that the rescue of the phenotypes was differentially pleiotropic.

Materials and Methods

***Drosophila husbandry*:** Flies were maintained at 23°C and 60% humidity and were reared in 20ml vials and 300ml milk bottles containing corn meal media [1% (w/v) Drosophila-

grade agar, 6% (w/v) sucrose, 10% (w/v) cornmeal, 1.5% (w/v) yeast and 0.375% (w/v) 2-methyl hydroxybenzoate]. To collect eggs, embryos, and first instar larvae, flies were allowed to lay eggs on apple juice plates [2.5% (w/v) Drosophila-grade agar, 6% (w/v) sucrose, 50% apple juice and 0.3% (w/v) 2-methy hydroxybenzoate] and the progeny aged to the appropriate stage for a given analysis. All genotypes were assembled with standard *Drosophila* crossing schemes.

Genetic screens for phenotypic non-specificity: A total of 13 TFs were used as part of this study: Labial (LAB), Deformed (DFD), Antennapedia (ANTP), Sex combs reduced (SCR), Doublesex male (DSX^M), Apterous (AP), Bric a bac 1 (BAB1), Eyeless (EY), Squeeze (SQZ), Forkhead box subgroup O (FOXO), Disco (DISCO), Broad Z1 (BR.Z1), Broad Z2 (BR.Z2). In each experiment, the resident TF and at least 12 non-resident transcription factors were screened for rescue of the seven TF phenotypes. All of these TFs were expressed from UAS constructs inserted on the second chromosome (*UAS-X*). The genotypes used in the six screens are:

Labial screen, *y w*; *P{UAS-X, w+}* or *P{UAS-X, w+}/CyO*; *lab[14] /TM6B*, *Tb*, *P{walLy}* X *y w*; *P{labGAL4, w+}/CyO*; *lab[4] / TM6B*, *Tb*, *P{walLy}*;

Deformed screen, *y w*; *P{UAS-X, w+}* or *P{UAS-X, w+}/CyO*; *Dfd [12]/TM6B*, *Tb*, *P{walLy}* X *y w*; *P{DfdGAL4, w+}*, *Dfd [16] e/ TM6B*, *Tb*, *P{walLy}*;

Sex combs reduced screen, *y w*; *P{UAS-X, w+}* or *P{UAS-X, w+}/CyO*; *Scr[4] e /TM6B*, *Tb*, *P{walLy}* X *y w*; *P{ScrGAL4, w+}*; *Scr[2] cu p[p]/ TM6B*, *Tb*, *P{walLy}*;

Ultrabithorax screen, *y w*; *P{UAS-X, w+}* or *P{UAS-X, w+}/CyO*; *Ubx[abx1, bx3, 61d, pbx1] /TM6B*, *Tb*, *P{walLy}* X *y w*; *Ubx[9.22]*, *P{UbxGAL4, w+} / TM6B*, *Tb*, *P{walLy}*;

Doublesex screen, *y w*; *P*{*UAS-X*, w^+} or *P*{*UAS-X*, w^+}/*CyO*; dsx^1 /*TM6B*, *Tb*, *P*{*walLy*} X *w*; *dsxGAL4* / *TM6B*, *Tb*;

Fruitless screens, *y w*; *P*{*UAS-X*, w^+} or *P*{*UAS-X*, w^+}/*CyO*; $fru^{4\text{-}40}$ /*TM6B*, *Tb*, *P*{*walLy*} X *w*; *fruGAL4A*/ *TM6B*, *Tb*, *P*{*walLy*} or *y w*; *fruGAL4B* / *TM6B*, *Tb*, *P*{*walLy*}.

The two control stocks for the *Ubx* and *dsx* screens with *UAS* insertions on the third chromosome are: *y w*; *P*{*UAS-Ubx*, w^+}, $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}$ /*TM6B*, *Tb*, *P*{*walLy*}, and *y w*; *P*{*UAS-*dsx^F, w^+}, dsx^1/ *TM6B*, *Tb*, *P*{*walLy*}.

The sources of the different lines used in the experiment are listed below:

TABLE-1 Source of GAL4 and mutant allele lines

Allele Name	Stock # and Source
lab-GAL4	43652 Bloomington
lab^{4}	2084 Bloomington
lab^{14}	2092 Bloomington
Scr-GAL4	43656 Bloomington
Scr^{2}	2185 Bloomington
Scr^{4}	2188 Bloomington
Dfd-GAL4	48844 Bloomington
Dfd^{12}	2315 Bloomington
Dfd^{16}	2325 Bloomington
dsx-GAL4	66674 Bloomington
dsx^{1}	1679 Bloomington
fru-GAL4 (A)	66696 Bloomington
fru-GAL4 (B)	30027 Bloomington
$fru^{4\text{-}40}$	Obtained from Amanda Moehring (Chowdhury *et al.*, 2020)
rn-GAL4	76179 Bloomington
Ubx-GAL4	48137 Bloomington
$Ubx^{9.22}$	3474 Bloomington
$Ubx^{abx1,\ bx3,\ 61d,\ pbx1}$	101566 Kyoto

Source of UAS lines

Allele Name	Stock # and Source
UAS-lab	7300 Bloomington
UAS-Scr	7302 Bloomington
UAS-Dfd	7299 Bloomington
UAS-Antp	7301 Bloomington
UAS-ap	42223 Bloomington
UAS-bab1	6939 Bloomington
UAS-br.Z1	51190 Bloomington
UAS-br.Z2	51380 Bloomington
UAS-disco	6846 Bloomington
$UAS\text{-}dsx^{F}$	44223 Bloomington
$UAS\text{-}dsx^{M}$	44224 Bloomington
UAS-ey	6294 Bloomington
$UAS\text{-}fru^{MC}$	66695 Bloomington
UAS-foxo	9575 Bloomington
UAS-sqz	36497 Bloomington
UAS-Ubx	911 Bloomington

"Bloomington" stands for Bloomington Drosophila Stock Center (Bloomington, USA). "Kyoto" stands for Kyoto Stock Center (Kyoto, Japan).

Phenotypic analysis: For the *Labial*, *Deformed*, and *Sex combs reduced* screens, the Hox^{null} genotypes were marked independently of the *Hox* phenotype with *yellow* (Hyduk & Percival-Smith, 1996). In all crosses, the parents were *y*; Hox^{null}/*TM6B*, *Tb*, *P*{*walLy*}, and therefore, the *y*; Hox^{null}/Hox^{null} progeny were yellow (y) because all other progeny have *TM6B*, *Tb*, *P*{*walLy*}, which were y^{+}. At 20-32h AEL, first instar larvae were dechorionated with bleach and devitellinized by shaking in a 1:1 heptane/methanol mixture. The larvae were mounted in Hoyer's mounting and viewed under bright field, phase contrast and dark field optics (Wieschaus & Nusslein-Volhard, 1986). The bright field images of the head skeletons were processed with the extended focus function of the software, Zen (Zeiss, Oberkochen, Germany). The phase contrast images were processed with the software ImageJ (NIH, Bethesda, USA). The proportion of Hox^{null} hatched larvae relative to the total numbers of eggs laid was determined. For *labial* morphometric analysis, the distance between mouth hooks and the length of the head skeleton were measured using Openlab software (Leica, Wetzlar, Germany). For *Scr* morphometric analysis, the number of T1 beard setae were counted manually in dark field images taken in Openlab software (Leica, Wetzlar, Germany). To count the number of sex combs in the *Scr* screen, the first pair of legs of *y w*; *P*{*ScrGAL4*, w^{+}}/*P*{*UAS-X*, w^{+}}; Scr^{4} *e*/ *TM6B*, *Tb*, *P*{*walLy*} were collected and mounted in Hoyer's mounting.

For the *Ultrabithorax* screen, the *yw*; *P*{*UAS-X*, w^{+}}/+; $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}$ /$Ubx^{9.22}$, *P*{*UbxGAL4*, w^{+}} genotypes were identified as adults or pupae by their non-humeral or non-tubby phenotypes, respectively (lacking *TM6B*, *Tb*, *P*{*walLy*}). Eclosed or pharate adults of the correct genotype were critical point dried, sputter coated and imaged with

scanning electron microscopy at the Biotron Integrated Microscopy Laboratory (London, ON).

To count sex combs for the *doublesex* screen, the first pair of legs were mounted in Hoyer's mounting. The dorsal abdomens were imaged with a dissecting microscope and the images processed with the extended focus function of the software, Zen. To image the genitals, eclosed flies were critical point dried, sputter coated, and imaged with scanning electron microscopy at the Biotron Integrated Microscopy Laboratory (London, ON).

For the *fruitless* screen, freshly eclosed male flies were placed in separate vials and aged for 3-5 days in all assays. For the fertility assay, individual males were mated with two wild type, two-day-old virgin females, and the females allowed to lay eggs for 3 days. The vials were assessed for larvae at 7 days after mating. For the courtship assays, males were introduced to either two-day-old virgin females or two-day-old males (marked on the wings with a Sharpie marker) and their behavior recorded on a video for 10 minutes. The videos were scored for orienting, male follows female, male wing extension, genital licking, attempted copulation and copulation behaviors that are associated with mating activity and a courtship index (CI) subsequently determined.

Statistical analysis: For the *lab* head skeleton lengths and *Scr* / *dsx* sex combs number data (assessed as normal and of equal variance using QQ plots and plotting residuals), ANOVA was performed and followed with a Tukey's or Dunnett's post hoc analysis. For the rescue of *lab*, *Dfd*, and *Scr*, as well as the fertility and CI of *fru*, ANOVA on ranks was performed

followed by a Dunn's pos hoc analysis. Rescue of *lab*, *Dfd*, and *Scr* was assessed using chi-squared tests. The t-test (one-tail) was performed on the sex combs number data between WT and FOXO ectopic expression flies.

Results

***Overview of phenotypic non-specificity screens*:**

Phenotypic non-specificity is observed with ectopic expression of both HOX and non-HOX TFs (Banreti *et al.*, 2014; Greig & Akam, 1995; Hirth *et al.*, 2001; Lelli *et al.*, 2011; Percival-Smith, 2017; Percival-Smith *et al.*, 2013; Percival-Smith *et al.*, 2005). However, the only example of functional complementation with a non-HOX TF is the rescue of the *pb* maxillary palp phenotype (Percival-Smith, 2017). Functional complementation of a phenotype has a more straight-forward interpretation than the induction of a phenotype by ectopic expression. Furthermore, the rescue of pleiotropic phenotypes allows the assessment of differential/uniform pleiotropy.

For these reasons, we screened for phenotypic non-specificity using a strategy based on functional complementation. In these experiments the UAS/GAL4 system was used to assess the phenotypic rescue of six *TF* loci, four *Hox* loci (*lab*, *Dfd*, *Scr* and *Ubx*) and two non-*Hox* loci (*dsx* and *fru*), by the expression of at least 12 non-resident TFs. Since we used functional complementation to study phenotypic non-specificity and pleiotropy, the resident TF we chose were pleiotropic and had obvious, clear phenotypes. HOX proteins establish embryonic segment identities along the AP axis of bilaterian bodies and *Hox* genes are pleiotropic (Carroll *et al.*, 2004). The flies with amorphic or hypomorphic

mutations of *Hox* genes exhibit reproducible, severe developmental defects. Therefore, the four *Hox* loci (*lab*, *Dfd*, *Scr* and *Ubx*) were chosen as the resident *TF* loci. Similar reasoning applies for *dsx* and *fru* as well. The *dsx* gene encodes a transcription factor required for both male and female sex determination of *Drosophila* and the male isoform of FRU is responsible for male fertility and courtship behavior. Their complex functions make them good candidates for the resident *TF* loci.

For non-resident TFs, TFs within the HD-superfamily and outside HD-superfamily were selected for the study of phenotypic non-specificity. Using this reasoning, available UAS construct lines were ordered.

For the four *Hox* loci (*lab*, *Dfd*, *Scr* and *Ubx*), we used drivers composed of *Hox* regulatory elements fused to *GAL4*. Genetic backgrounds were created that carried the driver and were hemizygous for null *Hox* alleles. For the two non-*Hox* loci, *dsx* and *fru*, we used driver stocks with insertion of GAL4 into the loci that created both a loss-of-function allele and expressed GAL4 from the regulatory sequences of these loci. In these genetic backgrounds we screened for rescue of the phenotype by resident and non-resident TFs expressed from *P*{*UAS-TF*} insertions. The identity of the correct constructs in the 13 UAS-TF lines was verified using PCR (**Figure 1**).

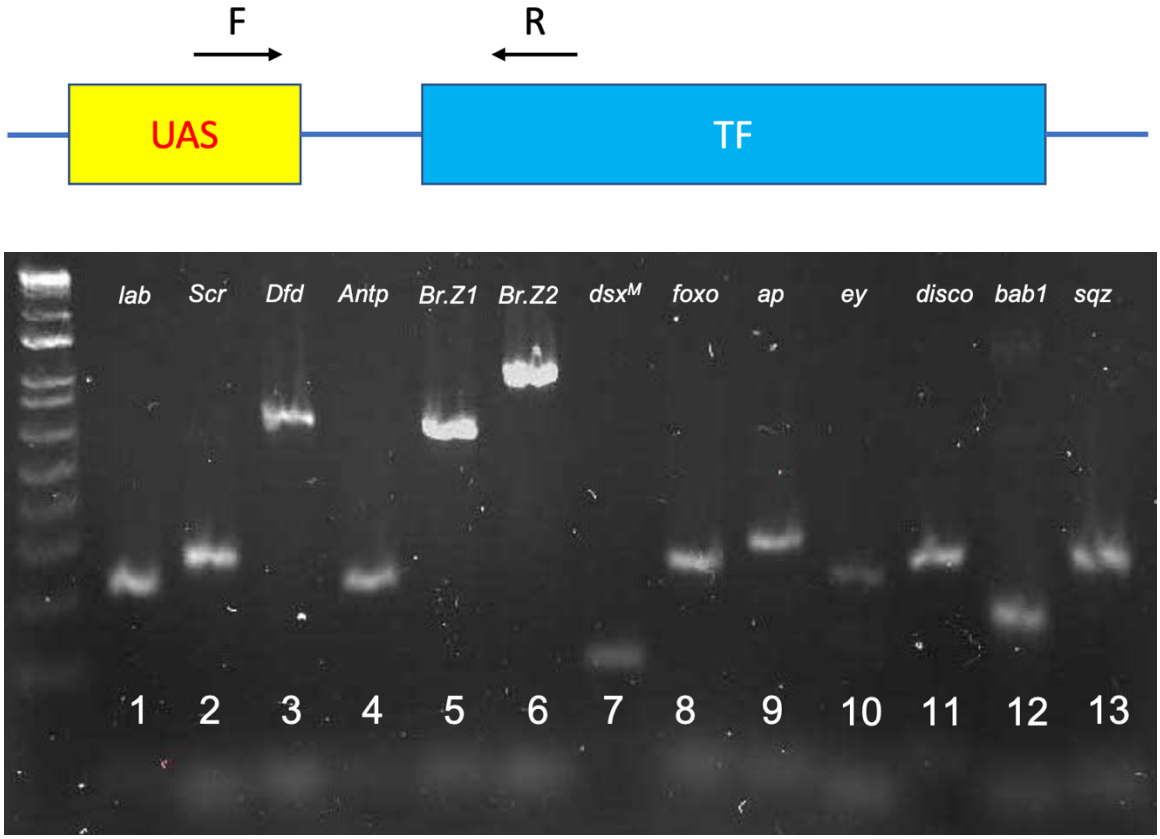


Figure 1. PCR verification of *UAS-TF* lines. The schematic (top) indicates the positions of the forward (F) and reverse (R) primers used to verify the *UAS-TF* lines. PCR amplicons were analyzed by gel electrophoresis (bottom). Lane 1*: UAS-lab*. Lane 2: *UAS-Scr*. Lane 3: *UAS-Dfd*. Lane 4: *UAS-Antp*. Lane 5: *UAS-br.Z1*. Lane 6: *UAS-br.Z2*. Lane 7: *UAS-dsx^{M}*. Lane 8: *UAS-foxo*. Lane 9: *UAS-ap*. Lane 10: *UAS-ey*. Lane 11: *UAS-disco*. Lane 12: *UAS-bab1*. Lane 13: *UAS-sqz*. The ladder is 1kb+ DNA ladder (Thermo Fisher, Massachusetts, USA).

Phenotypic rescue by non-resident TFs:

The results of the respective screens described in the previous section are presented sequentially below, beginning with *labial*.

labial: *Drosophila* lab^{4}/lab^{14} larvae lack the H-piece, including the bridge and the lateral bar (**Figure 2 A1**). Furthermore, the two mouth hooks are widely separated as opposed to being close together due to failure of head involution (**Figure 2 A1**). Since both the *lab GAL4* and *UAS-lab* insertions were heterozygous, the expected frequency of embryonic rescue among the yellow cuticles was 25%. Of the 50 yellow cuticles (lab^{4}/lab^{14}) examined in experiments assaying LAB expression, 11 were rescued (**Figure 2 A2**). This was not different from the expected 12.5 (χ^2 (1, N = 50) = 0.18, P = 0.7). The expression of LAB from *UAS-lab* rescued embryonic head involution and head skeleton defects (H-piece lateral bar and the dorsal bridge but not the H-piece bridge and median tooth) (**Figure 2 A2**). Out of 238 hatched embryos examined, eight yellow larvae hatched; however, these larvae did not survive to the pupal stage. No examples of pupal (non-Tb)/adult rescue were observed.

	lab rescue		*Dfd* rescue		*Scr* rescue	
Null	A1	0/50	B1	0/50	C1	0/50
Protein expressed						
LAB	A2	11/50 *P*=0.002	B2	0/50 *P*=1	C2	0/50 *P*=1
DFD	A3	0/50 *P*=1	B3	13/50 *P*<0.001	C3	0/50 *P*=1
SCR	A4	0/50 *P*=1	B4	0/50 *P*=1	C4	10/50 *P*=0.003
ANTP	A5	0/50 *P*=1	B5	0/50 *P*=1	C5	0/50 *P*=1
AP	A6	0/50 *P*=1	B6	1/50 *P*=1	C6	0/50 *P*=1
BAB1	A7	0/50 *P*=1	B7	0/50 *P*=1	C7	0/50 *P*=1
BRZ1	A8	0/50 *P*=1	B8	0/50 *P*=1	C8	0/50 *P*=1
BRZ2	A9	0/50 *P*=1	B9	0/50 *P*=1	C9	0/50 *P*=1
DISCO	A10	0/50 *P*=1	B10	0/50 *P*=1	C10	0/50 *P*=1
DSX^M	A11	9/50 *P*=0.03	B11	0/50 *P*=1	C11	0/50 *P*=1
EY	A12	0/50 *P*=1	B12	0/50 *P*=1	C12	0/50 *P*=1
FOXO	A13	0/50 *P*=1	B13	0/50 *P*=1	C13	0/50 *P*=1
SQZ	A14	0/50 *P*=1	B14	0/50 *P*=1	C14	0/50 *P*=1
Wild-type	A15	50/50 *P*<0.001	B15	50/50 *P*<0.001	C15	50/50 *P*<0.001

Figure 2. Screen for the rescue of lab (A), Dfd (B) and Scr (C) phenotypes.

All images are brightfield images of the head skeleton. The lab^{null} genotype is lab^{4} / lab^{14}; the Dfd^{null} genotype is Dfd^{12}/Dfd^{16}; and the Scr^{null} genotype is Scr^{2} / Scr^{4}. For all assays of rescue, 50 *y* larvae were examined, and the frequency of rescue indicated in the column to the right. For *Dfd*, the number of rescued mouth hooks was assessed. The rescue data was analyzed using an ANOVA on ranks [*lab H* (14) = 561 $P<0.0001$; *Dfd H* (14) = 624 $P<0.0001$; *Scr H* (14) = 640 $P<0.0001$]. The *P* values of a Dunn's post hoc test relative to the control null mutant are indicated below the frequency of the rescue. The scale bar in A1 indicates 100 μm and is the same in all other images. The red arrow indicates rescue of *lab* by expression of LAB. The green arrow indicates rescue of *lab* by expression of DSX^{M}. The blue arrow indicates rescue of *Dfd* by expression of DFD. The purple arrow indicates rescue of *Scr* by expression of SCR.

