# The X chromosome as a unique environment for gene expression in *Drosophila melanogaster*

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München


Aleksei Belyi

München 2021

Diese Dissertation wurde angefertigt unter der Leitung von

Prof. Dr. John Parsch an der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. John Parsch

Zweitgutachter: Prof. Dr. Richard Merrill

Tag der Abgabe: den 19.05.2021

Tag der mündlichen Prüfung: den 8.09.2021

## Erklärung

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. John Parsch betreut. Ich erkläre hiermit, dass die Dissertation nicht einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung unterzogen habe.

München, 19. Mai 2021

_____

Aleksei Belyi

## Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist.

München, 19. Mai 2021

_____

Aleksei Belyi

# Contents

**Contents**

# Declaration of Author's Contribution

In my thesis, we studied the two X chromosome-wide regulatory mechanisms in *D. melanogaster*: dosage compensation in male somatic tissues and X-chromosome suppression in the male germline. The first part of my project involved generating a large number of random insertions of a ubiquitously expressed reporter gene on the X chromosome and measuring their expression level in different tissues of males and females. By calculating the correlation between the reporter gene expression in males, females, the male-to-female expression ratio, as well as the distance to the nearest dosage compensation complex (DCC) binding sites, we aimed to test the effects of the chromosomal context on the expression of X-linked genes in *D. melanogaster* males and females. The second part of my project involved a forward genetic screen of flies carrying a reporter gene specifically expressed in *Drosophila* testis. With this approach, we aimed to detect candidate genes associated with X chromosomal suppression in the male germline.

This study was conducted between October 2017 and March 2021 with the following contributions from other scientists:

1. The conception and design of the study were done by John Parsch and me. The manuscript was written by me with the support and revision of my supervisor, John Parsch.
2. The results of the first part of this study are based on the publication in *Genome Biology and Evolution*:

   Belyi A, Argyridou E, Parsch J. 2020. The influence of chromosomal environment on X-linked gene expression in *Drosophila melanogaster*. Genome Biol. Evol. 12:2391–2402. https://academic.oup.com/gbe/article/12/12/2391/5940000.

In my thesis, I used the original data, text and figures from the published paper with minor changes without providing references to the published paper. The generation of 82 new reporter gene insertions, mapping the insertion sites, measuring reporter gene expression, collecting all data, and statistical analysis were done by myself. The *CMV-lacZ* reporter gene construct was generated by John Parsch. Initial insertions of this reporter gene at 20 unique X-linked locations were generated by Eliza Argyridou and John Parsch.

3. The genetic screen of mutant flies was performed by me in collaboration with Eliza Argyridou, Hildegard Lainer, Anna-Lena Schumacher and two supervised master students: Ingo Müller and Derek Roberts. All genetic crosses, the generation of reporter gene expression data, and the statistical analysis for all candidate lines were performed by me. Mapping the chromosomal location of causative mutations using genetic crosses and analysing the association of target mutations and mutant phenotype were carried out by myself. PCR assays for the association analysis were performed by myself with the help of Hildegard Lainer. The design of the primers for the association analysis was done by John Parsch. Mapping the genomic location of all induced mutations using whole-genome sequencing and searching for candidate mutations were done by John Parsch and Eliza Argyridou. Expression of endogenous X-linked genes in the testis of the *INXS1* line was measured using the RNA sequencing approach with library preparation and data analysis performed by Eliza Argyridou. The knockdown of the candidate genes using RNA interference (RNAi) was performed by myself. The expression of endogenous X-linked genes in the testis of the *INXS2* mutants and flies with knockout genes was measured using the quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) approach by myself with the help of Hildegard Lainer. I performed fertility tests and morphological examination of the testis and germline of both mutant lines.

# Abstract

The evolution of genetic sex determination in many organisms has led to the development of heteromorphic sex chromosomes. Such chromosomes often acquire specific regulatory mechanisms and gene content due to the absence of recombination between them. In *D. melanogaster*, two major chromosomal regulatory mechanisms, X chromosome-specific dosage compensation and X suppression in the male germline, create a unique environment for gene expression in different sexes and across tissues.

In the male soma, X-linked genes are dosage compensated by having their expression up-regulated, a process mediated by the binding of the dosage compensation complex (DCC). Previous studies of X-linked gene expression found a negative correlation between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site in somatic tissues, including head and brain, which suggests that dosage compensation influences sex-biased gene expression. However, a limitation of the previous studies was that they focused on endogenous X-linked genes and, thus, could not disentangle the effects of chromosomal position from those of gene-specific regulation. In the first part of my thesis, we addressed this limitation by examining the expression of a *CMV-lacZ* reporter gene construct consisting of the *Escherichia coli lacZ* gene under the control of the minimal human cytomegalovirus (CMV) promoter inserted at many locations spanning the X chromosome. By doing so, we were able to test the effect of a reporter gene's proximity to a DCC binding site on its expression in males, females and the male-to-female expression ratio across different tissues. We observed a negative correlation between the male-to-female expression ratio of the reporter gene and its distance to the nearest DCC binding site in somatic tissues but not in gonads. A reporter gene's location relative to a DCC binding site had a greater influence on its expression than the local regulatory elements of neighbouring endogenous genes, suggesting that intra-chromosomal variation in the strength of dosage compensation is a major determinant of sex-biased gene expression. Average levels of sex-biased expression did not

differ between head and brain, but there was greater positional effect variation in the brain, which may explain the observed excess of endogenous sex-biased genes located on the X chromosome in this tissue.

In the male germline, however, there is no DCC-mediated dosage compensation, and the expression of X-linked genes is suppressed through a mechanism analogous to the meiotic sex chromosome inactivation that occurs in mammals. In contrast to the latter, in *Drosophila* the X suppression is not complete and its extent is correlated with the gene's expression level in testes. As the genetic and molecular mechanisms behind this X suppression are unknown, in the second part of my dissertation, we used a forward genetic screen to discover the genes responsible for the X suppression. For this, we performed chemical mutagenesis on males with an X-linked copy of the *lacZ* reporter gene controlled by the *ocnus* promoter, which has testis-specific expression and previously had shown a strong effect of X suppression.

With this approach, we detected two mutant lines that we named *INvolved in X Suppression* (*INXS1* and *INXS2*), which showed a strong increase of reporter gene expression relative to the control line. Analysis of endogenous X-linked gene expression confirmed a general increase in expression in both mutants. Also, males of both *INXS* mutants were sterile and displayed complete sperm immobility. Taken together, these phenotypes indicate suggest that there is partial or complete relaxation of X suppression in the male germline. We found four top candidate genes on the X (*INXS1*: *CG13003*; *INXS2*: *CG1314*) and third chromosomes (*INXS1* and *INXS2*: *CG31525*; *INXS1*: *CG42654*), with all genes except *CG13003* having high testis-specific expression. The genes *CG13003*, *CG31525* and *CG42654* showed a partial association with increased reporter gene expression in *INXS1*, whereas, in *INXS2,* both *CG1314* and *CG31525* showed a complete association. However, with a preliminary functional analysis of the *CG13003*, *CG1314* and *CG31525* genes using targeted RNA interference (RNAi) knockdown in different germline cell types and whole body, we were not able to find direct evidence for their role in the X suppression. This study generated the first mutant lines with partial or complete disruption of X suppression and identified several promising targets for further functional testing.

# List of Abbreviations

| | |
|---|---|
| **ANCOVA** | analysis of covariance |
| **bp, kb, Mb** | base pair, kilo base, Mega base |
| **ChIP-chip** | chromatin immunoprecipitation followed by microarray hybridization |
| **ChIP-seq** | chromatin immunoprecipitation followed by high-throughput sequencing |
| **CV** | coefficient of variation |
| **DCC** | dosage compensation complex |
| **EMS** | ethyl methanesulfonate |
| **F1/2/3** | the first/second/third filial generation |
| **FDR** | false discovery rate |
| **G4** | G-quadruplex |
| **HAS** | high-affinity sites |
| ***INXS*** | *INvolved in X Suppression* (transgenic mutant line) |
| **MRE** | MSL recognition elements |
| **mRNA** | messenger RNA |
| **MSCI** | meiotic sex chromosome inactivation |
| **PBS** | phosphate-buffered saline |
| **PCR** | polymerase chain reaction |
| **qRT-PCR** | quantitative reverse transcriptase polymerase chain reaction |
| **RNAi** | RNA interference |
| **TAD** | topologically associating domains |
| **UAS** | upstream Activating Sequence |
| **UTR** | untranslated region |

# List of Figures

**List of Figures**

# List of Tables

# Chapter 1

## Introduction

The origin of sexual reproduction in the common ancestor of eukaryotes is one of the most significant steps in evolutionary history. Sexual reproduction is a process specific to eukaryotes in which two individuals equally provide genetic material for the development of progeny. It involves the fusion of haploid germ cells (syngamy) followed by meiosis and chromosome recombination. Although sexual reproduction cannot provide as rapid an increase in the number of offspring as asexual reproduction and is associated with inevitable costs, today almost all eukaryotic species reproduce sexually (Otto and Lenormand, 2002; Ligrone, 2019).

What advantages of sexual reproduction have ensured its prevalence in living organisms? As suggested over 100 years ago by August Weismann, sex provides greater genetic variability, so natural selection can work faster (Weismann et al., 1889). As a result of recombination, the offspring can obtain new combinations of alleles with favourable mutations in the new genetic background (Lane, 2019). On the other hand, sexual reproduction prevents the accumulation of harmful mutations, which, as described in the "Muller's ratchet" model, occurs over generations of clones in asexual populations (Muller, 1964). Another benefit of sexual reproduction was described by Bell in his now-classic "Red Queen hypothesis" for host-parasite systems (Bell, 1982). It suggests that the maintenance of a high diversity of genotypes by recombination allows the host population to resist (to stay in place by constantly running away) constantly adapting parasites to the common genotypes. In general, several conditions must be met for a population to benefit from sexual reproduction.

For example, the population must be of limited size and experience the effect of genetic drift, while being under spatially and temporally changing relatively strong selection (Hörandl, 2009; Ligrone, 2019).

Throughout eukaryotic evolution, various reproductive strategies, molecular mechanisms of reproduction, and behavioural patterns have emerged (Bell, 1982; Speijer et al., 2015). A minority of eukaryotes, mainly unicellular organisms, have isogamy, with phenotypically and, in some cases, genetically identical gametes. However, most organisms have developed anisogamy, where two types of gametes differ genetically and phenotypically (Perrin, 2012). Anisogamy appeared in different species independently due to disruptive selection to maximize the likelihood of mating, with one type of gamete loses motility and increases its size due to the accumulation of energetic resources, while the other type becomes highly motile without a large accumulation of resources. This results in a substantial increase in the number of male gametes produced and makes it easier for them to find female gametes, also through the development of a chemotaxis mechanisms (Bulmer and Parker, 2002). Furthermore, in the case of oogamy, the variation in the starting material for zygote development is reduced (Radzvilavicius et al., 2016). It is believed that the separation into two sexes (called gonochorism in animals and dioecy in plants) arose due to the oppositely directed selection of males and females (sexual antagonism) or as a mechanism to avoid self-fertilization and "inbreeding depression", that is, the accumulation of deleterious recessive alleles in the homozygous state, leading to a general decrease in the fitness of individuals (Falconer and Mackay, 1996; Delph, 2009; Leonard, 2018).

There are two major groups of sex determination systems: environmental and genetic, which include a great variety of forms (Mankiewicz et al., 2013; Holleley et al., 2015; Sabath et al., 2016; Capel, 2017; Pennell et al., 2018; Kossack and Draper, 2019). Environmental sex determination means that an organism retains phenotypic plasticity during its development, and under certain conditions, it switches to one of the sexes. It is observed in many cold-blooded organisms, such as fish, reptiles, molluscs and arthropods, and can be triggered by environmental temperature, social factors, photoperiodism and pathogens (Bouchon et al., 1998; McNair et al., 2015; Capel, 2017; Li and Gui, 2018; Geffroy and Wedekind, 2020). Genetic mechanisms for sex determination are no less widespread and are common in mammals, birds, reptiles, amphibians, insects, and many plants. Genetic factors include haploidy, specialized sex chromosomes, and one or several sex-determining loci (Bachtrog et

al., 2014; Capel, 2017; Li and Gui, 2018; Charlesworth, 2019; Coelho et al., 2019). In general, the evolution of sex determination is related to the environmental characteristics and sex specificity in male and female adaptation. Thus, in unpredictable environmental conditions or its high homogeneity, genetic sex determination will prevail. When conditions have very different effects on the adaptation of the different sexes, environmental sex determination will prevail (Charnov and Bull, 1977; Pennell et al., 2018).

## 1.1 Sex determination in *Drosophila*

Morphological and functional segregation of sex chromosomes evolved independently in different groups of organisms, so now we can observe various sex chromosome determination systems. In many insects, including *Drosophila*, mammals and some plants, females carry two X chromosomes (XX). In contrast, males carry one X chromosome and one Y chromosome (XY) or only one X chromosome (XO) (Bachtrog et al., 2014). In broad groups of vertebrates (birds, some reptilians, amphibians, fishes) and invertebrates (molluscs, crustaceans, arachnids, insects), females have heteromorphic sex chromosomes (ZW) or only one Z chromosome (ZO), while males have homomorphic sex chromosomes (ZZ) (Bachtrog et al., 2014). The primary triggers for subsequent sex determination signalling pathways include sex-determination alleles on the Y(W) chromosomes (humans and silkmoth), the ratio of sex chromosomes to autosomes (*Caenorhabditis elegans*, *Drosophila*), or the total number of sex chromosomes (birds) (Bachtrog et al., 2014).

In *D. melanogaster*, the sex determination signalling cascade begins with four X-linked signal elements, varying in expression level depending on the dosage of the X chromosome. Sufficiently high levels of these proteins, which can only be achieved normally in females, activate the *Sex-lethal* promoter, resulting in the formation of a functional Sxl protein, whose constant level in the cells is further maintained through a positive feedback mechanism (Salz and Erickson, 2010; Robinett et al., 2010). Sxl protein activates the *transformer* gene (*tra*), whose product acts as a splicing factor for the female-specific *doublesex* (*dsx*) and *fruitless* (*fru*) genes (Salz and Erickson, 2010; Robinett et al., 2010). These genes with *Sxl* regulate the development of major sex-specific characteristics (Demir and Dickson, 2005; Robinett et al., 2010; Suzuki, 2018). The *Sxl* gene also represses *male-specific-lethal-2* (*MSL2*) activity, thereby preventing the formation of the dosage

compensation complex (DCC) in females (Kelley et al., 1997; Gebauer et al., 1998; Salz and Erickson, 2010).

## 1.2 Evolution of sex chromosomes

Despite the diversity of genetic sex determination systems in different groups of organisms, sex chromosomes share common processes during their evolution. It is believed that X and Y (Z and W) chromosomes developed from initially homologous and recombining autosomes (Bull, 1983; Charlesworth et al., 2005). According to this classical theory, the proto-Y chromosome acquired a dominant male determining locus. At the same time, its homologue could obtain another allele, which triggers females' development when homozygous (e.g. *Sxl* in *D. melanogaster*). These sex-determining genes would affect selection on other closely linked genes and lead to the emergence of alleles that benefit one sex but are harmful to the other (sexually antagonistic genes) (Bull, 1983; Charlesworth et al., 2005). The sex-determining gene and sexually antagonistic mutations will lead to the loss of the recombination between homologous chromosomes (Charlesworth et al., 2005). The lack of recombination results in the accumulation of loss-of-function mutations, transposable elements, and repetitive DNA sequences and eventually in the erosion of genes on the Y chromosome over evolutionary time (Charlesworth et al., 2000; Bachtrog, 2013).

However, this classical Y chromosome evolution model is only partially applicable to *D. melanogaster* (Carvalho et al., 2009). The Y chromosome is almost entirely heterochromatic and contains only 13 protein-coding genes with no homology to the X chromosome (Vibranovski et al., 2008; Carvalho et al., 2015; Thurmond et al., 2019). All of the active genes found on the Y chromosome emerged on it relatively recently as a result of duplication from autosomal genes (Koerich et al., 2008). It also lacks the sex-determining gene, leading to the hypothesis that, due to further degradation of the Y chromosome, this gene, along with all the ancestral X chromosomal genes, has been lost, and the Y chromosome more recently gained new genes from autosomes (Bachtrog, 2013). However, according to another hypothesis, after complete loss of the Y chromosome, it was functionally replaced by supernumerary B chromosomes or the fourth chromosome (Carvalho et al., 2009).

## 1. Introduction

Despite its extreme paucity of genes, the Y chromosome is essential for male fertility. All Y chromosome genes have testis-specific expression, and six of them are male fertility factors (Carvalho et al., 2015; Hafezi et al., 2020). The Y chromosome's role extends beyond male reproduction and involves many other biological aspects, such as sex-specific ageing, regulation of gene expression on different chromosomes, and chromatin modification (Chippindale and Rice, 2001; Lemos et al., 2010; Brown et al., 2020).

At the same time, the X chromosome has not undergone such dramatic changes in chromatin structure and gene composition because it has a homolog for recombination in females. The X chromosome contains over 2,000 genes, which is proportionate to its size (about 16% of the genome). However, as a result of its monosomy in males, the X chromosome has evolved several differences from the autosomes in its regulation, gene content, and its role in speciation (Meisel and Connallon, 2013; Huylmans and Parsch, 2015; Grath and Parsch, 2016; Kuroda et al., 2016; Charlesworth et al., 2018).

### 1.3 Aneuploidy and generic dosage compensation in *Drosophila*

Aneuploidy, that is, the variation in the number of copies of chromosomes relative to the entire genome, causes genomic instability and severe disruptions in the cell and the entire organism, often causing lethality. Aneuploidy often triggers cell proliferation and transformation, leading to tumorigenesis (Gerlach and Herranz, 2020). However, it can also activate buffering mechanisms and mediate a stress response (*JNK* and *p53* pathways), thereby mitigating detrimental effects for the organism (Stenberg and Larsson, 2011; Lundberg et al., 2012).

Several levels of dosage compensation function in *Drosophila* to prevent the harmful effects of monosomy by balancing the expression of the X chromosome and autosomes in males and the single X chromosome in males with the two copies in females. First, the generic basal dosage compensation mechanism increases gene expression on the X chromosome in males or other monosomic regions of the genome by, on average, about 1.5-fold (Zhang et al., 2010a). This is a general form of buffering and gene-specific dosage compensation. It is thought that buffering is related to the passive property of the cell's complex biochemical environment to absorb changes in gene dosage (Zhang et al., 2010a). However, it depends on the size of the genomic region, the level of tissue specificity, and the

expression level of the particular target gene in the monosomal state (McAnally and Yampolsky, 2010; Lundberg et al., 2012). Gene-specific regulation operates at the transcriptional or post-transcriptional level and is based on a feedback system, often providing full compensation, especially for dose-sensitive genes (Zhang et al., 2010a). Gene-specific dosage compensation probably played a major role in mitigating the genomic instability that occurred during the evolution of gene composition on the Y chromosome (Gu and Walters, 2017).

Dosage compensation on autosomes with aneuploidy or gene copy number variation is most likely accomplished through the buffering mechanism and only, to a lesser extent, gene-specific regulation (Stenberg et al., 2009). The fourth chromosome differs from the other autosomes as it consists mainly of heterochromatin and only contains about 1% of the total genes. Still, the density of active genes and their expression level are comparable to those for autosomes (Riddle and Elgin, 2018). To balance its transcription levels with other chromosomes in both sexes, it has a unique dosage compensation mechanism mediated by the targeting protein Painting of fourth (POF) (Larsson et al., 2004). Interestingly, the genomic analysis of *Drosophila* and other species suggests that the fourth chromosome may have been the *Drosophila* genus' ancestral X chromosome (Riddle and Elgin, 2018).

## 1.4 X chromosome-specific dosage compensation

At a certain point in the progressive degeneration of the Y chromosome, a transition from compensation for individual genes to chromosome-wide dosage compensation may have become advantageous (Gu and Walters, 2017). This led to the appearance of X chromosome-specific somatic dosage compensation, which further increases the expression of X-linked genes in males to achieve complete two-fold compensation (Hamada et al., 2005; Deng et al., 2009; Zhang et al., 2010a).

Somatic dosage compensation is mediated by the X chromosome-specific binding of the DCC, which is composed of the protein products of three *male-specific lethal* genes (MSL1, MSL2, MSL3), the *maleless* gene (MLE), the *males absent on the first* gene (MOF), and two long non-coding RNAs (*roX1* and *roX2*) (Samata and Akhtar, 2018) (Figure 1.1). The DCC initially binds to about 250 genomic regions, called chromosomal entry sites or high-affinity sites (HAS) (Alekseyenko et al., 2008; Straub et al., 2013). All HAS are defined

by the binding of a core complex, which includes MSL1 and MSL2 (Dahlsveen et al., 2006; Alekseyenko et al., 2008; Straub et al., 2013). DCC activity in males is determined by the absence of the active SXL protein, which suppresses the expression of MSL2, the only male-limited component (Kelley et al., 1995; Beckmann et al., 2005).

Several key steps in the assembly of the complex can be distinguished. MSL1 forms a stable complex with MSL2, MSL3, and MOF followed by independent binding of MSL1 and MSL2 to HAS. MLE and MSL2 facilitate integrating *roX* RNAs to the DCC (Villa et al., 2016; Samata and Akhtar, 2018) (Figure 1.1). At each of the binding sites, MOF, as the functional component of the DCC, directs acetylation of histone H4 on lysine 16 (H4Ac16), which is enriched at X-linked gene bodies and promotes elongation of RNA polymerase II. This results in an open chromatin structure and hyper-transcription of genes in the exposed region (Ferrari et al., 2013; Kuroda et al., 2016). Additionally, DCC binding requires the CLAMP protein, which is not a DCC component and is ubiquitous in the genome of both sexes (Hallacli et al., 2012). The role of CLAMP is probably related to DNA release from the nucleosomes, enabling DCC binding (Urban et al., 2017).

## 1.5 Precise targeting and distribution of the DCC

The question that remains open is which molecular processes ensure the fine-tuning of DCC mediated dosage compensation. Therefore, it is crucial to study how the DCC recognizes the initial binding sites along the X chromosome and further spreads to neighbouring genes. The first contact with the X chromosome occurs at HAS sites, two of which carry the genetic loci *roX1* and *roX2* and have the highest priority for binding (Kelley et al., 1999; Meller et al., 2000). This contact is achieved by MSL2 and CLAMP proteins, which identify and interact with the MSL recognition elements (MRE), 21-base-pair GA-rich consensus motifs found at all HAS (Alekseyenko et al., 2008; Straub et al., 2008; Villa et al., 2012; Albig et al., 2019) (Figure 1.1).

However, MREs are not specific for the X chromosome, suggesting some other factors facilitate the X chromosome recognition. First, a small fraction of HAS carries a sequence adjacent to the 5' GA-rich motif, called PionX (Villa et al., 2016). Second, the X chromosome has an abundance of 1.688 repeats, which synthesize small interfering RNA (siRNA) (Alekseyenko et al., 2008). It was found that the 1.688 repeats also contribute to an increase

in the transcription of neighbouring genes and chromatin topology changes (Menon and Meller, 2012; Menon and Meller, 2015). It is also plausible that *roX* RNAs direct the DCC on the X chromosome and prevent its nonspecific targeting of heterochromatin, such as on the fourth chromosome or in chromocenters (Demakova et al., 2003; Figueiredo et al., 2014).

Once a functioning DCC is assembled, it spreads to nearby genomic regions by the activity of MSL1 and MSL3, which direct the complex from HAS to nearby genes since they have an affinity for the active gene sites (Kind and Akhtar, 2007; Larschan et al., 2007; Kim et al., 2010; Moore et al., 2010; Hallacli et al., 2012; Chlamydas et al., 2016) (Figure 1.1). It was shown that in the absence of functioning MLEs and *roX* RNAs, there was a disruption in the spread of the complex from HAS (Figueiredo et al., 2014). However, it is not evident whether DCC spreading on the X chromosome is achieved mainly by the actual relocation of DCC, or if it reaches other genomic sites due to the spatial organization of the chromatin, or if genomic sites recruit additional DCC.

The DCC distribution is highly dependent on the spatial architecture of chromatin, which has a hierarchical structure. At the highest level, active and inactive chromatin compartments further comprise topologically associating domains (TAD), which are self-organized contact-rich chromatin structures isolated by different types of insulator elements (Ramírez et al., 2015; Ulianov et al., 2016). The DCC binding sites span the entire X chromosome, with most HAS located in the gene-rich regions and at the TAD boundaries (Alekseyenko et al., 2012; Ramírez et al., 2015; Schauer et al., 2017). The location of HAS in these regions with high connectivity allows DCC to spread over long distances and promote its assembly at other X-linked sites with lower affinity (Ramírez et al., 2015; Schauer et al., 2017). However, the chromatin structure may also impose a limit on the maximum distance DCC can spread, which to some extent may be due to the local concentration of elements of the complex (Demakova et al., 2003; Dahlsveen et al., 2006; Wood et al., 2011; Sexton et al., 2012). Thus, the degree of compensation depends on the distance to the binding sites with the most substantial transcriptional activation of genes within 50 kb of the closest HAS (Schauer et al., 2017).

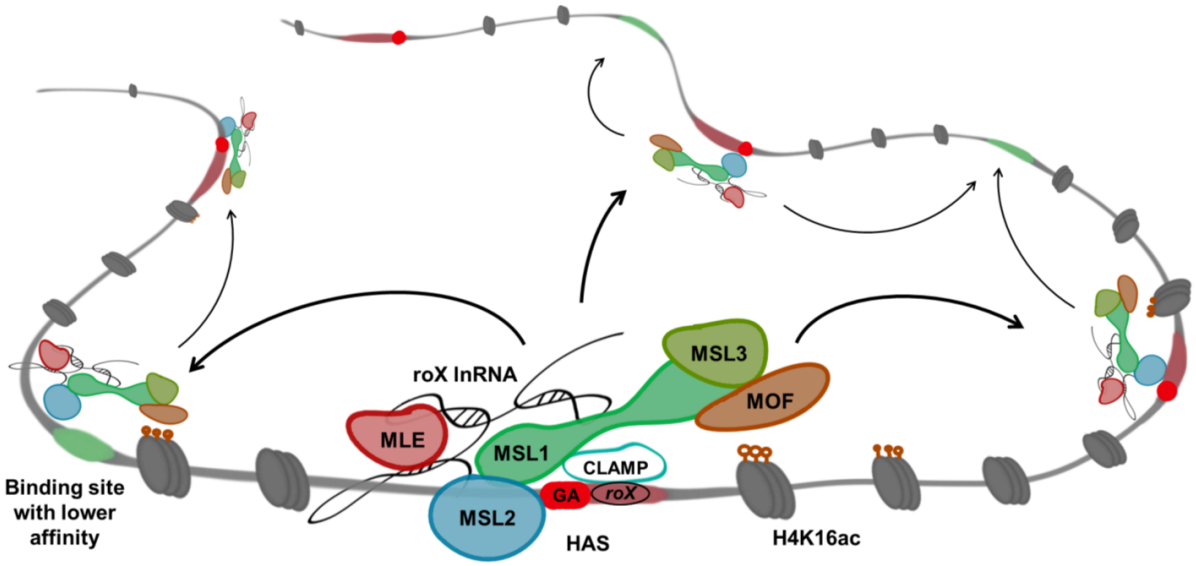


**Figure 1.1.** Schematic representation of the DCC and its distribution on the X chromosome (modified from Samata and Akhtar, 2018). The DCC, which consist of five proteins (MSL1, MSL2, MSL3, MLE and MOF) and two *roX* RNAs, recognizes and binds to HAS through the activity of MSL1 and MSL2. In this process, MSL2 interacts with MREs identified by a GA-rich motif (marked in red) assisted by the CLAMP. Two HAS having genes encoding *roX* RNAs have the highest priority for binding. At the binding site, MOF directs acetylation of H4Ac16, leading to hyper-transcription of genes in this region. After the initial assembly of a functioning DCC on HAS, it spreads to spatially close genomic sites with lower affinity (indicated by arrows). The spreading involves MSL1 (with affinity to gene promoter sites), MLE (helicase activity), *roX* RNAs and MSL3 (with affinity to the epigenetic marker of active genes H3K36me3) (Samata and Akhtar, 2018).

One of the potential factors constraining the DCC's distribution is G-quadruplexes (G4s) which are four-stranded secondary structures formed by guanine-rich DNA sequences (Burge et al., 2006; Bochman et al., 2012; Qian et al., 2019). G4s are enriched on gene promoters and are targeted by transcription factors and chromatin remodelling enzymes, suggesting their regulatory role (Siddiqui-Jain et al., 2002; Hänsel-Hertsch et al., 2016; Chen et al., 2018; Makowski et al., 2018). It was noted that the density of these structures on the X chromosome in the HAS and H4K16ac regions was significantly lower compared with the autosomes. In contrast, in the flanking regions, the density was higher than on the autosomes, suggesting a role of G4 as a limiter of DCC spreading (Qian et al., 2019).

## 1.6 Dosage compensation during embryonic development

Studies of gene expression changes during embryonic development provide further insights into the dosage compensation mechanisms and their effect on nearby genes' expression. During embryonic development of *D. melanogaster*, up to about stage 5 (cellularization of the blastoderm), mainly maternal transcripts are present, and the transcriptional activity of the zygote only begins to emerge (Tadros and Lipshitz, 2009). From this point onwards, females show balanced expression between sex chromosomes and autosomes, while males show a paucity of fully compensated genes, suggesting a lack of a functioning DCC (Paris et al., 2015; Prayitno et al., 2019). Complete dosage compensation gradually develops towards the end of embryonic development, which is associated with a progressive increase in the acetylation of H4K16 (Prayitno et al., 2019). The first genes to reach full compensation are the genes encoding components of the DCC and housekeeping genes that tend to be located close to HAS (Lott et al., 2011; Prayitno et al., 2019). This gradual increase in the degree of compensation is possibly due to a change in the chromatin structure and the gradual assembly of a functioning DCC (Franke et al., 1996; Prayitno et al., 2019). However, even from the early stages of embryo development, individual genes show close to full compensation, suggesting the activity of other dosage compensation mechanisms (Franke et al., 1996; Paris et al., 2015; Prayitno et al., 2019).

## 1.7 X chromosome suppression in the male germline

In contrast to the soma, X chromosome-specific dosage compensation does not occur in the male germline (Meiklejohn et al., 2011; Meiklejohn and Presgraves, 2012; Argyridou and Parsch, 2018). Thus, gene expression on the X chromosome is, on average, 1.5-fold lower than autosomal expression in testis, which is similar to the results obtained for somatic cells with a dysfunctional DCC (Hamada et al., 2005; Stenberg et al., 2009; Zhang et al., 2010a; Meiklejohn et al., 2011; Meiklejohn and Presgraves, 2012; Argyridou and Parsch, 2018). The male germline lacks expression of all DCC components except MLE, which does not bind to the X chromosome. However, despite this, gene expression is likely still compensated to some degree by general dosage compensation (Meiklejohn et al., 2011).

In the germline of mammals and many organisms with heteromorphic chromosomes, transcription from the X (Z) chromosome is blocked during meiosis by the phenomenon

called meiotic sex chromosome inactivation (MSCI) (Lifschytz and Lindsley, 1972; Bean et al., 2004; Cabrero et al., 2007; Hense et al., 2007; Vibranovski et al., 2009; Meiklejohn et al., 2011). MSCI is a specific case of meiotic silencing of unsynapsed chromatin, and its principal characteristic is a massive chromatin conformation change leading to the formation of completely heterochromatic sex chromosomes (Turner et al., 2005; Baarends et al., 2005).