The expression of 12 non-resident TFs were screened for rescue by carefully examining 50 yellow cuticles. Evidence of rescue in 9 larval cuticles expressing DSX^M were found. This frequency of rescue was not different from the expected 12.5 (χ^2 (1, N = 50) = 0.98, P = 0.3). One hatched larva expressing DSX^M was found but did not survive to the third instar larval stage (1/172). No examples of pupal/adult rescue were observed.

The rescues with LAB and DSX^M were differentially pleiotropic. LAB rescued the head involution phenotype such that the mouth hooks were in close proximity and rescued development of the H-piece lateral bar and dorsal bridge, and DSX^M only rescued the mouth hooks phenotype (**Figure 2 A11**). With morphometric analysis measuring the mouth hooks distance (distance between two tips of mouth hooks) and the head length (distance between the anterior end of the head and the posterior end of ventral arms) (**Figure 3 A - D**), we found a clear difference between rescued and mutant larvae (**Figure 3 E and F**). Using the rescue of head involution as an indication of genotype (*UASlab* or *UASdsx^M*), the length of the head skeleton was rescued relative to the *lab* null mutant with LAB and DSX^M, but the rescue observed with DSX^M relative to LAB was not as strong ($P < 0.0001$).

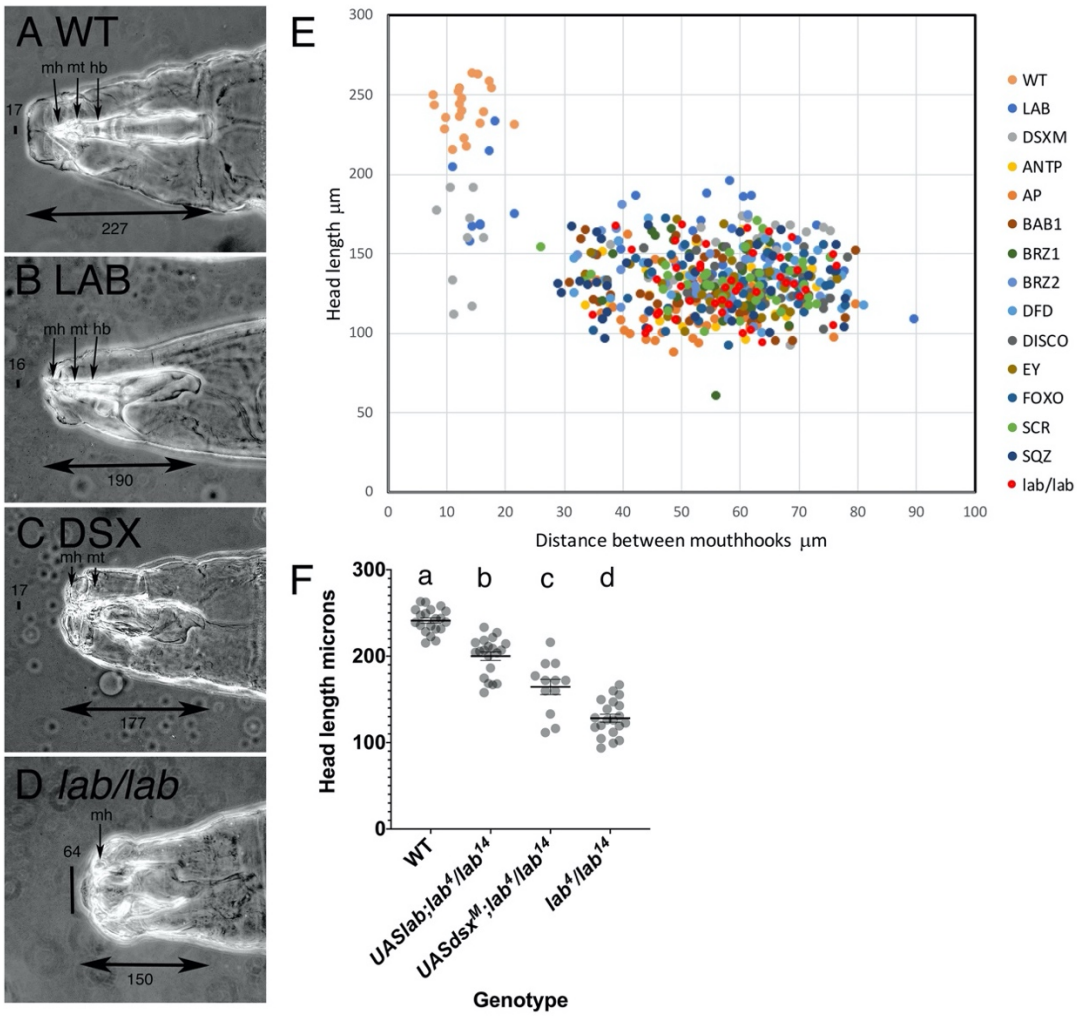


Figure 3. Characterization of the rescue of the *lab* phenotype.

Panels A-D are phase contrast images of the larval head skeleton. Panel A is wild type (*yw*); Panel B is *yw; UAS-lab/labGAL4; lab^{14}/lab^{4}*; Panel C is *y w; UAS-dsx^{M}/labGAL4; lab^{14}/lab^{4}*; and Panel D is *y w; lab^{14}/lab^{4}*. The vertical line is the measurement between mouth hooks and the horizontal line is the measurement of the length of the head skeleton. The arrows indicate mh: mouth hooks; mt: medium teeth; hb: H-piece bridge. Panel E is a plot of the distance between mouth hooks versus the length of the head skeleton for all larvae. Panel F is a plot of the length of the head skeleton of the rescued larvae. An ordinary

ANOVA was performed ($F_{3, 68} = 98$, $P < 0.0001$) followed by a Tukey's pair-wise comparison; the same letter indicates no difference ($P > 0.0001$). The mean and SEM are indicated.

Deformed: *Dfd* deficient (*Dfd* ^{12}/*Dfd* ^{16}) larvae lack the mouth hooks (**Figure 2 B1**) and cirri. Because the third chromosome carried both the *Dfd GAL4* insertion and the *Dfd* null allele and the *UAS-Dfd* insertion was heterozygous, the expected maximum frequency of rescue is 50%. Of the 50 yellow cuticles (*Dfd* ^{12}/*Dfd* ^{16}) examined in experiments assaying DFD expression, 13 had one or two rescued mouth hooks and 3 had rescued cirri (**Figure 2 B3**). This was less than the 25 expected (χ^2 (1, N = 50) = 5.76, P = 0.02). Out of 248 hatched embryos examined, one yellow larva hatched but the larva did not survive to the pupal stage. Allowing the progeny to develop to adulthood, no pupal or adult rescue was observed. The expression of 12 non-resident TFs were screened for rescue by carefully examining 50 yellow cuticles for each TF expressed. Although one cuticle expressing AP exhibited rescue of a mouth hook, the frequency of rescue was not significant (**Figure 2 B6**).

Sex combs reduced: During embryogenesis and metamorphosis SCR is required for head and thorax development (Sivanantharajah & Percival-Smith, 2009). *Scr*^{2}/*Scr* ^{4} embryos develop into larvae missing the medium tooth structure, and the anterior portion of the H-piece structure is curved (toward ventral side) (**Figure 2 C1**). Because both the *ScrGAL4* and *UAS-Scr* insertions were heterozygous, the expected maximum frequency of embryonic rescue was 25%. Of the 50 yellow cuticles (*Scr*^{2}/*Scr* ^{4}) examined in experiments assaying SCR expression, 10 were rescued (**Figure 2 C4**), which was not different from the expected 12.5 (χ^2 (1, N = 50) = 0.5, P = 0.5). Out of 162 eggs/embryos examined, no yellow larva hatched, and no non-Tb pupae were observed when the progeny were allowed to develop to adulthood. Head skeleton defects, T1 beard formation and duplication of the

antennal sense organ were rescued with the expression of SCR (**Figure 4 A – C**). Counting the number of setae in the T1 beard showed strong rescue of the T1 beard in larvae with rescue of the head skeleton relative to the Scr^{2}/Scr^{4} mutant ($P < 0.0001$); however, the number of setae in the T1 beard was less than that observed in wild type controls ($P < 0.0001$). In addition, expression of SCR was found to increase the number of male sex combs by about 2 bristles in a Scr^{4}/ Scr^{+} heterozygote (**Figure 4 E**; $P < 0.0001$). The number of sex combs is linearly associated with the dose/activity of SCR (Sivanantharajah & Percival-Smith, 2009); therefore, the increase of 2 bristles suggests that the expression of SCR from *UAS-Scr* by *ScrGAL4* is 20% of wild type levels. The T1 beard and sex comb data suggest that the level of SCR expression using the *ScrGAL4* driver and *UAS-Scr* was significantly less than wild type levels.

The expression of 12 non-resident TFs were screened for rescue by carefully examining 50 yellow cuticles for each TF. No rescue of the head skeleton defects was observed. The number of T1 beard setae were counted on at least 12 yellow larvae for each TF and no rescue of beard formation was observed (not a single larva had more than 80 setae) (**Figure 4 D**). The number of sex combs were counted on *P{UAS-TF}/P{ScrGAL4};Scr⁴/TM6B* adult males. The expression of FOXO increased the number of sex combs by about 2 bristles ($P < 0.0001$). The rescues with SCR and FOXO were differentially pleiotropic. The expression of SCR rescued larval head skeleton, T1 beard formation and increased the number of sex combs, whereas the expression of FOXO only increased the number of sex combs.

To test whether expression of FOXO induces ectopic sex combs like SCR, the *rnGAL4* driver was used to drive *UAS-Scr* and *UAS-foxo* expression in all three pairs of legs of *y* flies (**Figure 5**) (Sivanantharajah & Percival-Smith, 2014). Ectopic expression of SCR increases the number of sex combs on the first leg (**Figure 5 A**), induces ectopic transverse rows and sex combs on the second leg (**Figure 5 C**) and induces ectopic sex combs on third legs (**Figure 5 E**). Ectopic expression of FOXO increases the number of sex combs by 3 bristles on the first leg from 10 (WT) to 13.1 ($t(13) = -7.03$, $P < 0.00001$) (**Figure 5 B**) but does not induce ectopic sex combs on the second and third legs (**Figure 5 D, F**).

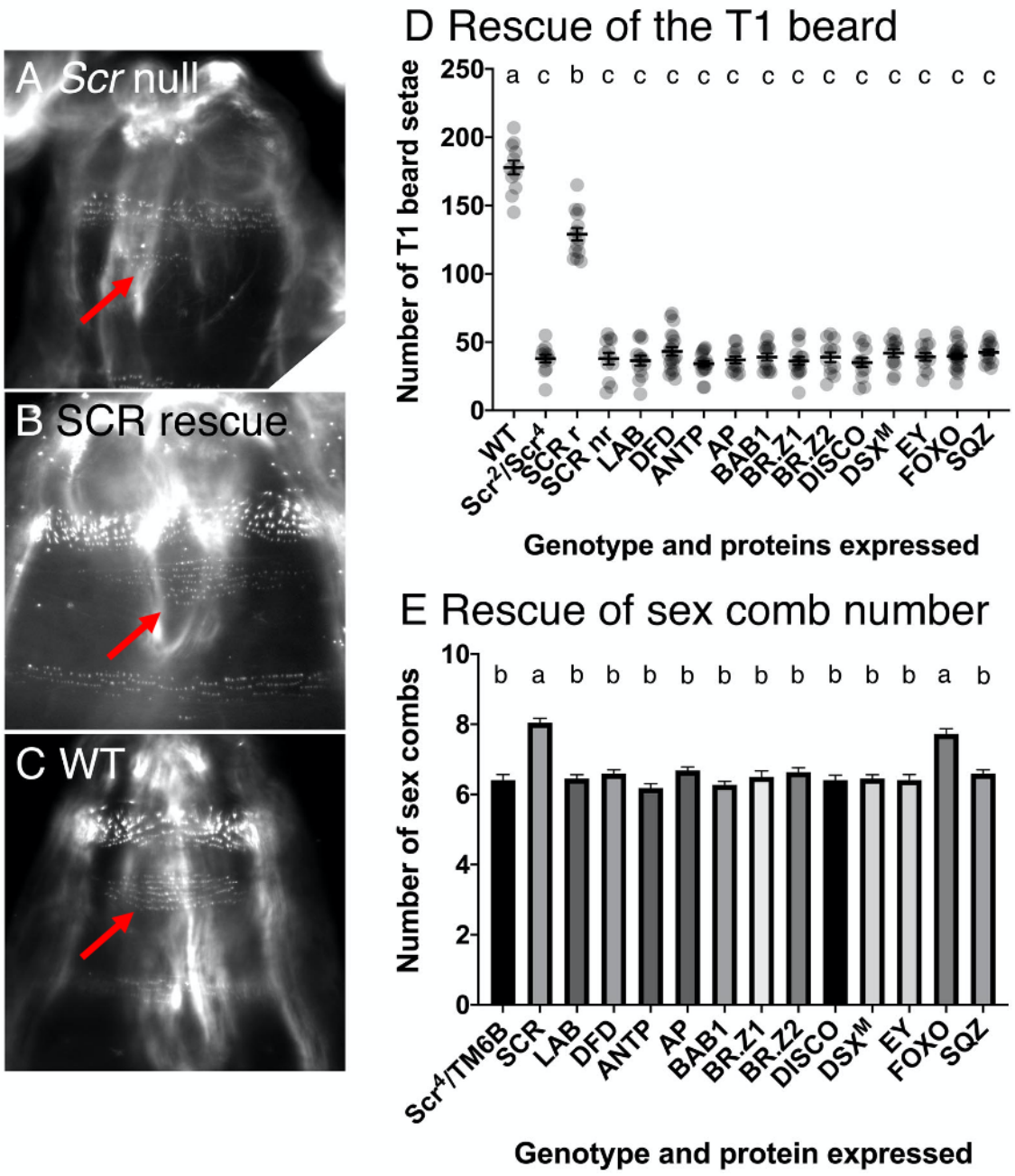


Figure 4. Characterization of the rescue of the *Scr* phenotypes.

Panels A – C are dark field micrographs of larval T1 segments. The red arrows point to T1 beards of a Scr^{2}/Scr^{4} larva (A), a *ScrGAL4/UAS-Scr;*Scr^{2}/Scr^{4} larva (B) and a wild type larva (C). Panel D is a scatter plot of the number of T1 setae in various genotypes. SCRr and SCRnr refers to the number of setae on larval cuticles that have rescued head skeletons

and the number of setae on larval cuticles that exhibit no rescue, respectively. Analysis of the data with an ordinary ANOVA ($F_{15, 212}$=148; P<0.0001) detected differences, and data that are not different (P>0.05) have the same letter using Tukey's post hoc pair-wise comparisons. Panel E is a bar graph of the number of sex combs on *ScrGAL4*; *Scr* ^{4}/ + adults expressing no protein or the indicated protein. An ordinary ANOVA detected differences ($F_{13, 294}$=17, P<0.0001) and the pair-wise comparisons using Tukey's post hoc analysis that were not different are indicated with the same letter (P>0.05). The mean and SEM are indicated in Panels D and E.

Figure 5. Ectopic expression of SCR and FOXO in all three pairs of legs of *y* male flies.

Using *rnGAL4* driver to express SCR and FOXO ectopically in all three pairs of legs of *y* flies. “1” are first legs; “2” are second legs and “3” are third legs. Panels A, C and E are images of flies’ legs with SCR ectopic expression. Panels B, D and F are images of flie’s legs with FOXO ectopic expression.

Ultrabithorax: Taking genomic DNA fragments from the *Ubx* locus and screening them for enhancer activity when fused to *GAL4* identified a fragment that reproduced the *Ubx* embryonic expression pattern. This driver is expressed throughout the haltere imaginal disc and is expressed ectopically in the notum and wing pouch of the wing imaginal disc (Jenett *et al.*, 2012). The third chromosome carrying the *UbxGAL4* insertion also carried a y^+ allele; therefore, rescue of the *Ubx* larval cuticular phenotype could not be assessed. The genotype $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}/Ubx^{9.22}\ P\{UbxGAL4,\ w^+\}$ was used to assess rescue of the adult viable $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}$ hypomorphic allele. This allelic combination gives the classic four-winged fly (Rivlin *et al.*, 2001), where the third thoracic segment (T3) and haltere are transformed into the likeness of the second thoracic segment (T2) (**Figure 6 A and B**). Expression of UBX in this mutant background resulted in partial rescue of the haltere to wing transformation in T3; the scabellum and pedicellus are wild type in appearance (of particular note are the transverse rows of campaniform sensilla specific to the haltere) (**Figure 6 C, D and E**). Although the wing in T3 is drastically reduced, the capitellum is not rescued (the capitellum has characteristic short trichomes) (**Figure 6 D and E**). In addition, expression of UBX in T3 suppresses the T2-like notum such that it has a wild type appearance. The ectopic expression of UBX in the wing imaginal disc resulted in a reduction of the wing and partial transformation to a haltere and partial suppression of the T2 notum (**Figure 6 C**). The partial transformation of the wing to a haltere includes transformation of wing campaniform sensilla on the dorsal proximal radius to haltere-like sensilla (**Figure 6 F and G**). In addition, 2/15 flies lacked the third legs; 7/15 lacked one third leg, and the remainder had six legs (A role of UBX is suppression of leg development

on the abdominal segments) (Lewis, 1985; Vachon *et al.*, 1992). This phenotype is the opposite of hypomorphic *Ubx* mutant combinations where an ectopic abdominal leg form.

Screening expression of 13 non-resident TFs identified one very clear example of phenotypic non-specificity: ANTP. Expression of DFD was either embryonic or larval lethal, and the expression of BR.Z1, BR.Z2 and FOXO in $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}$ /$Ubx^{9.22}$ *P*{*UbxGAL4*, w^{+}} flies caused failure to develop into pharate or eclosed adults during metamorphosis. The eclosed adults that expressed DISCO and BAB1 had four wings (**Figure 6 H and I**). Flies expressing AP, and SQZ did not eclose but the pharate adults had the four-winged phenotype (**Figure 6 J and L**). Flies expressing DSX^{M} did not eclose but the pharate adults still had the four-winged phenotype and an extensive deletion of the notum in T2 and T3 (**Figure 6 K**). Flies expressing ANTP and EY did not eclose and the four wings were reduced (**Figure 6 M and N**). The reduced wings on T2 and T3 of ANTP expressing flies had campaniform sensilla characteristic of a haltere indicating a transformation toward a haltere (**Figure 6 Q**). T3 expressing ANTP was not rescued to wild type and some of the notum of T2 was absent. Flies expressing LAB eclosed with the four wings transformed into tissue with micro and macrochaetes (**Figure 6 O and S**). Flies expressing SCR did not eclose and only had two wings plus a reduction of the T2 notum (**Figure 6 P and T**). The two-winged phenotype is not due to the rescue of T3 to wild type with a haltere but is a deletion of the derivatives of the haltere imaginal disc (**Figure 6 T**).

The rescue with UBX, EY and ANTP were differentially pleiotropic. UBX, EY and ANTP reduced the wings and the wings expressing UBX and ANTP had campaniform sensilla

that are haltere-like, but EY only reduced the wings. In addition, UBX rescues the T3 notum to wild type, but ANTP and EY do not.

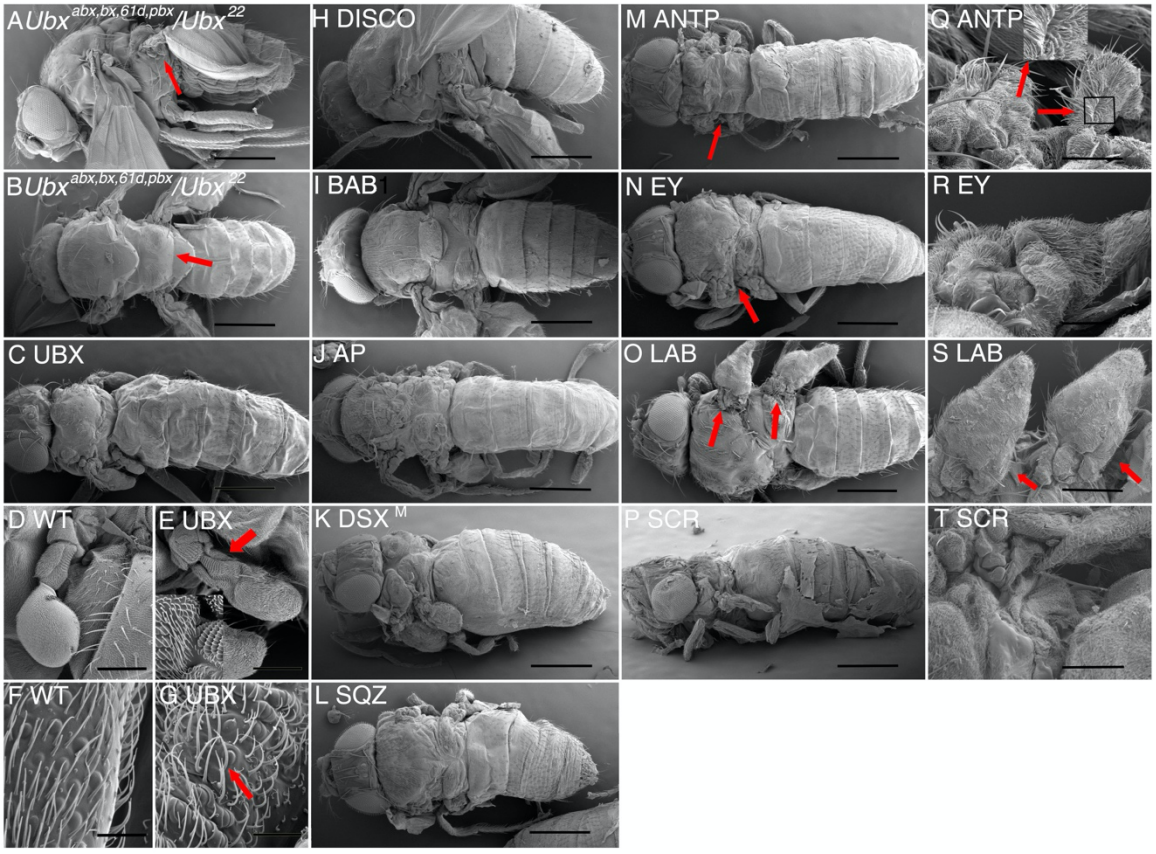


Figure 6. Screen for rescue of adult *Ubx* phenotypes.

Panels A and B are lateral and dorsal images of $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}/Ubx^{9.22}$ flies. The second pair of wings indicate the T3 to T2 transformations (red arrows). Panels C, E, G are the expression of UBX in $Ubx^{abx1,\ bx3,\ 61d,\ pbx1}/\ Ubx^{9.22}$ flies. In Panel C the reduced wings indicate the suppression of T2 notum development. In Panel E the restored scabellum and pedicellum (red arrow) indicate rescue of the haltere, and the insert is a close-up of the haltere specific transverse rows of campaniform sensilla. In Panel G the haltere-like sensilla (red arrow) indicate a wing to haltere transformation due to ectopic UBX expression. Panels D, F are a wild type haltere and wing for comparison with panels E and G. Panels H and I are eclosed adults expressing DISCO and BAB1, respectively. Panels J-L are pharate adults expressing AP, DSX^{M} and SQZ, respectively. Panels M-P are pharate

adults expressing ANTP, EY, LAB and SCR, respectively. Panels Q-T are close ups of T2 and T3 of ANTP, EY, LAB and SCR, respectively. In Panel M the wings are reduced and some of the notum of T2 was absent (red arrow). In Panel N the wings are reduced (red arrow). In Panel R the wing is reduced. In Panel O and S the four wings are transformed into tissue with micro and macrochaetes (red arrows). In Panel Q the campaniform sensilla characteristic of a haltere indicates a transformation toward a haltere (red arrows). The insert in panel Q is a close-up of the haltere-like campaniform sensilla. The bars in panel A, B, C, H, I, J, K, L, M, N, O and P indicate 500 μm; the bars in panel Q, R, S and T indicate 100 μm; the bars in Panel D and E indicate 100 μm; the bars in Panel F and G indicate 10 μm.

Doublesex: The *dsx* locus encodes two TFs with distinct activities: DSX^M suppresses the formation of female genitals and DSX^F suppresses the formation of male genitals (Cho & Wensink, 1997; Rideout *et al.*, 2010; Robinett *et al.*, 2010; Waterbury *et al.*, 1999). The external somatic secondary sexual characteristics examined were male sex combs, abdominal pigmentation, and genitalia. The male sex combs are a vertical row of about 10 to 12 darkly pigmented thick bristles with rounded tips; in females two horizontal rows of approximately 5 lightly-pigmented, spike-like bristles are the equivalent bristles (Tanaka *et al.*, 2009). In a dsx^{null} mutant, the 5 lightly-pigmented, spike-like female bristles are organized into a single row that is partially rotated towards the vertical. The important male specific phenotypes of sex combs are an increase in bristle number, a vertical orientation, a change in morphology (rounded tips instead of spike-like) and dark pigmentation (**Figure 7 A1**). The A5 and A6 segments of the male abdomen and the dsx^{null} mutant are fully pigmented (**Figure 7 A2**); whereas, only the posterior portion of tergite 5 and most of tergite 6 are pigmented in females (**Figure 7 B2**). The male genitalia has a genital ridge wrapped round the anus and characteristic claspers (**Figure 7 A3**); whereas, the female genitalia is a vaginal plate decorated with a single row of distinctive bristles, the vaginal teeth, located on each side of the vagina and under the anus (**Figure 7 B3**) (True *et al.*, 1997). In a dsx^{null} mutants, the genitals are rotated 90 degrees relative to the dorsal ventral axis and both male and female genitalia form (**Figure 7 D3**).

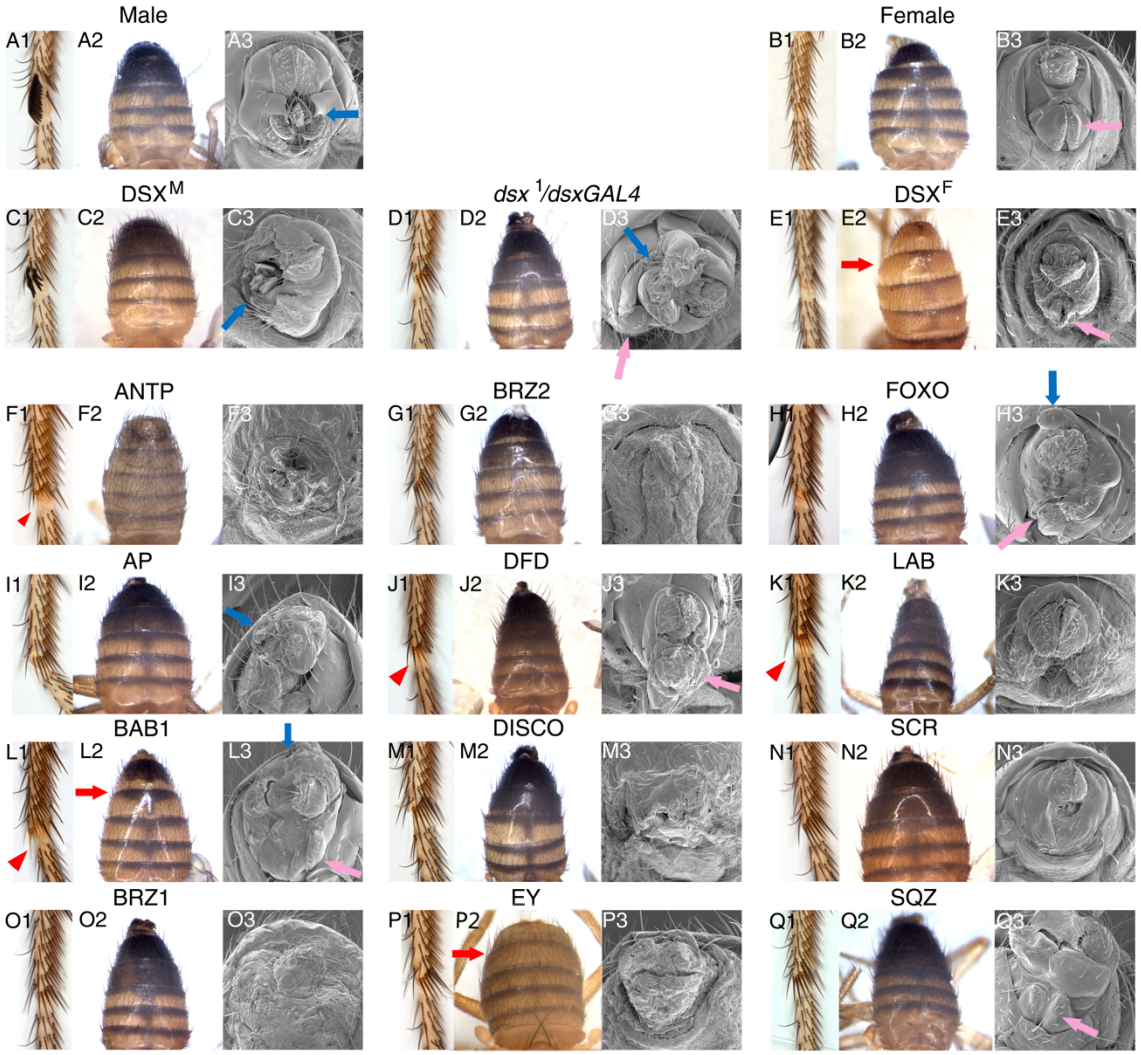


Figure 7. Screen for rescue of *dsx* phenotypes of 12 TFs.

Each panel is composed of three images: first legs (1), abdomen (2) and genitals (3). Panels A and B are wild type male and female flies, respectively. Panel D is a *dsx*^{1}/*dsxGAL4* mutant flanked by panels C and E which are *dsx*^{1}/*dsxGAL4*flies expressing either DSX^{M} or DSX^{F} protein, respectively. Panels F-Q are *dsx*^{1}/*dsxGAL4*flies expressing one of 12 TFs indicated above the panel. Red arrows indicate female pigmentation of abdomen. Blue arrows indicate male genitals and pink arrows indicate female genitals. Red arrowheads indicated depigmented sex combs.

We used the targeted insertion of GAL4 in the *dsx* locus, which is also a dsx^{null} allele, to express TFs in a $dsxGAL4/dsx^{1}$ mutant background (Robinett *et al.*, 2010). Expression of DSX^{M} from a UAS promoter rescues the vertical orientation, morphology, and pigmentation of sex combs; however, only 4.2 shortened sex combs form indicating partial rescue (**Figure 7 C1**). Tergite 5 and 6 of the abdomen are pigmented (**Figure 7 C2**), and development of the female plate is suppressed but the male genitalia is rotated (**Figure 7 C3**). Expression of DSX^{F} from a UAS promoter rescues the morphology of the female genitalia. Female vaginal plates formed with each plate having a single row of vaginal teeth (**Figure 7 E3**). Tergite 5 has female-like pigmentation (**Figure 7 E2**), the pigmentation is restricted to the very posterior edge of the segment, and most of tergite 6 is depigmented unlike in wild type females (**Figure 7 E2**).

We screened the expression of 12 TFs for masculinization or feminization of the dsx null phenotype. The observed rescues exhibited differential pleiotropy. **Masculinization**. Expression of ANTP and AP increased the number of sex comb bristles from 5.4 to 6.3 ($P<0.0001$) (**Figure 8 C**). Expression of AP suppressed vagina formation. **Feminization**. The major phenotypes associated with feminization are suppression of the vertical orientation, number, pigmentation and morphology of sex comb bristles, the pigmentation of the abdomen and the suppression of male genitalia. Expression of ANTP, BAB1, DFD and LAB depigmented the sex combs (**Figure 7 F1, L1, J1 and K1**). LAB repressed the vertical orientation of the sex combs (**Figure 7 K1**); the sex combs have a horizontal rather than vertical orientation and are shorter. Expression of ANTP, BAB1 and EY suppressed abdominal pigmentation in the anterior portion of tergite 5 and 6; although ANTP also

suppressed abdominal pigmentation overall. Expression of BAB1 partially suppresses male genitalia and expands the vaginal plate (**Figure 7 L3**). DFD, LAB and SCR suppress male genitalia and DFD transforms the vaginal plate; whereas the vagina was not observed with expression of LAB and SCR. Expression of FOXO suppresses male genitalia (**Figure 7 H3**). Expression of SQZ increases the number of rows of vaginal teeth (**Figure 7 Q3**). **Non-specific**. The genitalia are lost or unrecognizable with expression of ANTP, BR.Z1, BR.Z2, DISCO and EY.

The *dsx* dominant mutation alleles, dsx^{dom}, constitutively express DSX^{M} (Baker & Ridge, 1980; Nagoshi & Baker, 1990). Expression of DSX^{M} in females by these dominant gain-of-function alleles results in an intersex phenotype similar to the *dsx* null phenotype; fewer sex combs with a changed morphology and orientation, development of both male and female genitals. The intersex phenotype in females is hypothesized to be due to DSX^{M} inhibiting the function of DSX^{F} (Rideout *et al.*, 2010; Waterbury *et al.*, 1999; Yang *et al.*, 2008). Likewise when DSX^{F} is ectopically expressed in males the intersex phenotype is also expected (Waterbury *et al.*, 1999). When *dsxGAL4* was used to express DSX^{M} in females, 2-4 sex combs formed on the first leg that were shorter than normal but were rotated, pigmented and had rounded tips; the abdomen was pigmented in tergites 5 and 6, and the female genitals were absent and rotated male genitals form (**Figure 8 B and Figure 9 b1-b3**). Unexpectedly, when *dsxGAL4* was used to express DSX^{M} in males, only 2-3 sex combs formed on the first leg that were shorter than normal but were rotated, pigmented, and had rounded tips. The abdomen was pigmented in tergites 5 and 6, and the male genitals were affected with the genital ridge and claspers not fully formed (**Figure 8 A and Figure**

9 B1-B3). When *dsxGAL4* was used to express DSX^F in males, the bristles on the first leg were female like, male genital formation was suppressed, and the vaginal plate was present but lacked vaginal teeth (**Figure 8 A and Figure 9 A1-A3**). Expression of DSX^F in females reduced abdominal pigmentation (**Figure 9 a1-a3**). In summary, expression of DSX^M with *dsxGAL4* in females resulted in a male-like phenotype and not an intersex phenotype. Expression of DSX^F with *dsxGAL4* in males resulted in a female-like phenotype and not an intersex phenotype.

Screening the 12 TFs for affects in males and females detected an array of interactions. Expression of ANTP, AP, BAB1, BRZ1, DISCO, EY, LAB, and SQZ in males suppressed the number of sex combs that form as was observed with both DSX^M and DSX^F (**Figure 9 C1, D1, E1, F1, I1, J1, L1, N1 and Figure 8 A**). Expression of ANTP, BAB1, BRZ1, DFD, DISCO, LAB, SCR, SQZ in males depigmented the sex combs as was observed with the expression of DSX^F (**Figure 9 C1, E1, F1, H1, I1, L1, M1 and N1**). Expression of ANTP and LAB in males, the sex combs are not rotated toward the vertical as was observed with the expression of DSX^F (**Figure 9 C1, L1**). Expression of ANTP, BRZ1 and LAB shorten the sex combs as was observed with expression of DSX^M in males (**Figure 9 C1, F1 and L1**). Expression of BRZ1 feminized the morphology of the sex combs from rounded tips to spikey tips (**Figure 9 F1**). Expression of AP, BAB1, SCR and SQZ in males rotates the male genitals (**Figure 9 D3, E3, M3 and N3**). Expression of ANTP, BRZ1, DISCO and EY in males deleted the male genitals (**Figure 9 C3, F3, I3 and J3**). Expression of DFD, FOXO, LAB and SCR in males reduced the male genitals (**Figure 9 H3, K3, L3 and M3**). Expression of ANTP and EY in males depigmented the abdomen

overall and tergites A5 and A6 had a female pattern of pigmentation (**Figure 9 C2**). Expression of ANTP, BAB1 and EY in females depigmented the anterior of tergite 5 and most of tergite 6 (**Figure 9 c2, e2 and j2**). Expression of ANTP, AP, BRZ1, DISCO and EY in females resulted in the female genitals not forming (**Figure 9 c3, d3, f3, i3 and j3**). Expression of DFD, LAB and SCR in females did not suppress vagina formation but the morphology is not wild type (**Figure 9 h3, l3 and m3**). Expression of SQZ in females increased the number of rows of vaginal teeth (**Figure 9 n3**).

The rescue of *dsx* and effects on male and female development by expression of the 12 TFs exhibits extensive differential pleiotropy because not all somatic sexual phenotypes are affected to the same extent by expression of a non-resident TF. As an example, expression of SCR in males depigments sex combs but does not reduce the number, rotation or change the morphology and suppresses male genital formation. **Table 2** is a summary of phenotypes of TF expression in males and females.

Table 2. Summary of phenotypes of TF (*UAS-X*) expression in males and females driven by *dsxGAL4*.

The names of TFs are listed in the column of "Protein". The term "depig" stands for depigmentation.