Since MSCI is a widespread phenomenon in various taxa, ideas have arisen about its existence in *Drosophila* (Lifschytz and Lindsley, 1972). However, it was shown that the chromatin conformation of the X chromosome and the content of the active transcription markers on it do not differ significantly from autosomes during spermatogenesis (McKee and Handel, 1993; Rastelli and Kuroda, 1998). Studies using reporter genes with originally autosomal or X-linked testis-specific promoters showed that X-linked genes typically show three-to-eight times lower expression in the male germline than their autosomal counterparts. The suppression effect was chromosome-wide with no regions of the X chromosome escaping suppression (Hoyle et al., 1995; Hense et al., 2007; Kemkemer et al., 2011). Similar results were obtained in studies of transpositions and translocations, where the expression of both testis-specific and housekeeping genes was compared between the X chromosome and autosomes (Landeen et al., 2016). Such a high difference in expression could not be explained only by the absence of dosage compensation in the germline, and there is no evidence that the suppression effect is limited to meiotic cells only. Moreover, the degree of X suppression was dependent on the maximal expression level of the gene in testis (Landeen et al., 2016; Argyridou and Parsch, 2018). One of the possible causes of this expression effect is an oppositely directed compensation process functioning at the gene-specific level and associated with sequence motifs in promoter regions of X-linked testis-specific genes (Landeen et al., 2016).

Thus, it can be assumed that *Drosophila*'s X inactivation is an evolutionarily analogous mechanism to MSCI and shares similar functions, although it differs in the details of its molecular mechanism. MSCI might be advantageous in preventing accidental recombination of completely or partially non-homologous sex chromosomes during meiosis. Such recombination could lead to ectopic exchanges causing rearrangements and aneuploidy (McKee and Handel, 1993). The suppression of X-linked expression may also be important in avoiding the effect of "selfish genes" that may lead to segregation distortion and alter the sex ratio of the population (Namekawa and Lee, 2009; Meiklejohn and Tao, 2010). Finally, MSCI

could be a protective mechanism against the harmful effects of sexual antagonism (Wu and Yujun Xu, 2003). Sexual antagonism is defined by the presence of traits favourable for one sex and detrimental for the other. Therefore, genes associated with these traits may have a different optimum expression level in males and females (Rice and Chippindale, 2001; Ellegren and Parsch, 2007; Grath and Parsch, 2016). The presence of two copies in females results in the X chromosome spending twice as much evolutionary time in females than in males. This can lead to a conflict of optimal gene expression levels between the sexes and, depending on the degree of dominance, result in the accumulation of mutations on the X chromosome favourable to females but detrimental to males, or *vice verse* (Ellegren and Parsch, 2007; Fry, 2010; Parsch and Ellegren, 2013). Males, in this case, could down-regulate the expression of these genes during germ cell development to reduce their adverse effects (Wu and Yujun Xu, 2003). However, to what extent these functions can be attributed to the process of X suppression in *Drosophila* is not yet clear.

## 1.8 Sex-biased gene expression underlies sexual dimorphism

In addition to chromosome-wide differences in expression regulation between females and males in *Drosophila* and other species with sexual dimorphism, there are also many gene-specific differences in expression regulation between the sexes (Arnqvist and Rowe, 2005). These genes, known as sex-biased genes, probably explain most phenotypic differences between the sexes because the gene composition of males and females is the same except for the Y chromosome, which carries very few genes in many species (Ellegren and Parsch, 2007).

In *Drosophila*, genes with consistent differential expression between the sexes represent a large proportion of autosomal and X-linked genes (Gnad and Parsch, 2006; Grath and Parsch, 2016). Depending on the statistical approach and confidence interval levels, the proportion of sex-biased genes can vary from 30% to more than 60% of the total number of genes in *Drosophila* (Gnad and Parsch, 2006; Grath and Parsch, 2016). Such a substantial number of genes most likely requires the coordinated interaction of several expression regulation levels, both at the level of transcription, through *cis- and trans-* factors, epigenetic modification and chromatin conformation changes, and at the level of post-transcriptional precursor messenger RNA (mRNA) modification and translation. Moreover, the sex-biased expression of the same gene can vary depending on environmental and biological factors,

such as tissue type, developmental stage of the organism, and its genotype (Jin et al., 2001; Arbeitman et al., 2002; Wyman et al., 2010; Mank et al., 2010; Huylmans and Parsch, 2015). Therefore, determining the mechanisms involved in such extensive expression changes is by no means a trivial task.

Very high variation in the abundance of genes with sex-biased expression and the amplitude of male-to-female expression ratio can be seen among different tissues, which is partially related to the degree of sexual dimorphism in the tissue (Andrews et al., 2000; Parisi et al., 2004; Chang et al., 2011; Assis et al., 2012; Catalán et al., 2012; Perry et al., 2014; Khodursky et al., 2020). Moreover, in different tissues, sex-biased genes are not equally distributed among the X chromosome and the autosomes. The X chromosome is enriched for genes with female-biased expression ("feminization") in whole flies and various somatic tissues, including brains, heads, Malpighian tubules, and gonads (Huylmans and Parsch, 2015). In contrast, a significant paucity of genes with male-biased expression ("demasculinization") has been observed in studies of whole flies and gonads (Parisi et al., 2003; Ranz et al., 2003). This demasculinization pattern does not hold for all tissues, however, as a significant over-representation of male-biased genes has been reported in studies of brain and head expression (Chang et al., 2011; Catalán et al., 2012; Huylmans and Parsch, 2015; Khodursky et al., 2020).

## 1.9 Evolution of sex-biased gene expression

Sexual antagonism is probably the driving force behind this unequal distribution of sex-biased genes among chromosomes, since different chromosomes can contribute unequally to sexually conflicting traits (Rice, 1984; Patten, 2018). This can lead to the accumulation of recessive male-beneficial mutations on the X chromosome, which will be immediately exposed to selection in hemizygous males. In females, on the contrary, homozygous X chromosomes will mask recessive beneficial mutations and favour the accumulation of dominant female-beneficial mutations on the X chromosome, since it spends twice as much time in females compared to males (Connallon and Knowles, 2006; Dean et al., 2012). Evidence of the involvement of sexual antagonism in the evolution of the X chromosome can be inferred from its enrichment with loci showing sexually antagonistic expression (Rice, 1984; Gibson et al., 2002; Fry, 2010; Innocenti and Morrow, 2010; Ruzicka et al., 2019).

In addition, the evolution of sex-biased expression could be driven by such gene properties as gene age and expression breadth (Waxman and Peck, 1998; Larracuente et al., 2008). Most X-linked male-biased genes are evolutionarily young genes that could be related to the sexual antagonistic effect. In contrast, on autosomes, male-biased genes are often older (Zhang et al., 2010b; Long et al., 2012). Expression breadth is thought to serve as a constraint for the evolution of sex-biased expression, as it makes it difficult to find a balance in the expression of genes with high pleiotropy (Mank et al., 2008; Perry et al., 2014; Dean and Mank, 2016; Campos et al., 2018; Khodursky et al., 2020). Many male-biased genes are expressed specifically in the testis, the tissue where the highest number of male-biased genes have been found (Mikhaylova and Nurminsky, 2011; Meisel et al., 2012). It has also been shown that there is a lack of genes with narrower expression breadth on the X chromosome (Mikhaylova and Nurminsky 2011; Meisel et al. 2012; but see Campos et al. 2018).

In addition to the specific properties of X-linked genes, the X chromosomal regulatory mechanisms can also affect sex-biased expression. It has been hypothesized that the enrichment of sex-biased genes on the X chromosome in the brain and head is related to dosage compensation, with genes located near DCC binding sites having greater up-regulation of expression in males than genes located far from DCC binding sites (Huylmans and Parsch, 2015) (Figure 1.2). Consistent with this model, a negative correlation between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site has been observed for the brain and head. In whole flies or gonads, which typically show a much greater degree of sex-biased expression, a positive correlation has been observed, suggesting that gene-specific regulation is the predominant driver of sex-biased expression and that DCC binding may interfere with sex-specific regulation (Bachtrog et al., 2010; Huylmans and Parsch, 2015). Furthermore, it has been proposed that the head and brain may be more sensitive to dosage compensation than other tissues, as they show higher expression of the DCC components MSL2 and MLE than other tissues (Straub et al., 2013; Vensko and Stone, 2015; Huylmans and Parsch, 2015). Thus, endogenous gene expression studies do not allow us to disentangle the influence of X chromosome-wide regulatory mechanisms and the local chromatin environment from gene-specific regulation evolutionarily related to gene properties such as tissue specificity, age, and their effect on male and female fitness.

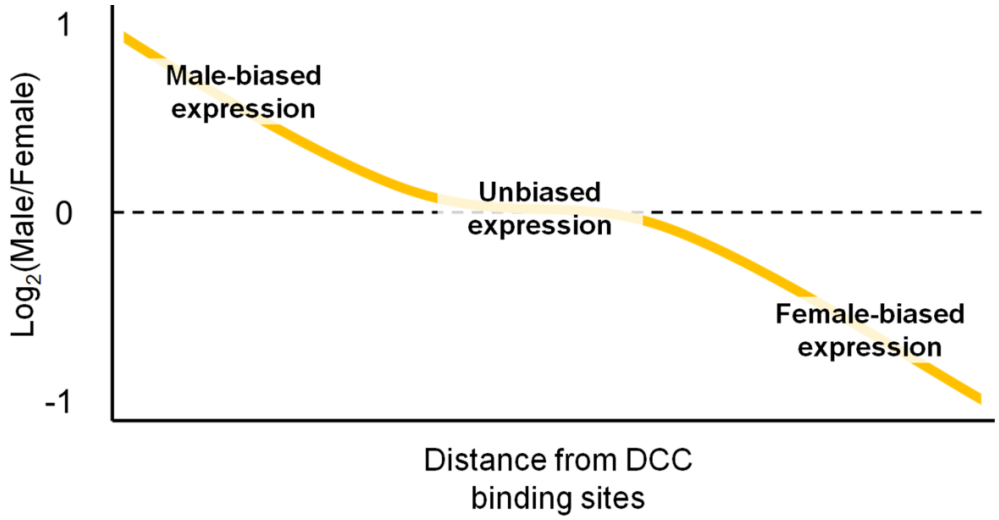


**Figure 1.2.** An illustration of the overcompensation model for brains and heads (Huylmans and Parsch, 2015). The X chromosomal position of genes closer to DCC binding sites results in their male-biased expression (shown as $\log_2$ of the male/female expression ratio) due to their up-regulation in males by more than 2-fold. Genes located further away from the DCC binding sites show a weaker level of up-regulation (< 2-fold), resulting in their female-biased expression.

## 1.10 The role of *Drosophila* as a model organism in genetic research

*Drosophila* is one of the most popular model organisms in modern biology. Besides such advantages of using *Drosophila* in genetics as the convenience and rapidity of breeding, the absence of meiotic recombination in males, the compactness of the genome and its similarity to the mammalian genome, a wide variety of classical genetic tools have been created over a hundred years of research on this model organism. These include marker and reporter genes, cytogenetic maps, gene cloning, transposable element vectors, forward and reverse genetic screens (Ashburner, 2005). Other commonly used *Drosophila* tools are balancer chromosomes, specially designed chromosomes with multiple inversions to prevent crossing over with additional chromosomes. Balancer chromosomes also often include recessive lethal alleles and dominant marker genes (Ashburner, 2005). Additionally, recent advances in genomic sequencing and the annotation of many functional elements of *Drosophila* have allowed the study of evolutionary processes, development and functioning of the organism at the genetic level (Rubin, 2000).

Reporter genes are a standard and frequently used tool in *Drosophila* genetic studies, both *in vitro* and *in vivo*. They are used to study the regulation of transcription, mRNA

processing, and translation by detecting their products directly or indirectly when the genomic environment, cellular processes, or external conditions are changed (Serebriiskii and Golemis, 2000). The *lacZ* reporter gene encodes β-galactosidase, an enzyme that degrades β-glycosidic linkages in D-lactose. It is most commonly used in the enzymatic reaction with 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-gal) or o-nitrophenyl-b-D-galactopyranoside, resulting in 5,5'-dibromo-4,4'-dichloro-indigo (blue product) or o-nitrophenol (yellow) respectively (Serebriiskii and Golemis, 2000). The main advantages of using the *lacZ* system in *Drosophila* studies is that there is no background expression of this gene, and cells can tolerate very high concentrations of β-galactosidase, so it is possible to detect extreme changes in transcription. Also, the reaction's expression products are very stable and can be easily detected by eye or spectrophotometrically in homogenized samples or using a microscope in *Drosophila* tissues, which is especially important in screening experiments (Serebriiskii and Golemis, 2000).

Often, reporter genes are used as part of transposable *P* elements that do not have a sequence encoding transposase enzyme but can themselves be mobilized by a second *P* element, which in contrast is responsible for encoding transposase but cannot be displaced. Such a reporter gene construct contains a marker gene to detect the construct's presence in the genome and may also contain a promoter (Cooley et al., 1988). This system makes it possible to quickly obtain new construct insertions using genetic crosses that can be used for studying the influence of the genetic environment on expression.

There are two main approaches to discovering new genes and matching them with certain biological processes in *Drosophila* research. The first is a forward screen, in which flies are mutagenized using ethyl methanesulfonate (EMS), X-rays, or *P* element insertions (Lewis and Bacher, 1968; Gao et al., 1999; Spradling et al., 1999). Next, a series of crosses with a fly line carrying balancer chromosomes can be carried out to obtain offspring with homozygous and heterozygous mutant chromosomes, which will allow the identification of both dominant and recessive mutations. Such flies are screened for the phenotype of interest using tests of reporter gene expression, developmental process, or behaviour followed by mapping of causal mutations (Nüsslein-Volhard and Wieschaus, 1980; Singh and Heberlein, 2000; Brenman et al., 2001; St Johnston, 2002). Depending on the size of the genome, the biological process under study, and the degree of redundancy of the genes responsible for it, the number of screened genomes needed to identify the target phenotype can range from

several hundred to tens of thousands (Gao et al., 1999; Singh and Heberlein, 2000; Neukomm et al., 2014; Deal and Yamamoto, 2019; Gevedon et al., 2019).

The UAS/Gal4 system is often used for misexpression and reverse screens. A misexpression screen utilizes a *P*-element transformation vector (EP element) that carries an Upstream Activating Sequence (UAS). UAS is a 17 base pairs (bp) sequence in the yeast *Saccharomyces cerevisiae* genome that activates *GAL10* and *GAL1* genes' transcription. UAS, in turn, is a specific target for the yeast transcription factor, Gal4, which can exhibit its activity in various organisms as a part of the transgene construct (Duffy, 2002). When activated by Gal4 in a particular tissue or a particular developmental stage, UAS induces nearby genes' expression, making it one of the most widely used research tools (Rørth, 1996) (Figure 1.3). Another approach might be investigating the functions of already known candidate genes or targeted gene deactivation of a broad set of genes. In this so-called reverse screen, it is possible to turn off the gene of interest in the target tissue by taking advantage of RNA interference (RNAi), followed by a phenotype test (Brand and Perrimon, 1993) (Figure 1.3).

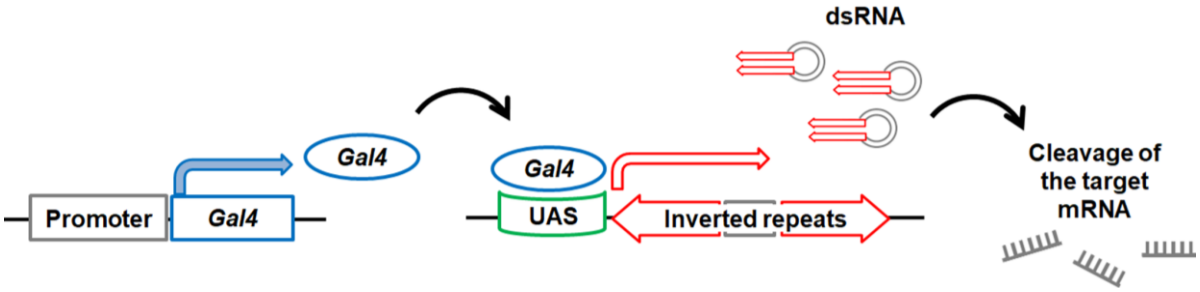


**Figure 1.3.** The UAS/Gal4 expression system used to suppress the expression of the gene of interest by RNAi (modified from Nishihara, 2007). A so-called Gal4 driver fly line expressing *Gal4* in the target tissue under the control of an endogenous promoter is crossed with a line carrying a UAS and a downstream short inverse repeat sequence corresponding to the gene of interest. In the F1, after being combined with Gal4, the UAS activates the expression of the inverse repeat sequence in the target tissue. Expression of the inverse repeat leads to the formation of double stranded RNA (dsRNA) hairpin structures. After a series of biochemical reactions, single stranded RNAs are formed, which are incorporated in the RNA-induced silencing complex (RISC). RISC cleaves the mRNA of the target gene, leading to its degradation in a spatiotemporal manner (Duffy, 2002; Kavi et al., 2008).

In my dissertation, I leveraged classical *Drosophila* research methods such as a genetic screen and reporter gene expression analysis by combining them with new genome mapping and expression level measurement techniques such as next-generation sequencing. This allowed me to carry out a large-scale reporter gene analysis to study the two regulatory mechanisms on the X chromosome (dosage compensation and X suppression) in a relatively short amount of time.

## 1.11 Aims and objectives

The first aim of my thesis was to test the effects of the chromosomal context on the expression of X-linked genes in *D. melanogaster* males and females. The evolution of sex chromosomes in *Drosophila* has led to specific regulatory mechanisms on the X chromosome, creating a unique environment for gene expression across tissues and shaping the chromosome's gene content (Ellegren and Parsch, 2007).

My first objective was to test the effect of a reporter gene's proximity to a DCC binding site on its sex-biased expression in somatic tissues. It has been hypothesized that in the brain and head, genes located near DCC binding sites have greater male-biased expression than genes located far from DCC binding sites as a result of "overcompensation" (Huylmans and Parsch, 2015) (Figure 1.2). Consistent with this model, a negative correlation between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site has been observed for the brain and head (Huylmans and Parsch, 2015). Because previous studies of the effect of X-chromosomal location on gene expression were limited to endogenous genes, it was not possible to disentangle the influence of chromosomal location from that of gene-specific regulation. To overcome this limitation, we inserted an exogenous reporter gene at many unique locations across the X chromosome. The reporter gene consisted of the *lacZ* reporter gene under the control of the minimal human cytomegalovirus (CMV) promoter, which was shown to drive expression in multiple *D. melanogaster* tissues (Parsch, 2004). By measuring reporter gene expression in both sexes and multiple tissues, we could determine the effects of chromosomal context on gene expression, while avoiding the effects of gene-, tissue-, and sex-specific regulation that are common to endogenous genes.

My second objective was to compare sex-biased reporter gene expression between the brain, head, and other somatic tissues. Since previous studies have shown that the brain has an

excess of both male-biased and female-biased genes, as well as a strong negative correlation between the distance of a gene to the nearest DCC binding sites and its male-to-female expression ratio, it was suggested that the brain differs from other somatic tissues with regard to the effects of DCC binding on expression (Huylmans and Parsch, 2015; Khodursky et al., 2020).

The third objective was to examine the possible influence of endogenous regulatory elements on reporter gene expression. Firstly, as all individual insertions were located near or within the active gene bodies, we tested whether there might be an influence of local regulatory factors on the reporter gene expression. For that, we compared the male-to-female expression ratio of the reporter genes with that of the endogenous genes in which they were located. Secondly, because the primary analysis of sex-biased expression was performed using flies of both sexes with a single copy of the reporter gene, we also examined the effect of gene dose by measuring expression in homozygous females. Thirdly, we tested the possible effect of G4s, which presumably have a local function in transcriptional regulation, on reporter gene expression (Qian et al., 2019).

The second aim of my thesis was to discover genes responsible for the chromosome-wide suppression of X-linked genes in the *D. melanogaster* male germline. In many species, the X chromosome is silenced during spermatogenesis through a mechanism known as MSCI (Lifschytz and Lindsley, 1972; Bean et al., 2004; Cabrero et al., 2007; Hense et al., 2007; Vibranovski et al., 2009; Meiklejohn et al., 2011). However, there is no evidence in *Drosophila* that X suppression is associated with global chromatin modification and/or silencing of the whole X chromosome as is seen in other species, which suggest that there is an alternative mechanism operating in *Drosophila* (McKee and Handel, 1993; Rastelli and Kuroda, 1998; Landeen et al., 2016). In *Drosophila*, the expression of many X-linked genes is suppressed in the male germline by a poorly understood chromosome-wide mechanism, through which the strength of suppression is dependent on the gene's expression level in testes (Landeen et al., 2016; Argyridou and Parsch, 2018).

My first objective was to perform chemical mutagenesis of males carrying an X-linked copy of the *ocn-lacZ* (*wol*) reporter gene. This reporter gene has previously been shown to have a seven-fold decrease in expression when located on the X chromosome as compared to an autosome, making it a suitable indicator of the presence of the X suppression (Hense et al.,

2007; Kemkemer et al., 2011). Using a forward genetic screen, we attempted to find mutant lines with increased expression of this reporter gene, which could be further tested for the general expression up-regulation of X-linked endogenous genes in males.

The second objective was to determine the genomic location of all induced mutations associated with the observed mutant phenotype. This could be done by performing genetic crosses and whole-genome re-sequencing, followed by selecting the top candidate mutations within annotated genes. To verify the results obtained from genome sequencing, we performed additional genotyping to test the association of the candidate mutations with the increased expression of the reporter gene. Further, we performed a functional analysis of the top candidate genes using targeted gene expression knockdown. The identification of candidate genes that are potentially involved in X suppression is the first step in understanding the genetic basis of this previously unexplored chromosome-specific regulatory mechanism.

# Chapter 2

# Materials and Methods

## 2.1 Reporter gene constructs and *D. melanogaster* strains

To test for the effect of a gene's proximity to a DCC binding site on its expression, we generated transgenic lines with individual reporter gene insertions that span the X chromosome. We used a *CMV-lacZ* reporter gene construct, which consists of two copies of the *Escherichia coli lacZ* gene with the minimal human cytomegalovirus (CMV) promoter (Parsch, 2004) (Figure 2.1*A*). Having two copies of these elements provides a higher level of expression, which is particularly important when analyzing tissues with small size and low overall expression, such as testes. Both the reporter gene and the promoter are foreign to the *Drosophila* genome. Therefore, reporter gene expression is not expected to be influenced by sex-, or tissue-specific regulatory factors, making it possible to compare the expression in different tissues in males and females. The construct is flanked by the terminal repeat sequences of a *P*-element. The *P*-element repeats are recognized by the transposase, the enzyme that induces excision of the construct and its insertion into a new random genomic location. To detect an insertion in flies during crosses and to maintain the fly stocks, the *D. melanogaster mini-white* eye colour marker gene was used. To generate new fly lines with insertions of the reporter gene on the X chromosome, we started with previously obtained transgenic flies carrying an X-linked insertion at the following positions: 3,451,351 (Line ID: *X20*), 14,222,240 (Line ID: *X25*) and 6,998,869 (Line ID: *X27*) (Argyridou and Parsch, 2018). The transgenic line *yw; Δ2-3, Sb/TM6,* which contained a source of transposase (*Δ2-3*) on the third chromosome linked to the stubble bristles (*Sb*) phenotypic marker, was used to mobilize

the insertion into new genomic sites. A line containing no reporter gene, with a *yellow*, *white* (*yw*) background, was used as a negative control for measuring β-galactosidase activity.

A forward genetic screen was carried out using a fly line with the X-linked *P[wFl-ocnlacZ]* (*wol*) reporter gene construct generated in a previous study, with the genomic position at 16,677,891 (*wol20X*) (Hense et al., 2007). This construct contains a testis-specific promoter of the *ocnus* (*ocn*) gene, the *mini-white* gene as a selective marker and the *lacZ* reporter gene (Figure 2.1*B*). The *ocn* promoter, as previously shown, induces a high level of expression in testes and the reporter gene displays the X chromosome suppression effect with an average 7-fold decrease in expression when it is located on the X chromosome compared to the autosomal locations. Two other lines with insertions on the X chromosome, one with an undetermined genomic position (*wol5X*) and one at position 7,231,447 (*wol12X*) were used for the additional crosses of mutant lines in the mapping of causative mutations (see the additional cross 2 in section 2.6). As control of β-galactosidase activity in testis without the X-suppression effect, we used an additional fly line with the same reporter gene construct located on the third chromosome (*wol16*). After carrying out chemical mutagenesis (see section 1.10), to establish and maintain mutant fly stocks, we performed several genetic crosses with a line carrying four balancer autosomes with different phenotypic markers (*yw; CyO/ScO; Ubx/Sb*). Individual copies of the mutant lines were also maintained using the *FM7i* balancer X chromosome with the dominant mutation in the gene *Bar* as a phenotypic marker.

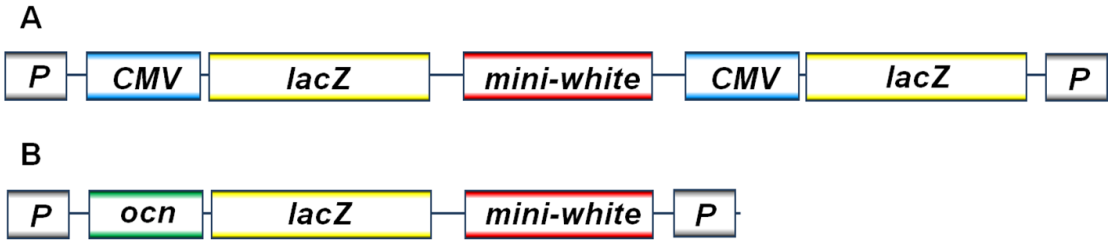


**Figure 2.1.** Schematic diagram of the reporter gene constructs. Schematic diagram of (*A*) the *CMV-lacZ* reporter gene construct which contains two copies of the *E. coli lacZ* gene fused to the CMV promoter flanked by the terminal repeat sequences of a *P* transposable element. The *D. melanogaster mini-white* gene is included as a visible eye colour marker gene. Schematic diagram of (*B*) the *wol* reporter gene construct containing the testis-specific *ocnus* promoter, the *lacZ* reporter gene and the *mini-white* eye colour marker gene. *P*-element terminal repeats flank all elements of this construct.

## 2. Materials and Methods

Functional tests of candidate genes were performed by crosses between flies carrying specific UAS-RNAi and Gal4 insertions. In the progeny carrying both of these elements, Gal4 activity enabled controlled activation of the UAS in various tissues. Three UAS-RNAi lines provided by the Vienna *Drosophila* Resource Center (VDRC, Vienna, Austria) were used, in which the UAS is fused with the inverse repeat sequence of one of the three target genes: i) *CG13003* (VDRC ID: *107843*. Second chromosome insertion; referred to as *RNAi-CG13003*), ii) *CG1314* (VDRC ID: *41502*. Third chromosome insertion; referred to as *RNAi-CG1314*), iii) *CG31525* (VDRC ID: *107890*. Second chromosome insertion; referred to as *RNAi-CG31525*) (Dietzl et al., 2007). For RNAi tests, Gal4 lines available in the group of Prof. Parsch were used in which Gal4 was under the control of four different promoters: i) *Bam*, which drives Gal4 expression in late spermatogonia and early spermatocytes (Line ID: *UAS Dicer1; Bam-Gal4/TM3; Sb*. Third chromosome insertion. Referred to as *Bam-Gal4*); ii) *Tj*, which drives *GAL4* expression in cyst stem cell and cyst cells (Line ID: *Tj-Gal4*. Third chromosome insertion. Referred to as *Tj-Gal4*); ii) *Nanos*, which drives Gal4 expression in germline stem cell and spermatogonia (Line ID: *nanos-Gal4*. Second chromosome insertion. Referred to as *Nanos-Gal4*); iv) *Act5C*, which drives ubiquitous Gal4 expression (Line ID: *Act5c-Gal4/CyO*. Second chromosome insertion against *CyO* balancer. Referred to as *Act5C-Gal4*) (White-Cooper, 2012; Yu et al., 2016). For each RNAi/Gal4 test, we used control lines containing a specific Gal4 with a background of each UAS-RNAi line but without the inverse repeat of a part of the target gene sequence (line with VDRC ID: *60100* for *RNAi-CG13003* and *RNAi-CG31525*, and the line with VDRC ID: *60000* for *RNAi-CG1314*). The absence of the inverse repeat sequences means that there is no hairpin RNA expression to bind and degrade the target gene's mRNA.

All fly stocks were kept on a standard cornmeal-molasses food at a constant temperature of 21 °C with a 14/10 light/dark cycle. For quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis, genetic crosses and fly ageing were performed at 25 °C with a 14/10 light/dark cycle.

## 2.2 Generation of new insertions of the *CMV-lacZ* reporter gene and the forward genetic screen using transgenic flies carrying the *wol* reporter gene

### 2.2.1 Mobilizing the *CMV-lacZ* reporter gene to new chromosomal locations

Each of the new insertions was obtained by mobilization following the genetic crossing scheme described by Hense et al. (2007). This crossing scheme includes several main steps described below (Figure 2.2*A*). First, we crossed the initial transgenic females with the X-linked reporter gene, which is marked by the *mini-white* gene, to *yw; Δ2-3, Sb/TM6* males, marked by the *Stubble* (*Sb)* phenotype. This cross was performed using 2–3 males and 2–3 virgin females per vial, and all subsequent crosses were performed using one male and 2–3 virgin females per vial in parallel to generate a large number of replicates. Male offspring with red eyes and stubble bristles were crossed to females of the *yw* strain. If in the subsequent generation red eyes and stubble bristles were consistently inherited together, we could assume that both the transposase coding gene and our reporter gene construct were located on the third chromosome. Males with this phenotype were used for further crosses with *yw* females to mobilize the transgene from the third chromosome marked by *Sb*. In this case, we collected offspring that have red eyes but wild-type bristles. Flies with this phenotype carry new X-linked or autosomal insertions and can be used for further mapping. To identify the flies carrying the transgene on the X chromosome, crosses were made between transgenic males with red eyes and *yw; Δ2-3, Sb/TM6* females. If all female offspring, but no male offspring, inherit the *mini-white* marker, it would indicate that the male has an insertion on the X chromosome (Figure 2.2*B*).

## 2. Materials and Methods

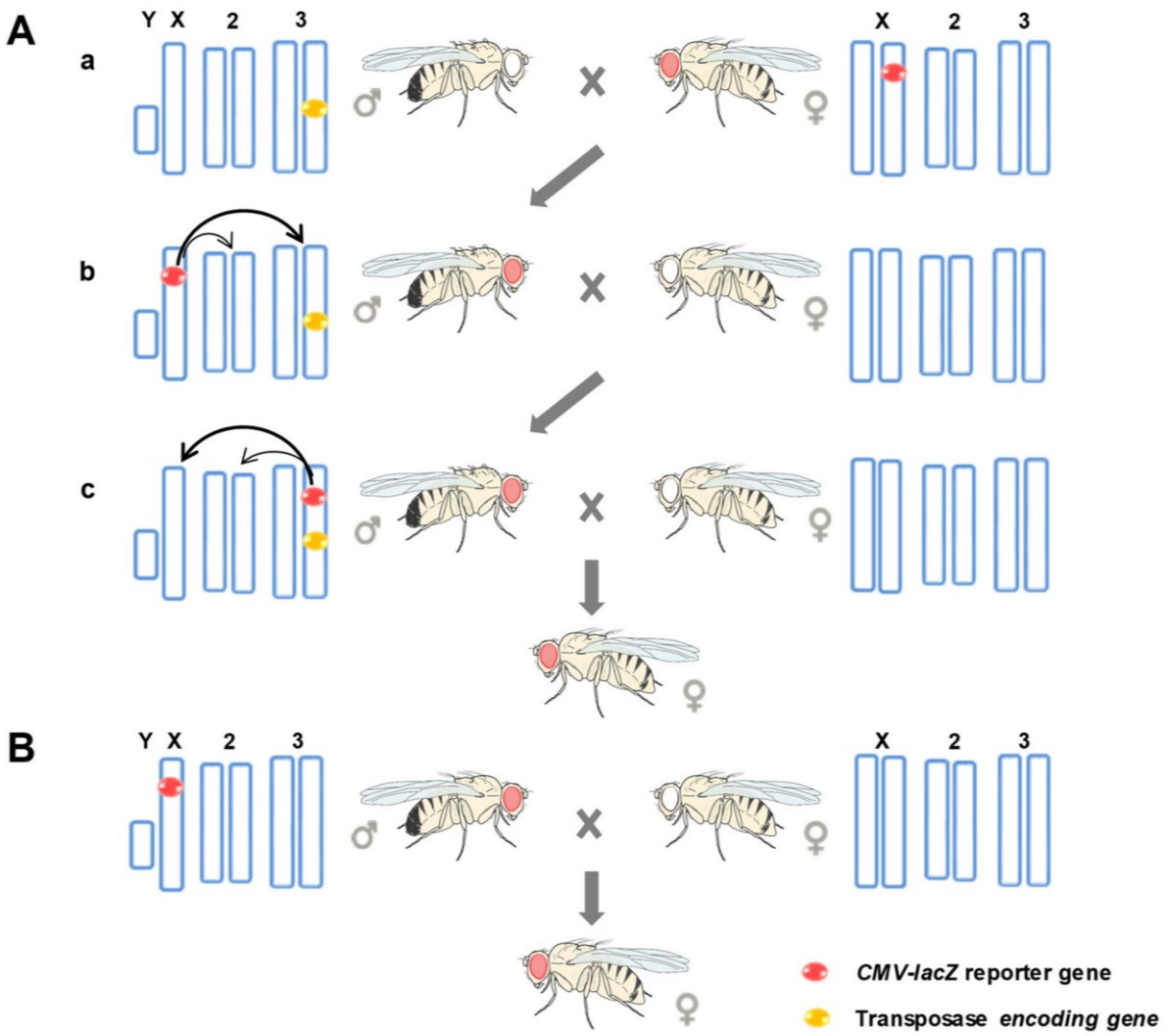


**Figure 2.2.** Fly crosses for mobilizing and mapping reporter gene insertions. (*A*) Schematic diagram of crosses designed to generate new X-linked insertions of the CMV-lacZ reporter gene. (*a*) Crosses between males carrying the original insertion on the X chromosome and females carrying the transposase encoding gene. (*b*) Due to transposase activity in the germline of the first filial generation (F1) males, the reporter gene was mobilized to a new genomic position, including a third chromosome, which carried the transposase encoding gene. (*c*) In the germline of the second filial generation (F2) males, the insertion could be mobilized away from the third chromosome. In this case, only the red-eyed marker was observed in the third filial generation (F3) females. (*B*) Schematic diagram of the mapping crosses. The selected males were mated with *yw* females. The inheritance of the red eye marker exclusively by females, whereas all males had white eyes was indicative of an X-linked reporter gene insertion.

### 2.2.2 Mutagenesis and genetic crosses of flies with the X-linked *wol* transgene

Forward mutagenesis was carried out using EMS, which has been shown to cause mainly point mutations at random DNA sites (Pastink et al., 1991; Blumenstiel et al., 2009; Phadnis et al., 2015). For this, newly-emerged *wol20X* males were aged for one day and, after starvation for 6 hours, placed in vials with a cotton pad soaked in a 6% sucrose solution containing 25 mM EMS overnight. After a period of recovery on a standard medium for one day, males carrying random mutations were crossed with virgin females of the *yw; CyO/ScO; Ubx/Sb* line. One male and 1–2 females were crossed per vial with about 50 replicates for each mutagenesis round. The flies were allowed to lay eggs for about one week and were then placed in new vials. Because EMS has the highest effect on mature spermatozoa and spermatids and the period of renewal of sperm in testes is about a week, all offspring in these vials are expected to have unique random mutations.

From each vial, 5–20 virgin females of the F1 generation containing one mutagenized copy of the X, second and third chromosomes were crossed to the *yw; CyO/ScO; Ubx/Sb* males with one female and 2–3 males per vial (Figure 2.3). Thus, all offspring will contain only one variant of the mutant chromosomes. Some of the red-eyed males of the F2 generation that were hemizygous for the mutant X chromosome and, when available, heterozygous for the mutant second and third chromosomes were preliminarily assayed for β-galactosidase activity. The remaining flies of this generation with different combinations of mutant and balancer chromosomes were randomly crossed with each other by transferring them into new vials *en masse* (Figure 2.3). Male offspring with the red-eye marker that were hemizygous for the mutant X chromosome and homozygous for both mutant autosomes were tested for β-galactosidase activity. Therefore, it may have been possible to detect recessive mutations affecting X suppression on each chromosome. However, often we could not find males with homozygous second or third autosomes (or both), in which case males with at least one homozygous or heterozygous mutagenized autosome were tested.

## 2. Materials and Methods

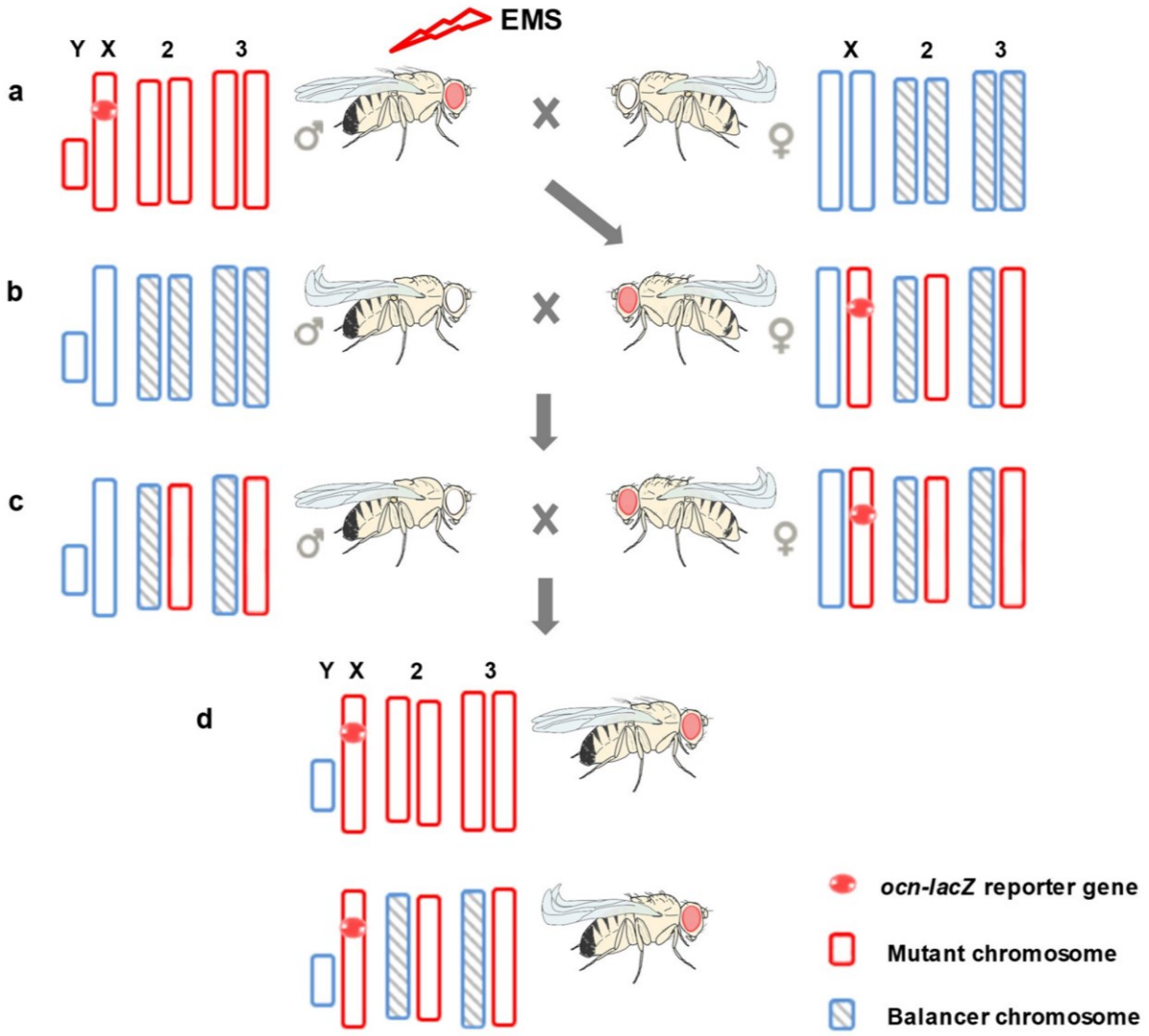


**Figure 2.3.** Schematic diagram of genetic crosses for mutagenesis screening. (*a*) Cross between males with the X-linked *wol* reporter gene subjected to EMS mutagen and *yw; CyO/ScO; Ubx/Sb* females. (*b*) Males from the *yw; CyO/ScO; Ubx/Sb* line were crossed with mutant females from the F1 generation carrying the *wol* reporter gene, second (*CyO* or *ScO*) and third (*Sb* or *Ubx*) balancer chromosomes. (*c*) F2 offspring with different combinations of mutant and balancer chromosomes were allowed to mate randomly. However, we expected that males with a mutant X chromosome could be sterile. (*d*) F3 male offspring in each mutant line with different combinations of homozygous or heterozygous autosomes and a mutant X chromosome were tested for β-galactosidase activity.

## 2.3 Mapping insertion locations by an inverse PCR approach

To find the exact chromosomal location of each reporter gene insertion, we used an inverse polymerase chain reaction (PCR) technique (Bellen et al., 2004). The whole bodies of five males and four females were homogenized, and DNA was extracted with MasterPure DNA Purification Kit (Epicentre, Madison, WI, USA) (Protocol C1, Appendix C). DNA concentration and quality were confirmed with a NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA). A restriction digest was then performed using the *HinP*I or *Hpa*II enzymes (Protocol C2) with subsequent self-ligation with T4 DNA ligase (New England Biolabs, Ipswich, MA, USA)(Protocol C3). The DNA fragments containing the transgene and the adjacent unknown sequence were PCR-amplified from the *pP[wFl]* transformation vector with primers: Plac1-Plac4 and EY.3.F-EY.3.R (Table B1, Protocol C4). PCR products were used for agarose gel electrophoresis to identify the number of insertions. The subsequent sequencing of PCR products was performed using BigDye v1.1 chemistry on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA) with two primers: Sp1 and EY.3.F (Table B1, Protocol C5). Sequencing was run on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA) by the Genomics Service Unit (LMU Munich, Faculty of Biology, Division of Genetics, Planegg-Martinsried, Germany). The genomic locations of the insertions were determined by a BLAST search (Altschul et al., 1990) using the genomic sequences flanking the transgene as the query and the *D. melanogaster* genome (Release 6.31) as the reference.

## 2.4 Characterizing the genomic environment of reporter gene insertions

The locations of DCC component binding sites (MLE, MSL2 and MSL3), as well as the HAS (defined by the co-localization of MLE and MSL2), were taken from previously published ChIP-chip (Alekseyenko et al., 2006) and ChIP-seq (Straub et al., 2013) studies. The distance between the binding sites of each DCC component and the reporter gene insertions was calculated as the minimum number of base pairs between their starting (or ending) genomic coordinates. Because the regulatory effect of the DCC is thought to be limited to active chromatin compartments of approximately 50 kb (Schauer et al., 2017), we limited our main analysis to reporter genes located within this distance of a DCC binding site.

The location of each reporter gene insertion relative to the annotated genes of *D. melanogaster* was determined using FlyBase release 6.31 (Thurmond et al., 2019). We classified each insertion based on its position relative to the closest gene (5' flanking or 3' flanking for intergenic locations) or functional element of a gene (5' UTR, coding sequence, intron, or 3' UTR for intragenic locations). Eight of the insertions were in locations overlapping the transcriptional units of multiple genes. Four of these were cases in which one gene was embedded within a long intron of another gene. In these cases, we considered the insertion to be in the inner gene. The other four insertions were in locations where two genes had (partially) overlapping transcriptional units, depending on the mRNA isoform. In these cases, we considered the insertion to be in the gene that had its coding sequence closer to the insertion site.

To determine the sex bias of the endogenous genes in which the reporter genes were located, we used the $\log_2$(male/female) expression values compiled by Huylmans and Parsch (2014). This included expression data for the brain (Catalán et al., 2012), head (Meisel et al., 2012), whole fly (Gnad and Parsch, 2006), and gonads (Brown et al., 2014). Only genes with expression data in the above studies were included. For carcass, this amounted to 59 genes, while for both head and gonad, it was 60 genes.

## 2.5 Reporter gene expression assays

### 2.5.1 Somatic and germline expression of the *CMV-lacZ* reporter gene

The expression of the *CMV-lacZ* reporter gene was measured with a β-galactosidase enzymatic activity assay (Protocol C11). The activity was measured in the carcass (the whole fly with the gonads and head removed), head, testis, and ovary. Five hemizygous males or heterozygous females were used for protein extraction. In addition, homozygous females were assayed for a subset of 15 of the transgenic lines. For a subset of 32 transgenic lines, heads were dissected into the brain and head case (the remaining head after brain extraction). Brains and head cases from 10 hemizygous males and heterozygous females were used for protein extraction. The β-galactosidase activity was measured as the change in absorbance per minute (mOD/min) for the linear range of the reaction curve. For each transgenic line, as well as a *yw* negative control line, 2–4 biological replicates were tested, with the activity of each biological replicate corresponding to the mean of its two technical replicates.

To standardize enzymatic activity in samples derived from different tissues and sexes, the total soluble protein concentration was determined using the Lowry assay (Lowry, 1951) (Protocol C13). For each technical replicate, 10 µl (carcass, ovaries, and whole head) or 20 µl (brain, head case, and testis) of protein extract was used for the analysis. Standardized activity (units/mg) was calculated as the enzyme activity divided by the total protein amount in 1 ml. For subsequent analyses of the *CMV-lacZ* reporter gene, standardized enzyme activity was used unless otherwise indicated.

Using two cutoffs for the activity difference between males and females (1.25-fold and 1.5-fold), we classified the reporter gene insertions by their male-to-female expression ratio. Reporter genes with an expression ratio greater than 1.25 and 1.5 were classified as 1.25-fold and 1.5-fold male-biased, respectively. Reporter genes with an expression ratio lower than 0.8 and 0.67 were classified as 1.25-fold and 1.5-fold female-biased, respectively. Reporter genes with an expression ratio between 0.8 (0.67) and 1.25 (1.5) were classified as unbiased.

### 2.5.2 Forward genetic screen

To test for changes in the *wol* reporter gene expression in testes of mutagenized *wol20X* males, a modified β-galactosidase activity assay was performed (Protocol C12). For this, mutant males were collected from each line for two days and aged five days under standard conditions. Enzyme activity in either whole males or abdomens was measured using the same method as described above, with at least two replicates per mutant line. As a negative control, we used *yw; CyO/ScO; Ubx/Sb* males, which lack the reporter gene. Males with a homozygous autosomal insertion of the reporter gene (*wol16*) were used as a positive control to estimate the reporter gene expression in testes in the absence of suppression (homozygous expression refers to 100% expression) (Figure 2.4). To estimate 50%, 25% and 12.5% β-galactosidase activity, proteins extracted from positive control males were diluted 2-, 4- and 8-fold. Since the *wol20X* males have only one copy of the reporter gene on the X chromosome, we also compared activity in males from candidate mutant lines to the activity in *wol16* males heterozygous for the autosomal reporter gene insertion.

Most screened samples are expected to show activity levels similar to X-linked controls, which on average show a 7-fold lower β-galactosidase activity than males with a heterozygous autosomal insertion. Thus, we aimed to find samples with the activity increased

by up to 7-fold, suggesting that a mutation occurred in a gene required for the X suppression (Figure 2.4). We also searched for samples with reduced activity, which would suggest local changes in expression regulation or mutations that disrupt the reporter gene's function. Because only one male was tested per assay, additional replicates were performed on all lines showing unusually high or low activity. For a set of lines showing increased expression, we additionally measured β-galactosidase activity in the dissected testes to determine that the increased reporter gene expression is not the result of a loss of tissue-specific regulation. We also standardized the β-galactosidase activity by measuring the total soluble protein concentration in samples derived from the mutant and control lines to account for variations in a tissue amount. For this, β-galactosidase activity assay was performed on ten dissected testes per sample (Protocol C11), and the total soluble protein concentration was measured in 20 µl of protein extract by the Lowry assay (Protocol C13).

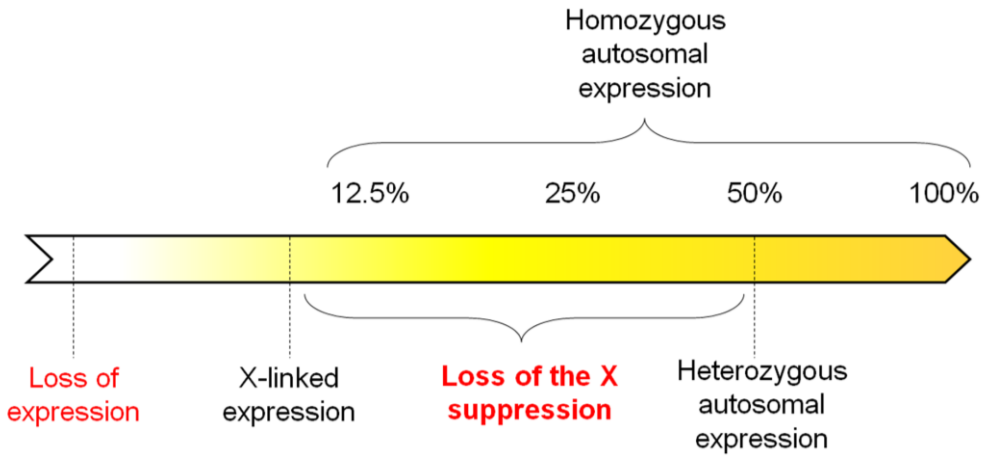


**Figure 2.4.** Mutant screening design. The yellow bar illustrates the β-galactosidase activity gradient expected for various samples. The range of activity levels expected for mutants with low or high reporter gene expression is marked in red. The control X-linked activity level is that of the *wol20X* line, while the autosomal activity levels are from the *wol16* line.

## 2.6 Maintaining the candidate mutant strains and mapping of causative mutations by genetic crosses

After the changes in reporter gene activity had been confirmed in the *Low* and *INXS* mutant lines, it was necessary to maintain these stocks without losing the mutant X chromosome or either of the autosomes 2 or 3. From the initial screening tests of the *Low 1/2* males, we were able to map the location of the causative mutations on the X chromosome. Furthermore, as we

had not found any significant variation in reporter gene expression, these lines were kept with homozygous autosomes simply by transferring each new generation into vials with fresh food.

In *INXS 1/2* mutant lines, we found that β-galactosidase activity had high variation among samples, which could be explained, at least partially, by the segregation of mutations among individuals. Therefore, we performed a series of crosses to establish mutant lines with stable overexpression and to narrow down the potential chromosomes carrying a causative mutation. Since males carrying the reporter gene construct on the X chromosome of both *INXS 1/2* lines were sterile, we crossed mutant virgin females with red eyes with *yw; CyO/ScO; Ubx/Sb* males. Males of the subsequent generation carrying the reporter gene were assayed for β-galactosidase activity. These tests were performed using males with different sets of balancer chromosomes, so we could infer the presence of causative mutations on a mutant autosome if its replacement by a balancer resulted in a restoration of control activity level. Next, when the increased expression was detected in a large number of males collected from the same parental vial, females with red eyes and mutant autosomes were crossed with their white-eyed siblings. Each replicate strain of *INXS 1/2* mutants generated by this method was maintained independently by mating females carrying the reporter gene on the X chromosome and heterozygous autosomes maintained over balancers with their white-eyed siblings or *yw; CyO/ScO; Ubx/Sb* males. To avoid recombination of the mutant X chromosome and *yw* chromosome, independent replicate stocks of both mutant lines were kept with the mutant X chromosome over the *FM7i* balancer. For this purpose, virgin females of each of the mutant lines with the mutant X chromosome and the balancer autosomes were crossed with males of the line carrying the *FM7i* balancer.

Two additional crosses between the mutant and transgenic lines with the X-linked reporter gene were performed to map chromosomes carrying causative mutations. In the first additional cross, we used mutant females with red eyes and the *CyO* and *Sb* balancer markers over mutant chromosomes and *wol20X* males. The F1 male offspring with red eyes were assayed for β-galactosidase activity. If the increased activity was observed only in males without the *CyO* (or *Sb*) marker, this could suggest a causative mutation on the second (or third) chromosome. In the second additional cross, mutant females with no reporter gene insertion on the X chromosome carrying the *CyO* and *Sb* balancer markers were crossed with males from one of the three transgenic lines with an X-linked reporter gene insertion (*wol20X*, *wol12X*, *wol5X*). Female offspring with red eyes and with no balancers were then crossed

with males of the corresponding *wol* line. The male offspring were then tested for β-galactosidase activity. Increased activity in these males could suggest that the causative mutation is not located on the X chromosome. For all additional crosses, one female and 2–3 males were used per vial.

## 2.7 Detecting *de novo* mutations by next-generation sequencing

After the preliminary genomic mapping of the causative mutations, we performed a whole-genome sequencing analysis of the *Low 1/2* and *INXS 1/2* lines to detect the EMS-induced mutations. To determine the background (non-mutant) genome sequence, we used six samples of *wol20X* and six samples of the *yw; CyO/ScO; Ubx/Sb* lines used in our crosses. For each line, four samples were of individual males, and the other two were pooled samples of seven males each. For the *Low* lines, we used one sample per line, which consisted of an individual male carrying a mutant X chromosome and both autosomes (over balancers for *Low1* and homozygous for *Low2*). For the *INXS1* sequencing, we used ten samples, each consisting of a single male carrying a mutant X chromosome and mutant autosomes maintained over different balancers. For the *INXS2* sequencing, we used nine samples, each consisting of a single male carrying a mutant X chromosome and different combinations of mutant autosomes and balancers. Prior to sequencing, the abdomen of each individual male was assayed for the β-galactosidase activity to verify the level of reporter gene expression. The rest of the bodies (head and thorax) were stored in -80 °C in 95% alcohol until DNA extraction. Genomic DNA was extracted with the MasterPure DNA Purification Kit (Epicentre, Madison, WI, USA) (Protocol C1). DNA concentration and quality were confirmed with a NanoDrop 1000. Sequencing was performed by an external sequencing facility using the Illumina NovaSeq 6000 platform (Novogene, Beijing, China) to generate approximately 30 million 150-base-pair (bp) paired-end reads per library (about 50X coverage of the *D. melanogaster* genome), which is sufficient for the detection of point mutations and short insertions/deletions (Haelterman et al., 2014).

Next-gen sequence data analysis described by Haelterman et al. (2014) was conducted by John Parsch and Eliza Argyridou. Raw sequence reads were mapped to the *D. melanogaster* reference genome (Release 6 to generate a list of single nucleotide variants and short indel variants. Further, mutations that occur in the genomes of wild populations of *D. melanogaster* (used references) or have been found in the genomes of parental lines (*wol20X*

and *yw; CyO/ScO; Ubx/Sb*) were excluded. Our focus was mainly on variants that affected the genome's functional parts and were found on chromosomes associated with up-regulated expression of the reporter gene.

## 2.8 Endogenous gene expression analysis with RNA-sequencing and qRT-PCR

In order to test the general increase in gene expression on the X chromosome, RNA-sequencing analysis of the *INXS1* mutant was performed. For this, pooled samples from 30 testes in 3 biological replicates of the *INXS1* mutant and the parental *wol20X* line were used. RNA extraction using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA) and analysis of sequencing data was performed by Eliza Argyridou. Sequencing was performed by an external sequencing facility using the Illumina NovaSeq 6000 platform (GENEWIZ, Leipzig, Germany) to generate approximately 36 million 150bp paired-end reads with poly-A enrichment. Mapping of reads and subsequent gene expression analysis were carried out with NextGenMap (Sedlazeck et al., 2013) and DEseq2 packages (Love et al., 2014). For the normalization of reads, the external RNA control consortium (ERCC) spike-in controls with a minimum gene count of 50 were used (Jiang et al., 2011). Increased gene expression in testes was determined for all genes with a minimum normalized gene count of 50 or greater. Normalized gene counts of *INXS1* mutant and *wol20X* control lines were further used for differential gene expression analysis (see below).

The general increase in expression on the X chromosome in *INXS2* was tested using qRT-PCR. For this purpose, six X-linked genes with high testis-specific expression and showing a high fold-change expression difference in the *INXS1* RNA-seq analysis were selected (*CG12689*, *CG3323*, *CG15306*, *CG15892*, *CG11068*, *CG7349*). A total of 5–10 biological replicates were analysed for the mutant and parental *wol20X* lines and each pair of amplification primers (Table B1). A total of 10 testes from 4–6 day-old males were collected per sample in 1X phosphate-buffered saline (PBS) and stored in liquid nitrogen. RNA extraction was carried out on the same day with the MasterPure RNA Purification Kit (Epicentre, Madison, WI, USA) (Protocol C8). RNA concentration and quality were confirmed with a NanoDrop 1000. This was followed by cDNA generation from 650 ng of isolated RNA with SuperScript III reverse transcriptase (Life Technologies) (Protocol C9). qRT-PCR analysis was carried out on a Real-Time thermal cycler CFX96 (Bio-Rad, Hercules, CA, USA) with 2–3 technical replicates for each primer pair amplifying the target and

reference (*RpL32*) gene sequences (Protocol C10). Fold-change expression difference was calculated using the *ΔΔ*Ct method, where the ratio of the mean expression of the target and reference genes (*Δ*Ct) was applied to normalise the expression and then the ratio of the mean *Δ*Ct of the mutant and control X-linked lines was calculated (*ΔΔ*Ct) (Livak and Schmittgen, 2001). The fold-change for expression was then found as 2^-(*ΔΔ*Ct) with a 95% confidence interval.

## 2.9 Functional test of *CG13003*, *CG1314* and *CG31525* candidate genes

To test whether the candidate mutated genes were associated with the observed lack of X chromosome suppression in the male germline, we took advantage of the targeted gene knock-down by the *UAS-Gal4* system (Figure 1.3). The aim was to obtain an increase in the reporter or endogenous genes expression to the level of mutant lines and male sterility. However, as there may be variation in the silencing efficiency in different cell types and different with different *UAS-RNAi* constructs, we also expected variation in the manifestation of these traits in individual males. A preliminary RNAi test was performed using four different Gal4 promoters (*Bam*, *Tj*, *Nanos*, *Act5C*) and the *CG13003* target gene. For this purpose, Gal4 and UAS-RNAi constructs were introduced into a wild-type background of the parental *wol20X* line by genetic crosses followed by testing for the resulting phenotype. Females of the *wol20X* line were crossed with *yw; CyO/ScO; Ubx/Sb* males. F1 males with red eyes and the *CyO* balancer on the second chromosome were collected and crossed with *RNAi-CG13003* females to obtain female offspring carrying both the reporter gene on the X chromosome and *UAS-CG13003* over the *CyO* balancer on the second chromosome. As a final step, these females were crossed separately with males of each Gal4 lines. Male offspring with the reporter gene, *UAS-CG13003*, and Gal4 were assayed for their β-galactosidase activity. Males without *UAS-CG13003* (with the *ScO* balancer) were used as controls. Only males with β-galactosidase activity higher than that in males without a reporter gene insert (*yw*) were considered in subsequent analysis.

We also used qRT-PCR to test if the knockdown leads to up-regulation of endogenous X-linked genes with confirmed increased expression in *INXS 1/2* mutants (*CG12689*, *CG11068*). This was done by crossing *Bam-Gal4* and *Act5C-Gal4* females with *RNAi-CG13003*, *RNAi-CG1314* and *RNAi-CG31525* males and *Tj-Gal4* females with *RNAi-CG13003* males. As a control, males lacking the UAS-RNAi constructs were used. These

control males were obtained by crossing females of each of the Gal4 lines with males of the *60100* (for tests of *RNAi-CG13003* and *RNAi-CG31525*) and *60000* (for tests of *RNAi-CG1314*) lines. One male and 2–3 females per vial were used for all crosses. Then, 10 testes from 4–6 day-old males were dissected and placed in DNA/RNA Shield (Zymo Research, Irvine, CA) and stored at -80 °C until RNA extraction. RNA extraction was performed on the same day using the MasterPureTM DNA/RNA Purification Kit (Protocol C8). This was followed by cDNA generation from 650 ng (1.5 μg for samples with 7 whole flies when expression of *CG13003* was tested) of isolated RNA with SuperScript III reverse transcriptase (Invitrogen™ Life Technologies, Carlsbad, CA, USA) (Protocol C9). qRT-PCR analysis and expression fold-change calculation were performed according to the methodology described above for primer pairs amplifying sequences of target and reference (*RpL32*) genes (Table B1, Protocol C9). The RNAi knockdown efficiency was tested by measuring the target RNAi gene's expression in samples with *UAS-Gal4* and control samples. The efficiency of the knockdown was calculated as (1 - fold-change)*100, where the fold change was measured for genes targeted by RNAi (*CG13003, CG1314, CG31525*).

## 2.10 Association analysis of mutations and increased reporter gene expression in *INXS 1/2* mutants

To examine which genomic variants contribute to increased reporter gene expression in both mutants, we performed a series of β-galactosidase activity measurements in males, followed by genotype testing. Males of the *INXS 1/2* mutant lines with different combinations of balancer chromosomes and males obtained from the additional crosses (see section 2.6) were collected, and their abdomens were tested for β-galactosidase activity (Protocol C12). The rest of the body was frozen at -80 °C in 95% alcohol until the time of DNA extraction (Protocol C1). We tested 12 different genomic variants identified by whole-genome sequencing (10 in *INXS1* and 2 in *INXS2*). Six of the variants were located in candidate genes (*INXS1*: *CG13003*, *CG17344*, *CG42654*, *CG31525*; *INXS1*: *CG1314*, *CG31525*), and six were selected as markers of chromosome regions in order to test for recombination between mutant and wild-type chromosomes (*INXS1*: *Kua*, *LRR*, *sgg*, *Mcm3*, *Usp4A*, *Ipp*) (Table B1).

Point mutations were tested by PCR (Protocol C6) followed by restriction digest (Protocol C7) if the mutation caused a DNA restriction site polymorphism. Thus, mutant and wild-type amplification products after incubation with restriction enzyme could be separated

by size using agarose gel electrophoresis (Table B1). In the absence of restriction site polymorphism, Sanger sequencing was performed (Protocol C5). The mutant and control DNA sequences were analysed with QIAGEN CLC Main Workbench 21.0.3 (Qiagen, Hilden, Germany). The short deletion in the *CG13003* gene was screened using PCR followed by gel electrophoresis. Two pairs of primers were used to specifically amplify sequences of different lengths of the mutant and wild-type genotypes (Table B1, Protocol C6).

## 2.11 Imaging of the *wol* reporter gene expression and male fertility test

We visualised the reporter gene expression in the testes of 4–6 day-old males of *INXS 1/2*, *wol20X* and *wol16* lines. The intact testes were dissected in 1X PBS and further incubated in an assay buffer containing 1 mg/ml ferric ammonium citrate and 1.8 mg/ml of S-GAL sodium salt (Sigma-Aldrich, Munich, Germany) for 6 hours at 37 °C. The stained testes were then rinsed in PBS solution for 20 minutes and imaged with a Levenhuk M300 Base camera (Tampa, United States) at 3.5X magnification.

Male fertility tests were performed using 71 *INXS1* and 79 *INXS2* males. All males had the red-eye marker and different combinations of balancer chromosomes. Additionally, 23–29 males of the knockdown lines (*RNAi-CG13003*, *RNAi-CG1314* and *RNAi-CG31525*) and males of the *wol20X* control line were analysed. After the virgin males were collected, they were placed in separate vials with 2–3 virgin *wol20X* females to allow for mating and egg-laying. Flies were kept at 25 °C, and then on the tenth day, vials with live larvae and tubes without a single larva was counted.