	Males						Females	
Protein	**Male sex combs**				**Male genitals**	**Back pigmentation**	**Vagina**	**Back pigmentation**
	Number	pigmentation	rotation	Morphology				
WT	WT	WT	WT	WT	WT	WT	WT	WT
DSX^F	Fewer/ None	depig	No	female	reduced, vagina transformed	female	WT	female
DSX^M	Fewer	WT	WT	shorter	affected	male	male	male
ANTP	Fewer	depig	Not vertical	shorter	deleted	depig, female	deleted	depig, female
AP	Fewer	WT	WT	WT	rotated	male	deleted	female
BAB1	Fewer	depig	WT	WT	rotated	female	WT	female
BRZ1	Fewer	depig	WT	pointed	deleted	male	deleted	male
BRZ2	WT	WT	WT	WT	deleted	male	deleted	female
DFD	WT	depig	WT	WT	reduced	male	transformed	female
DISCO	Fewer	depig	WT	WT	deleted	male	deleted	female
EY	Fewer	WT	WT	WT	deleted	depig, female	deleted	female
FOXO	WT	WT	WT	WT	reduced	male	transformed	female
LAB	Fewer	depig	Not vertical	shorter	reduced	male	transformed	female
SCR	WT	depig	WT	WT	reduced rotated	male	transformed	female
SQZ	Fewer	depig	WT	WT	rotated	male	more teeth	female

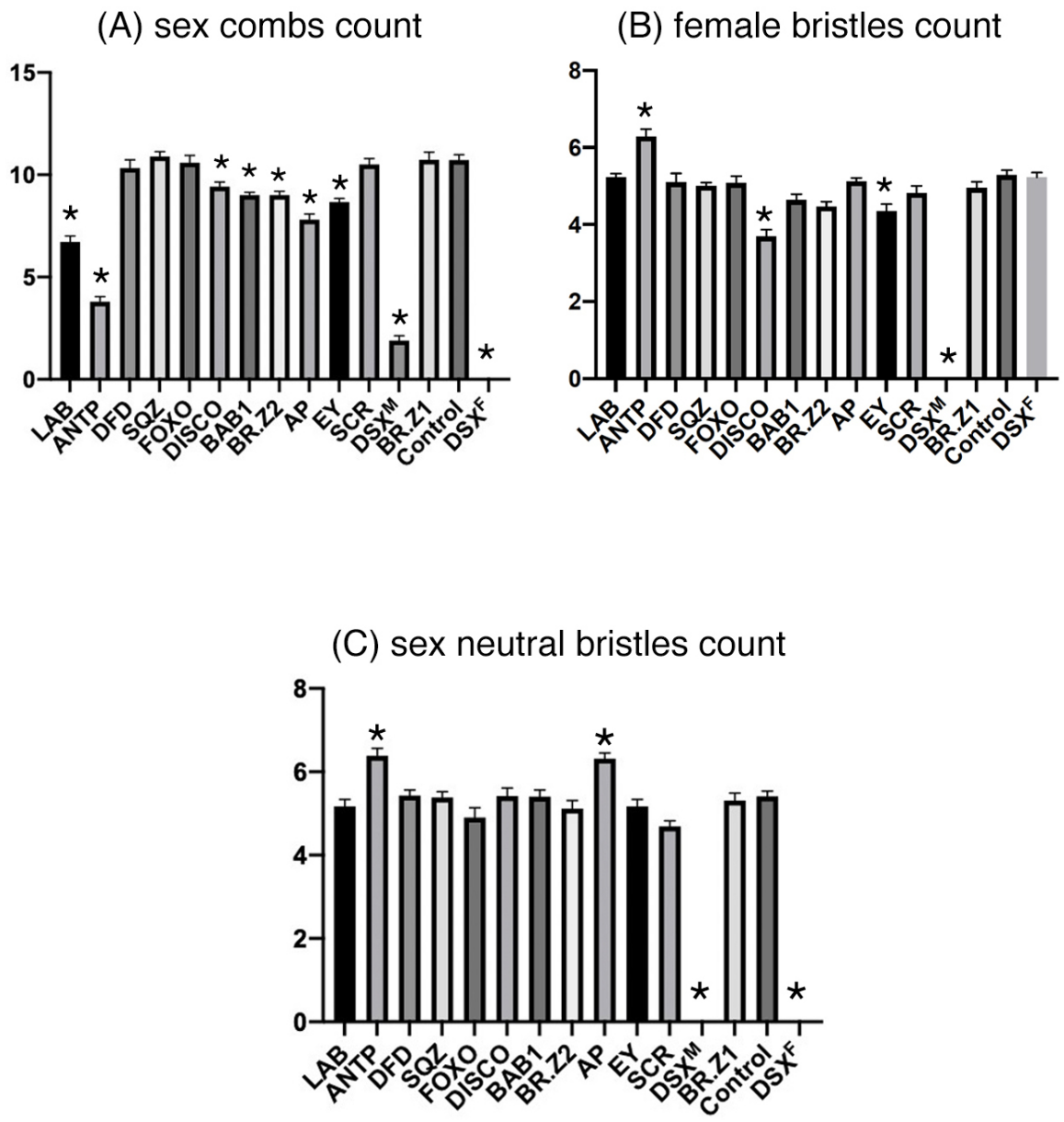
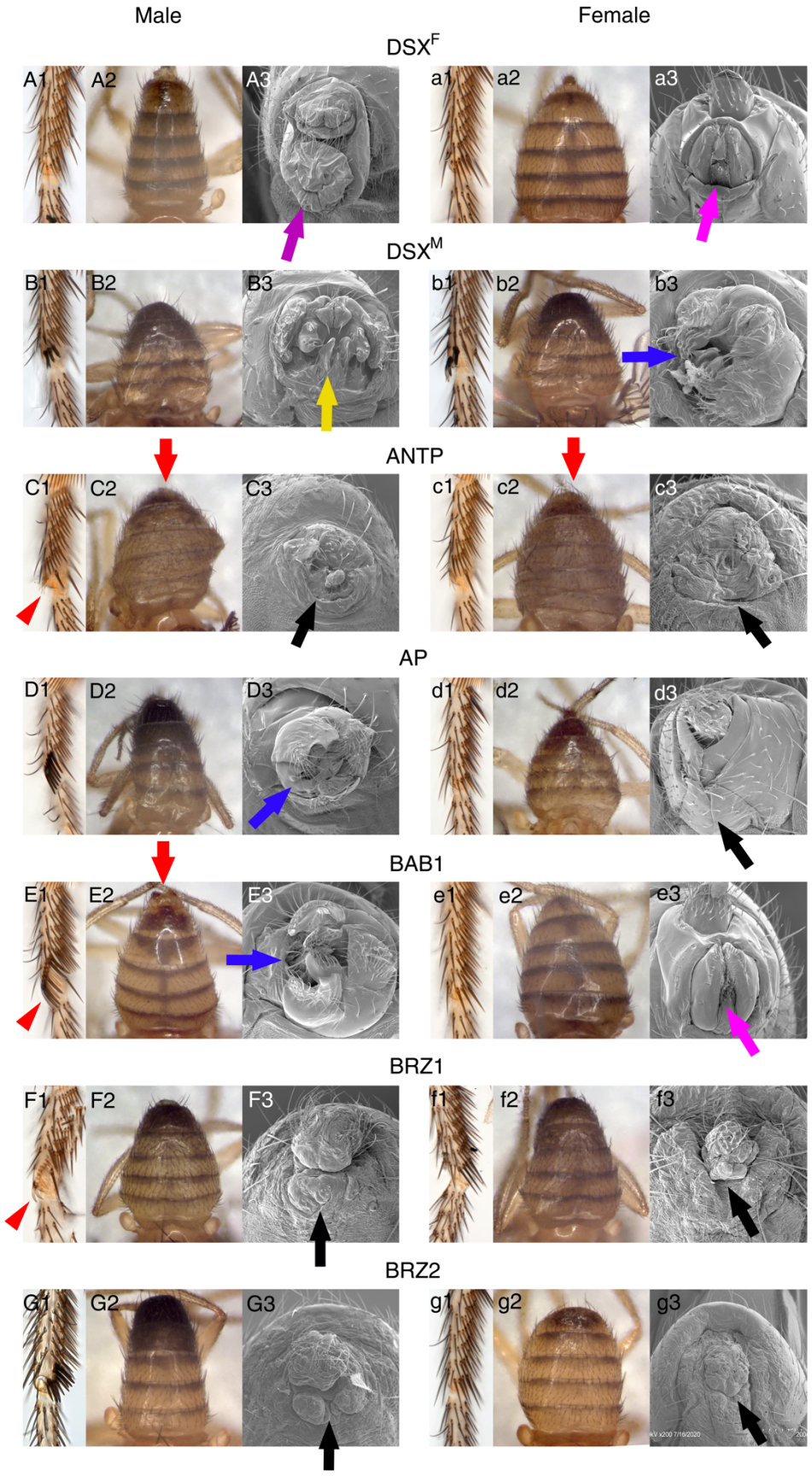


Figure 8. Summary of the bristle counts on the first legs.

Panel A is a bar graph of the number of sex combs on *dsxGAL4*/+ adults expressing no protein or the indicated protein. The control column in the graph is the wild type male. Analysis of the data with an ordinary ANOVA ($F_{14,\ 174}$=143.5; P<0.0001) detected differences, and data that are different from the control (P<0.05) are indicated with an asterisk (Dunnett's multiple comparisons). Panel B is a bar graph of the number of female

transverse bristles of first leg on *dsxGAL4*/+ adults expressing no protein or the indicated protein. The control column in the graph is a wild type female. Analysis of the data with an ordinary ANOVA ($F_{14, 199} = 59.32$; $P<0.0001$) detected differences, and data that are different from the control ($P<0.05$) are indicated with an asterisk (Dunnett's multiple comparisons). Panel C is a bar graph of the number of sex neutral bristles on *dsxGAL4/dsx*^{1} adults expressing no protein or the indicated protein. The control column in the graph is a sex neutral fly, *dsxGAL4/dsx*^{1}. Analysis of the data with an ordinary ANOVA ($F_{14, 191}=127.3$; $P<0.0001$) detected differences, and data that are different from control ($P<0.05$) are indicated with an asterisk (Dunnett's multiple comparisons). DSX^{M} masculinized the bristles and DSX^{F} feminized the bristles, the bristles are not sex neutral. The data for DSX^{M} and DSX^{F} is zero.


Male
Female
DSX^F
A1
A2
A3
a1
a2
a3
DSX^M
B1
B2
B3
b1
b2
b3
ANTP
C1
C2
C3
c1
c2
c3
AP
D1
D2
D3
d1
d2
d3
BAB1
E1
E2
E3
e1
e2
e3
BRZ1
F1
F2
F3
f1
f2
f3
BRZ2
G1
G2
G3
g1
g2
g3

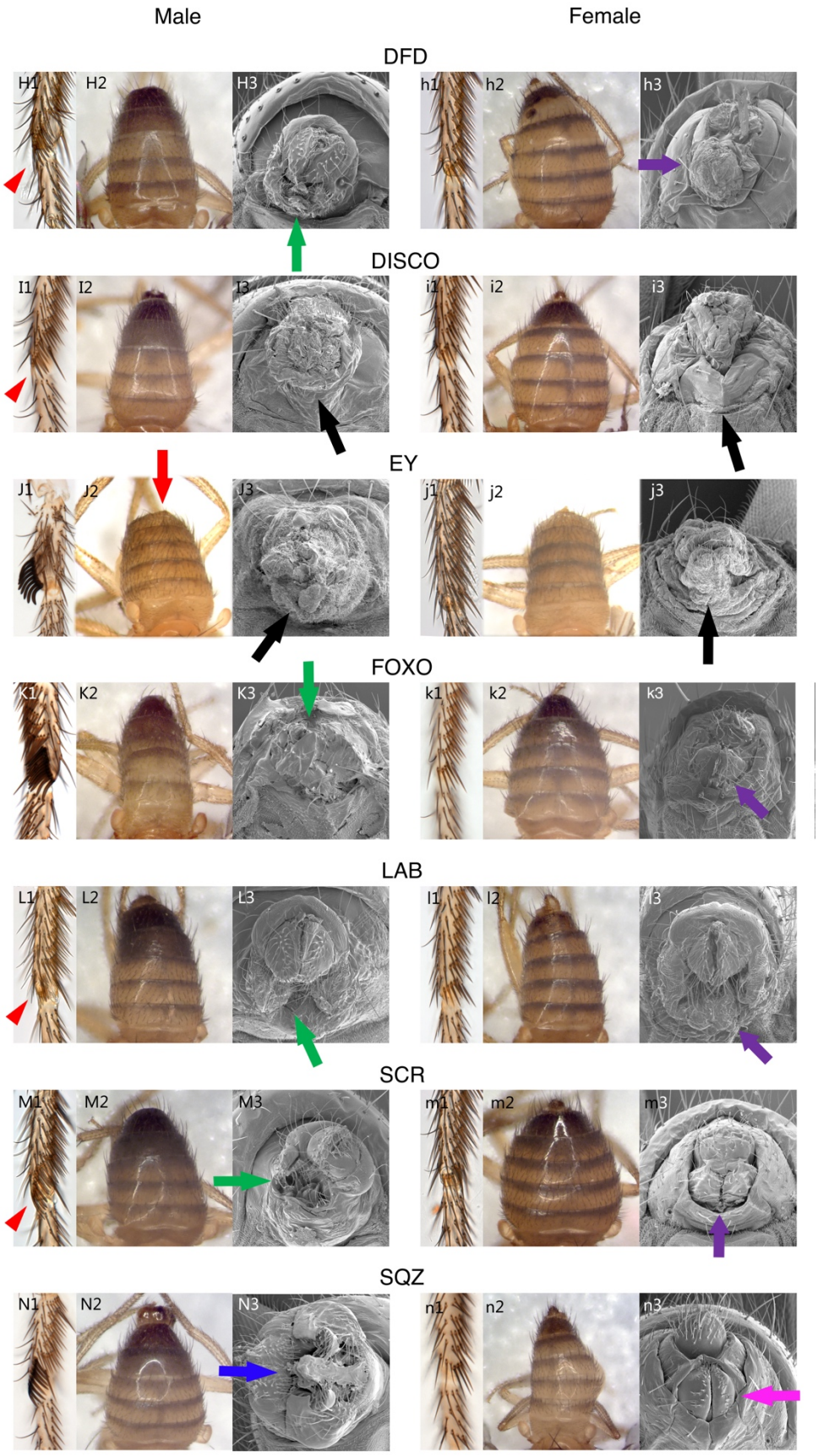

Male
Female
DFD
H1
H2
H3
h1
h2
h3
DISCO
I1
I2
I3
i1
i2
i3
EY
J1
J2
J3
j1
j2
j3
FOXO
K1
K2
K3
k1
k2
k3
LAB
L1
L2
L3
l1
l2
l3
SCR
M1
M2
M3
m1
m2
m3
SQZ
N1
N2
N3
n1
n2
n3

Figure 9. Screen for the suppression of DSX^M and DSX^F in males and females of 12 TFs.

The panels with a label starting with a capital letter are male fly images. The panels with a label starting with a lowercase letter are female fly images. Each panel is composed of three images: first legs (1), abdomen (2) and genitals (3). Panels A and a are *UAS-dsx^F, dsxGAL4*/*TM6B* male and female flies, respectively. Panels B and b are *UAS-dsx^M, dsxGAL4*/*TM6B* male and female flies, respectively. Panels C-N are *dsxGAL4*/*TM6B* flies expressing one of 12 TFs indicated above the panel. Red arrows indicate the depigmented abdomen. Blue arrows indicate rotated male genitals and pink arrows indicate female genitals. Yellow arrow indicates underdeveloped male genitals. Black arrows indicate deleted genitals, green arrows indicate reduced and rotated male genitals and purple arrows indicate transformed female genitals. Red arrowheads indicated depigmented sex combs.

***fruitless*:** The *fru* locus is structurally complex expressing multiple protein isoforms. Of these isoforms, those expressed from transcripts initiated at the P1 promoter are male specific and required for male fertility and courtship. Phenotypic non-specificity was assessed with two insertions of GAL4 in the *fru* locus (**Figure 10 A**). The *fruGAL4A* allele is a targeted insertion that fuses GAL4 to the N-terminus of the male specific isoforms, and results in a decrease in male fertility and courtship (**Figure 10 B**, $P < 0.0001$). The *fruGAL4B* allele is an insertion of a GAL4 enhancer detector that strongly reduces male fertility and courtship (**Figure 10 C**, $P < 0.0001$). In the screen with *fruGAL4A*, expression of DISCO rescues male fertility, and DISCO is the only protein that has fertility using *fruGAL4B*. The *fruGAL4A* allele reduced the fertility to 50%, and expression of DISCO increased the fertility to 90% ($P < 0.0001$) (**Figure 10 B**). The fertility rescued by expression of DISCO was not different from wild type or the fru^{4-40} heterozygote (**Figure 10 C**). The *fruGAL4B* allele reduced fertility to zero and expression of DISCO was the only protein that increased the fertility, although not significantly, to 14% ($P = 0.8$) (**Figure 10 C**). Repeating this expression of DISCO showed a significant increase in the fertility when *fruGAL4B*/fru^{4-40} is set as the control ($P < 0.0001$) **(Figure 10 F**). The increased fertility observed with expression of DISCO was less than the fertility of wild type and the fru^{4-40} heterozygote ($P < 0.0001$) (**Figure 10 C**).

To characterize the rescue of the *fru* phenotype by DISCO, male-female (M/F) and male-male (M/M) courtship indices (CI) were determined (**Figure 10 D, E**). The M/F CI was lower in *fruGAL4A*/fru^{4-40} and *y*; *fruGAL4B*/fru^{4-40} males than with wild type and fru^{4-40}/+ males ($P < 0.0001$). The M/F CI of $P\{UAS\text{-}disco, w^{+}\}$; *fruGAL4A*/$fru^{4-40}$ males was not

different from wild type (P = 0.7) and *fru*$^{4\text{-}40}$/+ heterozygous males (P > 0.9999), but significantly higher than *fruGAL4A*/*fru*$^{4\text{-}40}$ males (P = 0.0007) indicating that DISCO rescues the *fruGAL4A/fru*$^{4\text{-}40}$ courtship phenotype. *Fruitless* mutants are reported to have a higher M/M CI (Demir & Dickson, 2005). Although we observed that expression of DISCO reduces the M/M courtship of *y; fruGAL4B*/*fru*$^{4\text{-}40}$ (P = 0.0159) the reduction is not observed with *fruGal4A*/*fru*$^{4\text{-}40}$ (P = 0.4822). In addition, the M/M CI for *y; fru*$^{4\text{-}40}$/+ was the same as *y; fruGAL4B*/*fru*$^{4\text{-}40}$ (P > 0.9999) which is not expected. Therefore, it is difficult to determine the effect of expression of DISCO on the M/M CI.

There are three FRU^{M} protein isoforms, isoform A (FRU^{MA}), isoform B (FRU^{MB}) and isoform C (FRU^{MC}). To test whether FRU^{MC} rescued fertility and whether DISCO could be a TF required for fertility and activated by FRU^{M}, the fertility of eleven genotypes were assessed (**Figure 10 F**). The expression of *UAS-fru*^{M} (expressing FRU^{MC}) failed to rescue the fertility of hemizygous *fruGAL4*/*fru*$^{4\text{-}40}$ males (**Figure 10 F**). Knocking down DISCO expression using *disco* RNAi did not decrease fertility indicating that DISCO is not downstream of FRU in the male behaviour pathway (**Figure 10 F**).