Additionally, sperm motility in the testes was tested for both mutants. For this, six pairs of testes from 6–8 day-old males of the *INXS 1/2* and control *wol20X* lines were used. Dissection was performed in 1X PBS solution on ice. The pairs of testes were immediately transferred to a microscope slide, covered by a cover glass and slightly compressed. The testes were then imaged at 40X magnification with a 31.25 ms exposure time, using a Leica DM750 microscope (Leica Microsystems, Wetzlar, Germany) and a Levenhuk 300 Base camera (Levenhuk Inc., Tampa, USA). For each pair of testes, the spermatozoa were considered immobile if no spermatozoa showed the typical oscillatory movements.

## 2.12 Statistical analysis

### 2.12.1 Proximity effects of DCC binding sites on sex-biased reporter gene expression

Correlation analyses for a gene's proximity to the nearest DCC binding site and its expression in males, females and male/female expression were performed using both the Spearman rank correlation (*rho*) and linear regression. For the main data set of 83 insertion locations, we compared β-galactosidase activity and $\log_2$(male/female β-galactosidase activity) in the different tissues by a paired *t*-test using the insertion locations as replicates. For smaller subsets of 15 and 32 transgenic lines (dosage and brain/head case analyses), we compared β-galactosidase activity and $\log_2$(male/female β-galactosidase activity) in the different sexes and tissues with a Wilcoxon signed-rank test using the independent transformed lines as replicates. The comparison of the groups of insertions with different proximity to the nearest DCC binding sites was carried out with an unpaired Wilcoxon signed-rank test. One-way analysis of covariance (ANCOVA) was performed for 64 insertions located within transcribed regions of genes to determine the effect of endogenous genes' sex-biased expression (covariate) on $\log_2$(male/female β-galactosidase activity) in relation to the distance of each insertion to the nearest DCC binding site (independent variable). Two levels of distances were used: insertions within 25 kb ("close") and insertions within the range of 25–50 kb ("distant") to the DCC binding sites. One-way ANCOVA was performed using expression data from homozygous females, heterozygous females, and hemizygous males for 15 transgenic lines. For this, we first analyzed the influence of the distance to the nearest DCC binding site (covariate) on $\log_2$(male/female β-galactosidase activity) among homozygous and heterozygous females (independent variable). Second, we analyzed the influence of sex-biased expression of endogenous genes (covariate) on $\log_2$(male/female β-galactosidase activity) among these two groups of females (independent variable). To compare sources of variation among insertions in brain and head case, we used an *F*-test, a variance component analysis (Schützenmeister and Piepho, 2012), and an asymptotic test for the equality of the coefficient of variation (CV) (Feltz and Miller, 1996).

### 2.12.2 Identifying genes involved in the X suppression

Up-regulation of β-galactosidase activity in *INXS 1/2* mutant males and males with RNAi-mediated *CG13003* gene knockdown was tested with a *t*-test or, in case of small sample size, with the Wilcoxon signed-rank test. Associations between mutations in *INXS 1/2* mutants and

increased reporter gene expression were found by calculating proportions of males carrying the mutation in the categories of males with low and high expression. The difference in proportions between these groups for each of the target genes was tested with a Fisher's exact test. The fertility test of males with RNAi targeting three candidate genes was performed by calculating the proportions of sterile males in the UAS-RNAi group and the UAS control group. The difference in proportions between these groups was tested with a Fisher's exact test.

For the differential gene expression analysis of *INXS1* mutant and *wol20X* control lines, log2(fold-change for expression difference) was determined using DEseq2 (Love et al., 2014) and the significance was assessed using the Wald test. *P*-values were adjusted for multiple testing using the false discovery rate (FDR) approach (Benjamini and Hochberg, 1995). Then we filtered genes by considering only those with sufficient expression to calculate FDR. Genes were considered differentially expressed if they had FDR <5% and fold-change >1.5. Proportions of up-regulated genes for different chromosome arms were tested with Pearson's chi-squared test with the Yates' continuity correction. General upregulation of X-linked genes in *INXS2* mutants and flies with RNAi-mediated knockdown of *CG13003*, *CG1314* and *CG31525* candidate genes was tested using qRT-PCR. The fold change for expression in the test and control lines was calculated. The significance of fold-change for expression of the X-linked genes was tested using Welch's two sample *t*-test.

# Chapter 3

# Results

## 3.1 X-chromosome specific dosage compensation

### 3.1.1 Chromosomal location and expression of reporter genes

We obtained a total of 102 transgenic lines with unique X chromosomal insertions of the *CMV-lacZ* reporter gene (Table B2). Some of these insertions were very close to each other. In order to avoid pseudoreplication in our analyses, we did not treat insertions within 500 bp of each other as independent replicates of different locations. Instead, they were treated as replicates of the same location. In total, there were 32 such insertions, which were present at 13 different genomic locations. After combining these 32 insertions as replicates of 13 locations, a total of 83 unique insertion locations remained (Figure 3.1, Table B2). The CV for reporter gene expression among replicates of the combined insertions was not greater than that among biological replicates of individual insertions (Figure A1), indicating that short-range (within 500 bp) chromosomal effects on expression are negligible.

All of the 83 distinct insertions of the *CMV-lacZ* reporter gene were located in active chromatin regions and in close proximity to genes, with 64 of them being located within transcribed regions (Table 3.1).


**Figure 3.1.** Reporter gene insertion locations on the X chromosome. Vertical lines show the locations of the 83 reporter gene insertions. Bold lines indicate locations where two insertions are within 50 kb of each other.

**Table 3.1.** The number of reporter gene insertions in different gene regions.

| Gene region | Insertions |
|---|---|
| 5' flanking | 17 |
| 5' UTR | 28 (7)[a] |
| Coding | 1 |
| Intron | 30 |
| 3' UTR | 5 (1)[a] |
| 3' flanking | 2 |

[a]Numbers in parentheses indicate insertions that could also be classified as intronic, depending on the transcript isoform.

For each insertion location, reporter gene expression was measured in the head, gonad, and carcass (here defined as the whole fly with the head and gonads removed) using a β-galactosidase activity assay (Figure A2). To control for differences in X chromosome gene dose between males and females, activity was measured in females that were heterozygous for the reporter gene insertion (i.e., both sexes had only one copy of the reporter construct). To account for potential differences in enzyme activity due to variation in body or tissue size, activity was standardized by the total amount of protein in the sample (Figure A3). The total protein yield was higher in females than in males for all tissues, reflecting the larger body/gonad size of females.

Somatic tissues showed very different patterns of standardized β-galactosidase activity in males and females. In the carcass, standardized β-galactosidase activity level was significantly higher in males than in females (paired *t*-test, $P = 2.9 \times 10^{-7}$), which may be the result of partial dosage compensation of X-linked reporter genes in males (Figure 3.2, Table B3). It should be noted that since the reporter genes are present as a single copy in both sexes,

X chromosome dosage compensation is expected to lead to higher expression in males. But the difference in expression between sexes is only about 1.2-fold, which is well below the two-fold difference expected under complete dosage compensation. On the contrary, in the head, there was significantly higher expression in females than in males (paired *t*-test, $P = 1.2 \times 10^{-13}$). In the gonads, the reporter genes also had significantly higher expression in females than in males (paired *t*-test, $P < 3.7\times 10^{-13}$).

In the testis, β-galactosidase activity was much lower than in somatic tissues (Figure 3.2, Table B3), which could potentially affect the accuracy of the results. Since there was both low levels of non-normalized activity and total protein in the testis, we examined whether these low values led to increased variance in the reporter gene expression measurements. There were no significant differences in the level of variation of β-galactosidase activity between the testis and other tissues (asymptotic test for the equality of CV, $P = 0.139$ for testis *vs.* head and $P = 0.094$ for testis *vs.* carcass), indicating that reporter gene expression can be measured reliably in this tissue.

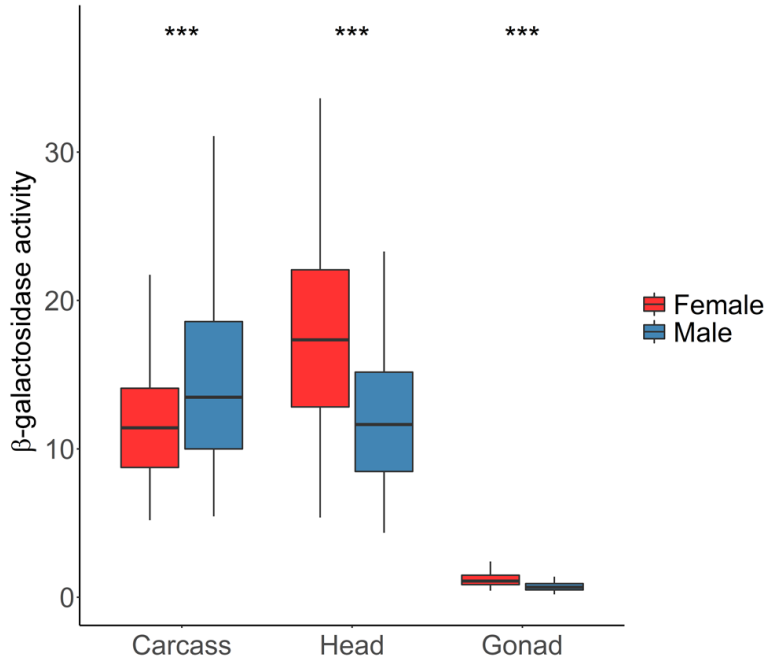


**Figure 3.2.** β-galactosidase activity in different sexes and tissues. For each tissue, differences between the sexes were tested using a paired *t*-test. ***$P < 0.001$.

### 3.1.2 Sex-biased expression and the influence of endogenous regulatory elements

We determined the degree of sex-biased expression for each of the reporter gene insertions by measuring the ratio of reporter gene activity between males and females (Figure 3.3). In general, the results were consistent with the levels of activity seen for the two sexes separately (Figure 3.2). In the carcass, there were more male-biased than female-biased genes (63 vs. 20). In contrast, for both the head and gonad, there was an excess of female-biased genes (6 vs. 77 in head, 8 vs. 75 in gonad).

To test for a tissue-independent effect of insertion location on reporter gene expression, we calculated Spearman's correlation coefficient of β-galactosidase activity for each tissue pair. The male-to-female expression ratio of the insertions was positively correlated between carcass and head (*rho* = 0.20, *P* = 0.034), between carcass and gonad (*rho* = 0.28, *P* = 0.002), and between head and gonad (*rho* = 0.15, *P* = 0.103).

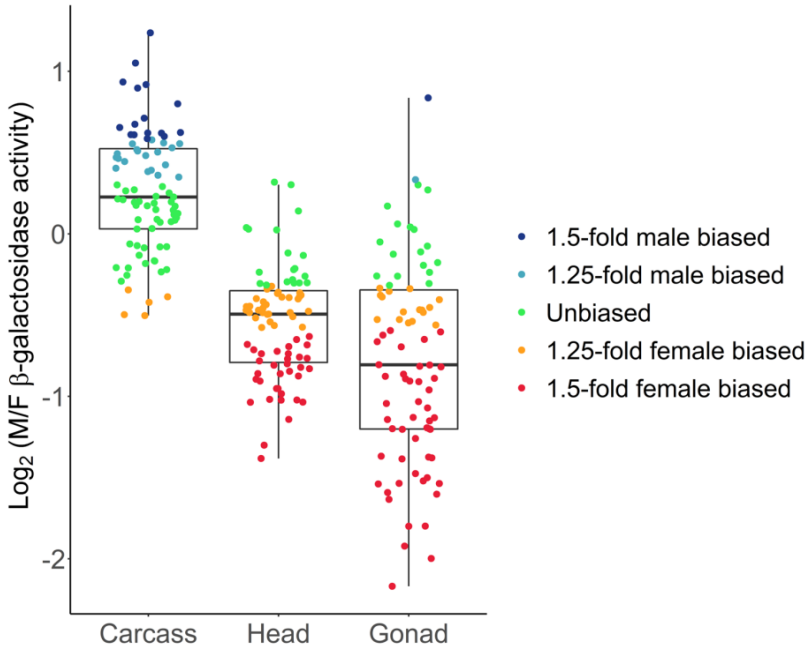


**Figure 3.3.** Male-to-female expression ratio (M/F) of reporter genes in different tissues. Coloured points indicate the level of sex-biased expression for the individual *CMV-lacZ* reporter gene insertion locations.

As mentioned above, 64 of the reporter gene insertions were located within the transcriptional units of genes (Table 3.1). Thus, sequence elements regulating the expression of these genes may also influence the expression of the embedded reporter genes. To test this

possibility, we compared the male-to-female expression ratio of the reporter genes with that of the endogenous genes in which they were located using previously published sex-biased expression data from the different tissues (Gnad and Parsch, 2006; Meisel et al., 2012; Catalán et al., 2012; Brown et al., 2014; Huylmans and Parsch, 2015) (Figure 3.4). There was not a significant correlation between the sex-biased expression of the reporter genes and the endogenous genes in any of the tissues (Spearman rank correlation, $P = 0.062$ in the carcass, $P = 0.593$ in the head, $P = 0.277$ in the gonads).

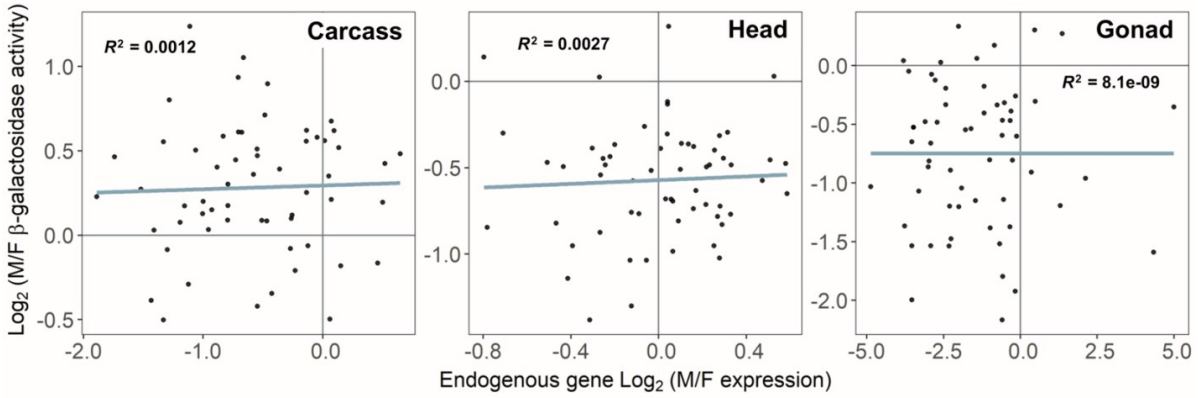


**Figure 3.4.** Correlation between the male-to-female expression ratio of each reporter gene and the endogenous gene in which it is located. Lines represent the least-squares linear regression.

### 3.1.3 The effect of DCC binding site proximity on reporter gene expression

To test if the mechanism of dosage compensation influences the sex-biased expression of X-linked genes, we examined the correlation between the male-to-female expression ratio of each reporter gene insertion and its distance to the nearest DCC binding site (Table B4, B5).

A previous study of the association between chromatin spatial structure and the size of TADs on DCC spreading showed that the highest number of genes were activated within 50 kb from HAS (Schauer et al., 2017). Therefore, we focused our analysis on *CMV-lacZ* insertions located within this range from the nearest DCC binding sites, which included the vast majority of the insertion locations (Figure 3.5). The reporter gene insertions had different average distances to the nearest DCC binding sites (HAS: 17,924 bp; MLE: 13,546 bp; MSL2: 20,450 bp; MSL3: 9,391 bp), which is associated with the frequency of binding sites of the different DCC components on the X chromosome (HAS: 244; MLE: 509; MSL2: 208; MSL3: 1,384).

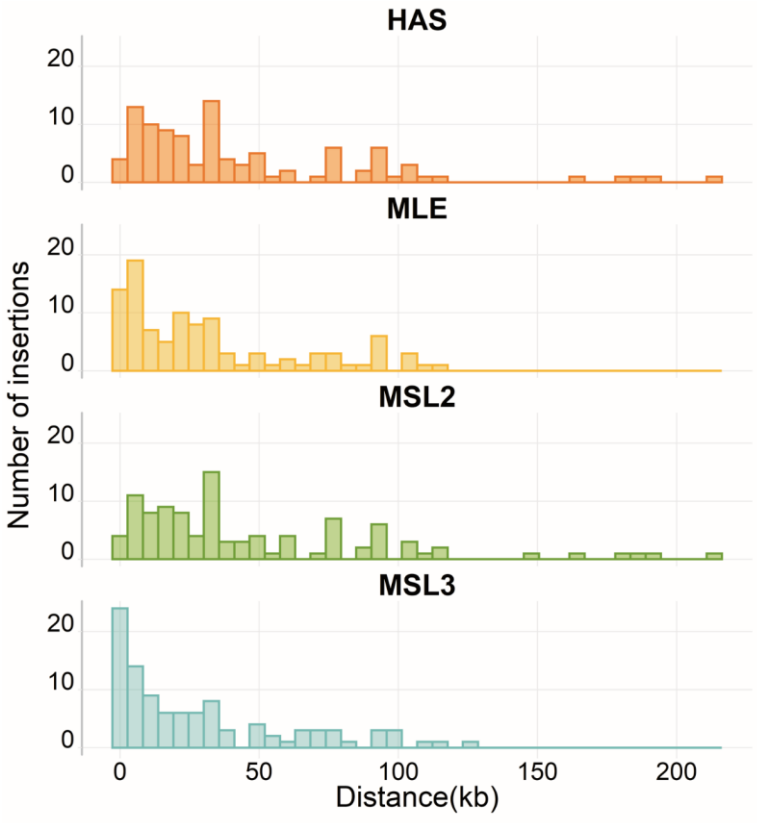


**Figure 3.5.** Histograms of the minimum distance between reporter gene insertion sites and the binding sites of different DCC components.

The correlation was consistently negative for all DCC components for the somatic tissues (head and carcass) (Figure 3.6*,* 3.7*A*, Table B4). This pattern was not observed for the gonad, where the correlation was close to zero and not significant (Figure 3.6*,* 3.7*A*, Table B4). When the expression is considered only in males, similar negative correlations are seen for all tissues (Figure 3.7*B*, A4*A,* Table B4), which is consistent with dosage compensation affecting the expression of the X chromosome in males. However, this negative correlation was weaker in the gonad. In females, there was no significant correlation between expression and distance to the binding site of any DCC component (Figure 3.7*C*, A4*B,* Table B4). We found no significant correlation for the whole dataset when we did not impose a maximum distance of 50 kb, except for the female carcass and head, where expression was negatively correlated with the distance to MSL3 binding sites (Table B5).

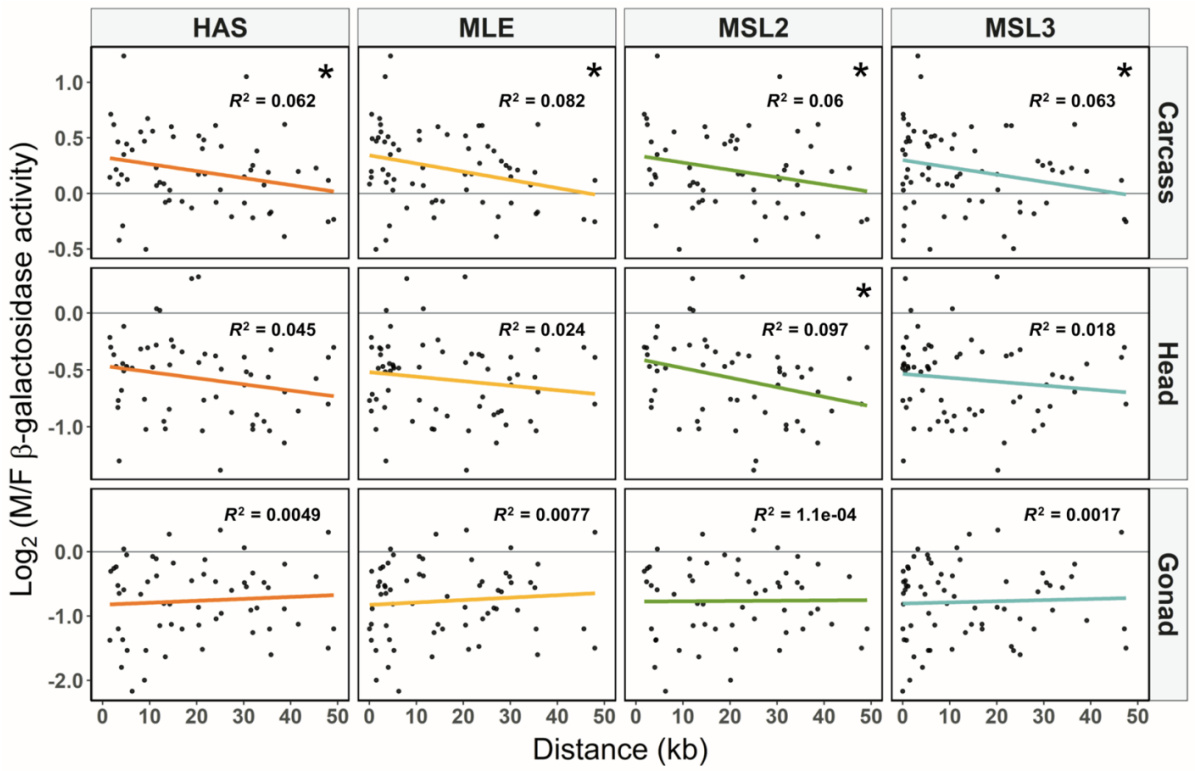


**Figure 3.6.** Male-to-female expression ratio (M/F) and distance to the nearest DCC binding sites for different tissues. The binding sites of different DCC components are shown in columns. Colored lines represent the least squares linear regression. *$P < 0.05$.

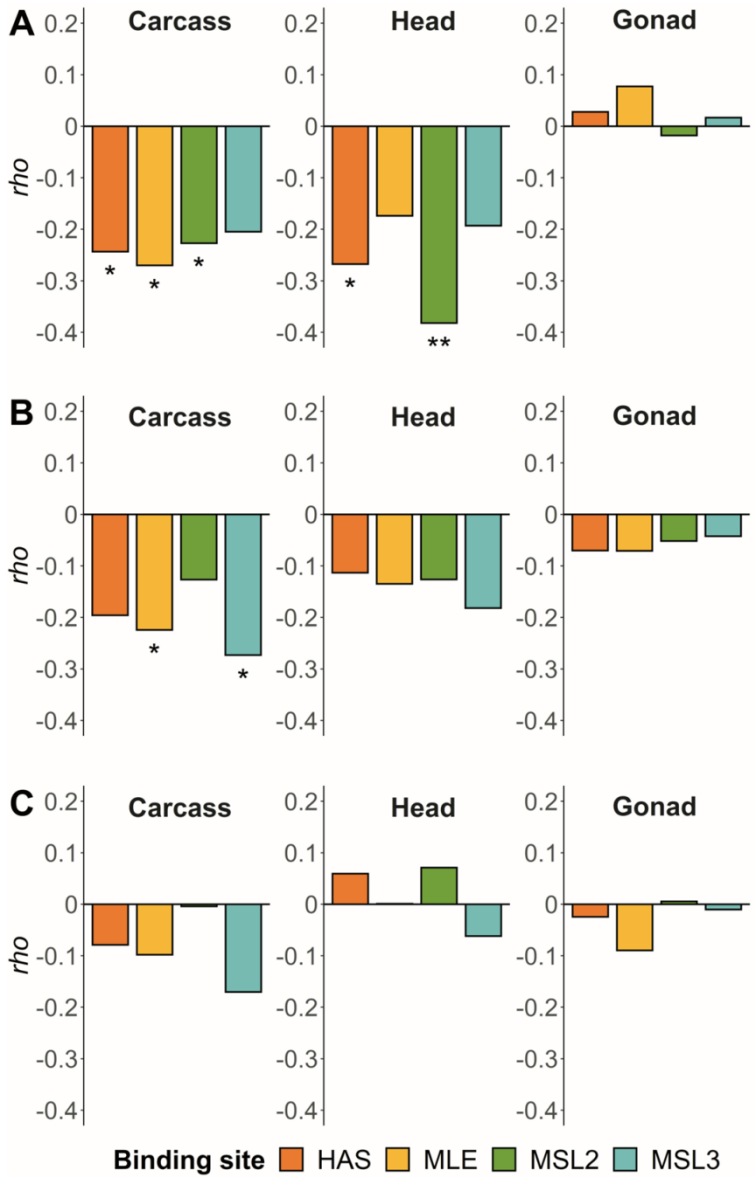


**Figure 3.7.** Relationship between reporter gene expression and distance to the nearest DCC binding sites for different tissues. Spearman's correlation coefficient (*rho*) was calculated for the correlation between distance to the nearest DCC component binding sites and (*A*) male/female expression, (*B*) male expression, and (*C*) female expression. *$P < 0.05$, **$P < 0.01$.

A one-way ANCOVA was performed to examine the effect of the proximity of an insertion to the nearest DCC binding site on the male-to-female reporter gene expression ratio, after controlling for the sex-biased expression of the endogenous genes (Figure 3.8, Table B6). In carcass, insertions located close to MLE and MSL2 binding sites had significantly higher male-to-female reporter gene expression (ANCOVA, $F = 3.9$, df =1, 41, $P = 0.027$ for MLE; $F = 3.2$, df = 1, 37, $P = 0.042$ for MSL2). In head, the same pattern was

found for HAS (ANCOVA, $F = 3.6$, df = 1, 39, $P = 0.032$) and MSL2 (ANCOVA, $F = 7.3$, df = 1, 38, $P < 0.01$) binding sites.

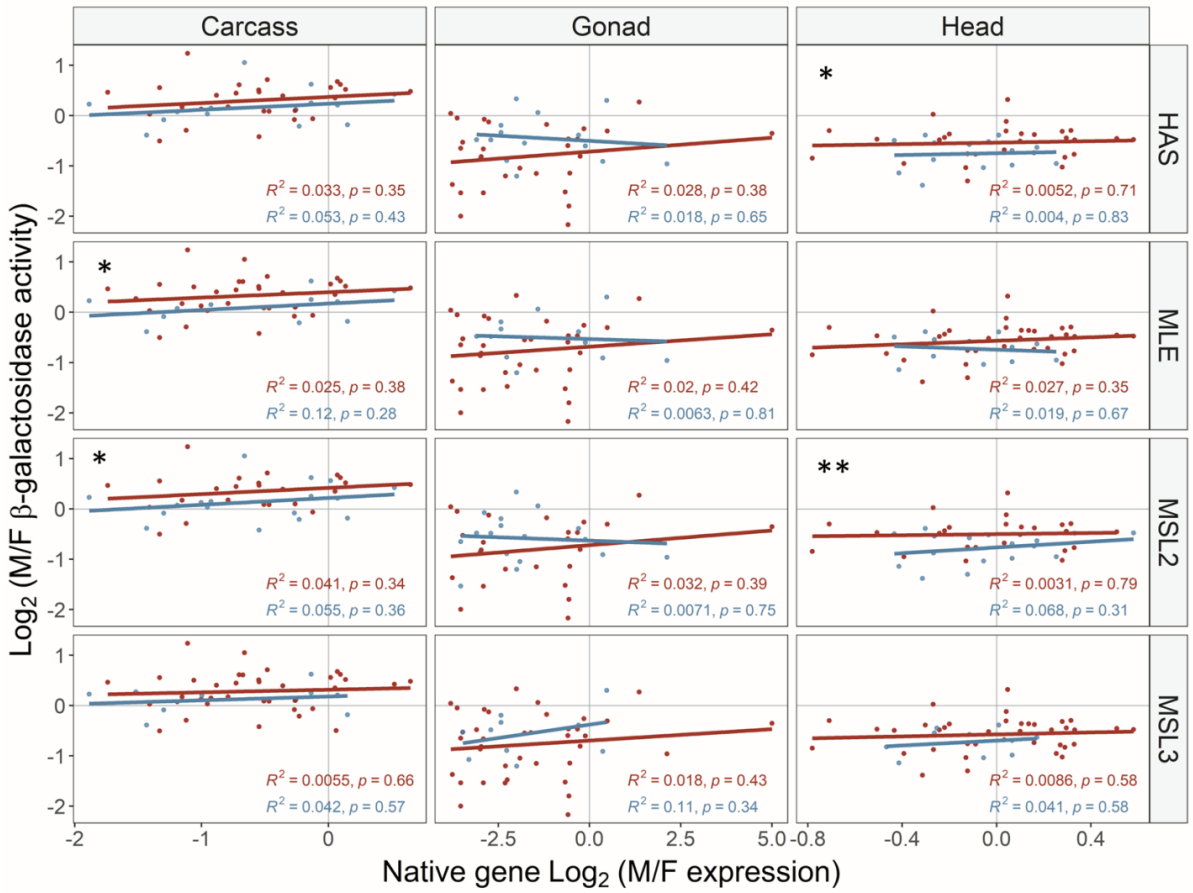


**Figure 3.8.** Male-to-female expression ratio of each reporter gene and the endogenous gene in which it is located for groups of their proximity to the nearest DCC binding site in different tissues. Lines represent the least-squares linear regression. Dots represent the individual *CMV-lacZ* reporter gene insertions. The colour indicates "close" (red) and "distant" (blue) groups of proximity to the binding sites of different components of the DCC (rows). *$P <$ 0.05, **$P < 0.01$ for the effect of the distance on the male-to-female expression ratio of the reporter gene.

Since each reporter gene insertion on the X chromosome is flanked by some number of G4s, we additionally tested whether it affects the expression of the reporter gene serving as an insulator of the spread of the complex. Interestingly, in all tissues, we found a positive correlation between the reporter gene's proximity to a nearby G4 location and its expression in females and males (Figure 3.9, Table B7). The correlation was slightly weaker in the heads and gonads for the male-to-female expression ratio, and close to zero in the carcass. In all

tissues, there was no significant difference in reporter gene male-to-female expression ratio, expression in males or expression in females between the following two groups: i) when the reporter gene is flanked only by G4s, and ii) when the reporter gene is flanked by at least one DCC binding site (Figure A5).

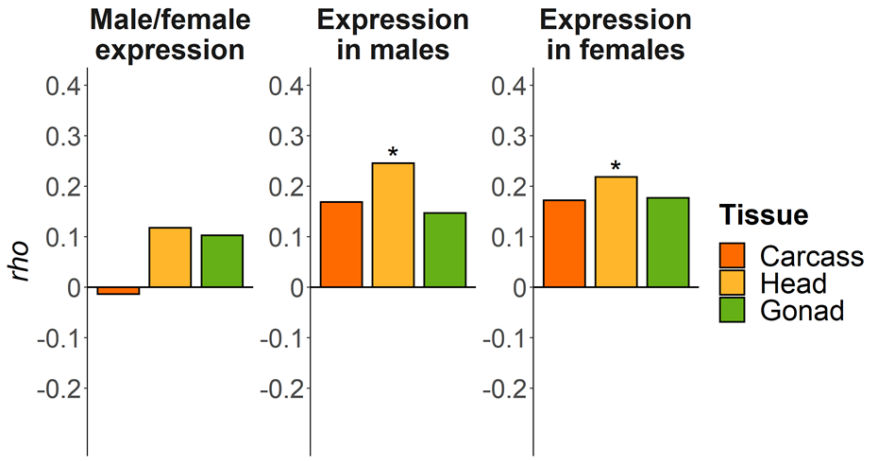


**Figure 3.9.** Spearman's correlation coefficient (*rho*) for the correlation between distance to the nearest G4s sites and male/female expression, male expression, and female expression. *$P < 0.05$.

### 3.1.4 Reporter gene expression in heterozygous and homozygous females

Although our primary analysis controlled for X chromosome copy number differences between the sexes by comparing females heterozygous for the reporter gene insertion to hemizygous males, we also examined the effect of gene dose by measuring expression in homozygous females of 15 independent insertion lines (Figure 3.10*A,* Table B8). As expected, homozygous females had higher expression than heterozygous females in all tissues (Figure 3.10*A*). The ratio of homozygous to heterozygous female expression (95% CI) was 1.7 (1.45 – 2.08) in carcass, 1.5 (1.25 – 1.85) in head, and 2.0 (1.56 – 2.61) in ovary. In the carcass and gonad, the 95% CI included 2, suggesting a nearly linear relationship between expression and gene dose. However, in the head, the expression ratio was below 2, suggesting that there is not a simple linear relationship between gene dose and expression in this tissue. The homozygous females also had higher expression than hemizygous males (Figure 3.10*A*), indicating that exogenous reporter genes introduced onto the X chromosome are not fully dosage compensated.

To test if there is some influence of dosage compensation on this subset of genes, we separated them into two distance categories on the basis of their proximity to each binding site, with 7 (8 for MSL3) insertions being considered "close" (15–15,061 bp) and 7 (MSL2 and MSL3) or 8 (HAS and MLE) "distant" (24,987–47,502 bp). We then compared the expression ratios between hemizygous males and homozygous females. In all tissues and for all DCC binding sites, the "close" insertions showed higher relative expression in males than the "distant" insertions, although the difference between categories was only significant for the head (Figure 3.10*B* for HAS, Figure A6 for MLE, MSL2, and MSL3).

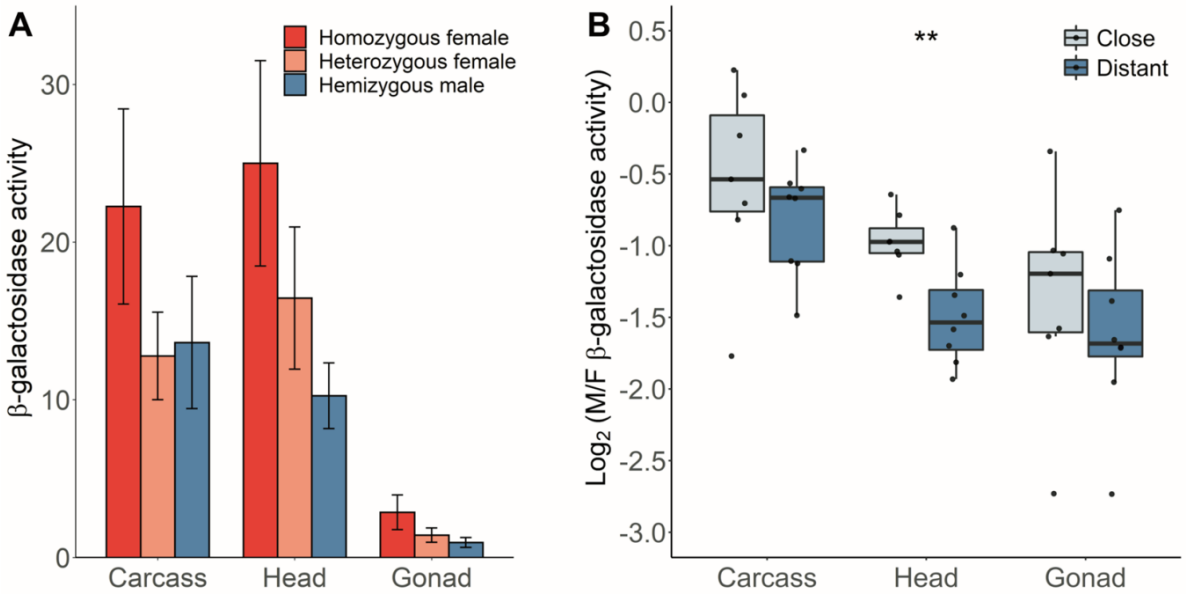


**Figure 3.10.** Effect of gene dose on β-galactosidase activity in various tissues. (*A*) Mean β-galactosidase activity for heterozygous females, homozygous females, and hemizygous males. Error bars indicate the standard deviation across insertions at different locations. (*B*) Male-to-female expression ratio (M/F) of reporter genes grouped by their proximity to the nearest HAS in different tissues for homozygous females and hemizygous males. Differences between the groups were tested with a Wilcoxon signed-rank test. *$P < 0.05$.

For this subset of 15 reporter gene insertions, we tested how the dosage of the reporter gene in females, the expression pattern of the surrounding endogenous gene, and the proximity to the nearest DCC binding site affect the sex-biased expression of the reporter gene (Figure A7, A8, Table B6). In carcass, there was a significant positive correlation between the male-to-female expression ratio of the reporter genes and that of their surrounding endogenous genes (ANCOVA, $F = 4.7$, df = 1, 20, $P = 0.043$), and this pattern was similar for the groups with homozygous and heterozygous females (Figure A7, Table

B6). In all three tissues, after adjusting for the effect of endogenous genes there was a significant difference in the reporter gene sex-biased expression between these two groups (ANCOVA, $F = 28.5$, df = 1, 20, $P < 0.001$ for the carcass; $F = 19.5$, df = 1, 20, $P < 0.001$ for the head; $F = 16.8$, df = 1, 20, $P < 0.001$ for the gonad). After adjusting for the effect of the distance to the nearest DCC binding site, there was also a significant difference in the reporter gene sex-biased expression between the groups of homozygous and heterozygous females in all tissues (Figure A8, Table B6). In the carcass, there was a significant effect of the distance to HAS, MLE, and MSL2 binding sites on the sex-biased expression of the reporter gene. In the head, this pattern was even stronger for all DCC binding sites (Figure A8, Table B6). In gonad, we also found a negative correlation between the sex-biased expression of the reporter gene and the proximity to only HAS (ANCOVA, $F = 4.7$, df = 1, 20, $P = 0.043$).

### 3.1.5 Reporter gene expression in the brain and head case.

To test for differences in sex-biased expression between the brain and the rest of the head, we measured reporter gene expression in the brain and head case (the whole head with the brain removed) for a subset of 32 of our transgenic lines (Figure 3.11*A*, Table B9). Similar to the whole head (Figure 3.2), both tissues showed higher reporter gene activity in females (paired Wilcoxon signed-rank test, $P = 0.015$ and $P = 4.7 \times 10^{-10}$ for brain and head case, respectively), although the degree of female bias was higher in the head case than the brain (paired Wilcoxon signed-rank test, $P = 0.017$) (Figure 3.11*A*).

When considering the male-to-female expression ratio of the individual reporter gene insertions, the median values were similar in the brain and head case (Figure 3.11*B*). However, there was much greater variance among the insertion locations in the brain (*F*-test, $P = 5.6 \times 10^{-10}$), with some showing strong male- or female-biased expression (Figure 3.11*B*). The difference in variance between the tissues could have two causes. First, it could be that there is greater technical variation among repeated activity measurements in the brain (intralocus variation), possibly because it is a smaller tissue with relatively low levels of reporter gene expression. Second, it could be that there is a greater position-effect variation (interlocus variation) among the X chromosomal insertion locations in the brain.

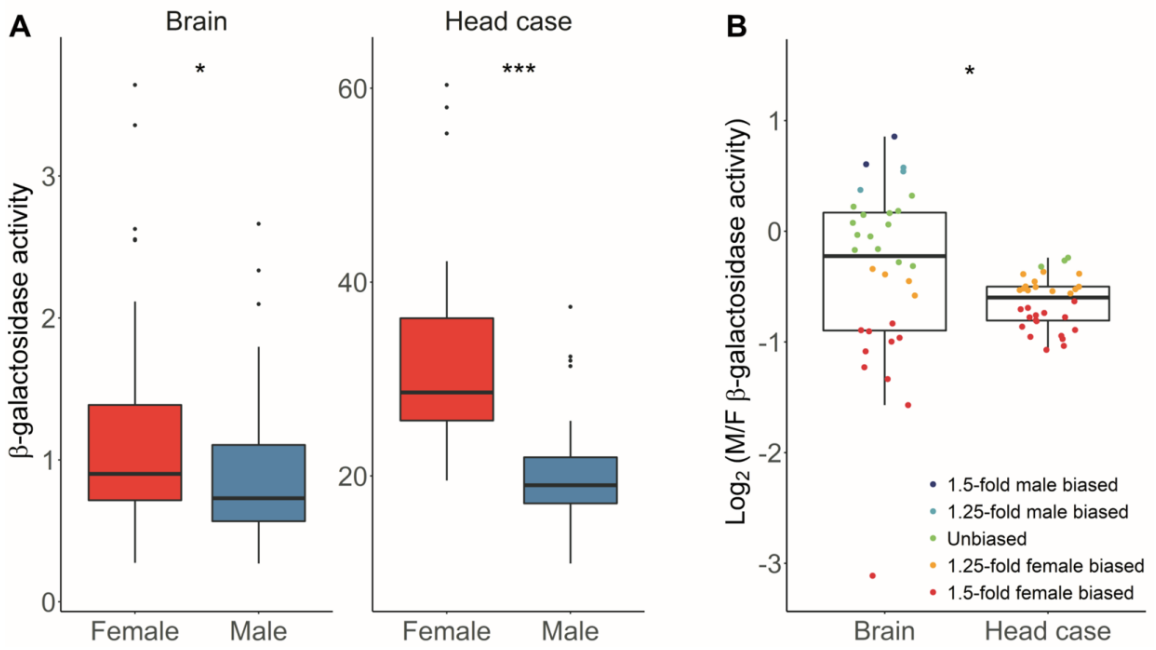


**Figure 3.11.** Reporter gene expression in brains and head cases. (*A*) β-galactosidase activity in males and females. (*B*) The ratio of male/female expression. Coloured points indicate the level of sex-biased expression for the individual *CMV-lacZ* reporter gene insertion locations. Differences between sexes and tissues were tested with a paired Wilcoxon signed-rank test. *P < 0.05, ***P < 0.001.

To determine the contributions of these two factors to the overall variance, we carried out a variance component analysis (Schützenmeister and Piepho, 2012) (Figure 3.12*A*, *B*). Although the brain displayed higher intralocus variation than the head case in both sexes (asymptotic test for the equality of CV, $P = 0.021$ and $P = 4.8 \times 10^{-4}$ for females and males, respectively), it also displayed significantly higher interlocus variation (asymptotic test for the equality of CV, $P = 1.7 \times 10^{-6}$ and $P = 7.8 \times 10^{-4}$ for females and males, respectively). Thus, gene expression appears to be more sensitive to the chromosomal location in the brain than in the rest of the head. However, for the head case and brain, we found no significant correlation between the male-to-female expression ratio, expression in males and females of each reporter gene insertion and its distance to the nearest DCC binding site, possibly due to small sample size (Table B10).

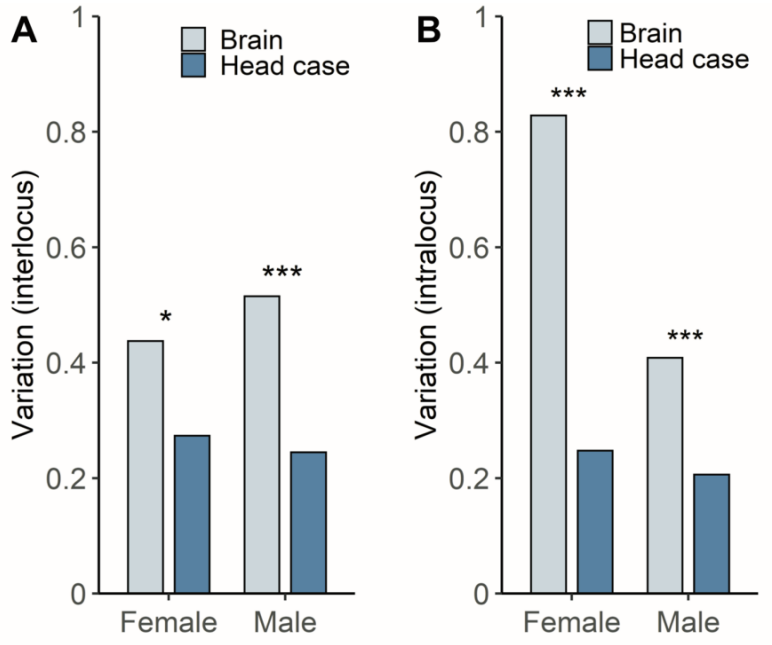


**Figure 3.12.** Sources of variation in β-galactosidase activity in brain and head case. (*A*) The interlocus component reflects variation among insertions at different X-chromosomal locations. (*B*) The intralocus component reflects variation among biological replicates of insertions at the same location. In all plots, variation is in units of standard deviation divided by the mean (CV). Differences between tissues were tested with an asymptotic test for the equality of CV. **P* < 0.05, ****P* < 0.001.

## 3.2 X chromosome suppression in the male germline

### 3.2.1 Forward genetic screen and reporter gene expression analysis of candidate mutant lines

We performed chemical mutagenesis of males containing an X-linked copy of the *wol* reporter gene. After a series of crosses (Figure 2.3), we tested β-galactosidase activity in males carrying a mutagenized X chromosome and autosomes in a homozygous state or heterozygous with the balancer chromosomes. In total, more than 5,000 independent mutant lines were screened. Using an average of 2–3 replicate assays for each mutant line, we were able to test both dominant and recessive mutations on all chromosomes. However, since each assay was performed using the testis of one fly per sample, the activity measurements often had a high variation due to differences in tissue size or sample preparation error. Therefore, for approximately one-tenth of all lines studied, additional replicates were performed to confirm an initially observed up-regulation or down-regulation of activity level.

We found four candidate lines with the mean β-galactosidase activity in testis of 2.15±0.44 mOD/min, which is slightly above the X-linked control level of 1.72±0.24 mOD/min (Figure 3.13*A*). Nevertheless, we could not confirm a consistent increase in reporter gene expression for these lines after normalization by total protein amount using pooled samples (Figure 3.13*B*). This may indicate the presence of mutations affecting fly development, leading to greater variation in gonad size, or mutations driving broader expression of the reporter gene. These lines were not considered further.

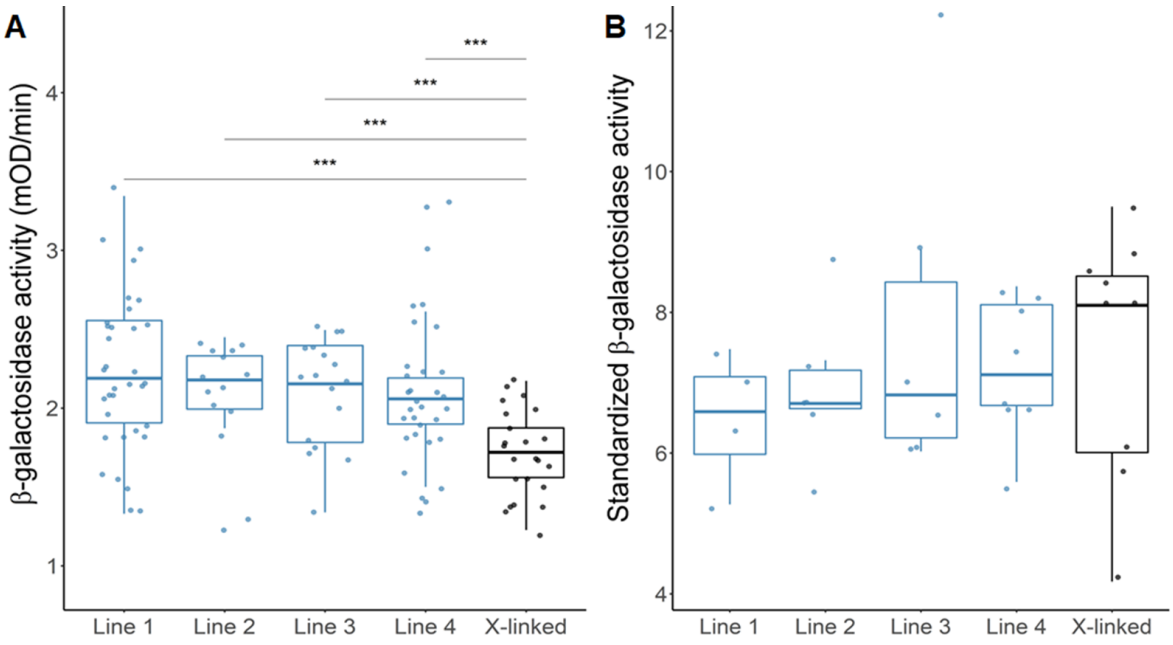


**Figure 3.13.** Reporter gene expression in testes of four candidate mutant lines. The *wol* reporter gene expression is measured as (*A*) β-galactosidase activity in single fly or (*B*) standardized β-galactosidase activity in pooled samples. For each variant, differences between the mutant and control X-linked lines were tested using a Wilcoxon signed-rank test. ***$P <$ 0.001.

We found two lines with decreased β-galactosidase activity in testis, *Low1* (mean activity: 0.35±0.13 mOD/min) and *Low2* (0.14±0.03 mOD/min) (Figure 3.14). These expression levels correspond to the background expression in tissues of wild-type flies (0.16±0.16 mOD/min), suggesting a complete loss of reporter gene activity.

Our primary interest was focused on two lines with increased β-galactosidase activity in the testis that we named *INvolved in X Suppression*: *INXS1* (3.86±1.6 mOD/min) and *INXS2* (4.63±1.61 mOD/min) (Figure 3.14). Both *INXS* mutant lines showed a similar

dispersion of the reporter gene expression among individually-tested flies, although *INXS2* had a significantly higher mean expression (*t*-test, $P < 0.001$). The activity of both lines was significantly higher than β-galactosidase activity of the X-linked control and lower than the activity of heterozygous autosomal control (7.61±1.63 mOD/min) (*t*-test, $P < 0.001$) (Figure 3.14). At the same time, the expression in testis of the *INXS* lines partially overlapped with both control lines due to a high variation in expression among individuals. This is expected, as the lines are not isogenic and the mutations will segregate among individuals. Furthermore, some of the variations could be attributable to the flies having different combinations of balancer chromosomes for chromosomes 2 and 3.

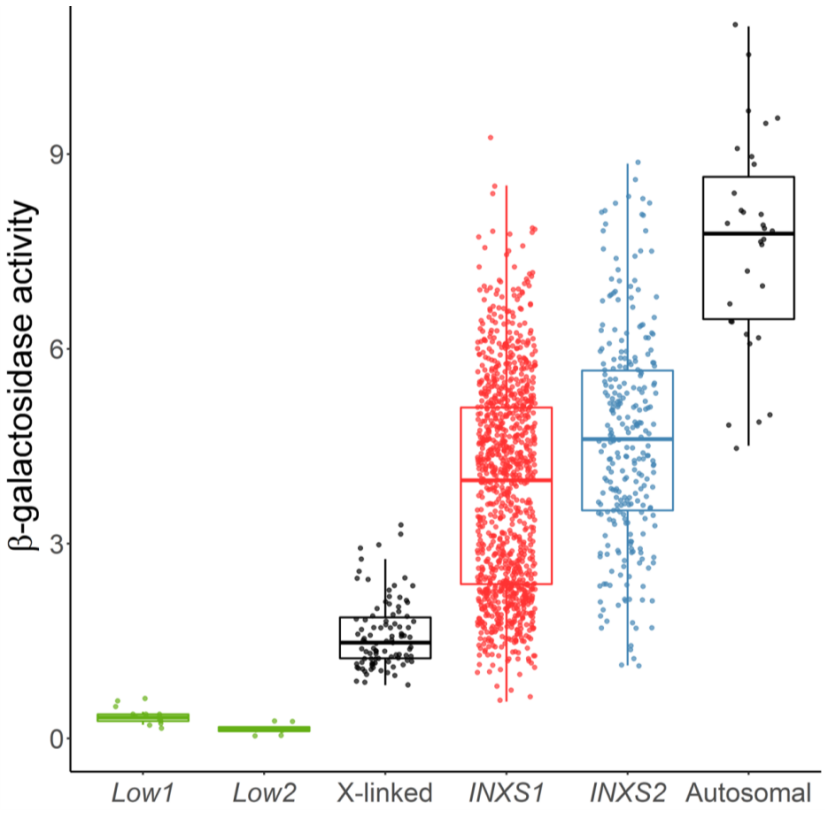


**Figure 3.14.** Reporter gene expression in testis of *Low 1/2* and *INXS 1/2* compared with the X-linked (*wol20X*) and heterozygous autosomal (*wol16A*) controls. Points indicate the level of β-galactosidase activity in one fly. Differences between both mutant lines and the autosomal or X-linked control lines were tested using a *t*-test. All comparisons between both *INXS* and between *INXSs* and controls showed *P*-value $< 0.001$.

In order to determine the chromosomal locations of causative mutations and their dominance, we attempted to test flies with each of the mutant chromosomes replaced by a balancer or wild-type chromosome in a variety of combinations. For this purpose, we

searched for flies with the desired combinations of balancer chromosomes in the *INXS* mutant stocks and additionally performed a series of mapping crosses with flies carrying neither the balancer chromosomes nor the X-linked reporter gene (see crossing protocols in section 2.6). Assuming no recombination, this approach would allow us to exclude the possibility that the causative mutations were on the X chromosome or one of the autosomes if they could be replaced with non-mutant chromosomes and a high level of reporter gene expression was still observed.

We did not see a decrease in the reporter gene expression in the absence of only the mutant second chromosome in both *INXS* mutants, which had similar average expression levels in the testis (Figure 3.15). This suggests it is unlikely that the second chromosome carries the causal mutations. In the absence of only the mutant third mutant chromosome, *INXS1* showed lower expression than *INXS2* (*t*-test, $P < 0.001$). Although most *INXS1* samples had expression patterns in testis close to the X-linked level, the average expression was significantly higher (*t*-test, $P < 0.001$). A slightly different pattern was observed when both autosomes were substituted. In this case, *INXS2* had significantly lower reporter gene activity comparing with *INXS1* (*t*-test, $P < 0.001$) and almost complete recovery of the X-linked expression (*t*-test, $P = 0.052$). Activity in *INXS1* and *INXS2* was equally decreased when only the X chromosome is substituted (*t*-test, $P = 0.150$). Still, it was higher than the activity in the X-linked control (*t*-test, *INXS1*: $P < 0.001$, *INXS2*: $P < 0.01$). Interesting results were obtained for males with all mutant chromosomes, which showed a very high variation in β-galactosidase activity in both *INXS* mutants. In this case, *INXS1* had a slightly lower expression than *INXS2* (*t*-test, $P < 0.001$), possibly because a high proportion of the samples had expression similar to the X-linked control (Figure 3.15).

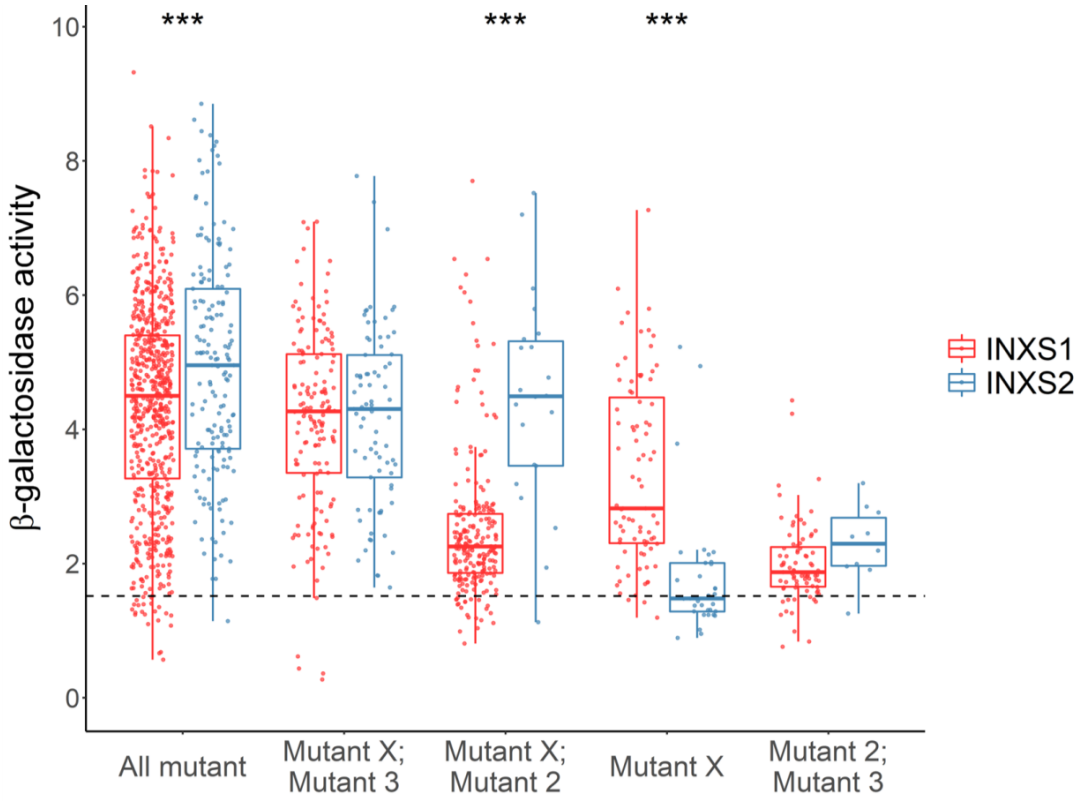


**Figure 3.15.** Reporter gene expression in testis of *INXS1* and *INXS2* with different combinations of mutant chromosomes replaced by balancer or wild-type chromosomes. Points indicate the level of β-galactosidase activity in one fly. The dashed line represents the activity level of the X-linked control. Differences between the two mutant lines were tested using a *t*-test. ***$P < 0.001$.

The reporter gene expression was equally decreased in the absence of the mutant X chromosome in *INXS1* and *INXS2* (Wilcoxon signed-rank test, $P = 0.150$), but it was still higher than expression in the X-linked control (Figure 3.16). To rule out the possibility that the effect of reporter gene up-regulation is specific to the wol20X reporter gene, females of *INXS1* without the mutant X chromosome were crossed to other, independent lines carrying X-linked reporter genes (*wol12X* and *wol5X*)(see the additional crosses in section 2.6). In both cases, *INXS1* showed a moderate increase in β-galactosidase activity compared to controls, suggesting that the target mutation is on the X chromosome (Figure 3.16). However, after crossing with all three X-linked lines, we still saw high expression levels in some samples so we could assume there was some recombination of the genomic region carrying the target mutation between the mutant and non-mutant X chromosomes.

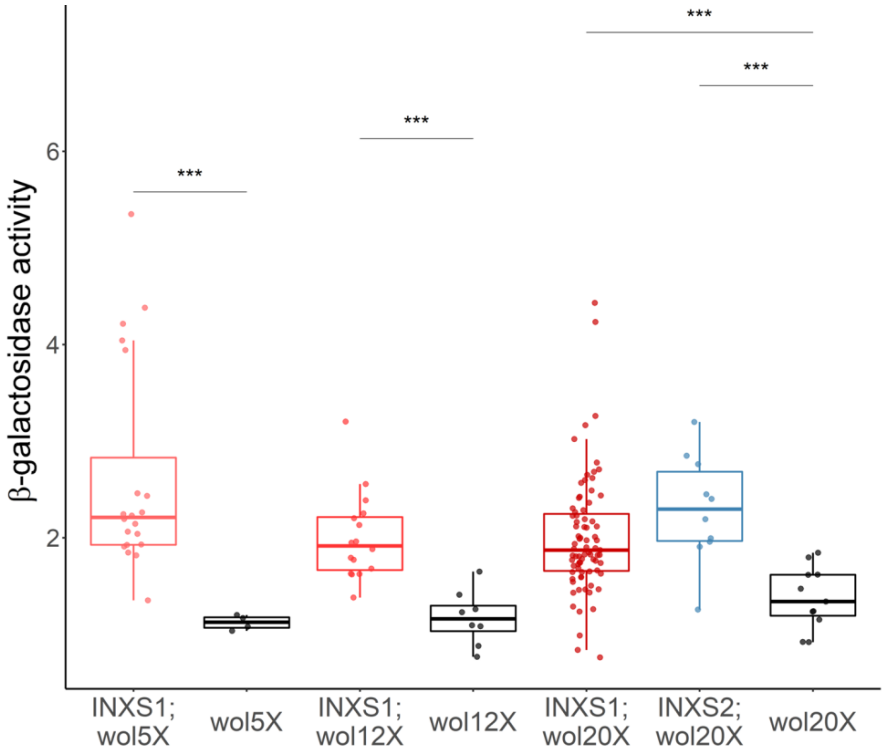


**Figure 3.16.** Reporter gene expression in testis of *INXS1* and *INXS2* male offspring after the additional crosses with non-mutant lines carrying X-linked reporter gene (*wol5X*, *wol12X* and *wol20X*). Points indicate the level of β-galactosidase activity in one fly. Differences between the two mutant lines were tested using a Wilcoxon signed-rank test. ***$P < 0.001$.

A very high variation in reporter gene expression level was typical for both *INXS* mutants and different combinations of chromosomes (Figure 3.15). Moreover, a high variation was observed both between individual replicate cultures of each mutant line and within replicate cultures, where the phenotype of interest may have been lost in successive generations that were particularly pronounced for *INXS1*. However, we could narrow down the possible chromosomes to the X and third chromosomes using these chromosome mapping results.

In addition to the increased expression of the reporter gene, we found almost complete sterility of males in the *INXS1* and *INXS2* lines, whereas we did not find a reduction in female fertility. Of 71 *INXS1* males tested, sterility was observed in 65, and of 79 *INXS2* males tested, 72 males were sterile. The average level of β-galactosidase activity in testis for sterile males were substantially higher than for fertile males (*INXS1*: 2.14 vs. 4.32 mOD/min, *INXS2*: 1.51 vs. 4.65). Thus, it is possible that mutations associated with increased expression of the reporter gene also lead to sterility in males. Microscopic examination of dissected testes

and spermatozoa revealed complete immobility of spermatozoa in all flies examined in both mutants. In both mutants, we did not observe significant changes in the different stages of spermatogenesis and there were clearly visible individual elongated spermatids, which however, did not result in the mature spermatozoa capable of entering the seminal vesicle.

In additions, to visualize β-galactosidase activity in testis of *INXS 1/2* flies and determine whether the activity is ubiquitously increased in the testis or concentrated only in individual germ cell populations, we stained dissected testes using the S-gal reagent. We did not detect any distinctive β-galactosidase activity pattern in the *INXS 1/2* mutants compared with X-linked and autosomal controls (Figure 3.17). However, we found that all dissected testes of mutant males had markedly reduced volume of the seminal vesicles, probably due to the absence of fully-developed motile spermatozoa. Also, the total volume of the testes in the apical region i.e. the approximate area of the testes adjacent to the apical tip, was greater than in control males (Figure 3.17).