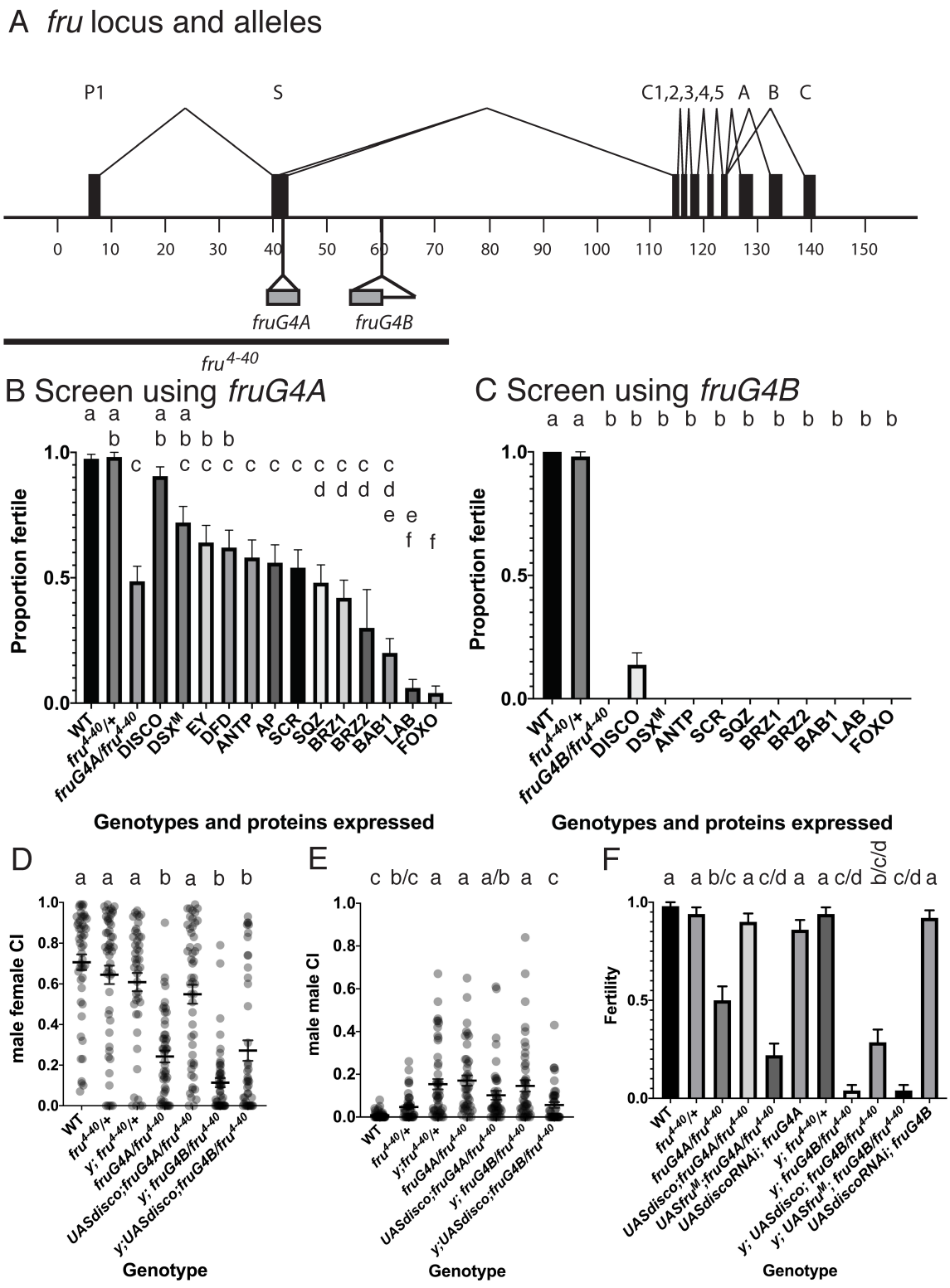

A fru locus and alleles
P1
S
C1,2,3,4,5 A B C
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
fruG4A
fruG4B
fru4-40
B Screen using fruG4A
Proportion fertile
1.0
0.5
0.0
a a b c a b a a b b b c c c c c c c c d d d c d e e f f
WT
fru4-40/+
fruG4A/fru4-40
DISCO
DSXM
EY
DFD
ANTP
AP
SCR
SQZ
BRZ1
BRZ2
BAB1
LAB
FOXO
Genotypes and proteins expressed
C Screen using fruG4B
a a b b b b b b b b b b b
Proportion fertile
1.0
0.5
0.0
WT
fru4-40/+
fruG4B/fru4-40
DISCO
DSXM
ANTP
SCR
SQZ
BRZ1
BRZ2
BAB1
LAB
FOXO
Genotypes and proteins expressed
D
a a a b a b b
male female CI
1.0
0.8
0.6
0.4
0.2
0.0
WT
fru4-40/+
y; fru4-40/+
fruG4A/fru4-40
UASdisco;fruG4A/fru4-40
y; fruG4B/fru4-40
y;UASdisco;fruG4B/fru4-40
Genotype
E
c b/c a a a/b a c
male male CI
1.0
0.8
0.6
0.4
0.2
0.0
WT
fru4-40/+
y;fru4-40/+
fruG4A/fru4-40
UASdisco;fruG4A/fru4-40
y; fruG4B/fru4-40
y;UASdisco;fruG4B/fru4-40
Genotype
F
a a b/c a c/d a a c/d b/c/d c/d a
Fertility
1.0
0.5
0.0
WT
fru4-40/+
fruG4A/fru4-40
UASdisco;fruG4A/fru4-40
UASfruM;fruG4A/fru4-40
UASdiscoRNAi; fruG4A
y; fru4-40/+
y; fruG4B/fru4-40
y; UASdisco; fruG4B/fru4-40
y; UASfruM; fruG4B/fru4-40
UASdiscoRNAi; fruG4B
Genotype

Figure 10. Summary of the screen for rescue of the *fru* fertility and courtship phenotypes.

Panel A: The *fru* locus with the insertion site of *fruGAL4A* (*fruG4A*) and *fruGAL4B* (*fruG4B*) and deletion of $fru^{4\text{-}40}$ indicated. P1 is one of the alternative promoters of *fru* gene, S is the sex-specifically spliced exon found only in P1 transcript. C1-C5 are common exons and A-C are alternative 3' exons. Panel B is a bar graph of the fertility of $fruGAL4A/fru^{4\text{-}40}$ adult males expressing no protein or the indicated protein. An ANOVA on ranks detected differences (H (15) = 272.8, $P < 0.0001$) and the pair-wise comparisons using Dunn's multiple comparisons analysis that were not different are indicated with the same letter ($P > 0.05$). Panel C is a bar graph of fertility of $fruGAL4B/fru^{4\text{-}40}$ adult males expressing no protein or the indicated protein. An ANOVA on ranks detected differences (H (12) = 620.2, $P < 0.0001$) and the pair-wise comparisons using Dunn's multiple comparisons analysis that were not different are indicated with the same letter ($P > 0.05$). Panel D and E are scatter plots with means and SEM indicated of male female (M/F) CI and male (M/M) CI for various genotypes (indicated on the x-axis), respectively. An ANOVA on ranks (For M/M CI: H (6) = 85.53, $P < 0.0001$; for M/F CI: H (6) = 123.1, $P < 0.0001$) detected differences, and data that are not different ($P > 0.05$) have the same letter after Dunn's multiple comparisons. Panel F is a bar graph with SEMs of male fertility in various genotypes indicated on the x-axis. An ANOVA on ranks detected differences (H (10) = 315.6, $P < 0.0001$) and the pair-wise comparisons using Dunn's multiple comparisons analysis that were no different are indicated with the same letter ($P > 0.05$).

Discussion

The frequency of phenotypic non-specificity

The hypothesis that proposes a limited specificity of TF function predicts that phenotypic non-specificity should be observed frequently (Percival-Smith, 2018). In this project, I have assessed functional complementation of loss-of-function alleles in six *TF* loci with at least 12 non-resident TFs and found many examples of phenotypic non-specificity. Five of six TF loci were rescued by non-resident TFs: *lab* was rescued by expression of DSX^{M}; *Scr* was rescued by expression of FOXO; *Ubx* was rescued by expression of ANTP and EY; *dsx* was rescued to some extent by the expression of a majority of non-resident TFs; and *fru* was rescued by expression of DISCO. In these screens of *TF* loci, 74 non-resident situations were assessed, and 18 examples of rescue were observed. Thus, ¼ of TFs on average rescued TF phenotypes; therefore, to have a 95% confidence of detecting phenotypic non-specificity only about 12 TFs need to be screened. The frequency is affected by the number observed with rescue of *dsx*, which could be considered as encoding two independent TFs. If you take *dsx* out and add in data with *pb* then the frequency is about 1/12 of TFs on average rescued TF phenotypes; therefore, to have a 95% confidence of detecting phenotypic non-specificity about 36-40 TFs need to be screened (Percival-Smith, 2017). A more accurate number will require a larger analysis, but irrespective of the large range that can be proposed, the lower frequency of 1/12 is above any initial expectation. Phenotypic non-specificity is a frequent observation.

Differential pleiotropy of rescue

All the *TF* loci examined are highly pleiotropic (i.e., the TF encoded by the locus is required in multiple tissues and at multiple times during development). For example, SCR is required during embryogenesis for larval head skeleton development, number of T1 setae and salivary gland formation, as well as during metamorphosis for development of the labial palps and decoration of the first legs. The rescue observed exhibited differential pleiotropy where non-resident TFs did not rescue all the phenotypes of the TF locus. For example, FOXO ectopic expression is able to increase the number of sex combs but cannot rescue the embryonic head defect of *Scr* phenotype.

The rescue is not dependent on DNA sequence recognition

The DNA recognition sequences of the resident and non-resident TFs from the rescues are analyzed and listed in **Figure 11**. The DNA sequences recognized by the resident TFs are very different from the sequences recognized by the non-resident TFs. LAB, SCR, ANTP, PB and UBX are HOX proteins and belong to the HD-family (Carroll *et al.*, 2004). DSX and FRU are zinc-finger proteins (Dalton *et al.*, 2013; Erdman & Burtis, 1993). FOXO is a forkhead box protein (Tia *et al.*, 2018). The rescue is not dependent on DNA sequence recognition. For example, FRU protein has three male isoforms, FRU^{MA}, FRU^{MB} and FRU^{MC} and each of the isoform has a unique DNA recognition sequence (Dalton *et al.*, 2013) (**Figure 11**). The male fertility of *Drosophila* requires all 3 FRU^{M} isoforms but can be rescued by one TF DISCO which recognizes a distinct DNA binding site from the 3 FRU^{M} isoforms.

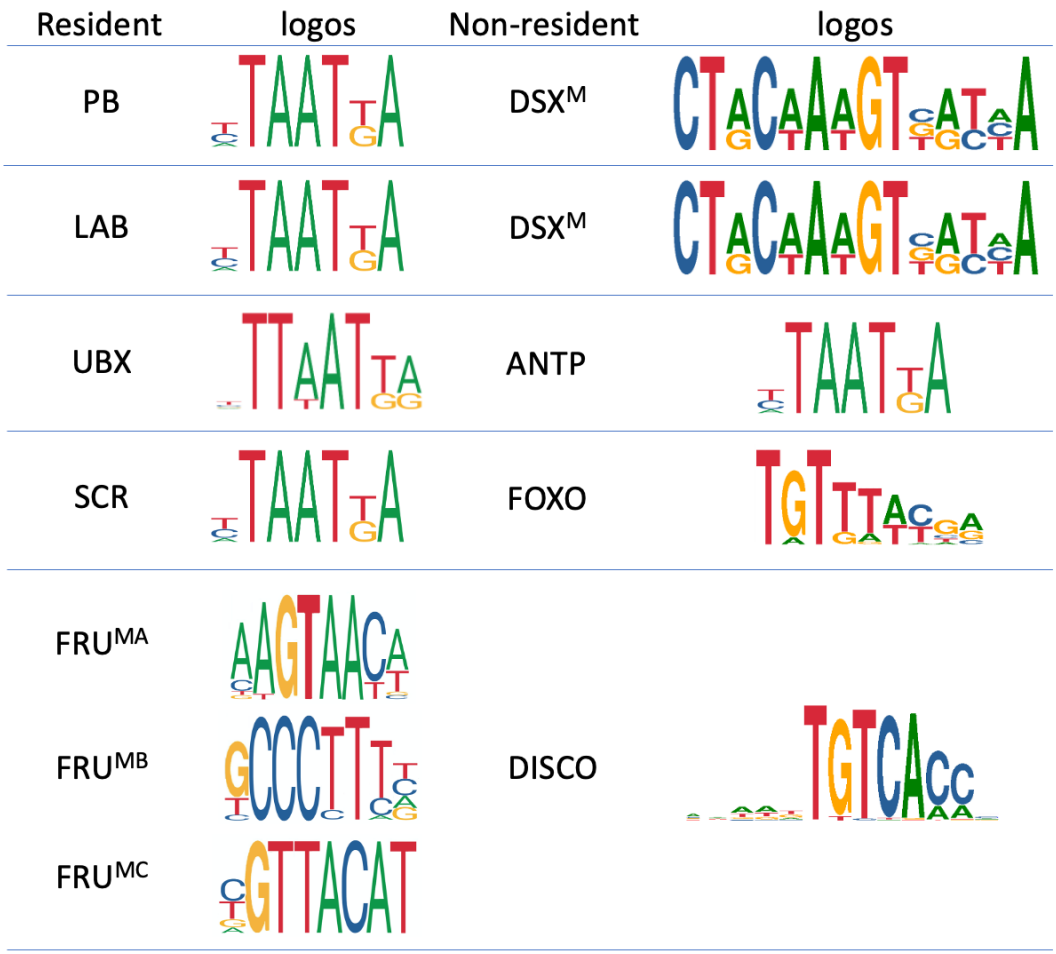


Figure 11. DNA recognition sequences of resident and non-resident TFs.

"logos" stand for the logo of the DNA recognition sequences. The recognition sequence data for PB, LAB, UBX, SCR and ANTP was obtained from JASPAR 2020 database (Fornes *et al.*, 2020). The recognition sequence data for FOXO, DSX^M and DISCO was obtained from Fly Factor Survey (Zhu *et al.*, 2011). The data for FRU is from Dalton *et al.* (2013).

Possible explanations of the results

There are two possible explanations of the rescue events: 1) the non-resident TF substitutes for the resident TF, or 2) the non-resident TF functions downstream of the resident TF. If the non-resident TF is downstream member of the resident TF function, then mutations of non-resident TFs will have similar phenotypes as the resident TF. In the case of DSX^M rescuing *lab*, *dsx* mutants do not affect larval mouthparts (data not shown) and therefore DSX^M does not function downstream of LAB. In the case of DISCO rescuing *fru*, knocking down DISCO expression using *disco* RNAi did not decrease fertility indicating that DISCO is not downstream of FRU in the male behaviour pathway. But as a counterpoint, BAB1 is required for suppression of abdominal pigmentation and does function downstream of DSX^F (Williams *et al.*, 2008). However, EY has no role in abdominal pigmentation. I have not assessed whether FOXO is required for sex comb formation and therefore potentially downstream of SCR.

Unexpected transformations with expression of DSX^M and DSX^F

Although the phenotypes of the *TF* loci are well studied, I found unexpected results with *dsx*. The DSX proteins are transcription factors that determine all aspects of male and female somatic sex determination (Robinett *et al.*, 2010). DSX^F activates female-specific genes promoting development of female somatic sexual characteristics and prevents the development of male somatic sexual characteristics (Ryner & Baker, 1991). Conversely, DSX^M promotes development by activating male-specific genes and preventing female development (Ryner & Baker, 1991; Salz *et al.*, 1989). Most studies of the *in vivo* function of DSX analyze loss-of-function mutants (dsx^-) and constitutive mutants (dsx^{dom}). The

dsx^{dom} alleles constitutively express *dsx*^{M} which results in intersex females due to inhibition of DSX^{F} function (Nagoshi & Baker, 1990). The phenotype of the *dsx*^{dom} alleles suggests that the DSX^{M} and DSX^{F} TFs inhibit one another (Nagoshi & Baker, 1990). In my results DSX^{M} expression transforms female flies towards males and DSX^{M} expression in males using *dsxGAL4* results in an unexpected reduction of sex combs and malformation of the male genitals. However, male *dsx*^{dom} heterozygotes (*dsx*^{dom}/+) have a phenotype associated with wild type males (Nagoshi & Baker, 1990). Also, expression of DSX^{F} with *dsxGAL4* in females results in an abdominal pigmentation pattern that is more restricted than observed in wild type females. This may suggest that *dsxGAL4* has an expression pattern that is slightly different from the wild type *dsx* locus resulting in the unexpected phenotypes as a result of ectopic expression of GAL4.

The differential pleiotropy of phenotypic non-specificity

The hypothesis of limited specificity of TF function emphasizes that although the properties of DNA sequence recognition and cooperative interaction are specific, the specificity has a relatively limited range and is not high enough to target only a small set of genes affecting certain phenotypes. There are three major expectations of the hypothesis of limited specificity of TF function. First, as shown in this chapter, that the observation of phenotypic non-specificity is frequent. Second, when the resident TF is substituted for another TF, the *cis*-regulatory sequences normally required for expression of a gene (leading to a given phenotype) will no longer be required and will be substituted by other *cis*-elements present in the promoter. This phenomenon is referred to as *cis*-element bypass and has yet to be described. Third, that the functional organization of TF functional

domains will lack modularity outside the DNA-binding domains and the important elements are short sequences that make a small and tissue specific contribution to overall TF activity. The genetic consequence of this organization of TF functional domains results in differential pleiotropy of mutations in the *TF* locus.

In this chapter, and with rescue of the *pb* phenotype by DSX^M (Percival-Smith, 2017), extensive differential pleiotropy is observed in the rescue by non-resident TFs. The differential pleiotropy observed with alleles within the *TF* locus and with rescue by a non-resident TF allows a speculative explanation of TF function. Most eukaryotic TFs have intrinsically disordered protein regions. These intrinsically disordered protein regions may mediate condensation of TFs into protein liquid droplets to form transcription hubs (Malik & Roeder, 2010). Tissues express distinct sets of TFs and what set of TFs are expressed may determine how the TFs partition between protein droplets, such that in a mutational analysis of function, alleles differentially affect partitioning of TFs in different tissues. In phenotypic non-specificity, the non-resident TF is able to enter a particular protein droplet in one tissue and rescue the phenotype in that tissue but is unable to do so in another tissue. This is distinct from the proposal that differential pleiotropy is the consequence of the ensemble nature of TF allostery (Sivanantharajah & Percival-Smith, 2015).

Differential pleiotropy provides a mechanism that facilitates evolution by alleviating the severity of mutations of crucial genes (Hittinger *et al.*, 2005; Merabet *et al.*, 2011). For example, many *TF* genes, such as *Hox* genes, serve different functions during the development of the organism, therefore, the mutation of the coding region of these genes

should be under strong purifying selection. However, the diversity of the morphological variations in nature raised an interesting question; how do these genes evolve? Differential pleiotropy along with gene functional redundancy and modular *cis*-regulatory elements are potential ways to protect a mutant allele from intense purifying selection (Carroll, 2005; Hittinger *et al.*, 2005; Mann & Carroll, 2002). The mutation of the coding region of the TF may only disrupt a small set of the TF:TF interactions impacting only a small subset of function of the pleiotropic TF. Therefore, the mutation would "survive" from purifying selection and the accumulation of these mutations would facilitates the evolution of the TF genes.

Phenotypic non-specificity and evolution and development

The "genetic tool kit" hypothesis is based on the observations of conservation of structure (amino acid sequence), expression, requirement and function of genes required for development (Carroll, 2005, 2008). The experiments designed to test conservation of TF function are based on an implicit presumption that TF function is specific for the regulation of the specific sets of genes required for the phenotype such that expression of an ortholog from another species would only regulate the same set of genes if the function of the two molecules is conserved (Halder *et al.*, 1995; Hunter & Kenyon, 1995; Lutz *et al.*, 1996; Malicki *et al.*, 1990; Percival-Smith & Laing Bondy, 1999; Zhao *et al.*, 1993). Functional non-specificity of TF function undermines this interpretation because there is the possibility that testing > 40 unrelated TFs would uncover examples of rescue showing that a protein of distinct structure and that recognizes a distinct DNA binding site can function in the process, and therefore, although rescue is observed with an ortholog from another

species, the rescue cannot be used to discriminate between conservation of function and phenotypic non-specificity. One of the clearest examples of conservation of function in the literature is the rescue of the *lab* phenotype by the expression of the chicken HOXb1 protein from *lab* regulatory sequences (Lutz *et al.*, 1996). However, I have found that an unrelated TF DSX^{M} is also able to rescue the labial cuticle phenotypes; indeed in addition, all *Drosophila* HOX proteins, with the exception of ABD-B, rescue a LAB dependent neurogenic phenotype (Hirth *et al.*, 2001). The observation of phenotypic non-specificity for rescue of labial phenotypes changes the interpretation of observations of rescue/induction of phenotypes by orthologous TFs to supporting rather than proving conservation of function.

Experiment limitations

I used the GAL4-UAS system for functional complementation assays to assess phenotypic non-specificity of TF function. I identified multiple phenotypic non-specificity events. However, there are several limitations that need to be addressed. First, the expression of LAB in for rescue of the *lab* phenotype, expression of DFD for the rescue of the Dfd phenotype and expression of SCR for the rescue of the *Scr* phenotype did not rescue all the way to a wild type phenotype. Even though the expression of LAB from *UAS-lab* rescued embryonic head involution and head skeleton defects of *lab* null mutants, the length of the rescued head structure was smaller than wild type. In experiments assessing DFD expression in *Dfd* null embryos, the frequency of the individuals which demonstrated a rescued phenotype is lower than expected. Expression of SCR did not rescue the number of T1 beard to wild type levels. The partial rescues are likely due to under expression of

the TF using the GAL4. Second, the expression pattern of *dsxGAL4* is unlikely to be the same as the wild type *dsx* locus. Expression of DSX^{M} in male flies unexpectedly reduced the number of sex combs and resulted in abnormal genitals. This result may be due to the expression pattern of *dsxGAL4* not recapitulating that of the wild type *dsx* locus. The insertion of *GAL4* sequences into the dsx locus may affect the expression pattern. Last, not all TF isoforms expressed from a locus were assessed for rescue. UBX protein has six isoforms and FRU^{M} protein has three isoforms. In the *Ubx* and *fru* rescue experiments, only one isoform of UBX and FRU^{M} were used as the control (UBX1 in *Ubx* experiment and FRU^{MC} in *fru* experiment). UBX1 was able to rescue the haltere phenotype induced by *Ubx* mutation but FRU^{MC} was not able to rescue the fertility of male *fru* flies. In order to observe complete rescue, multiple isoforms may need to be expressed.