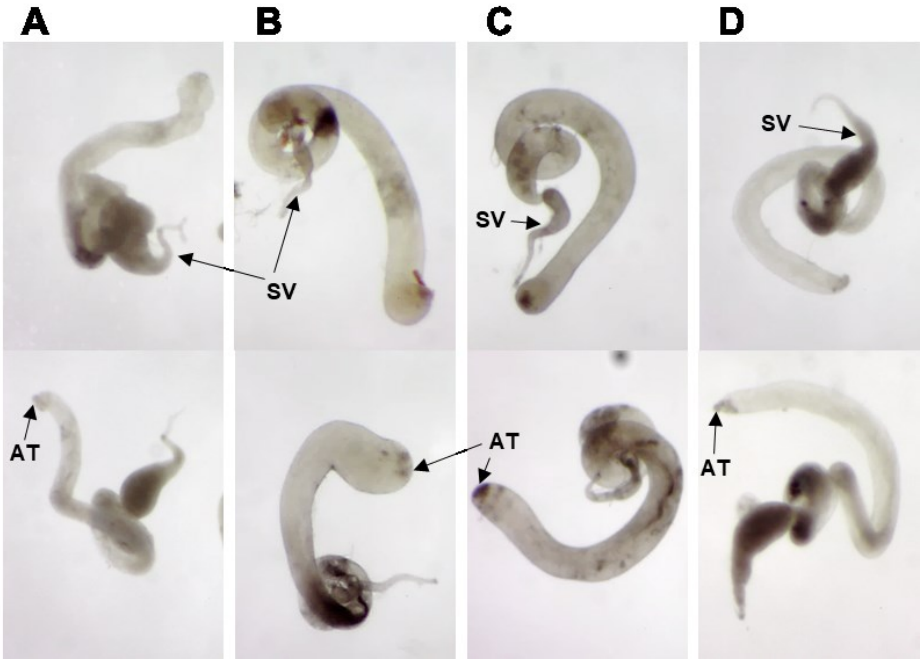


**Figure 3.17.** β-galactosidase activity in the testis. Expression of β-galactosidase in the testis of (*A*) X-linked control (*wol20X*), (*B*) *INXS1*, (*C*) *INXS2* and (*D*) autosomal control (*wol16A*) was visualized by staining with S-gal for six hours. For illustration, the testes of two individuals for each fly line (rows) are shown. All images are shown at the same scale and with the same imaging mode. SV refers to the seminal vesicle of a testis (shown only for the top row). AT refers to the apical tip of a testis (shown only for the bottom row).

### 3.2.2 Endogenous gene expression analysis in *INXS1* and *INXS2*

Although in *INXS 1/2* mutants we observed increased expression of the *wol* reporter gene in the testis and complete male sterility, it is still possible that the mutations associated with these traits do not lead to a general X chromosome upregulation. Therefore, we used RNA-seq analysis of gene expression in *INXS1* testis to measure endogenous X-linked genes' expression in the mutant background. It should be noted that we measured expression in *INXS1* testes using pooled samples. Thus, we could only determine the average level of expression in these fly cultures regardless of the presence of the target mutations. However, there was a strong increase in X-linked expression in the mutant background compared to the X-linked control (*wol20X*) (Figure 3.18*A*). Interestingly, the fourth chromosome also showed an increase in expression, although much less pronounced (Figure 3.18B).

We examined the difference in the number of up- and down-regulated genes between chromosomes with filtered out genes showing low expression, which presumably avoid the most substantial effect of the X suppression. The X chromosome showed the highest proportion and total number of up-regulated genes (Figure 3.19). Although the second and third chromosomes also showed a relatively large number of up-regulated genes, their proportion was more symmetrical to down-regulated genes (Figure 3.18*C*, *D*, *E*, *F*, Figure 3.19).

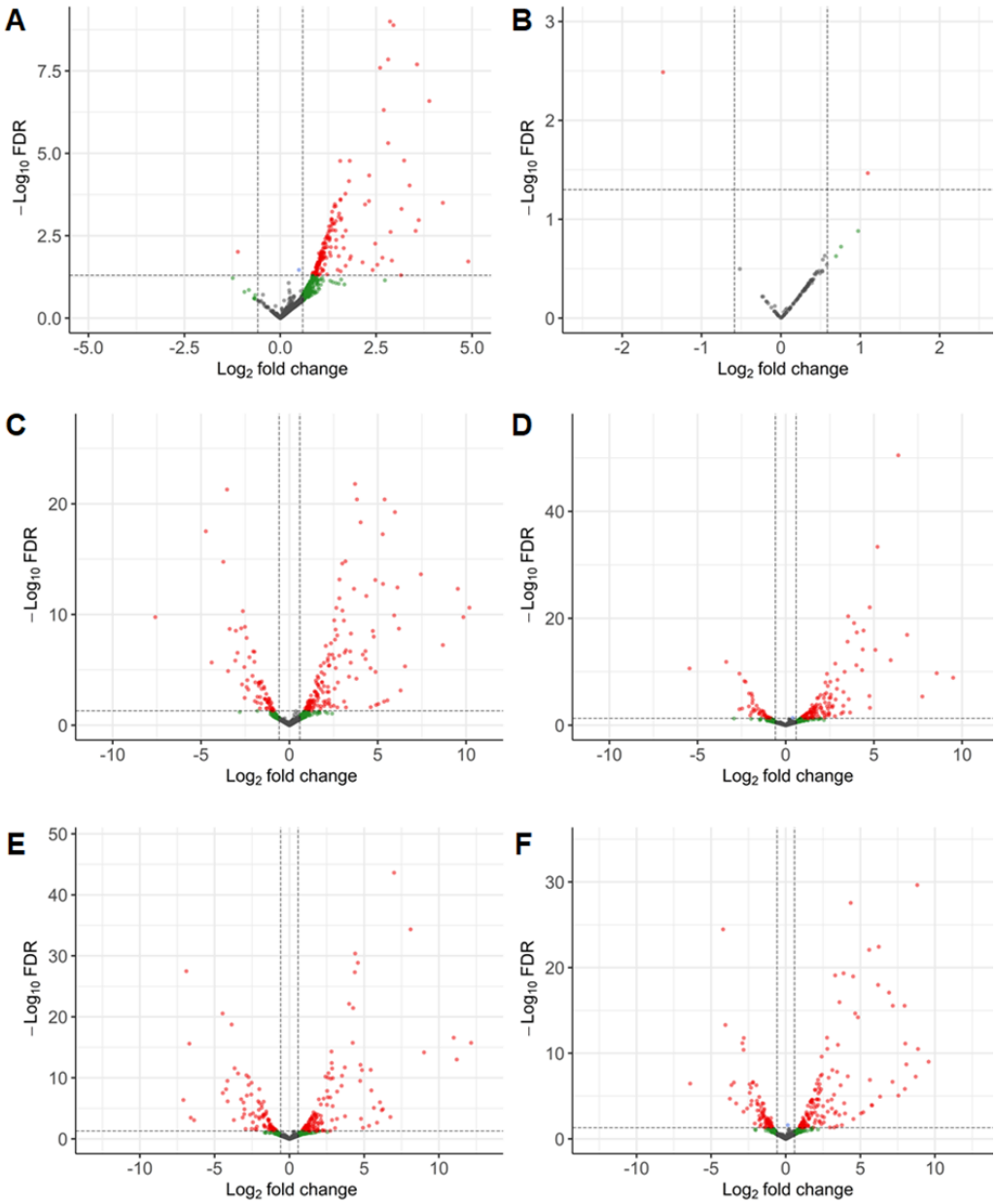


**Figure 3.18.** Expression of endogenous genes in the testis of *INXS1* compared with an X-linked control (*wol20X*). Differential expression analysis was performed for genes located on (*A*) the X chromosome, (*B*) fourth chromosome, and the arms of (*C* and *D*) the second and (*E* and *F*) third autosomes. Each point depicts a single gene with detectable expression in both tested lines. The $\log_2$ fold-change difference in expression between *INXS1* and *wol20X* samples is shown on the X-axis, and the negative $\log_{10}$ of false discovery rate (FDR) is shown on the Y-axis. Differentially expressed genes (red points) were determined based on an FDR α level of 0.05 (horizontal dashed lines) and the fold-change cutoff of 1.5 (vertical dashed lines).

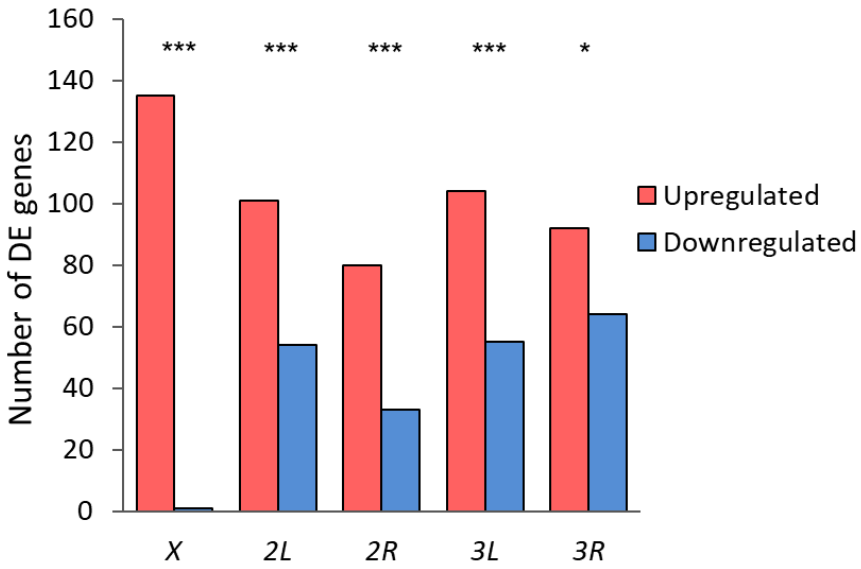


**Figure 3.19.** Differentially expressed genes in the testis of *INXS1* relative to an X-linked control. Differential expression analysis was performed for the X chromosome and the arms of the second and third autosomes. Differentially expressed (DE) genes were determined based on a FDR α level of 0.05, with a minimum fold-change of 1.5 and a minimum normalized read count of 50. Proportions of up-regulated genes were tested with an Pearson's chi-squared test. *$P < 0.05$, ***$P < 0.001$.

To test the effect of *INXS2* on X chromosomal expression, we chose a small set of X-linked testis-specific genes that showed strong up-regulation in *INXS1* and tested them with qRT-PCR in *INXS2* testis. All of them demonstrated a strong up-regulation similar to that observed in *INXS1* (Figure 3.20).

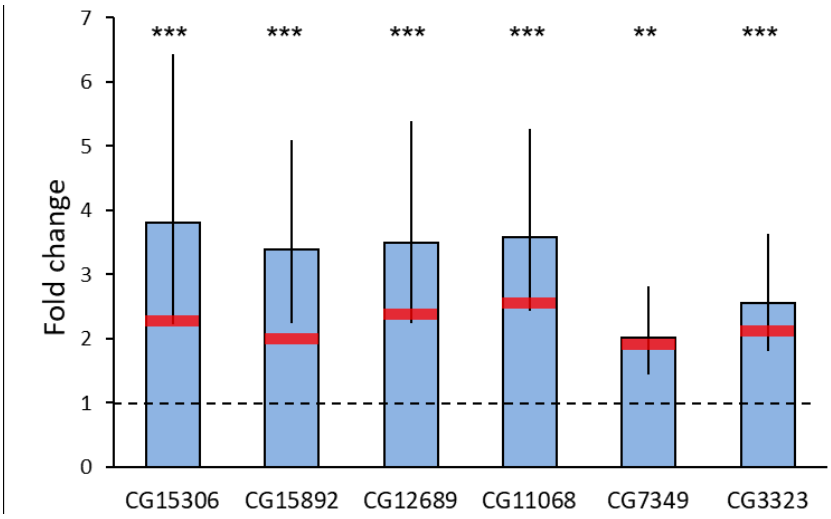


**Figure 3.20.** qRT-PCR analysis of the expression of six X-linked genes in *INXS2*. The fold-change expression difference between *INXS2* and the X-linked control was calculated as 2^-(*ΔΔ*Ct) with error bars indicating the 95% confidence interval. The dashed line represents the fold-change level of the X-linked control. The significance of fold-change was tested using Welch Two Sample *t*-test. ***P* < 0.01, ****P* < 0.001. Red bars indicate a fold-change expression difference for *INXS1* as determined by RNA-sequencing.

### 3.2.3 Mapping causative mutations with whole-genome sequencing

We performed a genome-wide re-sequencing analysis of the two *Low* lines and the two *INXS* lines to map all induced mutations. A single sample per line without balancing chromosomes was used for the *Low* lines, as mapping crosses showed that the causative mutation was on the X chromosome in both lines. We found that both *Low* mutants had nonsynonymous mutations in catalytic domains of the reporter gene, which likely knocked out β-galactosidase activity. *Low1* had a G/A missense mutation at position 1,697 in the *LacZ* gene of the *P[wFl-ocnlacZ]* construct, changing a Gly to an Asp. *Low2* had a G/A missense mutation at position 2,992 of the *LacZ* gene of the *P[wFl-ocnlacZ]* construct sequence, changing an Asp to an Asn.

The situation was somewhat different for the *INXS* mutants, since mapping using classical genetic crosses resulted in only an approximate estimate of the target mutation's location either on the X chromosome or on the third chromosome. Therefore, for genome-wide analysis of each of the *INXS* lines, we used a set of samples with flies carrying, in addition to the mutant X chromosome, also at least one variant of a balancer chromosome and mutant chromosome for each pair of autosomes. We also knew the reporter gene expression level for each sample: all *INXS1* samples were up-regulated, while for *INXS2*, we selected

both samples with increased expression and decreased expression. This enabled us to associate the observed mutations on a particular chromosome to the overexpression and determine whether there was recombination between the balancer and mutant chromosomes. Additionally, we used the available functional annotations from *D. melanogaster* to identify reasonable candidates, such as genes expressed in testis and genes known to cause male sterility when disrupted.

In *INXS1*, we found one candidate gene, *CG13003*, which carried a 12 bp in-frame deletion in an exon at position 16,847,455 and was located on a segment of the X chromosome that in all samples was not recombined with the wild type chromosome (Figure 3.21, Table B11). This mutation perfectly deletes four amino acids from the protein sequence. There was also a T/A point mutation in an intergenic region at position 16,239,180. This site was far from any annotated coding or functional regions of the genome. In *INXS2*, we found a total of 18 point mutations on the unrecombined part of the X chromosome shared by all lines. Of these, two mutations were located within a gene's coding sequence (*CG9784* and *CG1314*), five mutations were in introns (*CG11178*, *Septin4*, *Ggt-1*, *Dop2R*, *DIP-beta*) and three were in 3' UTRs (*ppk23*, *CG8188*, *CG32549*). Eight mutations were in intergenic regions, with three of them located close to coding genes (*CG11584*, *ppk23*, *RunxA*). However, one *INXS2* sample with low activity had the wild-type sequence and did not match the mutant lines with high activity after position 16,693,568. Accordingly, it is most likely that the causative mutation is located after position 16,693,568, leaving *CG1314* as the top candidate gene (Figure 3.21, Table B11). In this gene, there was a G/A mutation at position 20,869,041, changing a Gly to an Asp. We found no signs of differential expression in the testis on the X chromosome for both top candidate X-linked genes (based on FDR α level of 0.05 and 1.5 fold-change cutoffs).

On the second chromosome of *INXS1*, we found one candidate gene, *CG17344*, with an A/T mutation, changing a Ser to an Cys. This gene is located within the recombined region that spanned parts of 2L and 2R and was present in samples with and without mutant chromosomes. In *INXS2*, there was no genomic region with shared mutations between all high expression samples.

On the third chromosome in both *INXS1* and *INXS2*, we found a shared candidate gene *CG31525*, which had different nonsynonymous mutations in the two mutants. In *INXS1*, a

missense mutation is located at position 4,427,425, and in *INXS2* one is located at position 4,427,171. The former changes a Ser to a Phe. The latter disrupts the start codon and is, thus, likely to completely knock out gene function (Figure 3.22, Table B11). In both mutants, this gene was located on the chromosome region recombined with balancer chromosomes. Also, the third chromosome in *INXS1* showed signs of recombination with balancers in the region, carrying a G/A mutation at position 19,954,508 in a start-codon of the gene *CG42654* (Figure 3.22, Table B11).

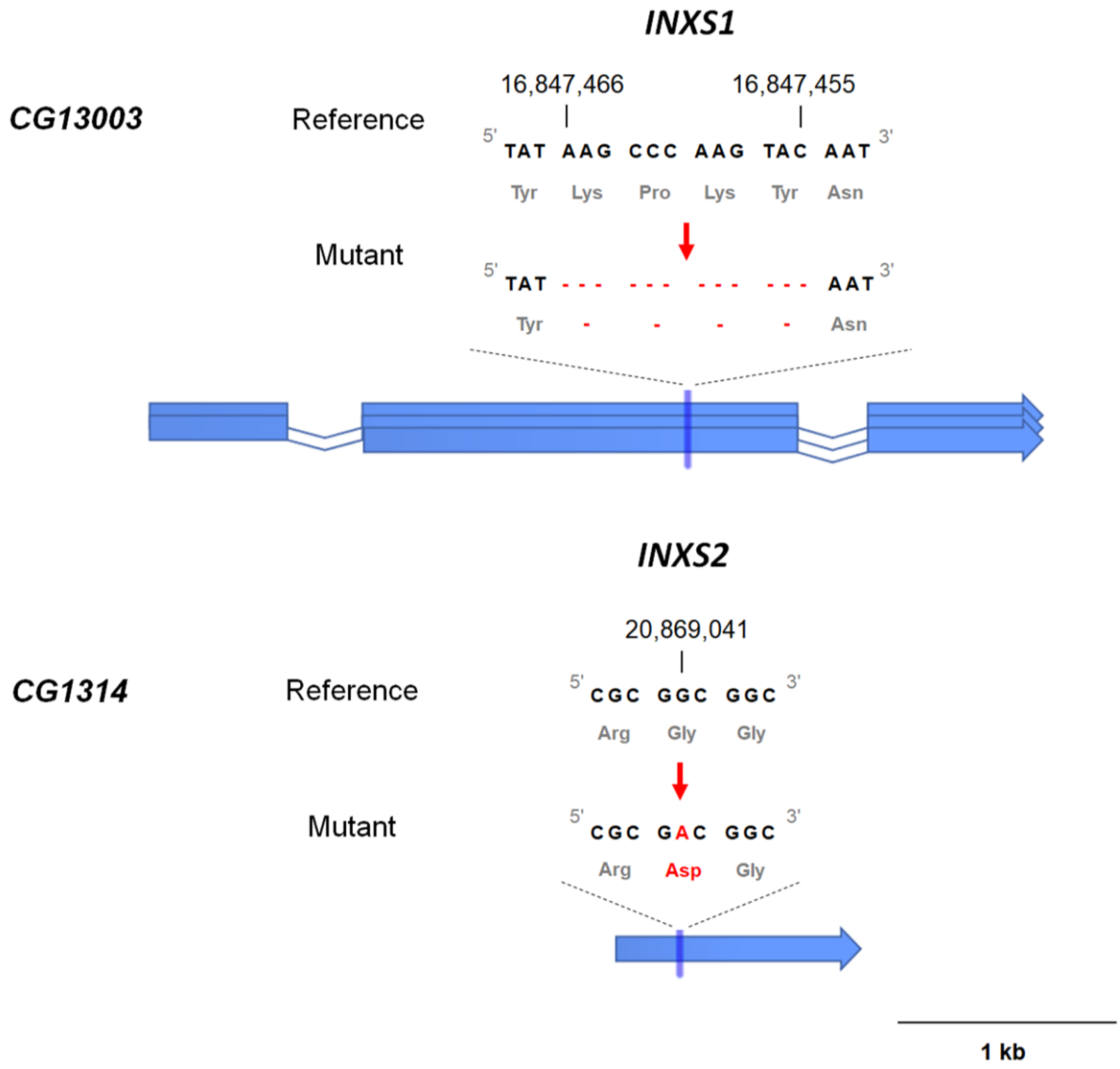


**Figure 3.21.** Schematic representation of the top candidate genes on the X chromosome (*CG13003*, *CG1314*). The nucleotide and amino acid sequences of the *D. melanogaster* reference genome and *INXS* 1/2 mutant genome are shown for each gene. Blue arrows represent coding regions of genes with mutation sites marked by blue vertical lines.

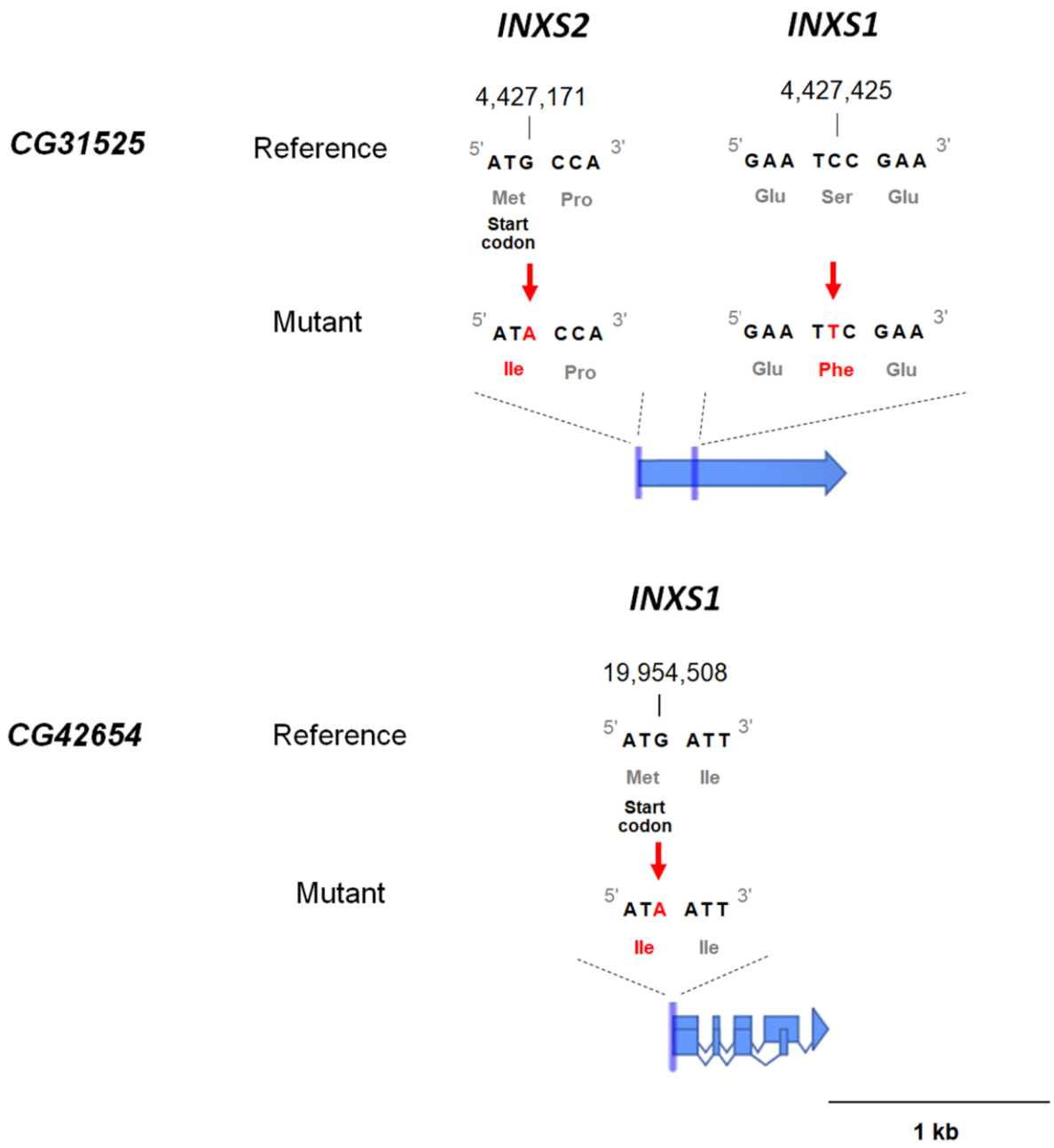


**Figure 3.22.** Schematic representation of the top candidate genes on the third chromosome (*CG31525*, *CG42654*). The nucleotide and amino acid sequences of the *D. melanogaster* reference genome and *INXS* 1/2 mutant genome are shown for each gene. Blue arrows represent coding regions of genes with mutation sites marked by blue vertical lines.

### 3.2.4 Associations between different mutations and increased reporter gene expression in *INXS 1/2* mutants

Sequencing results showed that several recombination events occurred in *INXS* lines between mutant chromosomes and wild-type chromosomes or balancers. Thus, causative mutations are likely to segregate in the mutant stocks. Using mutation markers in the top candidate genes and markers for recombination obtained from the whole-genome sequencing, we aimed to test the association of these mutations with increased reporter gene expression and narrow down the possible genomic locations of the causative mutations. For this purpose, we screened for the markers (Table B11) in males with different combinations of balancers and different levels of β-galactosidase activity in testes using a combination of PCR, restriction enzyme digestion and Sanger sequencing (Table 3.2).

In *INXS1*, on the X chromosome, we found associations between increased β-galactosidase activity in the testis and mutations in the candidate gene *CG13003* and another gene *Ucp4A*, which is located one mega base (Mb) away (Table 3.2). Recombination markers on the second chromosome showed no association with increased testis' β-galactosidase activity in the tested males. On the third chromosome, increased activity was linked to the mutations in the genes *CG42654* and *Ipp*, suggesting that the causative mutation is located in the region of these two genes. Interestingly, however, we could not see the same association pattern for our top candidate gene, *CG31525* (Table 3.2).

In *INXS2*, we found a relatively strong association between increased β-galactosidase activity and the mutations in the top candidate genes *CG1314* and *CG31525*, which had mutations in all males with high reporter gene expression (Table 3.2).

**Table 3.2.** Association between mutation markers in different genes and β-galactosidase activity in the testis.

| Mutant | Chr.[1] | Coord. (Mb)[2] | Gene[3] | N (Low)[4] | N (High)[4] | $f$(Low)[5] | $f$(High)[5] | $f$(High) - $f$(Low) | $p$-value[6] |
|---|---|---|---|---|---|---|---|---|---|
| *INXS1* | X | 2.64 | *sgg* | 7 | 17 | 0.29 | 0.18 | -0.11 | 0.874 |
| | | 5.33 | *Mcm3* | 7 | 17 | 0.29 | 0.18 | -0.11 | 0.874 |
| | | 16.85 | *CG13003* | 23 | 34 | 0.39 | 0.68 | 0.29 | 0.032 |
| | | 17.85 | *Ucp4A* | 24 | 27 | 0.33 | 0.78 | 0.44 | 1.6E-03 |
| | 2 | 7.68 | *LRR* | 5 | 20 | 0.40 | 0.35 | -0.05 | 0.770 |
| | | 19.1 | *CG17344* | 1 | 5 | 1 | 0.40 | -0.60 | 1 |
| | | 20.1 | *Kua* | 5 | 20 | 0.40 | 0.40 | 0 | 0.699 |
| | 3 | 4.43 | *CG31525* | 21 | 36 | 0.33 | 0.56 | 0.22 | 0.089 |
| | | 13.96 | *Ipp* | 7 | 17 | 0.57 | 0.82 | 0.25 | 0.215 |
| | | 19.95 | *CG42654* | 7 | 17 | 0.43 | 0.88 | 0.45 | 0.038 |
| *INXS2* | X | 4.43 | *CG31525* | 14 | 12 | 0.43 | 1 | 0.57 | 1.9E-03 |
| | 3 | 20.87 | *CG1314* | 13 | 20 | 0.46 | 1 | 0.54 | 4.0E-04 |

[1]Chromosome containing the mutation marker to be tested; [2]Chromosomal coordinate of the mutation markers; [4]Total count of tested males showing "Low" (less than 2.750 mOD/min) and "High" (greater than 2.750 mOD/min) β-galactosidase activity; [5]Ratio of number of males with mutant allele to total number of tested males in each group;[6]Proportions of males carrying the mutation in both group were compared using Fisher's Exact Test test. Cells with bold font indicate a significant difference in the proportions ($p$-value < 0.05).

### 3.2.5 Functional tests of the *CG13003*, *CG1314* and *CG31525* candidate genes

To test whether the candidate genes *CG13003*, *CG1314* and *CG31525* are responsible for the observed lack of X chromosome suppression in the male germline, we tried to knock down the expression of these genes using targeted RNAi with the UAS/Gal4 system (Figure 1.3). We tested promoters driving ubiquitous (*Act5C*) and germline Gal4 expression (*Bam*: late spermatogonia and early spermatocytes, *Nanos*: germline stem cell and spermatogonia, *TJ*: cyst stem cell and cyst cells) (White-Cooper, 2012; Yu et al., 2016).

We performed a preliminary test of this approach for *CG13003* gene by measuring β-galactosidase activity in testes of *RNAi-CG13003/Gal4* males and control males with only Gal4 (Figure 3.23). We found increased expression of the reporter gene when the target gene

expression was knocked down with the *Bam* and *TJ* Gal4 promoters (Figure 3.23). However, the large variation in activity makes it difficult to draw clear conclusions about the effect of this gene knockdown.

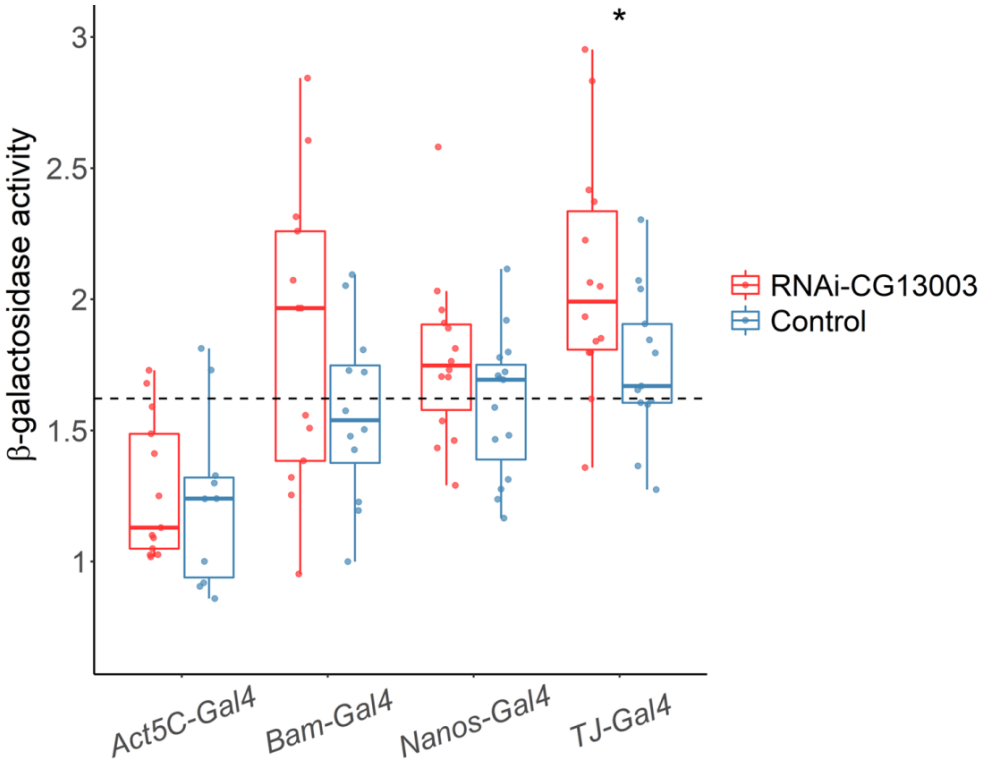


**Figure 3.23.** β-galactosidase activity in testis of flies with RNAi-mediated *CG13003* knockdown. β-galactosidase activity was measured in males carrying the Gal4 activator of UAS (RNAi-*CG13003*) with promoters specific for various stages of germline development (*Bam, Nanos, TJ*) and a promoter driving ubiquitous expression (*Act5C*). Males carrying only the Gal4 gene with the promoter of interest were used as a control. Points indicate the level of β-galactosidase activity in one fly. Differences between the knockdown and control lines were tested using the Wilcoxon signed-rank test. *$P < 0.05$.

Next, we used qRT-PCR to test if the expression of endogenous genes on the X chromosome was changed after each candidate gene's knockdown. For RNAi-mediated knockdown of the *CG13003* candidate gene, we measured the expression of two male-biased testis-specific genes (*CG12689* and *CG11068*), which showed significant up-regulation of expression in the *INXS 1/2* mutants (Figure 3.20). For knockdowns of the *CG1314* and *CG31525* candidate genes, we measured only *CG11068* expression. We also tested RNAi

efficiency by measuring each candidate gene's expression in male testis after the knockdown and comparing it with the control level (Table 3.3).