REFERENCE

Baker, B. S., & Ridge, K. A. (1980). Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. *Genetics, 94*(2), 383-423.

Banreti, A., Hudry, B., Sass, M., Saurin, A. J., & Graba, Y. (2014). Hox proteins mediate developmental and environmental control of autophagy. *Dev Cell, 28*(1), 56-69.

Carroll, S. B. (2005). Evolution at Two Levels: On Genes and Form. *PLOS Biology, 3*(7), e245.

Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell, 134*(1), 25-36.

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2004). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Wiley.

Cho, S., & Wensink, P. C. (1997). DNA Binding by the Male and Female Doublesex Proteins of Drosophila melanogaster. *Journal of Biological Chemistry, 272*(6), 3185-3189.

Chowdhury, T., Calhoun, R. M., Bruch, K., & Moehring, A. J. (2020). The fruitless gene affects female receptivity and species isolation. *Proceedings of the Royal Society B: Biological Sciences, 287*(1923), 20192765.

Dalton, J. E., Fear, J. M., Knott, S., Baker, B. S., McIntyre, L. M., & Arbeitman, M. N. (2013). Male-specific Fruitless isoforms have different regulatory roles conferred by distinct zinc finger DNA binding domains. *BMC Genomics, 14*(1), 659.

Demir, E., & Dickson, B. J. (2005). *fruitless* Splicing Specifies Male Courtship Behavior in *Drosophila*. *Cell, 121*(5), 785-794.

Elble, R., & Tye, B. K. (1991). Both activation and repression of a-mating-type-specific genes in yeast require transcription factor Mcm1. *Proceedings of the National Academy of Sciences of the United States of America, 88*(23), 10966-10970.

Erdman, S. E., & Burtis, K. C. (1993). The Drosophila doublesex proteins share a novel zinc finger related DNA binding domain. *Embo j, 12*(2), 527-535.

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W., & Mathelier, A. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research, 48*(D1), D87-D92.

Greig, S., & Akam, M. (1995). The role of homeotic genes in the specification of the *Drosophila* gonad. *Current Biology, 5*(9), 1057-1062.

Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science, 267*(5205), 1788.

Hirth, F., Loop, T., Egger, B., Miller, D. F., Kaufman, T. C., & Reichert, H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. *Development, 128*(23), 4781-4788.

Hittinger, C. T., Stern, D. L., & Carroll, S. B. (2005). Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development, 132*(23), 5261.

Hunter, C. P., & Kenyon, C. (1995). Specification of anteroposterior cell fates in Caenorhabditis elegans by Drosophila Hox proteins. *Nature, 377*(6546), 229-232.

Hyduk, D., & Percival-Smith, A. (1996). Genetic Characterization of the Homeodomain-Independent Activity of the Drosophila *fushi taruzu* Gene Product. *Genetics, 142*(2), 481.

Jenett, A., Rubin, Gerald M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, Barret D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, Joanna H., Peng, H., Trautman, Eric T., Svirskas, Robert R., Myers, Eugene W., Iwinski, Zbigniew R., Aso, Y., DePasquale, Gina M., Enos, A., Hulamm, P., Lam, Shing Chun B., Li, H.-H., Laverty, Todd R., Long, F., Qu, L., Murphy, Sean D., Rokicki, K., Safford, T., Shaw, K., Simpson, Julie H., Sowell, A., Tae, S., Yu, Y., & Zugates, Christopher T. (2012). A GAL4-Driver Line Resource for *Drosophila* Neurobiology. *Cell Reports, 2*(4), 991-1001.

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., & Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell, 131*(3), 530-543.

Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., & Desplan, C. (2018). Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. *Cell, 174*(3), 622-635.e613.

Kulcenty, K., Wróblewska, J., Mazurek, S., Liszewska, E., & Jaworski, J. (2015). Molecular mechanisms of induced pluripotency. *Contemporary oncology (Poznan, Poland), 19*(1A), A22-A29.

Lelli, K. M., Noro, B., & Mann, R. S. (2011). Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. *Proc Natl Acad Sci U S A, 108*(52), 21122-21127.

Lewis, E. B. (1985). Regulation of the genes of the bithorax complex in Drosophila. *Cold Spring Harb Symp Quant Biol, 50*, 155-164.

Ludwig, M. Z. (2002). Functional evolution of noncoding DNA. *Curr Opin Genet Dev, 12*(6), 634-639.

Lutz, B., Lu, H. C., Eichele, G., Miller, D., & Kaufman, T. C. (1996). Rescue of Drosophila labial null mutant by the chicken ortholog Hoxb-1 demonstrates that the function of Hox genes is phylogenetically conserved. *Genes & Development, 10*(2), 176-184.

Malicki, J., Schughart, K., & McGinnis, W. (1990). Mouse *Hox-2.2* specifies thoracic segmental identity in Drosophila embryos and larvae. *Cell, 63*(5), 961-967.

Malik, S., & Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics, 11*(11), 761-772.

Mann, R. S., & Carroll, S. B. (2002). Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev, 12*(5), 592-600.

Merabet, S., Litim-Mecheri, I., Karlsson, D., Dixit, R., Saadaoui, M., Monier, B., Brun, C., Thor, S., Vijayraghavan, K., Perrin, L., Pradel, J., & Graba, Y. (2011). Insights into Hox Protein Function from a Large Scale Combinatorial Analysis of Protein Domains. *PLOS Genetics, 7*(10), e1002302.

Nagoshi, R. N., & Baker, B. S. (1990). Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes & Development, 4*(1), 89-97.

Percival-Smith, A. (2017). Non-specificity of Transcription Factor function in Drosophila melanogaster. *Development Genes and Evolution*.

Percival-Smith, A. (2018). Phenotypic Nonspecificity as the Result of Limited Specificity of Transcription Factor Function. *Genetics research international, 2018*, 7089109-7089109.

Percival-Smith, A., & Laing Bondy, J. A. (1999). Analysis of murine HOXA-2 activity in Drosophila melanogaster. *Developmental Genetics, 24*(3-4), 336-344.

Percival-Smith, A., Sivanantharajah, L., Pelling, J. H., & Teft, W. (2013). Developmental competence and the induction of ectopic proboscises in Drosophila melanogaster. *Development Genes and Evolution, 223*(6), 375-387.

Percival-Smith, A., Teft, W. A., & Barta, J. L. (2005). Tarsus determination in Drosophila melanogaster. *Genome, 48*(4), 712-721.

Prince, F., Katsuyama, T., Oshima, Y., Plaza, S., Resendez-Perez, D., Berry, M., Kurata, S., & Gehring, W. J. (2008). The YPWM motif links Antennapedia to the basal transcriptional machinery. *Development, 135*(9), 1669.

Rideout, E. J., Dornan, A. J., Neville, M. C., Eadie, S., & Goodwin, S. F. (2010). Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster. *Nature Neuroscience, 13*(4), 458-466.

Rivlin, P. K., Gong, A., Schneiderman, A. M., & Booker, R. (2001). The role of Ultrabithorax in the patterning of adult thoracic muscles in Drosophila melanogaster. *Development Genes and Evolution, 211*(2), 55-66.

Robinett, C. C., Vaughan, A. G., Knapp, J.-M., & Baker, B. S. (2010). Sex and the Single Cell. II. There Is a Time and Place for Sex. *PLOS Biology, 8*(5), e1000365.

Ryner, L. C., & Baker, B. S. (1991). Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes & Development, 5*(11), 2071-2085.

Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W., & Schedl, P. (1989). The Drosophila female-specific sex-determination gene, Sex-lethal, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes & Development, 3*(5), 708-719.

Simpson, P. (2002). Evolution of development in closely related species of flies and worms. *Nature Reviews Genetics, 3*(12), 907-907.

Sivanantharajah, L., & Percival-Smith, A. (2009). Analysis of the Sequence and Phenotype of Drosophila Sex combs reduced Alleles Reveals Potential Functions of Conserved Protein Motifs of the Sex combs reduced Protein. *Genetics, 182*(1), 191-203.

Sivanantharajah, L., & Percival-Smith, A. (2014). Acquisition of a Leucine Zipper Motif as a Mechanism of Antimorphy for an Allele of the Drosophila Hox Gene Sex Combs Reduced. *G3: Genes|Genomes|Genetics, 4*(5), 829-838.

Sivanantharajah, L., & Percival-Smith, A. (2015). Differential pleiotropy and HOX functional organization. *Developmental Biology, 398*(1), 1-10.

Stern, D. L. (2000). Perspective: Evolutionary Developmental Biology and The Problem of Variation. *Evolution, 54*(4), 1079-1091.

Tanaka, K., Barmina, O., & Kopp, A. (2009). Distinct developmental mechanisms underlie the evolutionary diversification of Drosophila sex combs. *Proceedings of the National Academy of Sciences, 106*(12), 4764.

Tautz, D. (2000). Evolution of transcriptional regulation. *Current Opinion in Genetics & Development, 10*(5), 575-579.

Tia, N., Singh, A. K., Pandey, P., Azad, C. S., Chaudhary, P., & Gambhir, I. S. (2018). Role of Forkhead Box O (FOXO) transcription factor in aging and diseases. *Gene, 648*, 97-105.

Tour, E., Hittinger, C. T., & McGinnis, W. (2005). Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax. *Development, 132*(23), 5271.

True, J. R., Liu, J., Stam, L. F., Zeng, Z. B., & Laurie, C. C. (1997). Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila Simulans* and *Drosophila Mauritiana*. *Evolution, 51*(3), 816-832.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J., & Cohen, S. M. (1992). Homeotic genes of the bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene *Distal-less*. *Cell, 71*(3), 437-450.

Waterbury, J. A., Jackson, L. L., & Schedl, P. (1999). Analysis of the doublesex female protein in Drosophila melanogaster: role on sexual differentiation and behavior and dependence on intersex. *Genetics, 152*(4), 1653-1667.

Wieschaus, E., & Nusslein-Volhard, C. (1986). Looking at embryos. *Drosophila: A Practical Approach*, 199-227.

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*. *Cell, 134*(4), 610-623.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., & Romano, L. A. (2003). The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution, 20*(9), 1377-1419.

Yang, Y., Zhang, W., Bayrer, J. R., & Weiss, M. A. (2008). Doublesex and the Regulation of Sexual Dimorphism in Drosophila melanogaster: structure, function, and mutagenesis of a female-specific domain. *Journal of Biological Chemistry, 283*(11), 7280-7292.

Zhao, J. J., Lazzarini, R. A., & Pick, L. (1993). The mouse Hox-1.3 gene is functionally equivalent to the Drosophila Sex combs reduced gene. *Genes & Development, 7*(3), 343-354.

Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., Brasefield, J. A., Zhu, C., Asriyan, Y., Lapointe, D. S., Sinha, S., Wolfe, S. A., & Brodsky, M. H. (2011). FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic acids research, 39*(Database issue), D111-D117.

Chapter 4. Summary and Discussion

Summary: The objective of this thesis was to study phenotypic non-specificity of transcription factor function. The original experimental design was to study this using two different approaches:

(1) Integration of non-resident TFs into target *TF* loci (via recombinase mediated cassette exchange) subsequent to *attP* sites having been introduced into the *TF* locus by homologous recombination induced by CRISPR.

(2) Assessment of the functional complementation of loss-of-function alleles in several *TF* loci using the *UAS-GAL4* system to express non-resident TFs.

The induction of transposition of white by CRISPR/Cas9: The initial step of the first objective was to introduce *attP—y^+/w^+—attP* cassettes at loci encoding proteins required for *Drosophila* body plan determination. To insert the *attP—y^+/w^+—attP* cassette at a target locus, HDR between a repair template containing the *attP* flanked site and the endogenous locus was to be initiated by CRISPR-mediated DSBs. This modification of a locus would then allow the integration of non-resident TFs into the target *TF* locus using RMCE, thereby setting the foundation for further studies into the phenotypic non-specificity of transcription factor function.

Three w^+ transformants were collected from the HDR experiments: two when targeting *bcd* (*Bcd 4* and *Bcd 39*) and one when targeting *Scr* (*Scr-D1*). However, none of the three w^+ transformants were the result of homologous recombination. *Bcd 4* and *Bcd 39* were the result of the mobilization of the *white* gene on a transposon. When CRISPR was

injected into the w^{67c23} embryos for the experiment targeting *bcd*, the w^{67c23} allele flanked by FB elements migrated to intron 1 of the *osp* gene. The promoter of the *osp* gene drove transcription of w^{67c23} allele, creating chimeric *osp-white* mRNA where the reading frame initiated from the *osp* start codon was in frame with the *w* reading frame leading to expression of a chimeric protein that is functional for transport of eye pigment precursors. The third transformant (*Scr-D1*) from the experiment targeting *Scr* was the result of the insertion of *mini-white* gene of the *Scr* repair template into the genome. All three non-homologous recombination events are the result of activation of transposition.

The significance of the results presented in Chapter 2 are four-fold. First, through the analysis of w^+ transformants, I determined the sequence of the w^{67c23} allele. The w^{67c23} allele has a 130 kb DNA deletion upstream the *white* gene locus, including the first exon, start codon and promoter region of the *white* gene. The FB-NOF-FB element is in the deleted region next to a short piece of duplicated *w* sequence upstream of intron 1. My results show that w^{67c23} is not a stable *w* mutant allele. Second, I provide additional evidence supporting the idea that the FB-NOF element is a non-autonomous transposable element whose transposition is catalyzed by transposases of other transposons (Badal *et al.*, 2013). Third, as the w^{67c23} allele must be inserted into the intron of a gene which is expressed during eye pigmentation in an orientation and reading frame that will allow White expression, the frequency of the w^{67c23} allele migration is likely to be 24 to 48 times higher than what can be identified using w^+ as the marker of migration. Furthermore, the migration of w^{67c23} raises the possibility that all TEs in the Drosophila genome are activated by DSBs. This possibility raises important concerns regarding the unintended consequences of

CRISPR/Cas9 based genetic manipulations with DSBs potentially inducing mass transposition. The unintended activation of transposition by CRISPR induced DSBs may result in mutant alleles in genes closely linked to the locus being targeted for mutagenesis. Therefore, I suggest that the phenotype of at least two independent CRISPR alleles are assessed. Fourth, the w^{67c23} allele migration event is a perfect example of a well-known mechanism important in gene evolution, exon shuffling and transduction transposition, which refers to a phenomenon where exons from different genes are transposed into new genomic contexts and under control of new promoters (Gilbert, 1978; Moran *et al.*, 1999; Moschetti *et al.*, 2004). Therefore, there is the potential that DSB initiate exon migration by inducing transposition and facilitating the generation of novel genetic functions during the evolution.

Future directions for the activation of transposition by DSBs:. DSBs in the genome may activate endogenous TE transposition. This is based on the observation that both the injection of CRISPRCas9 and the transposon Bari1 cause transposition of the w^{67c23} allele. This proposal needs further investigation. This hypothesis could be supported by an experiment using other genetic editing techniques that induce DSBs, TALEN or Zinc-finger in $y^{1}w^{67c23}$ flies. In addition, w^{67c23} flies could be exposed to X-ray irradiation, which induces DSBs, and screened for the w^{+} phenotype. Lastly, the induction of P-element transposition results in DSBs and could be used to screen for migration of the w^{67c23} allele. If similar transposition events are observed, the hypothesis that DSBs activate transposition will be further supported.

I also suspect that other transposons that are not marked with a marker like *w* are also activated by DSBs which would be of a larger concern for analysis of CRISPR induced mutational changes. Experiments are required to determine the frequency of transposition of all TE in the *Drosophila* genome after DSBs are induced. If the activation of transposition by DSBs is high in *Drosophila*, it will need to be investigated in other insects, vertebrates, and humans in particular. The induction of transposition may be problematic in gene drive systems for controlling insect pest populations by inducing unintended resistant genotypes. The activation of transposition in humans may be problematic when CRISPR is used for gene therapy since TE insertions are carcinogenic (Zhang *et al.*, 2015).

Finally, how does the presence of DSBs in the genome transactivate transposition. I hypothesize that the phosphorylation of H2Av at the DSB sites overrides the TE silencing mechanisms thereby activating transposition. To test this hypothesis, activation of transposition in wild type, as well as mutants expressing $H2Av^{SA}$ that cannot be phosphorylated, and mutants expressing $H2Av^{SE}$ that mimic constitutive phosphorylation could be assessed. Also, the TE transcripts levels could be tested before and after CRISPR injection to show that TE expression levels are activated by DSBs. Furthermore, TE transcript levels are silenced by piRNAs and siRNAs systems. I speculate that the formation of DSBs inhibit the piRNAs and siRNAs silencing systems; therefore, it would be interesting to determine the level of piRNAs or siRNAs before and after CRISPR injections. Induction of transposition by DSBs may provide an experimental paradigm with which to study the mechanism of activation of TEs.

Differential pleiotropy of phenotypic non-specificity of TF function: The objective of Chapter 3 was to study phenotypic non-specificity of TF function using functional complementation (where the non-resident TFs are expressed using the *UAS-GAL4* system). I assessed functional complementation of loss-of-function alleles in six *TF* loci (*lab*, *Dfd*, *Scr*, *Ubx*, *dsx* and *fru*) with at least 12 non-resident TFs. I hypothesized that phenotypic non-specificity would be frequently observed and found that *lab* was rescued by expression of DSX^{M}; *Scr* was rescued by expression of FOXO; *Ubx* was rescued by expression of ANTP and EY; *dsx* was rescued to differing extents by the expression of a majority of non-resident TFs; and *fru* was rescued by expression of DISCO. All the *TF* loci examined are highly pleiotropic, that is, the TF is required in multiple tissues and different stages during development. Differential pleiotropy is defined as the distinct behavior of a set of alleles of a gene on two or more phenotypes or biological readouts (Sivanantharajah & Percival-Smith, 2015). In all rescues by the non-resident TFs, I found non-uniform rescue as opposed to rescue of all the phenotypes of the target TF. This suggests that the phenomenon of phenotypic non-specificity exhibits differential pleiotropy.

With respect to Chapter 3, three results emerge as being of particular significance. First, multiple examples of phenotypic non-specificity of TF function in the rescue of TF phenotypes were identified. The high frequency of phenotype non-specificity is an expectation of the model of limited non-specificity of TF function (Percival-Smith, 2018). Second, in the limited specificity of TF function model, differential pleiotropy is an expected outcome of the genetic functional dissection of TF function (Percival-Smith, 2018). Differential pleiotropy has been uncovered in HD containing TFs in *Drosophila*,

Human and yeast (Banreti *et al.*, 2014; Greig & Akam, 1995; Hirth *et al.*, 2001; Lelli *et al.*, 2011; Percival-Smith, 2017; Percival-Smith & Laing Bondy, 1999; Percival-Smith *et al.*, 2013; Percival-Smith *et al.*, 2005). The observation of differential pleiotropy is proposed to be the result of short peptide motifs making small tissue specific contributions to overall TF activity (Sivanantharajah & Percival-Smith, 2009, 2014). Therefore, observing differential pleiotropy in the rescue of phenotypes by non-resident TFs may suggest that the rescuing non-resident TF may only have a subset of the short peptide motifs of the resident TF (each making small contributions to TF activity such that rescue is observed for only some TF phenotypes and not others). Third, the conventional experiments designed to test TF functional conservation are based on the presumption that TF function is specific for a particular phenotype. However, the observation of phenotypic non-specificity undermines this initial assumption (Percival-Smith, 2018) and changes the interpretation of observations of rescue/induction of phenotypes by orthologous TFs to supporting rather than proving functional conservation.

Future directions for testing the predictions of the hypothesis of limited specificity of TF function: Five of the six *TF* loci were found to be rescued by a least one non-resident TF. A total of 18 examples of non-resident TFs rescuing phenotypes were identified. There are two possible explanations of the rescue events. First, the non-resident TF substitutes for the resident TF. Second, the non-resident TF functions downstream of the resident TF. We have shown that DISCO is not downstream of FRU^{M}, and that DSX^{M} is not downstream of LAB. However, this question needs to be answered for the other examples of phenotypic non-specificity observed. I have shown that phenotypic non-specificity is

observed frequently, which is one prediction of the model of limited specificity of TF function. In addition, I have shown that the rescue is differentially pleiotropic, which is an expectation of TFs in the model of limited specificity.

The model of limited specificity of TF function also predicts the phenomena of *cis*-element bypass and the phenomena that only a subset of TF regulated genes is required for the phenotype. The identification of phenotypic non-specificity with the *dsx* locus may allow testing these expectations. DSX^F is required for BAB1 expression in females to suppress abdominal pigmentation (Massey & Wittkopp, 2016). Furthermore, the regulatory element in *bab1* important for this sexually dimorphic expression has been identified (Williams *et al.*, 2008). This system may thus allow the testing of the expectation of *cis*-element bypass because I found that EY expression results in suppression of abdominal pigmentation. Does EY suppress expression of *bab1* as does DSX^F, and does EY bind EY DNA recognition sites in the regulatory element important for dimorphic BAB1 expression?

In addition, the genes regulated by DSX^M and DSX^F are well-characterized; therefore, sex determination provides a useful system to test whether this subset of genes important for male genital formation are suppressed by the TFs DSX^F, SCR, LAB and DFD that suppress male genital formation (Keyes *et al.*, 1992; Lucchesi & Kuroda, 2015; Robinett *et al.*, 2010; Ruiz & Sánchez, 2010). My identification of phenotypic non-specificity of TF function in the rescue of TF phenotypes may provide the reagents to test the expectations predicted by the model of limited specificity of TF function.

REFERENCE

Badal, M., Xamena, N., & Cabré, O. (2013). FB-NOF is a non-autonomous transposable element, expressed in Drosophila melanogaster and present only in the melanogaster group. *Gene, 526*(2), 459-463.

Banreti, A., Hudry, B., Sass, M., Saurin, A. J., & Graba, Y. (2014). Hox proteins mediate developmental and environmental control of autophagy. *Dev Cell, 28*(1), 56-69.

Gilbert, W. (1978). Why genes in pieces? *Nature, 271*(5645), 501-501.

Greig, S., & Akam, M. (1995). The role of homeotic genes in the specification of the *Drosophila* gonad. *Current Biology, 5*(9), 1057-1062.

Hirth, F., Loop, T., Egger, B., Miller, D. F., Kaufman, T. C., & Reichert, H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. *Development, 128*(23), 4781-4788.

Keyes, L. N., Cline, T. W., & Schedl, P. (1992). The primary sex determination signal of Drosophila acts at the level of transcription. *Cell, 68*(5), 933-943.

Lelli, K. M., Noro, B., & Mann, R. S. (2011). Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. *Proc Natl Acad Sci U S A, 108*(52), 21122-21127.

Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage Compensation in Drosophila. *Cold Spring Harbor Perspectives in Biology, 7*(5).

Massey, J. H., & Wittkopp, P. J. (2016). The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. *Current topics in developmental biology, 119*, 27-61.

Moran, J. V., DeBerardinis, R. J., & Kazazian, H. H. (1999). Exon Shuffling by L1 Retrotransposition. *Science, 283*(5407), 1530.

Moschetti, R., Marsano, R. M., Barsanti, P., Caggese, C., & Caizzi, R. (2004). FB elements can promote exon shuffling: a promoter-less white allele can be reactivated by FB mediated transposition in Drosophila melanogaster. *Molecular Genetics and Genomics, 271*(4), 394-401.

Percival-Smith, A. (2017). Non-specificity of Transcription Factor function in Drosophila melanogaster. *Development Genes and Evolution*.

Percival-Smith, A. (2018). Phenotypic Nonspecificity as the Result of Limited Specificity of Transcription Factor Function. *Genetics research international, 2018*, 7089109-7089109.

Percival-Smith, A., & Laing Bondy, J. A. (1999). Analysis of murine HOXA-2 activity in Drosophila melanogaster. *Developmental Genetics, 24*(3-4), 336-344.

Percival-Smith, A., Sivanantharajah, L., Pelling, J. H., & Teft, W. (2013). Developmental competence and the induction of ectopic proboscises in Drosophila melanogaster. *Development Genes and Evolution, 223*(6), 375-387.

Percival-Smith, A., Teft, W. A., & Barta, J. L. (2005). Tarsus determination in Drosophila melanogaster. *Genome, 48*(4), 712-721.

Robinett, C. C., Vaughan, A. G., Knapp, J.-M., & Baker, B. S. (2010). Sex and the Single Cell. II. There Is a Time and Place for Sex. *PLOS Biology, 8*(5), e1000365.

Ruiz, M.-F., & Sánchez, L. (2010). Effect of the gene transformer of Anastrepha on the somatic sexual development of Drosophila. *Int J Dev Biol, 54*(4), 627-633.

Sivanantharajah, L., & Percival-Smith, A. (2009). Analysis of the Sequence and Phenotype of Drosophila Sex combs reduced Alleles Reveals Potential Functions of Conserved Protein Motifs of the Sex combs reduced Protein. *Genetics, 182*(1), 191-203.

Sivanantharajah, L., & Percival-Smith, A. (2014). Acquisition of a Leucine Zipper Motif as a Mechanism of Antimorphy for an Allele of the Drosophila Hox Gene Sex Combs Reduced. *G3: Genes|Genomes|Genetics, 4*(5), 829-838.

Sivanantharajah, L., & Percival-Smith, A. (2015). Differential pleiotropy and HOX functional organization. *Developmental Biology, 398*(1), 1-10.

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*. *Cell, 134*(4), 610-623.

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids, 4*, e264.

Appendices

Appendix 1

w67c23 allele sequence

Red: duplication from exon 2 to exon 3.

Underlined bolded sequence is duplicated.

CTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGG
GCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGA
GCGACACATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTA
TTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATT
TGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGG
GCAGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATC
GCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAATGGCCAG
CATCTGCATCTGATATCTAGGTATCTTCGTGCGTATCTTGCTTCAAATTCTTAG
CACCTCGGCTTGTATAACAAAATAAATAAGTGAGTACGATATGCATATCTAG
CCCCGGGCTCTTTGAAACAATTTTGAAAAGTCTCAAAAAGTTATACAAGGAG
ATAAGAACTTTAATTCTTTTGGGAAGTAAGTAACGCAGTAAAGGTAACAAAG
TATTGAAAAATATGATATGTATGGAATATTTGAAGCCATCTTTAATTATATGT
TCGTTGCATATATGTACATATTGGGCCGTTTACGCTCTGATATTTCCCTAATCA
TATCGAGTGGTCGTCAGCTTTTCCGCTGAATAATTGCGCCTCCTTCAGTATTTC
GTTCTCATCAAAGTGAATGACTTTTGAGTGCACTCATTAGCCACTCGCACAAT
CGACAAAACGGGAGGAACAACAAATTGAGTCGCTTCTTTATTCAGTCATTTC
AATTTGTCAAGTTCATCACGATACTCCTTCGCCTCCCTGCACAGCGCCATCTA
GCGCCCACTCCTCGAGTTTCCTTCGTATTTCGTAAGTAATTCAAAAATAGCAC
GCAAAGTAAATAAAAAAAAAAGTAAGCCAAAAAATTAAATAAATATGAAAT
GTAGCATCTTTCGCTACTTATTTCAACAAAATACGAAAATAGGGAAAAGGTG
GGAGAGTAAACCAGAGGAGGCAGGGAGATAAGAAAACCCAGTGGGAAAGG
GACGGCGAAAACAGAAAAGAAAAGAAAACTTTTTAATAACATTTTTCGGCCT
CATTTGTTGTGAAAACAATTTCGCCATGTCCCTCTATCTCTTTCCCAGCCATAT
GGCCATCTCGCTCTGCTCGAGGAGTGCCTGAAAACGGATATTAAATTTTTTCG
CGTCTACCTCTCGTAACGTCACCAATTGTGCTTTTTTTTTTTGGTAACAGGCCA
AAAACTGGACACTGAGTGGCTAGAAGAATAAGCTGTCAAAGCATCGCAGCA
GAAGAGGGGCAGTGAAGCCTCGCTAAGGAGGAAATGCAACAGTTTTCAGGC
TTTGGTCTTTTAGCTAGACATACATATGTAGCAATGTAAACGATAAATCCAAA
TCCCCCAATCGTTTTGAAAAATAAATTTAATTTTAAATAGTTCAAAAAATGAA
AGTTCTAAGCAGGGCTCGCTGTAGCAAGGACCGCCAGGCGCTGAGTGCAAGG
TGGACATTATTTTTTCTTGCTTTTATTGAATGTCAGCTTTTTATTTGGCCCTCTT
CTTCGGATTTTGCCTTTCTCATTTGCTTGCATTCTTTTGCTTAATTTTTTTTAGG
ATTATTTATACAAATTATTATGAGAGCAGCGGCGGATAAAATGCTTTCCATAT
TGTGGCCATCATTTTGTCTTGTTTTTTCCATTTTATTTTTGCCTGGACGCCAATT
AGTGGATCCCTGTTATTTATGCTTTGCCGCAGGCCGCTGCGATGGCAAGAAAC
CCCACAAATGGAAAACGGAAAATTGTTTATGAAAATGAAAATATGCGCGAAC

GGAAGTGGTATGGGGCGAAAAGTTGGGGCGGGGAGGATGGCTTGATTTGAC
CATACCGCAAACGTCTTCGGTTACCGCCATTTCGAAAAGGGAAGCCCCACCT
GGGCGAAAACTGACAGCTCAGTGGGTTTGGAAATCACTTTACCCGTAAACAA
AATAACGTAAATAAGAATTTGTATTTCGGATTTATAAACTTAAACTCAATTGC
AGAAGTAGTTTGAACACTATTGCGTTCTTTAAATGAATATTCAACATAATTGG
ATCAAGTAAATGGACTGCTTTATACTGCTCATTGCACTTATCTACAAGAGATT
GAGTTTTCCCACAACCCAAAAAAAGCCAAAGTATCCCTAAAGTTTTATATTTA
ATAAAAACAAAACTGGCGAAGCTCAAAGTCCCCAAACGAAAGCCAAAGATG
CCAGCGATTTACACTCGCAATTGTCTTATGAGGTGGCAACATGGTGGAGGAA
CCTTGTCACTCCCACCACCCATCGTCCCATCGTCCCATCGTCCATCGCCCAAC
GACCTGTGACCCCGGCCCCGTCAACCGTCCACCGCCCAGCGCCCACTTTCAA
GGCCATCCACCAAGTGCCAGCCCTAATTTGCACATTTTAAATAGGCCTCGCCA
TCTCTTTTTTTGGGAAACTTGGAAAACTCGGATCTTGGGCATTTTTATTAAAC
AGATGGGCTCCTTCTTTATAAAGCNNNNNNNNNGGGTTGGGGTTAGTGCCCG
GAAACAGGGACGGCGGAGCAGGGATTAGCCAGGCTGGGCTAGATTTATGCA
CAGACGCCTTCATTTTTAGGGACGAACGCCGTGAAATTGAAAAGTTTTCAGCT
TGAAACCTTTTCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCG
TAGGATACTTCGTTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATA
TTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTG
CATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCGGCCACCC
AGTCCGCCGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAAC**CTC**
ACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGG
GCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAA
CGAGCGACACATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAG
TTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAA
TTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTT
GGCCGTGATGGGCAGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCC
CTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGC
GACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCG
CCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACA
CCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAG
CGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTC
AGCACACGATCATCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGA
AAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTG
ATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGT
CCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATT
CATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGC
CGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTT
CCTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCA
ACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGAC
TTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTG
ATCGGATCGCCAAGATATGCGACAATTTTGCCATTAGCAAAGTAGCCCGGGA
TATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCG
GAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCC
TGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCG
ACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCT

TACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCA
CACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAAT
AAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAG
ATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCT
TTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCAT
CTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATG
TAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTAATGA
ATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGC
CCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCG
AATTGCCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCG
ATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGT
CACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCA
GCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTC
CTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAA
ATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTA
ACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACA
CCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCC
GCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCG
GGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCG
ACATATATCCGAAATAACTGCTTGTTTTTTTTTTTTACCATTATTACCATCGT
GTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAAT
AAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCC
ACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATAC
CTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTA
TATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACC
ACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCA
CCTTTTGTTCGA

Appendix 2

Bcd 39 sequences at the insertion site

Red: *white* allele
Black: FB element
Purple: *osp* gene sequence
Bold Purple: duplicate sequence

Bcd 39 insertion site 5' junction sequence:

GGTTT**CTGACAGTGTG**AGCTCAAAGAAGCTGGGGTAGCTCAAAGAAGCT
GGGGTCGGAAAAATCGAATTTTTGAAATTTGAAAGCTGGAATCGTTTGCCCA
TTTTTTGCCCATGTTTGCCCACCAATTAGTTTTTTTTGCCCACGTCCAGTTTTT
GAGATATGGATTTTCGAAAAAGTTCGAAAATGTTCGAAAATCAAAAATTTCG
CTTTTTTCAAATTTTTTTTT----------

GGGCAAACGATTTACTCTCTGATTAGACCGAGGTAAACTTAAAAGCCTTATA
TTTTCTAAAGTATAAATTTTTTCAAAATTCTAAAGGGTGGGCAAACGTGGGCA
AACGATGTTATTGCGATTTAAAAAAAAAAATTTTGGAAAAAGCGAAATTTTT
GATTTTCGAAAATTTTCGAACTTTTTCGAAAATTCATATCTCAAAAACTGGAC
GTGGGCAAAAAAAACTAATTGGTGGGCAAACATGGGCAAAAAATGGGCAAA
CGATTCCAGCTTTCAAATTTCAAAAATTCGATTTTTCCGACCCCAGCTTCTTTG
AGCTCTCACCTATGCCTGGCACAATATGGACATCTT

Bcd 39 insertion site 3' junction sequence:

CGCATAAAAAGCAAAACGGGGGCATTGAAAAAGGTTTGTTTGTGCATTTTAA
AGCTCAAAGAAGCTGGGGTCGGAAAAATCGAATTTTTGAAATTTGAAAGCTG
GAATCGTTTGCCCATTTTTTGCCCATGTTTGCCCACCAATTAGTTTTTTTTGCC
CACGTCCAGTTTTTGAGATATGAATTTTCGAAAAAGTTCGAAAATTTTCGAAA
ATCAAAAATTTCGCTTTTTTCAAAATTTTTTTTTTTAAATCGCAATAACATCGT
TTGCCCACGTTTGCCCACCCTT----------

TTTCTAAAGTATAAATTTTTTCAAAATTCTAAAGGGTGGGCAAACGTGGGCA
AACGATGTTATTGCGATTTAAAAAAAAAAATTTTGAAAAAAGCGAAATTTTT
GATTTTCGAAAATTTTCGAACTTTTTCGAAAATTCATATCTCAAAAACTGGAC
GTGGGCAAAAAAAACTAATTGGTGGGCAAACATGGGCAAAAAATGGGCAAA
CGATTCCAGCTTTCAAATTTCAAAAATTCGATTTTTCCGACCCCAGCTTCTTTG
AG**CTGACAGTGTG**CCTTGGCCTGG

Appendix 3

Bcd 4 sequences at the insertion site

Red: *white* allele
Black: FB element
Purple: *osp* gene sequence
Bold Purple: duplicate sequence

Bcd 4 insertion site 5' junction sequence:

TTTAAAATTAAATAATTCT**TAGTTTGTT**GAAGCTGGGGTCGGAAA----------

ATGTTATTGCGATTTAAAAAAAAAAATTTTGGAAAAAGCGAAATTTTTGATTT
TCGAAAATTTTCGAACTTTTTCGAAAATTCATATCTCAAAAACTGGACGTGGG
CAAAAAAAACTAATTGGTGGGCAAACATGGGCAAAAAATGGGCAAACGATT
CCAGCTTTCAAATTTCAAAAATTCGATTTTTCCGACCCCAGCTTCTTTGAGCTC
TCACCTATGCCTGGCACAATATGGACATCTT

Bcd 4 insertion site 3' junction sequence:

TCAGCGTTTGATTTACGCATCGCACGGCGCATAAAAAGCAAAACGGGGGCAT
TGAAAAAGGTTTGTTTGTGCATTTTAAAGCTCAAAGAAGCTGGGGTCGGAAA
AATCGAATTTTTGAAATTTGAAAGCTGGAATCGTTTGCCCATTTTTTGCCCAT
GTTTGCCCACCAATTAGTTTTTTTTGCCCACGTCCAGTTTTTGAGATATGAATT
TTCG----------

ATCTCAAAAACTGGACGTGGGCAAAAAAAACTAATTGGTGGGCAAACATGG
GCAAAAAATGGGCAAACGATTCCAGCTTTCAAATTTCAAAAATTCGATTTTTC
CGACCCCAGCTTCTTTGAGC**TAGTTTGTT**AATGGTTTATTAAGATTAATCG

Appendix 4

Scr-D1 sequences at the insertion site.