To test the effect of *CG13003* knockdown, we used the *Bam* and *TJ* Gal4 promoters. For both promoters, the X-linked genes showed no significant increase in expression (Table 3.3). To test *CG31525* gene silencing, we used *Bam* promoter and for the gene *CG1314* silencing, we used *Bam* and *Act5C* promoter. Both candidate target genes also showed no significant increase in X-linked gene expression (Table 3.3). At the same time, the knockdown of all three candidate genes using the *Bam* promoter resulted in a significant decrease in X-linked gene expression (Table 3.3). However, it is interesting that the ubiquitous expression of RNAi for *CG13003* and *CG31525* genes driven by the *Act5C* promoter resulted in a lethal male phenotype. The efficiency of RNAi knockdown was relatively high for the *Bam* promoter but only 20–26% for the *TJ* and *Act5C* promoters (Table 3.3).

**Table 3.3.** Expression of the two endogenous X-linked genes in testes after RNAi-mediated knockdown of the *CG13003*, *CG1314* and *CG31525* candidate genes.

| RNAi target gene[1] | Gal4 promoter[2] | Tested gene[3] | Fold-change[4] | CI (Lower)[5] | CI (Upper)[5] | *p*-value[6] | Efficiency (%)[7] | *p*-value[8] |
|---|---|---|---|---|---|---|---|---|
| *CG13003* | *Bam* | *CG12689* | 0.90 | 1.32 | 0.62 | 0.095 | 95 | 1.9E-08 |
| | | *CG11068* | 0.81 | 1.14 | 0.57 | 0.025 | | |
| | *Tj* | *CG12689* | 0.91 | 1.32 | 0.62 | 0.164 | 26 | 0.051 |
| | | *CG11068* | 0.86 | 1.55 | 0.48 | 0.145 | | |
| *CG1314* | *Bam* | *CG12689* | 0.91 | 1.46 | 0.57 | 0.097 | 66 | 2.0E-04 |
| | *Act5c* | *CG12689* | 0.93 | 1.79 | 0.48 | 0.242 | 20 | 0.012 |
| *CG31525* | *Bam* | *CG12689* | 0.72 | 1.05 | 0.5 | 0.012 | 98 | 6.1E-07 |

[1]UAS-RNAi targeted at each of the candidate genes (*RNAi-CG13003*, *RNAi-CG1314*, *RNAi-CG31525*); [2]Gal4 promoters specific for various stages of germline development (*Bam, TJ*) and a promoter driving ubiquitous expression (*Act5C*). [3]The expression of two endogenous X-linked genes (*CG12689* and *CG11068*) in male testes was measured using qRT-PCR. [4]The fold-change of expression difference relative to the the X-linked control was calculated as 2^-(*ΔΔ*Ct); As a control, the male offspring after the crosses of females carrying the *Gal4* coding sequence linked to one of the promoters (*Bam, TJ, Act5C*) and males from the RNAi host strains (*60100*, *60000*) were used. [5]95% confidence intervals. [6]The significance of fold-change for expression of the X-linked genes was tested using Welch's two sample *t*-test; [7]RNAi knockdown efficiency was calculated as (1 - fold-change)*100; [8]The significance of fold-change for expression of the RNAi target genes was tested using Welch's two sample *t*-test.

Additionally, we tested male fertility for RNAi targeted to *CG13003*, *CG1314*, *CG31525* for all promoters of Gal4 (Table B12). In the test of *RNAi-CG13003*, a significant increase in the proportion of sterile males was observed only for the *Bam* driver. The *TJ* and *Nanos* drivers showed only a small proportion of sterile males. In the test of *RNAi-CG1314*, we also found a small number of sterile males when using the *Act5C* promoter (Table B12).

# Chapter 4

# Discussion

A striking feature of living organisms is that they show a remarkable diversity of phenotypes between species, populations, individuals and sexes. However, a large proportion of this diversity is caused not by differences in gene composition between organisms but by the differential expression of shared genes. The X chromosome of *D. melanogaster* has a distinct evolutionary history from the autosomes and has acquired specific regulatory mechanisms acting exclusively in males, such as X chromosome-specific dose-compensation in somatic tissues and X suppression in the germline. This creates a unique environment for gene expression on the X chromosome, leading to an interplay of gene-specific and chromosome-wide regulatory mechanisms. A good example of such interplay is the differential representation of sex-biased genes on the X chromosome, which differs from that on the autosomes in various tissues (Huylmans and Parsch, 2015).

## 4.1 Effect of an exogenous reporter gene's proximity to DCC binding sites on its expression in different sexes

The first part of my thesis is devoted to X chromosome dosage compensation mediated by the DCC in somatic tissues of *D. melanogaster* males and its effect on sex-biased gene expression. We found a negative correlation between a reporter gene's male-to-female expression ratio and its distance to the nearest DCC binding sites in all somatic tissues that we tested. Although the median values of the male-to-female expression ratio were similar in the brain and head case, there was greater positional effect variation in the brain, which may

contribute to the relative excess of X-linked sex-biased genes observed in this tissue. In contrast to somatic tissues, no significant effect of DCC binding sites on sex-biased reporter gene expression was found in the gonads, which include mainly germline cells, suggesting the absence of a functioning DCC. These results may provide the first evidence for a direct effect of DCC on sex-biased gene expression in all somatic tissues of *Drosophila*.

### 4.1.1 Proximity to a DCC binding site is the main determent of sex-biased reporter gene expression in male somatic tissues

We observed a negative correlation between the proximity of a reporter gene insertion to the DCC binding sites and its expression in males and the male-to-female expression ratio for head and carcass, which is consistent with dosage compensation affecting the expression of X-linked genes in males (Figure 3.6, 3.7*A*, *B*, A4*A*, Table B4) (Bachtrog et al., 2010; Chang et al., 2011; Huylmans and Parsch, 2015). The proximity effect was found for HAS and the binding sites of the individual DCC components (MLE, MSL2, MSL3), which increases the robustness of our results. It also confirms that the binding sites of different components of the complex are determinants of complex activity at a given locus (Straub et al., 2013). The effect of the proximity of the reporter gene to the DCC binding sites on its sex-biased expression in somatic tissues was also supported by testing the small set of insertions in homozygous females and hemizygous males (Figure 3.10*B*, A6).

Our results for the carcass differ from those observed in a previous study of endogenous gene expression in gonadectomized flies (Parisi et al., 2003), where a positive correlation between male-biased expression and the distance to the nearest DCC binding site was observed (Parisi et al., 2003). That is, the relationship between DCC distance and male-biased expression in the previous study was more similar to that seen in gonad than in somatic tissues. However, these expression data were from an early microarray study that detected only a few hundred sex-biased genes (Parisi et al., 2003), which may limit statistical power. Furthermore, the previous gonadectomized sample appears to have included residual gonadal transcripts, as the expression of several testis-specific genes was detected (Vensko and Stone, 2014). It is likely that the presence of gonad tissue may have a strong influence on the expression pattern in the whole body, as the gonads are enriched with sex-biased genes, especially those with a very high degree of sex bias. Indeed, this is what has been reported for whole-fly expression data, where the sex-biased expression pattern was similar to that seen in the gonad (Huylmans and Parsch, 2015). Therefore, in contrast to previous studies, we found

no significant difference in the pattern of gene expression in relation to its proximity to a DCC binding site in the head and/or brain compared to the carcass (Figure 3.6, 3.7*A*, *B*, A4*A*, Table B4, B10), so we believe that the proximity to a DCC binding site has a similar effect on a gene expression in all somatic tissues.

In the brain, we observed an increased level of expression variation among different insertions on the X chromosome (Figure 3.12). This may explain, at least partly, the strong enrichment of sex-biased genes, particularly male-biased genes, on the X chromosome compared to other tissues (Huylmans and Parsch, 2015; Khodursky et al., 2020). Because the sex-biased expression is calculated as the ratio of the expression measured independently in each sex, the variation will be amplified in the male/female ratio, leading to greater variation and more cases of sex-biased expression.

We found no significant correlations when we examined insertions with all possible distances (i.e., without imposing a 50 kb cutoff) except in female carcass and head, where expression was negatively correlated with the distance to the nearest MSL3 binding site (Table B5). This supports the idea that the major effect of the DCC on expression might be limited to short-range contacts within 50 kb in the active nuclear compartments (Schauer et al., 2017). However, the chromatin's spatial structure possibly facilitates long-range, interdomain DCC-chromatin interactions since HAS is more often located near the contact-reach TADs boundaries (Ramírez et al., 2015; Schauer et al., 2017). Thus it can be expected that a reporter gene may be strongly affected by spatially close HASs due to chromosome folding, but there may be no correlation with linear distance along the chromosome.

In females, we found no significant correlation between the reporter gene expression and its distance to the DCC binding sites in all tissues (Figure 3.7*C*, A4*B*, Table B4). This could be expected because the DCC does not assemble in females due to the suppression of MSL2 expression by SXL protein (Kelley et al., 1995). There was, however, a non-significant negative correlation for most DCC components in the female carcass, raising the possibility that these sites may also influence expression in females (Gallach and Betrán, 2016). The negative correlation in females may also be due to the tendency of DCC binding sites to be located in sites with active chromatin, where the chromatin environment itself can lead to the increased expression of genes (Ramírez et al., 2015).

### 4.1.2 No evidence for an effect of proximity to DCC binding sites on the reporter gene's sex-biased expression in the gonads

The pattern of reporter gene expression in the gonads was significantly different from that in somatic tissues. Firstly, the results of sex-biased expression in the gonads showed that the reporter genes had much higher expression in females than in males, which is consistent with the absence of dosage compensation in the male germline (Rastelli and Kuroda, 1998; Meiklejohn et al., 2011; Meiklejohn and Presgraves, 2012) and the general excess of female-biased gene expression observed in the ovaries (Huylmans and Parsch, 2015).

Secondly, there was no significant correlation between the male-to-female expression ratio and proximity to the DCC binding sites in gonads (Figure 3.6, 3.7, Table B4). The only exception was a negative correlation for HAS among homozygous and heterozygous females (Figure A8, Table B6). Interestingly, when looking at expression in males only, there was a weak negative correlation (Figure 3.7*B*, A4*A*, Table B4). This may be due to a small proportion of somatic cells in gonads, where dosage compensation may still occur. Furthermore, X chromosome-specific dosage compensation could also occur in somatic cells and certain types of germline cells in the testis, which may be mediated by an alternative dosage compensation mechanism utilising the DCC binding sites (Gupta et al., 2006; Hense et al., 2007; Witt et al., 2021). This is also supported by recent findings on dosage compensation in testes, where the somatic and early germ cell genes' position within 10 kb of HAS was associated with higher expression compared to that of genes located further away (Witt et al., 2021). Although previous studies of endogenous gene expression reported a positive correlation between male-biased expression and the distance to the nearest DCC binding site in the gonad, this was thought to be the result of selection to prevent interference between the dosage compensation machinery and the gene-specific regulation of testis-biased genes, which typically display strong tissue-specific regulation (Bachtrog et al., 2010; Huylmans and Parsch, 2015). Our results are consistent with this interpretation, as our reporter genes should not be affected by the sex- or tissue-specific regulation that affects the native *D. melanogaster* genes.

### 4.1.3 Dosage compensation and sex-biased gene expression in different taxa

We expect that the effects of dosage compensation on sex-biased gene expression reported here should be present only in taxa in which there is up-regulation of the hemizygous sex chromosome in the heterogametic sex. In mammals, where sex chromosome dosage

compensation occurs through inactivation of one of the X chromosomes in females, it is more likely that sex-biased expression will be influenced by variation in inactivation across the X chromosome, leading to female-biased expression of genes in regions that escape X-inactivation (Carrel and Willard, 2005). In some female heterogametic taxa, such as birds, sex-biased expression is more likely to be influenced by an absence of dosage compensation in females, leading to widespread male-biased expression of Z-linked genes (Itoh et al., 2007; Ellegren et al., 2007). It has recently been reported that a female-heterogametic Lepidopteran with a neo-Z chromosome (Monarch butterfly) displays two distinct modes of sex chromosome dosage compensation (Gu et al., 2019): the ancestral Z is down-regulated in ZZ males in a manner similar to that seen for the X chromosome in *Caenorhabditis elegans* females, while the neo-Z is up-regulated in ZW females in a manner similar to that seen for the X chromosome in *D. melanogaster* males. The molecular mechanism responsible for neo-Z chromosome up-regulation in Monarch females is not fully understood but, similar to *Drosophila*, it is associated with H4Ac16 (Gu et al., 2019). It is currently not known whether the degree of this up-regulation varies with Z-chromosomal location in a way analogous to that seen in *D. melanogaster* males.

Overall, we observed a substantial variation in reporter gene expression in all tissues attributable to the unique chromatin environment of the individual insertions. Therefore, we tested assumptions that our reporter gene been influenced by different local regulatory factors. We found that the male-to-female expression ratio was positively correlated between different tissues, which suggests that tissue-independent factors affect the sex-biased expression of the reporter genes. Furthermore, the reporter gene expression showed little evidence of being influenced by local sex-specific regulatory elements associated with endogenous *D. melanogaster* genes (Figure 3.4). Also, we found no significant change in the effect of DCC proximity on the expression pattern of the reporter gene if it was flanked by G4s, suggesting that there is no effect of G4s on DCC spreading (Qian et al., 2019). However, in all tissues, we found a positive correlation between the reporter gene's proximity to a nearby G4 location and its expression in females and males (Figure 3.9, Table B7). Our results support the putative regulatory role of G4 structures in gene transcription (Siddiqui-Jain et al., 2002; Hänsel-Hertsch et al., 2016; Chen et al., 2018; Makowski et al., 2018). Thus, by using the exogenous reporter gene, we were able to detect the major effect of DCC binding while avoiding the confounding factor of gene-specific regulation.

## 4. Discussion

To compensate for individual differences between the insertions in our study, we generated many individual insertions on the X chromosome. For mobilisation genetic crosses, we took advantage of transposable *P* elements as part of a reporter gene construct, as *P* elements are known to have an insertion bias to euchromatin and genic regions (Bellen et al., 2004). Thus, all reporter gene insertions we obtained are within 10 kb of the active genes, and a large proportion of them are located in genic regions. This increased the statistical power of our study, since DCC binding sites are known to be located in transcriptionally active parts of the chromatin and are located in active gene bodies (Alekseyenko et al., 2006; Gilfillan et al., 2006; Straub et al., 2008; Schauer et al., 2017). Indeed, the 50 kb cutoff for the distance between an insertion and the nearest DCC binding sites included about 70 percent of all insertion locations (Figure 3.5). Thus, our results can be compared with the results obtained for endogenous sex-biased genes, most of which were located within 50 kb from the DCC binding sites (Huylmans and Parsch, 2015).

It is worth noting some aspects of using the reporter gene with a CMV promoter in our study. Firstly, the reporter gene introduced to the X chromosome is not fully dosage compensated (Figure 3.10*A*), which has been reported previously for somatic tissues, suggesting that it does not experience the full effect of DCC upregulation (Laurie-Ahlberg and Stam, 1987; Parsch et al., 1997; Argyridou and Parsch, 2018). Moreover, most of the insertions showed female-biased expression in the head, supporting the moderate effect of DCC in this tissue (Figure 3.2). The reasons for this are unclear, and such a pattern was not observed in a previous study that used a smaller number of reporter gene insertions (Argyridou and Parsch, 2018). Secondly, the accuracy of our results could be affected by the very low levels of β-galactosidase activity in testes, even though there are two copies of the reporter gene coding sequence in the construct. Nevertheless, it was above that of negative controls (non-transgenic flies), for which the mean activity was zero (Argyridou and Parsch, 2018). Also, the absence of significant differences in the level of variation of β-galactosidase activity between the testis and other tissues suggests that reporter gene expression can be measured reliably in this tissue (Figure A1). Thirdly, the use of a CMV promoter may have been advantageous, as it drives low expression in testis and probably is not subjected to the effect of the X suppression (Landeen et al., 2016; Argyridou and Parsch, 2018).

## 4.2 Genetic screening for genes involved in the X chromosome suppression

In the second part of my thesis, we attempted to discover genes responsible for the chromosome-wide suppression of X-linked genes in the *D. melanogaster* male germline. Through chemical mutagenesis and genetic screening, we obtained two mutant lines, *INXS*1 and *INXS2*, showing an increase in testis-specific reporter gene expression, general up-regulation of X-linked gene expression in the male germline and male sterility, which may indicate a disruption of X chromosome suppression. In the genetic background of these mutant lines, we identified four genes with unknown functions on the X and third chromosomes potentially involved in the observed phenotype. Thus, we have successfully generated the first mutant lines defective in X suppression in the male germline and identified several X-linked and autosomal genes that may be involved in this mechanism.

### 4.2.1 The *INXS1* and *INXS2* mutant lines show defects in X suppression

We employed an EMS-based forward genetic screen to generate fly lines with random mutations that show an increase in the expression of an X-linked *wol* reporter gene. We were able to identify two independent lines with increased β-galactosidase activity in the testis that we named *INvolved in X Suppression*: *INXS1* and *INXS2*. Compared to the control, which was the original line with the X-linked reporter gene (*wol20X*) that was subjected to mutagenesis, we obtained a median increase in β-galactosidase activity of 2.2-fold for *INXS1* and 2.7-fold for *INXS2* (Figure 3.14). For both mutants, β-galactosidase activity in the upper quartile of all flies was similar to the expression in the autosomal line, suggesting a complete disruption of the X suppression.

In a preliminary mapping of the chromosomal position of causative mutations in which we replaced different mutant chromosomes with wild-type chromosomes or balancers, we found that both *INXS* mutants show a decrease in the reporter gene expression without the mutant X chromosome (Figure 3.15). A similar pattern was observed when the second and third chromosomes were replaced, with an expression decrease in *INXS1* and *INXS2*, and with an expression decrease in *INXS2* when only the third chromosome was replaced. Thus, we can assume that the causative mutant alleles are on the X chromosome and possibly on the third. We could exclude the second chromosome from subsequent analysis, as it did not show any significant changes when replaced.

## 4. Discussion

In almost all analyses, we observed a very high degree of variation in reporter gene expression in mutant males, which is particularly evident when all mutant chromosomes are present in the tested males (Figure 3.15). Also, in cases where the second and third chromosomes or only the third chromosome were replaced in *INXS1*, even when median expression decreased, a considerable proportion of samples still showed increased expression. This suggests segregation of causative mutant alleles in the mutant lines, which would be expected since both mutant lines have been maintained for multiple generations. Both *INXS* mutants may have experienced rare recombination events between balancer autosomes or the wild-type X chromosomes with mutant chromosomes (Miller et al., 2016), or possibly and spontaneous mutations (Alexander, 2015; Huang et al., 2016; Mérel et al., 2020), which could result in heterogeneous mutant and balancer chromosomes in individual replicate cultures due to genetic drift. This is supported by the low level of variation in the replacement of the mutant X chromosome, as the tested males were derived from two genetic crosses of several parental lines, which were most likely isogenic. Moreover, within each replicate culture, the phenotype of interest may have been lost in successive generations, something that was particularly pronounced for *INXS1*. Another possible reason for the high variation could be incomplete penetrance, i.e. a reduced frequency of the expression of a trait in the background of the determining genotype. Such variation in trait expression could be due to variation in the genetic background of the tested individuals and stochasticity in gene expression or chemical interactions of its expression products (Raj and van Oudenaarden, 2008; Koellner et al., 2020). Also, in most cases, when a mutant chromosome was replaced, leading to a decrease in reporter gene expression, the median level was still higher than the X-linked control. For example, this is apparent when the X chromosome in *INXS 1/2* or the third chromosome in *INXS1* is replaced (Figure 3.15), which suggest some undetermined effect of the mutant background.

We found 92% sterility in males of the *INXS1* line and 91% for *INXS2* with signs of defective spermatogenesis, whereas females retained full fertility. Microscopic examination of dissected testes showed that spermatogenesis, whose individual phases were distinct from the apical tip to the base of the testis, did not produce motile spermatozoa capable of transferring into the seminal vesicle. Morphologically, such testes were characterized by the reduced volume of the seminal vesicles lacking motile spermatozoa and an increased total volume of the testes almost along the entire length of the apical region (Figure 3.17). Presumably, both mutants have mutations affecting postmeiotic differentiation and elongation, with the germ

cells having gone through all early proliferation and meiosis but without forming mature motile spermatozoa (Castrillon et al., 1993). Mutations at this stage have been shown to be among the most common and result in male sterility and may be associated with the testes morphological traits we observed (Castrillon et al., 1993). Also, male sterility is often caused by mutations that disrupt chromosome segregation during meiosis I (McKee et al., 1998; Kanippayoor et al., 2020). Our results have shown that male sterility in both mutant lines is associated with the X chromosome. Although the X chromosome carries significantly fewer loci associated with male sterility and testis-specific expression genes (Parisi et al., 2003; Wakimoto et al., 2004), rearrangements of the X chromosome in about 75% of cases may lead to spermatogenesis disorders presumably as a result of improper chromosome segregation or defective chromosome-wide regulatory mechanisms, such as MSCI (Lifschytz and Lindsley, 1972; Kennison, 1983; McKee et al., 1998). Thus, since sterile males showed higher levels of reporter gene expression, it is likely that both mutant lines have a relaxation of X suppression. However, the basis for this may be either EMS-induced mutations in the genes or spontaneous X chromosomal rearrangements. It is also possible that the mutations responsible for sterility are associated with X suppression only due to genetic linkage.

We found an overall increase in the expression of endogenous genes on the X chromosome in the testis of the *INXS1* mutant line compared to the X-linked control (Figure 3.18). This expression pattern was significantly different from what was observed for the arms of the second and third chromosomes. When analysing genes with relatively high expression in testes on different chromosomes, we also found the highest total number of genes with increased expression and their highest proportion relative to genes with a decreased expression on the X chromosome compared to autosomes (Figure 3.19). This indicates that X suppression, which has been shown to have the greatest effect on genes that are highly expressed in testes, may be inactive (Landeen et al., 2016; Argyridou and Parsch, 2018). In *INXS2,* a similar level of increased expression was found for six endogenous genes with testis-specific expression that showed up-regulation in *INXS1* (Figure 3.20). A previous study of transpositions between the X chromosome and autosomes also showed increased expression for two of these genes when they are located on the X chromosome compared to the autosomal position (*CG12689*: 4-fold difference; *CG3323*: 5.6-fold difference) (Landeen et al., 2016). Since all six genes tested are spread out across the X chromosome (from the genomic position of 5.3 Mb to 18.9 Mb), it is likely that their increased expression is not the result of an X chromosomal-autosomal translocation, which was observed for one gene in a

previous study (Landeen et al., 2016). Thus, we propose that in *INXS1* and *INXS2*, a similar pattern of general overexpression is the result of a disruption of X suppression.

### 4.2.2 Identification of candidate mutant genes

Through genome-wide re-sequencing analysis of *INXS 1/2*, we were able to identify the top candidate mutant alleles associated with increased expression of the reporter gene (Figure 3.21, 3.22, Table B11). In *INXS1*, on the X chromosome, we found a 12-bp in-frame deletion in the coding region of *CG13003*. In *INXS2*, on the X chromosome, our top candidate gene was *CG1314*, which has a nonsynonymous mutation in the coding region. Interestingly, the *CG1314* promoter was used previously in a reporter gene study, where it showed up-regulation of expression when mobilised to the autosomes (Kemkemer et al., 2011). On the second chromosome, one candidate mutation was found in the gene *CG17344* in *INXS1*. On the third chromosome for both *INXS 1/2,* we found a common candidate gene *CG31525*, which had different nonsynonymous mutations in the two mutants (including one that knocked out the start codon in *INXS2*). Also, in *INXS2,* we found another mutation in a start-codon of the gene *CG42654*. All these genes have an unidentified function and, except for *CG13003,* have high testis-specific expression, making it possible for them to play a role in the X chromosome-specific regulatory mechanism in the testis. In addition to the top candidate genes listed above, we found mutations in other coding and non-coding regions, providing lower-priority candidates for further tests.

Further, for candidate genes and several other genes, which we used as markers of recombination between chromosomes (*sgg*, *Ucp4A*, *Mcm3*, *Kua*, *LRR*, *Ipp*), association tests for the presence of each mutant allele and increased expression of the reporter gene were carried out. It is worth noting that in this association analysis, we were more interested in males with increased expression, as reduced expression could be the result of incomplete penetrance. We found a complete association for the *CG13003* gene when testing *INXS1* males taken from the mutant stock, with only 3 of the 35 males tested showing the wild-type allele and low reporter gene expression (Table 3.2). However, when testing males from the additional crosses of the *INXS1* line (with *wol20X*, *wol5X* and *wol12X*), we observed a complete absence of the mutat allele even in males with overexpression. This could be expected since, in these crosses, the X chromosome was inherited from wild-type flies. Thus, the above-mentioned association is probably due to a genetic linkage between the reporter gene and the mutant allele, located only approximately 170 kb apart. It is also possible that if

the observed phenotype is a polygenic trait, the incomplete association may be due to one or more other mutant genes that we missed from the analysis. A similar association level was observed for another mutation in the gene *Ucp4A,* which is about 1 Mb from *CG13003* (Table 3.2). For the gene *Ucp4A,* several mutant males were found with high reporter gene expression obtained after additional crosses with *wol5X,* indicating recombination has occurred in this region. Also, in *INXS1*, we found a relatively high level of association on the third chromosome for the genes *CG42654* and *Ipp*. However, the top-candidate gene, *CG31525,* showed no significant association with increased expression (Table 3.2).

Interestingly, for *INXS2*, the mutation in the gene *CG31525* showed a complete association with high reporter gene expression, in contrast to the mutant allele of the same gene in *INXS1* (Table 3.2). This may be explained by the mutation being in the start codon in *INXS2* and presumably preventing protein production, whereas in *INXS1*, only a partial disruption of the product of this gene may have occurred. Moreover, we found that mutations in genes *CG31525* and *CG1314* were always found together in *INXS2* males with increased expression (Table 3.2). So far, it is difficult to explain the association of these two mutations on the different chromosomes. However, it is possible that either they both contribute to the up-regulated expression or perhaps there was a translocation between the X and third chromosomes, causing these two genes to be linked. Interestingly, for *CG31525,* we detected recombination events between the mutant and balancer chromosomes, which are expected to be very rare events (Miller et al., 2016; Miller et al., 2019).

In the final step of my study, we performed a functional test of the three top candidate genes (*CG13003*, *CG1314*, *CG31525*) using RNAi knockdown. We observed an increase in reporter gene expression in the testis for several samples when the gene *CG13003* was knocked down using testis-specific Gal4 promoters (Figure 3.23). However, the level of up-regulation was lower than in both mutant lines. We could not find any significant increase in the expression of endogenous X-linked genes with the RNAi knockdown of *CG13003*, *CG1314*, *CG31525* in the early stages of gametogenesis using the *Bam* and *TJ* promoter, where the onset of X suppression is expected to occur (Meiklejohn et al., 2011) (Table 3.3). For the *Act5C*-mediated ubiquitous activation of RNAi transcription targeted to the gene *CG1314,* we also found no up-regulation of endogenous X-linked gene expression. Ubiquitous knockdown of the *CG13003* and *CG31525* genes resulted in a lethal male phenotype, which might be expected for *CG13003*, which has moderately high expression

outside the testes (Leader et al., 2018). It is possible that the complete knockdown of the target mRNA molecules is also lethal in males, in contrast to the mutant alleles of these genes, which in the case of the gene *CG31525,* was always heterozygous in all tested *INXS* males. In addition, we found partial sterility in males when the *CG13003* gene was knocked down with the *Bam* promoter. However, several explanations for the lack of any effect of RNAi knockdown of the genes of interest are possible. Firstly, we do not know in which tissues or developmental stages the reduction in expression of these genes can lead to the phenotype of interest. Secondly, several genes may need to be simultaneously knocked out. Thirdly, for the *Tj* and *Act5C* promoters, we observed relatively low RNAi efficiency (Table 3.3). These issues will need to be addressed in future functional tests.

Our approach utilized mutant males of the F2 and F3 generations, which was a rather time-consuming process and significantly reduced the number of tested flies compared to screening F1 males. However, this approach allowed us to avoid possible mutations that would not be passed on to the next generation as a result of DNA repair mechanisms in F1 males and the genetic mosaicism (Nissani, 1977; Pastink et al., 1991; Ashburner, 2005). Our approach would also make it possible to identify both dominant and recessive mutations on autosomes for the majority of mutant lines, as most of the males tested had homozygous second or third autosomes. However, this also reduced the screening efficiency, as each line had to be tested in multiple replicates with different chromosomal combinations. Nevertheless, even screening of around 5000 mutant males allowed us to presumably mutagenize the entire genome effectively, as several genes in our analysis, such as the *wol* reporter gene and gene *CG31525,* were hit twice. Among the advantages of our screening was that we used the *ocn* promoter as part of the reporter gene construct, which has a high level of testis-specific expression and has previously been shown to have high sensitivity to the effects of X suppression (Hense et al., 2007). For the reporter strain we used, the expression level was 7-fold lower expression than that of autosomal insertions, which allowed us to detect even partial relaxation of the X suppression.

One of the biggest challenges was detecting the mutations responsible for the observed phenotypes in both mutants. Thus, the first round of causative mutation mapping using genetic crosses did not allow us to narrow down the possible position of causative mutations on the chromosomes with certainty due to possible spontaneous recombination of mutant chromosomes and incomplete penetrance of causative mutations. Although mapping mutant

alleles using whole-genome a sequencing approach has previously been shown to be highly effective in EMS mutagenesis (Blumenstiel et al., 2009; Haelterman et al., 2014; Phadnis et al., 2015), additional time-consuming steps will be required to determine the association between candidate mutations and the mutant phenotype, as well as for functional tests to validate the candidate genes.

## 4.3 Conclusions

In the first part of my thesis, we studied the effect of proximity to DCC binding sites on reporter gene expression in various tissues of *D. melanogaster* males and females. In this new approach, using an exogenous reporter gene with ubiquitous expression inserted at many X-chromosomal locations, we were able to avoid the confounding factor of gene-specific regulation that is common to studies of endogenous genes. Thus we could test the previously proposed model of the effect dosage compensation on sex-biased expression in somatic tissues (Bachtrog et al., 2010; Huylmans and Parsch, 2015). In particular, we examined whether genes located near DCC binding sites showed higher expression in males (and a higher male/female expression ratio) than those located far away. Our results suggest that this positive correlation is common for all somatic tissues, where X-chromosome dosage compensation is known to occur in males. Furthermore, the distance to a DCC binding site had a greater influence on reporter gene expression than local regulatory sequences affecting the native genes surrounding the insertion site. Although previous studies of the expression of endogenous genes reported a very strong positive correlation between male-biased expression and the distance to the nearest DCC binding site in the gonad (Huylmans and Parsch, 2015), we could not confirm these results by using the reporter gene, suggesting the absence of the X-chromosome-specific dosage compensation in this tissue.

Our results are consistent with the idea that in somatic tissues, male-biased expression results from the simultaneous activity of gene-specific regulation and chromosome-wide DCC mediated regulation (Huylmans and Parsch, 2015). As it is possible that these two mechanisms could interfere with each other, the genes are distributed at different distances from the DCC binding sites, depending on the degree of sex-biased expression required for each particular gene. Thus, the head and/or brain do not appear to be unusual regarding the DCC effect. However, these tissues do appear to be unusual in that they show strong enrichment of sex-biased genes, particularly male-biased genes, on the X chromosome

compared to other tissues. Our results suggest this may be explained, at least partly, by an increased level of expression variation among regions of the X chromosome, which we observe in both sexes.

It is likely that the effect of DCC distance on reporter gene expression occurs not only in a linear fashion but also in a highly complex 3D chromatin structure, which creates an opportunity for further studies. For example, the controlled relocation of the reporter gene relative to a specific DCC binding site and employing the latest techniques for studying chromatin structure, such as ATAC-seq, would make it possible to eliminate some of the random effects of the genomic environment on gene expression. Future studies could examine the effect of DCC binding and the evolution of sex-biased expression in more detail, including studying species with neo-X chromosomes that differ in dosage compensation from *D. melanogaster*.