Red: White allele
Black: plasmid sequence
Purple: 17.6 TE.
Bold Purple: duplicate sequence

Scr-D1 insertion site 3' (right junction)

GCTCTCCCAAAAACACTAACATATTCTTTAAGCAAGCACAG**AGGGTT**ATTG
CGCCTTCACTGTATGCCATGGCCCTAATTTTACAGTATCATCCAAGCATTTTCT
AAATTAAATGTATTCTTATTATTATAGTTGTTATTTTTGATATATATAAACAAC
ACTATTATGCCCACCATTTTTTTGAGATGCATCTACACAAGGAACAAACACTG
GATGTCACTTTCAGTTCAAATTGTAACGCTAATCACTCCGAACAGGTCACAAA
AAATTACCTTAAAAAGTCATAATATTAAATTAGAATAAATATAGCT……
……

Scr-D1 insertion site 5' (left junction)

TCAAAAAACAAACAAAAATAAGAAGCGAGAGGAGTTTTGGCACAGCACTTT
GTGTTTAATTGATGGCGTAAACCGCTTGGAGCTTCGTCACGAAACCGCTGAC
AAAGTGCAACTGAAGGCGGACATTGACGCTAGGTAACGCTACAAACGGTGTT
AGGGTTCAATGATAGTGCT**AGGGTT**CTCCTCATTTTCACTTTCATTTGATTT
TTAGTCTTAAGCTGAACGTTAATCAATAAACAACACAATCGATCCCGAAATTT
TGATTCGTTTTATTTTGGCAAAACTTAATTTTCAGCGTTGGTCTTAGTTCATAT
TCGGAACGGTCCATTTAATAGACTCAA

Appendix 5

Scr donor plasmid sequence

Black (Capital): pFUS_A Vector
Green: 5'HA (*Scr* Exon 2)
Purple: 3'HA (*Scr* Exon 3)
Yellow: *yellow*
Red (lowercase): *mini-white*
Blue: attP sites, (39*2)
Underlined: sequence transposed

TTGATGCCTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCCAGTCACGACG
TTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGC
TATACCTGGGGGCAAGTTTACAATATTTCCTTAATTTTTATTTTTGTTGTGTTTTCCCCGAGAA
CCCTGCTCATCAAATAAATCTCTGTCCAAAGTTGAGCATTCGCTCCTTGGCCAGTTTCGAATGG
CGTACGGCGTTTTAATTTAAGCCCAAGTTGAGAGCTCCTTTCATTTGGCCAACATGCTAAAGG
GTTAAATTGCCCACTGAATCAAAATTATTGGTTTCAAGCCTCTAAAAAGGGCAGGGAAGTGGG
GGGCTTGCATTTGCTGCATTGTTTTTAGCCAATGTTCTTCGTTCTCTGCTGTTGTTGCTTATTTC
GATACCCTTTAAGGTAGGTATTATCTACTCACATATACAGATAAGATATCTTACAACAGTTTTC
CGACTCTCGATACAATTTTCTCAGCCATGCCCAGCCAAGTTCATCTTCAAGTTCTACATATATG
GGGTAAACATAATATTTAGAACAGATCGAAAGGGTATTTTGTAATATATTGAAGCGTATTTTA
AATTCGGCATTTTGCCATCTTCGTTTTTAGTTTTGGGCATGTTCATGTGCAGCGGCCTCGCTTG
GCCTCTTTTGCCTTATTGCCTTTTTGCACTTTTTCGGCTCTGGGGCGATTCGAGAGCTCCTTTCC
TTCAGCTCCTTTGGAAGCGGAGGGTCAATGCTTTTGTTTGCTATCTGCGCAGCGGAGAATACG
AGTTTTCCGCCATTGAACAAACATGCCCAAGACCAAACCAAGACTGCAATATCTGCGATGGTT
GCACCGAGAGGAGTGGGTCTGGCCAGCAAGTGCAGCCGCAAGAAGTTTCCGCCAAGTGTGTC
CAGAAGAGTGGGCGGTCCCTCGGCATTCCTCGGCCAAAAGCTCCAGTACGTATATATATAGTA
TCCACTTCGGCGGGTAAAATTGCACGCTTAGGCGGCCCCAGCTCTCTCCACCAGCGGAAAATA
AGCTCAAAAAAAAAAAAGAAACGACTCTTTACCGTTCTCAAAGAAATAATAATAATAACCAA
GTAAAAGGCAAGTGAAATTATCAGCGGTTAAGAAAACACTTTGAGGGAGTAACAAATCAAGA
GGCCTGAGTGGGAAAGAGGCTTTTGGGCTAAGTTTCTATAAAATACTGAAAACAGCTCACTTA
GCTTGCCTGTTCAACAGGTATTCCTATTCTGTTGCTCATAGTCTTAAAAAAAAATGTCATCATA
TTTTATAAGATTCTCTTAAAATTATCATAGGCTTAAAAATGGATCATGCAAATAAATATACAT
GAGAATTTATGACCTCGGGCCTCGTATAGCGATCCCTTTCAAATAAGTTGGAGCCTGGAACTG
GCTTCCACTTGAGGAGGTATCTTTACTTAAATGGAAGTACTTAATCCCCCGCAGCTTTTCCAAT
ACAATGTTGCCAATCGCAAAGTGAGATCTCAACGCAAAAGTGAACAGCAAGAAAAATATATT
TGTATTTTCTGAGTCTTCTTATCAACTTCTGATCATTTTATTCAATTTGATTTTTGAATTTATGG
CAAAGGTATATAAGCCTTTAAAAAGTTTACAATTGCATTATATTTTGTTTTCAATAAGTTCCCA
TTTAATACCCCATCCTCTCATTTCCAGACTCCAATCCACGACAACGACCCCTGGCTTTACCCCA
ACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCATGCGACTATTAAATGATTATCGCCCGA
TTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACGATGGCTTAAGATAAAT
ACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCCTCTAAACGA
TTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGCACG
AGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCA
AGATTACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGC
ATCTAATTATTCCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCT
AGCGAGTTCCTTGGCGACATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTG
GAAATTGACTATTTTATTTATATTAATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGGGT
ACTAATAGTTTGAGCGCAGTGCATGTCATGGGGACATGTGCAATTGTGTGTAAGCGGGAAGTG
ATCGCGGCCTTCCGAATTTGGCCATGCCAAATAATCCCAGCTCGAAAGGAGGGGACCCagcactat
cattgaaccctaacaccgtttgtagcgttacctagcgtcaatgtccgccttcagttgcactttgtcagcggtttcgtgacgaagctccaagcggtttacgccatc
aattaaacacaaagtgctgtgccaaaactcctctcgcttcttatttttgtttgttttttgagtgattggggtggtgattggttttgggtgggtaagcaggggaaagtg
tgaaaaatcccggcaatgggccaagaggatcaggagctattaattcgcggaggcagcaaacacccatctgccgagcatctgaacaatgtgagtagtacat

gtgcatacatcttaagttcacttgatctataggaactgcgattgcaacatcaaattgtctgcggcgtgagaactgcgacccacaaaaatcccaaaccgcaatc
gcacaaacaaatagtgacacgaaacagattattctggtagctgtgctcgctatataagacaatttttaagatcatatcatgatcaagacatctaaaggcattcatt
ttcgactacattctttttttacaaaaaatataacaaccagatattttaagctgatcctagatgcacaaaaaataaataaaagtataaacctacttcgtaggatacttcg
ttttgttcgggggttagatgagcataacgcttgtagttgatatttgagatcccctatcattgcagggtgacagcggacgcttcgcagagctgcattaaccagggct
tcgggcaggccaaaaactacggcacgctcctgccacccagtccgccggaggactccggttcagggagcggccaactagccgagaacctcacctatgcc
tggcacaatatggacatctttggggcggtcaatcagccgggctccggatggcggcagctggtcaaccggacacgcggactattctgcaacgagcgacac
ataccggcgcccaggaaacatttgctcaagaacggtgagtttctattcgcagtcggctgatctgtgtgaaatcttaataaagggtccaattaccaatttgaaact
cagtttgcggcgtggcctatccgggcgaacttttggccgtgatgggcagttccggtgccggaaagacgaccctgctgaatgcccttgcctttcgatcgccgc
agggcatccaagtatcgccatccgggatgcgactgctcaatggccaacctgtggacgccaaggagatgcaggccaggtgcgcctatgtccagcaggatg
acctctttatcggctccctaacggccagggaacacctgattttccaggccatggtgcggatgccacgacatctgacctatcggcagcgagtggcccgcgtg
gatcaggtgatccaggagctttcgctcagcaaatgtcagcacacgatcatcggtgtgcccggcagggtgaaaggtctgtccggcggagaaaggaagcgt
ctggcattcgcctccgaggcactaaccgatccgccgcttctgatctgcgatgagcccacctccggactggactcatttaccgcccacagcgtcgtccaggtg
ctgaagaagctgtcgcagaagggcaagaccgtcatcctgaccattcatcagccgtcttccgagctgtttgagctctttgacaagatccttctgatggccgagg
gcagggtagctttcttgggcactcccagcgaagccgtcgacttcttttcctagtgagttcgatgtgtttattaagggtatctagcattacattacatctcaactcct
atccagcgtgggtgcccagtgtcctaccaactacaatccggcggacttttacgtacaggtgttggccgttgtgcccggacgggagatcgagtcccgtgatc
ggatcgccaagatatggcgacaattttgctattagcaaagtagcccgggatatggagcagttgttggccaccaaaaatttggagaagccactggagcagcc
ggagaatgggtacacctacaaggccacctggttcatgcagttccgggcggtcctgtggcgatcctggctgtcggtgctcaaggaaccactcctcgtaaaag
tgcgacttattcagacaacggtgagtggttccagtggaaacaaatgatataacgcttacaattcttggaaacaaattcgctagattttagttagaattgcctgatt
ccacaccttcttagttttttcaatgagatgtatagtttatagttttgcagaaaataaataaatttcatttaactcgcgaacatgttgaagatatgaatattaatgagat
gcgagtaacattttaatttgcagatggttgccatcttgattggcctcatctttttgggccaacaactcacgcaagtgggcgtgatgaatatcaacggagccatctt
cctcttcctgaccaacatgacctttcaaaacgtctttgccacgataaatgtaagtcttgtttagaatacatttgcatattaataatttactaactttctaatgaatcgatt
cgatttaggtgttcacctcagagctgccagtttttatgagggaggcccgaagtcgactttatcgctgtgacacatactttctgggcaaaacgattgccgaattac
cgctttttctcacagtgccactggtcttcacggcgattgcctatccgatgatcggactgcgggccggagtgctgcacttcttcaactgcctggcgctggtcact
ctggtggccaatgtgtcaacgtccttcggatatctaatatcctgcgccagctcctcgacctcgatggcgctgtctgtgggtccgccggttatcataccattcctg
ctctttggcggcttcttcttgaactcgggctcggtgccagtatacctcaaatggttgtcgtacctctcatggttccgttacgccaacgagggtctgctgattaacc
aatgggcggacgtggagccgggcgaaattagctgcacatcgtcgaacaccacgtgccccagttcgggcaaggtcatcctggagacgcttaacttctccgc
cgccgatctgccgctggactacgtgggtctggccattctcatcgtgagcttccgggtgctcgcatatctggctctaagacttcgggcccgacgcaaggagta
gccgacatatatccgaaataactgcttgttttttttttttaccattattaccatcgtgtttactgtttattgccccctcaaaaagctaatgtaattatatttgtgccaataaa
aacaagatatgacctatagaatacaagtatttcccctcgaacatccccacaagtagactttggatttgtcttctaaccaaaagacttacacacctgcatacctta
catcaaaaactcgtttatcgctacataaaacaccgggatatatttttttatatacatacttttcaaatcgcgcgccctcttcataattcacctccaccacaccacgttt
cgtagttgctctttcgctgtctcccacccgctctccgcaacacattcacctttgttcgacgaccttggagcgactgtcgttagttccgcgcgattcggttcgctc
aaatggttccgagtggttcatttcgtctcaatagaaattagtaataaatatttgtatgtacaatttatttgctccaatatatttgtatatatttccctcacagctatatttat
tctaatttaatattatgacttttaaggtaatttttgtgacctgttcggagtgattagcgttacaatttgaactgaaagtgacatccagtgtttgttccttgtgtagatgc
atctcaaaaaaatggtgggcataatagtgttgtttatatatatcaaaaataacaactataataataagaatacatttaatttagaaaatgcttggatgatactgtaaa
attagggccatggcatacagtgaaggcgcaataaccGGATCCGTTGACCTGCAGGTCGAGGACCCCCCAACTGAGA
GAACTCAAAGGTTACCCCAGTTGGGGGACGCGTGGCACTTTTCGGGTACTGGTTAGAGACTCA
ATCAGTTGTATCAGGAACCATATCAGACGGCGGCGGCGGCATCGGCGGCCAGTGGTTATCAG
TCGCAGGACGGCGGTCCAATTGGTGGCGGCTCGGTGGGAGTGGGTGGTGGTGCGGGGGGGCC
GGGCTCGCTGGCCAACGGTGGCAGCAATGGCAGCGGTCCCAACTCTCTCTTCGCCTCCGCCGC
CTCCAGCTCCCAGGCTCCCGACTGCATCAAGTATCCGCAGGAGTTCTGATCTCAGGTTATCAT
CAGGCAGCAGCATCAGCAGCAGCAGGATAACTCCATGGATCGGAATCAGGAGCGGGACCTAA
AACTGCTTTTGGAAACGGACTGTGAGCCCGATCCAGAGCTGCAACTGGAGTTCAAGGCGGAC
ATTGTCGAATGCAACTTGTTCTGCTGCTGAGCTATGTCCGCCTGCCACGCCCCTGCCGCCTCAT
GCCTGCCGCCTACCGCCCATCGCCTACCGCCTGTCGCCGCCCGTCGATCAACCAACATCCATA
GATGTTTCTTGTTTGCAACTGATTCGTGTTAAGCTAAGAAACGAGCCATATCCGCATGGTTCGT
TTGGGTTTATGATACTGGAGATGGAACCACGAGCAATCGACGTCACTCTTTAAGATATTAATC
ATCAACCTTTAACTAGAATTTGTGATACGCTGAGTCTAACATATCATGATCATATTATTCACGT
CGATTTGAATCAATGACATTAGTATGTGATGTTAGTTTGCTGTTGAATGATATGTATATGATTG
ATATTACTTGATAATGTCAAAGTTAGGTGTTATCCCTCACTCACTTTCTAAACCTTTCTATGCG
AATGCTCGAGGGTCCTTCCAGTATGATGAATCTCTTTCTCTAAGCAGTATTCGAAAAGCCGTT
AGTTTGTAAGCCTAGCATAATTTTAGCACGTAAGCTGCACGATGAGCAGACAAAGCTGTCGAG
TTGAACAAATCTAATATCTGCAATAATTTCAAAGTGATTTCTAATTAACATTAGGTCTTCGTTT
GTATCGCCCAATCTCCAACCTTGATCCCAACCAAACCCCCAACCAGCACCCTCCAGGGAGTCC
TAAAGAGTGTACTTAATTCCTAGCTAGTAGAGAATATAACTCTAGTTACCCACATAAGGCTTT
GTTTAGTTTGTAAATAGCAGAGCGCCCCAGATCGGGGATCGCCTTTTTATGTTTGTGTTTCGTC
TTGAAGAGAAAATTTGAAAGTATCTGCAACTGTTTAATTCAATTTATTTTGAGTGTTTGCGTTT

ATGTTGTTGAAATTGCTCTAGCAGCTTGGAAAATGCTTTACTAATTTTATCAAAAATCAATCGT
TTTTTATTGCCACTGAAGAACAAATTCGGAAGATAAACAGGAAGTAAAATAATTCAAAAATA
GCATAGACCTTGCTATAGTTAATCAGAGTTTTCTTTAGAACATCAGCCGATCCTTCCTCTTTAT
TTATTCAAAGGCCTCTCAACAATGTCTGTAAATTTAATTCGGTAGTTAATCGATACAGTGTAA
GCCAACGAAATTTGATTAAAGTGTGAATCGTTAAAGCCTAAAGCTAAAGGAAACCCATAACC
GTAATGTAAAGTAATTATTTATGCTGTACTTTTCGCTAAGCTAAGGTTAGTGCATTCTAGAACT
ATCGGTAGTATATGTCAACCTAGATCGTAAGCCTAAATTATGTATATCGAATTAGCAAGACAA
ATTTTAGAGAAAACAAATCGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGG
TGCTGTGCCAGGACCATGGCCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCG
CTGCTACCCAAGGTTGCCGGGTGACGCACACCGTGAAACGGATGAAGGCACGAACCCAGTGG
ACATAAGCCTGTTCGGTTCGTAAGCTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCC
AGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGAC
TGTTTTTTTGGGGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGT
CGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAA
GTTAAACATTATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTG
GCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGG
ATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGATG
AAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCG
AGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATC
CAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCG
AGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTT
GCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAG
GCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAA
TGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGA
TGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGC
TAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATCAGTTGGAAG
AATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAACCCTCGAGCCAC
CCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAA
AAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCC
ACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGC
TGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCT
ACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGA
AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC
CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGC
GCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTG
AGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGC
CATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCG
CCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGG
AGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCG
TTTTAT

Appendix 6

Figure permissions.

Chapter 1. Figure 1:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127450736456
Publication:	Confidential - Genetics
Title:	A Short History and Description of *Drosophila melanogaster* Classical Genetics: Chromosome Aberrations, Forward Genetic Screens, and the Nature of Mutations
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 3:

Dear Sheng Cheng,

I am well, thank you. Hope you are the same. I am happy for you to use that image from my thesis.

Good luck with the writing!

Lovesha

--
Lovesha Sivanantharajah, Ph.D.
Alzheimer's Society Junior Fellow
School of Biological Sciences
Bangor University

Brambell Building
Bangor University
Bangor, Gwynedd, Wales, UK
LL57 2UW

Dear Lovesha,

My name is Sheng Cheng, a PhD student from Dr. Percival-Smith's Lab, UWO. How are you doing ? I hope everything goes well for you. I am writing my thesis now and I am wondering whether I could use one of the figures from your PhD thesis. It is Figure 1.1 A which shows the hierarchy of the proteins expressed to control development. It will be great help if I can get your permission. Please let me know. Thank you.

Sincerely
Sheng Cheng

Chapter 1 Figure 5:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127451003130
Publication:	Developmental Biology
Title:	A genetic and developmental analysis of mutations in labial, a gene necessary for proper head formation in Drosophila melanogaster
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 6:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127460299047
Publication:	Confidential - Genetics
Title:	The Drosophila Genes *disconnected* and *disco-related* Are Redundant With Respect to Larval Head Development and Accumulation of mRNAs From Deformed Target Genes
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 7:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127460492405
Publication:	Confidential - Genetics
Title:	Analysis of the Sequence and Phenotype of Drosophila *Sex combs reduced* Alleles Reveals Potential Functions of Conserved Protein Motifs of the Sex combs reduced Protein
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 8:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127460768419
Publication:	Arthropod Structure & Development
Title:	The skeletomuscular system of the larva of Drosophila melanogaster (Drosophilidae, Diptera) – A contribution to the morphology of a model organism
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 9:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127460968521
Publication:	Development Genes and Evolution
Title:	The role of Ultrabithorax in the patterning of adult thoracic muscles in Drosophila melanogaster
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 10:
From Open access resource: PLOS Biology.

Chapter 1 Figure 11:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127461309332
Publication:	Confidential - Genetics
Title:	Analysis of the Doublesex Female Protein in Drosophila melanogaster: Role in Sexual Differentiation and Behavior and Dependence on Intersex
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 12:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127470340639
Publication:	Confidential - Genetics
Title:	Advances in Engineering the Fly Genome with the CRISPR-Cas System
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 13:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127470723948
Publication:	Confidential - Genetics
Title:	A Simple Polymerase Chain Reaction-Based Method for the Construction of Recombinase-Mediated Cassette Exchange Donor Vectors
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 1 Figure 16:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127470965770
Publication:	Molecular Genetics and Genomics
Title:	FB elements can promote exon shuffling: a promoter-less white allele can be reactivated by FB mediated transposition in Drosophila melanogaster
Type of Use:	Thesis/Dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Chapter 2 Figure 1:

Dear Mr. Sheng Cheng,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Sheng Cheng
Order Date:	Aug 14, 2021
Order Number:	5127471199086
Publication:	CELL REPORTS
Title:	Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 CAD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Curriculum Vitae

Name: Sheng Cheng

Post-secondary Education and Degrees:

University of Toronto
Toronto, ON, Canada
2006- 2011 Hon.B.Sc.

McMaster University
Hamilton, ON, Canada
2011- 2014 M.Sc.

University of Western Ontario
London, ON, Canada,
2015- 2021 Ph.D.

Honors and Awards:

Catherine Jane Stephenson Memorial Bursary (McMaster) 2012

McMaster Graduate Teaching Assistantship
2011- 2013

Western Graduate Research Scholarship (UWO)
2015- 2019

Related Work Experience:

Teaching Assistant (UWO)
2015- 2020

Teaching Assistant (McMaster)
2011- 2013

Publications:

Cheng, **S.,** Garofalo, L., & Percival-Smith, A. (2021). CRISPR activation of transposition in *Drosophila melanogaster*. G3 Genes | Genome | Genetics Submitted.

Selected Presentations:

Cheng, **S.** & Percival-Smith, A. (2019) Characterisation of CRISPR induced non-homologous recombination. Canfly 2019. Toronto, ON, Canada. Oral Presentation.

Advisory/Supervisory Role:

Research Project Advisor:

2019-2020: Tothong Sonpaveerawong, Honors thesis student, The University of Western Ontario

Thesis title: Identifying a Purine Transporter in *Staphylococcus aureus*