In the second part of my thesis, we performed a forward genetic screen to find genes responsible for the X suppression in the germline of *D. melanogaster*. We obtained two lines with EMS-induced mutations showing an increase in the X-linked testis-specific reporter gene expression: *INXS1* and *INXS2*. Both lines showed an overall increase in expression of endogenous X-linked genes and male sterility with complete sperm immobility, which could indicate the lack of the X chromosome suppression in the male germline (Lifschytz and Lindsley, 1972; Argyridou and Parsch, 2018). For both candidate mutants, we found four top candidate genes carrying mutations in coding sequences that are situated on the X (*CG13003*, *CG1314*) and third chromosomes (*CG31525*, *CG42654*). By analyzing the association of up-regulation and mutant alleles, we were able to show that the *CG13003* and *CG42654* genes may be responsible for the observed phenotype in *INXS1*. Since the association was equally high for the *CG1314* and *CG31525* genes in *INXS2*, the possibility of X–autosomal translocation cannot be excluded, which could also disrupt a general X-chromosomal regulatory mechanism, as it was proposed for MSCI (Lifschytz and Lindsley, 1972). Since among all of the candidate genes found, only *CG31525* was common for both *INXS* mutants, but there was an association between the mutation in this gene and increased expression only in *INXS2*, it can be assumed that several genes are involved in X suppression. Functional analysis of *CG13003*, *CG1314* and *CG31525* by targeted RNAi knockdown, however, showed no effect on the expression level of X-linked genes, suggesting that additional functional testing and/or the consideration of further candidate genes is necessary.

Several further tests could be considered to identify the genes responsible for the observed phenotype in both *INXS* mutants. First, using a long-read sequencing approach and polytene (or mitotic) chromosome imaging might help to identify the presence of any rearrangements between the X chromosome and autosomes in the *INXS* lines. Secondly, the rough mapping of causative mutations using the polymorphic marker chromosome method (Martin et al., 2001) or SNP mapping (Chen et al., 2008) could be continued, as it is more efficient and cost-effective than whole-genome sequencing. Thirdly, functional analysis of candidate genes should be continued, possibly using lines carrying knockout mutations or using additional tissue-specific promoters for RNAi, including simultaneous gene knockouts (Qiao et al., 2018).

# Appendix A

## Supplemental Figures

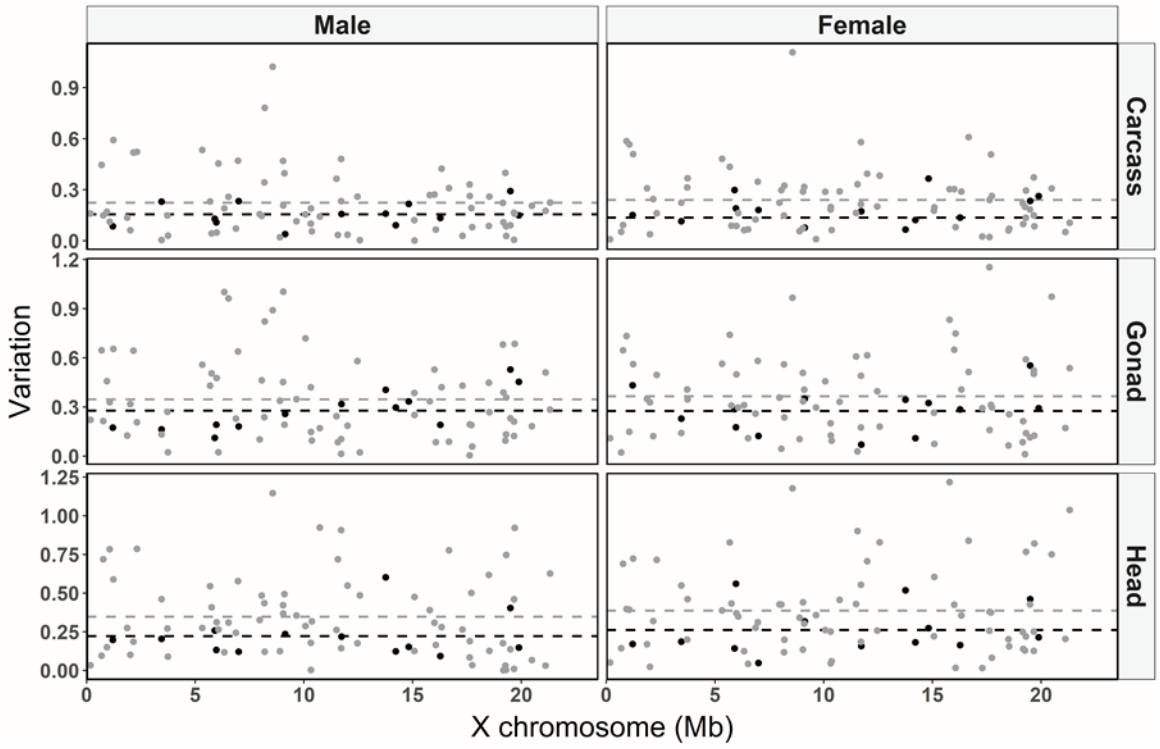


**Figure A1.** Variation of β-galactosidase activity among biological replicates in different sexes and tissues. Black dots represent 13 combined *CMV-lacZ* insertions, and grey dots represent 70 unique insertions. Dashed lines indicate the mean level of variation for 13 combined insertions (black) and 70 unique insertions (grey). In all plots, variation is in units of standard deviation divided by the mean.

## A. Supplemental Figures

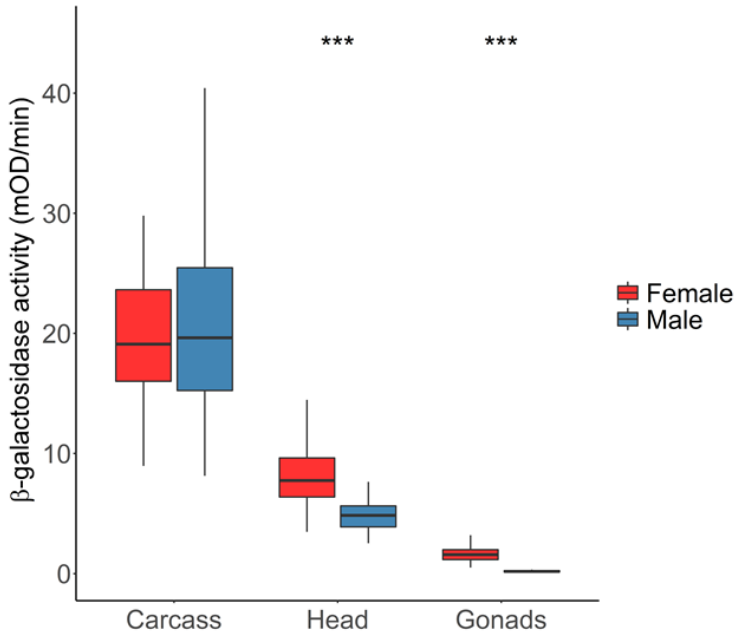


**Figure A2.** Non-standardized β-galactosidase activity in different sexes and tissues. For each tissue, differences between the sexes were tested using a paired *t*-test. ***$P < 0.001$.

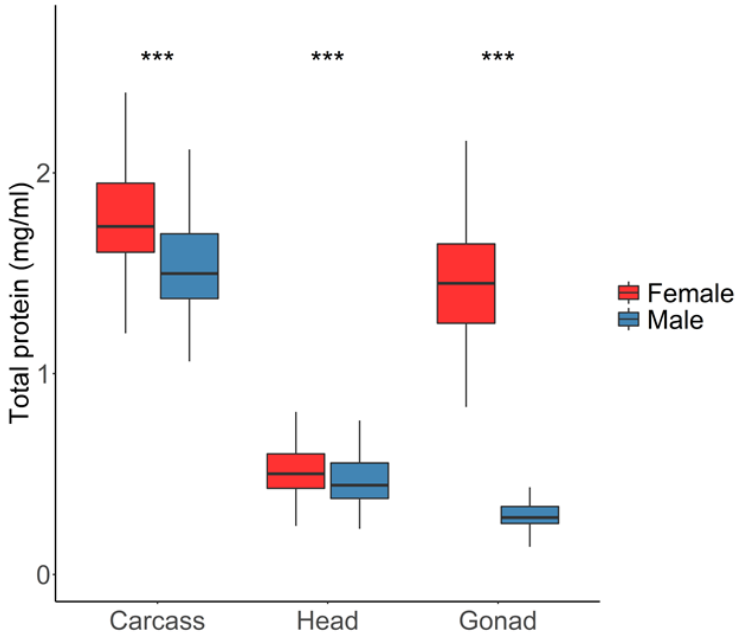


**Figure A3.** Total protein in different sexes and tissues for flies carrying an X-linked copy of the *CMV-lacZ* reporter gene. For each tissue, differences between the sexes were tested using a paired *t*-test. ***$P < 0.001$.

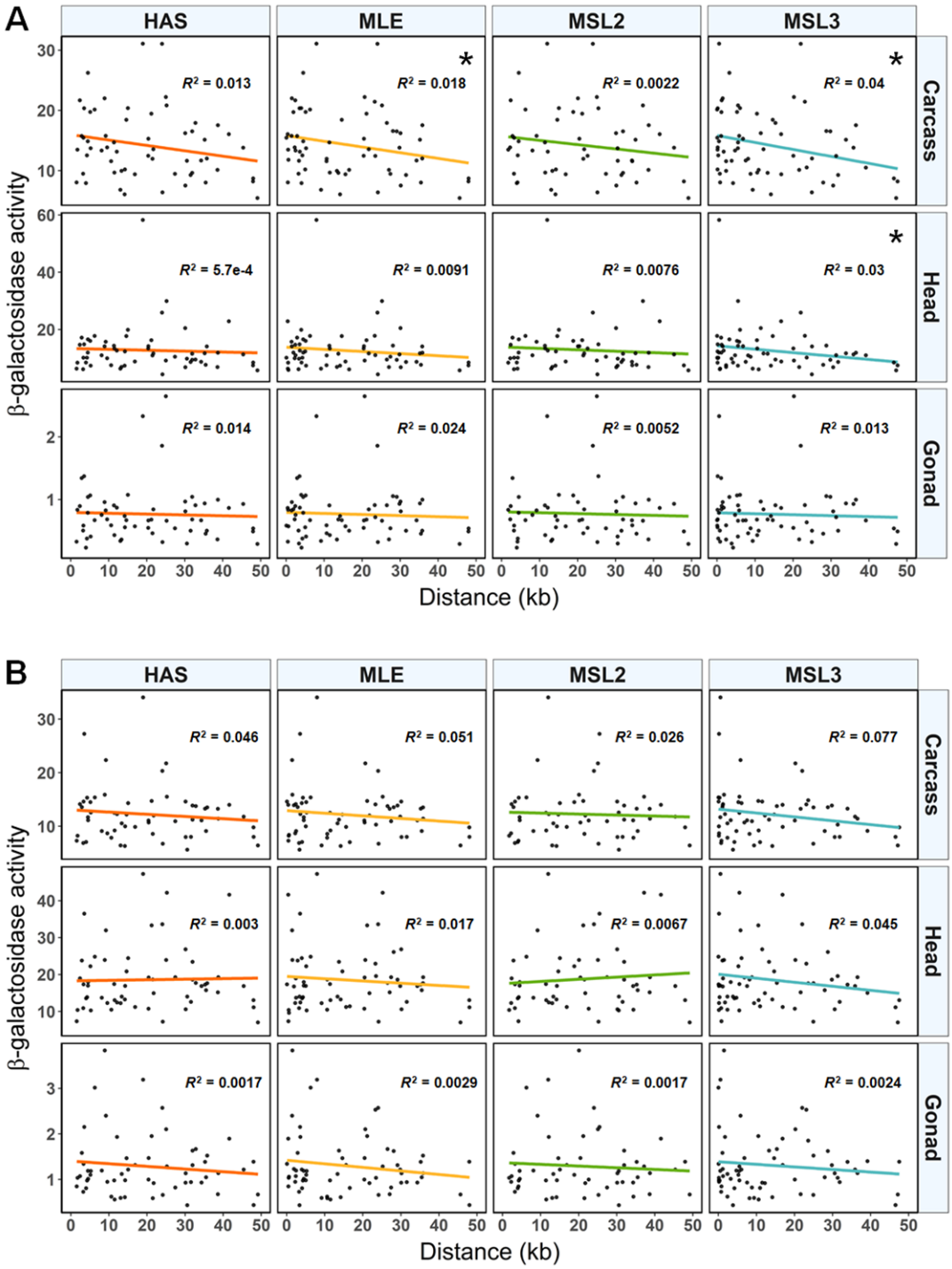


**Figure A4.** Reporter gene expression and distance to the nearest DCC binding site in different tissues. Reporter gene expression was measured in (*A*) males and (*B*) females. The binding sites of different DCC components are shown in columns. Coloured lines represent the least-squares linear regression. Dots represent the individual *CMV-lacZ* reporter gene insertions. **P* < 0.05.

## A. Supplemental Figures

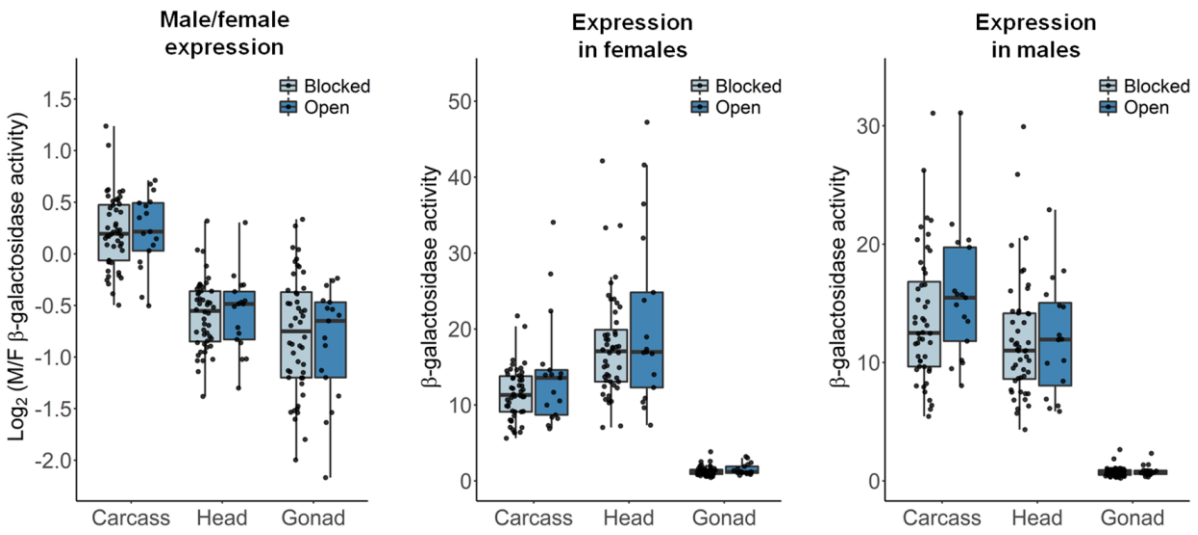


**Figure A5.** Male-to-female expression ratio (M/F) of *CMV-lacZ* reporter genes grouped by their location relative to G4 sites. "Blocked" indicates the reporter gene flanked only by G4s, "open" indicates the reporter gene flanked by at least one DCC binding site. Differences between tissues were tested with a Wilcoxon signed-rank test.

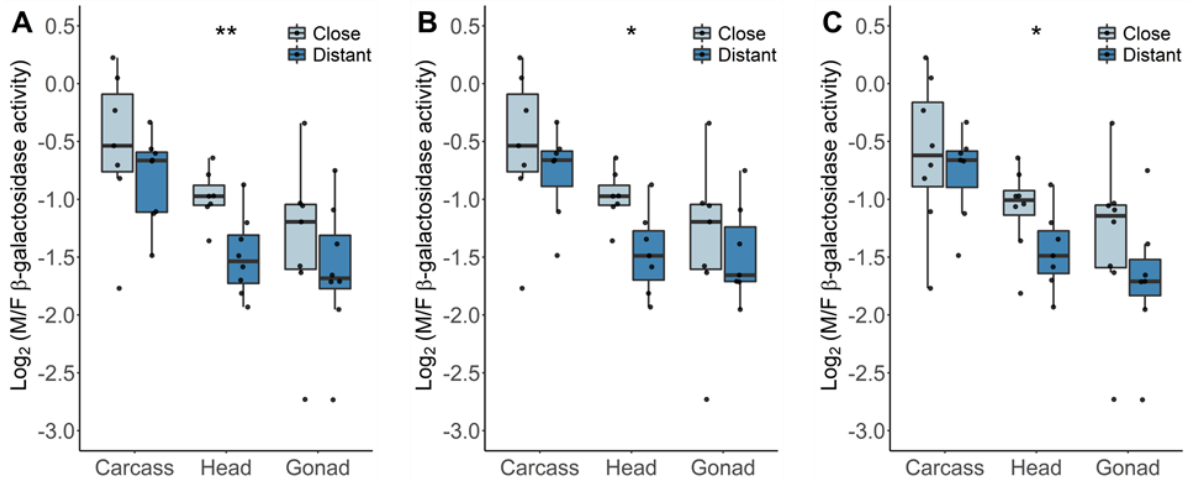


**Figure A6.** Male-to-female expression ratio (M/F) of *CMV-lacZ* reporter genes in different tissues for homozygous females and hemizygous males. Insertions are grouped by their proximity to the nearest (A) MLE, (B) MSL2, or (C) MSL3 binding sites. Differences between the groups were tested with a Wilcoxon signed-rank test. *P < 0.05, **P < 0.01.

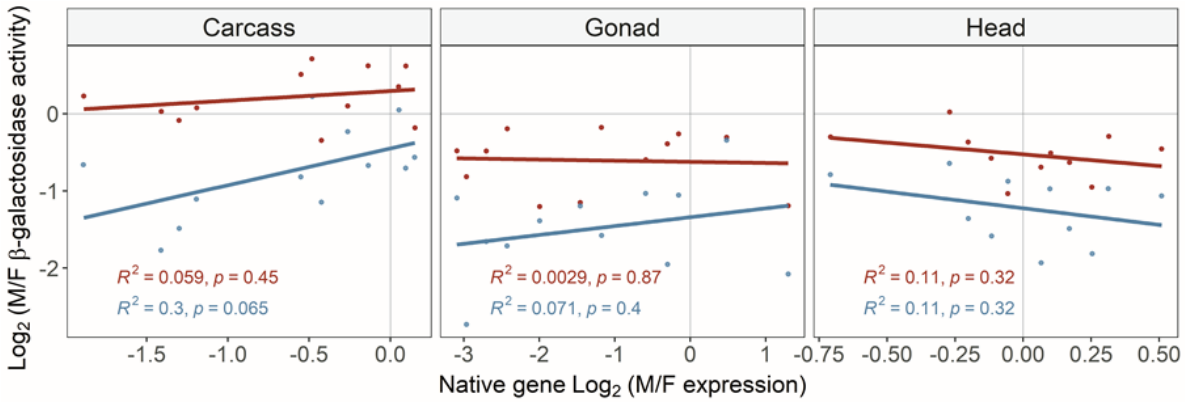


**Figure A7.** Male-to-female expression ratio of each reporter gene and the endogenous gene in which it is located for the groups of the dosages of the reporter gene (heterozygous and homozygous females) in different tissues. Lines represent the least-squares linear regression. Dots represent the individual *CMV-lacZ* insertions. Colour indicates the source of data on the reporter gene sex-biased expression; heterozygous females and hemizygous males (red); homozygous females and hemizygous males (blue).

## A. Supplemental Figures

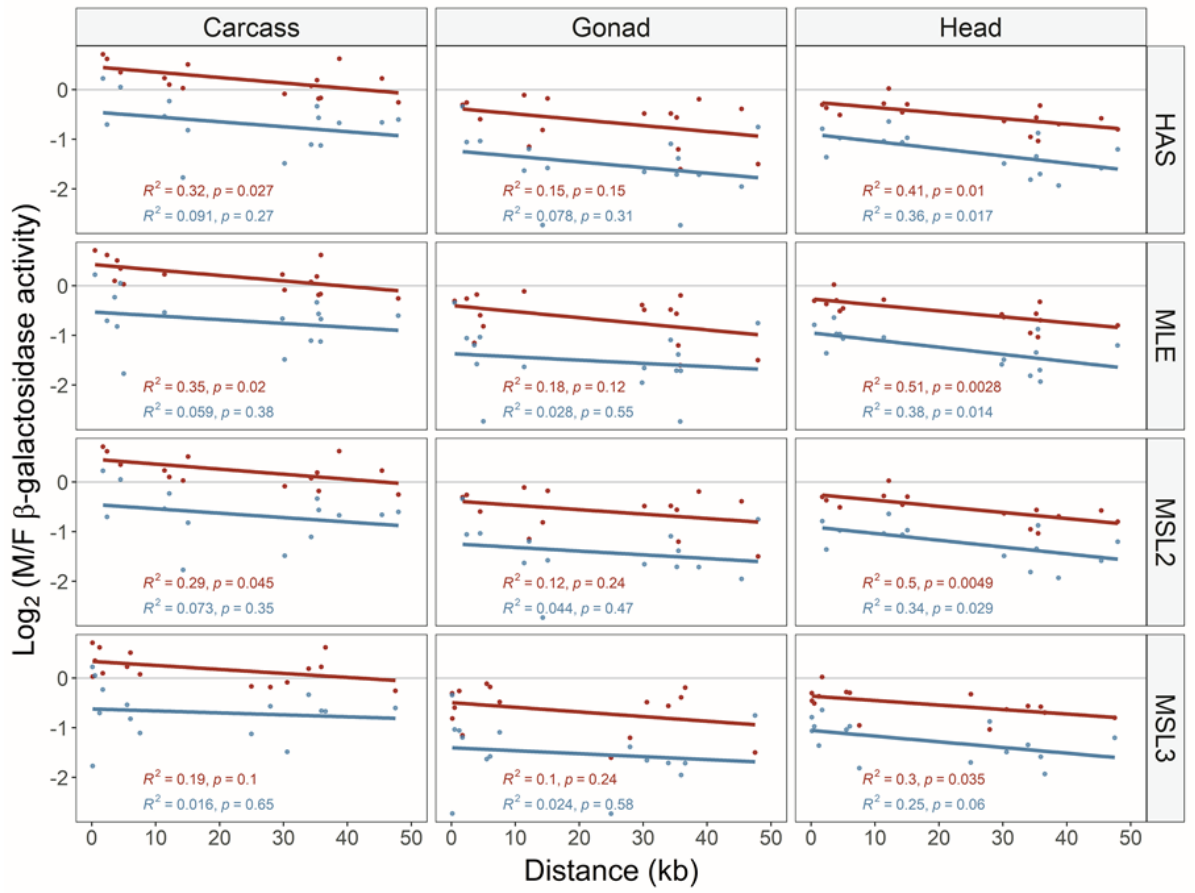


**Figure A8.** Male-to-female expression ratio of each reporter gene and distance to the nearest DCC binding sites for the groups of the dosages of the reporter gene (heterozygous and homozygous females) in different tissues. Binding sites of different components of the DCC are represented in rows. Lines represent the least-squares linear regression. Dots represent the individual *CMV-lacZ* insertions. Colour indicates the source of data on the reporter gene sex-biased expression; heterozygous females and hemizygous males (red), homozygous females, and hemizygous males (blue).

# Appendix B

## Supplemental Tables

## B. Supplemental Tables

**Table B1.** DNA oligonucleotides sequence used as primers for PCR assays

| Primer name[1] | Target | Sequence 5'->3' | Analysis | Restriction enzyme[2] |
|---|---|---|---|---|
| Plac1.F | *pP[wFl]* | CACCCAAGGCTCTGCTCCCACAAT | iPCR | - |
| Plac4.R | *pP[wFl]* | ACTGTGCGTTAGGTCCTGTTCATTGTT | iPCR | - |
| EY.3.F | *pP[wFl]* | CAATAAGTGCGAGTGAAAGG | iPCR/ Sanger sequencing | - |
| EY.3.R | *pP[wFl]* | ACAATCATATCGCTGTCTCAC | iPCR | - |
| Sp1.R | *pP[wFl]* | ACACAACCTTTCCTCTCAACAA | Sanger sequencing | - |
| CG31525-INXS2.F | *CG31525* | GGAACACATTTAGCATTCGTG | Sanger sequencing | - |
| CG31525-INXS2.R | *CG31525* | GACATCAATGGCACGCTGGAG | Sanger sequencing | - |
| CG1314.F | *CG1314* | ATGGAGAACAACAATTGCGAG | PCR with digestion | Hpy99I |
| CG1314.R | *CG1314* | CACAGTTGCTTTTGCCCCTG | PCR with digestion | Hpy99I |
| CG13003.F | *CG13003* | CTCGAACCGAACGGGCCGCT | PCR | - |
| CG13003.R.Ref | *CG13003* | GGTGATTCCATTGTACTTGGG | PCR | - |
| CG13003.R.Mut | *CG13003* | GGTGATTCCATTATACGACGT | PCR | - |
| Kua.F | *Kua* | CAGGACGCCCAGCGAGATACAG | PCR with digestion | StuI |
| Kua.R | *Kua* | GCAAGAACACCTCTTCGGTTCG | PCR with digestion | StuI |
| LRR.F | *LRR* | CCAACTTTGTTCATAATCTCGC | PCR with digestion | HindIII |
| LRR.R | *LRR* | GATGCTGAGCTAGCGCCCGTGC | PCR with digestion | HindIII |
| CG17344.F | *CG17344* | GAGTTTATCCGATTGAGATGGAG | Sanger sequencing | - |
| CG17344.R | *CG17344* | GCTGTATGCAATTAGATAGATCAC | Sanger sequencing | - |
| sgg.F | *sgg* | CGCAGCTCTCGTCGTCGTCTC | PCR with digestion | BsmAI |
| sgg.R | *sgg* | CATGGGTCTTCAAATGGGTGC | PCR with digestion | BsmAI |
| Mcm3.F | *Mcm3* | CACCGAGGTGTACGAAAAGTACG | PCR with digestion | NruI |
| Mcm3.R | *Mcm3* | CTATATCGTCGGCGACCATGATC | PCR with digestion | NruI |
| Ucp4A.F | *Ucp4A* | CTATCCGGCAACCCTGGCCTG | PCR with digestion | BsrDI |
| Ucp4A.R | *Ucp4A* | CGAATGGGCGGCTCCTTCGC | PCR with digestion | BsrDI |

Table B1 continued.

| Primer name[1] | Target | Sequence 5'->3' | Analysis | Restriction enzyme |
|---|---|---|---|---|
| CG42654.F | *CG42654* | CGTGAATTTAGTGACGTTGG | PCR with digestion | PsiI |
| CG42654.R | *CG42654* | CAGTGTAGAGAGCTTGTTCTGG | PCR with digestion | PsiI |
| CG31525-INXS1.F | *CG31525* | CAAGTTTACAAAATCACTTCC | PCR with digestion | BstB1 |
| CG31525-INXS1.R | *CG31525* | GCGGTTTGGGATTTCTAGCTAG | PCR with digestion | BstB1 |
| Ipp.F | *Ipp* | CAGACGCTACCGCCGAGTAC | PCR with digestion | BssHII |
| Ipp.R | *Ipp* | CTCGGCCAATTTGGCCAGCAG | PCR with digestion | BssHII |
| CG13003.F3 | *CG13003* | CTACACGCAGAGTACGAGGC | qRT-PCR | - |
| CG13003.R3 | *CG13003* | CAAAGTCGGGGGCCTGGAAT | qRT-PCR | - |
| CG12689.F | *CG12689* | AACACTAGTTTTAGCTCACACAAAA | qRT-PCR | - |
| CG12689.R | *CG12689* | CATTCTTGCCTGACTCGAAATCC | qRT-PCR | - |
| CG3323.F | *CG3323* | CCTGCCCACCGTTAACAAC | qRT-PCR | - |
| CG3323.R | *CG3323* | TGCTTAGGAACAACGGAGGG | qRT-PCR | - |
| CG15306.F | *CG15306* | TCCTGTACAACTTTCTTTGGCTTG | qRT-PCR | - |
| CG15306.R | *CG15306* | GAGAATATCGCCACGGGACC | qRT-PCR | - |
| CG15892.F | *CG15892* | AAGGTAAAGCCCTACGACGTG | qRT-PCR | - |
| CG15892.R | *CG15892* | CTGCTTGCGCTTCAGGAAAT | qRT-PCR | - |
| CG11068.F | *CG11068* | TGGAACGAGCGGCATTAACTT | qRT-PCR | - |
| CG11068.R | *CG11068* | CTGCGTGACGTTGCGAATC | qRT-PCR | - |
| CG7349.F | *CG7349* | TGGGTCATTGATTCCAGGGAC | qRT-PCR | - |
| CG7349.R | *CG7349* | TAGCCGGATTGAGGTGCTTG | qRT-PCR | - |
| CG1314.F | *CG1314* | TTGATGCCCGATCACCGTAG | qRT-PCR | - |
| CG1314.R | *CG1314* | GAAGCGGAAATGGCTTCTGC | qRT-PCR | - |
| CG31525.F | *CG31525* | AAAGAAGGCTTTTGACGCGG | qRT-PCR | - |
| CG31525.R | *CG31525* | TTTGGACATCAATGGCACGC | qRT-PCR | - |
| RpL32.F | *RpL32* | AGCATACAGGCCCAAGATCG | qRT-PCR | - |
| RpL32.R | *RpL32* | TGTTGTCGATACCCTCGGGC | qRT-PCR | - |

[1]Forward (F) and reverse (R) PCR primers. [2]Restriction enzymes used to separate PCR product sequences between mutants and control genotypes.

**B. Supplemental Tables**

**Table B2.** Locations of the X-linked reporter gene insertions.

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X83* | 1A1 | 175992 | 5' UTR/intron | *tyn* | - | - |
| *X210* | 1C2 | 687009 | 5' UTR | *sdk* | - | - |
| *X194* | 1C4 | 769982 | 5' UTR | *fz3* | *RpL22, CG5273* | - |
| *X232* | 1D2 | 935352 | 5' UTR, intron | *TRAM, CG32815* | *CG3706, CG3708, MED22, CG43867, Lztr1* | *CG3704, mus81, CG3703* |
| *X82* | 1E1 | 1063522 | intergenic | - | - | *CG3655* |
| *X54* | 1E4 | 1209204 | intergenic | - | *CG14625, CG11381, CG14624* | *CG11382, CG11398, CG3638* |
| *X29* | 1E4 | 1209204 | intergenic | - | *CG14625, CG11381, CG14624* | *CG11382, CG11398, CG3638* |
| *X213* | 1E5 | 1234390 | intron | *CG3638* | - | *CG11403, Atf3* |
| *X8* | 2B13 | 1873539 | 5' UTR | *Pgam5* | *CG14803, Pex5, MED18, CG14814* | *Vps26, CG14817, CG14817, CG14805, CG14818, CG14806, trr* |
| *X157* | 2C2 | 2011613 | intron | *east* | *mir-2496, PIG-K, Hr4* | - |
| *X188* | 2D5 | 2158967 | 5' UTR | *wapl* | *bcn92* | *Cyp4d1, CG3630* |
| *X53* | 2F5 | 2317609 | intron | *Raf* | - | - |
| *X192* | 3D3 | 3449374 | intergenic | - | *CG12535* | *CG16782* |
| *X20* | 3D3 | 3451351 | intergenic | - | *CG12535* | - |
| *X45* | 3D3 | 3451429 | intergenic | - | *CG12535* | *CG16782, CG10801* |
| *X60* | 3E7 | 3740658 | intron | *Tlk* | - | - |

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X187* | 4F5 | 5322715 | intergenic | - | *Rnp4F, Sas10, CG15930, CG4198, dhd* | *Mcm3, CG3309, XRCC1, CanB* |
| *X199* | 5A12 | 5679981 | intergenic | - | *Tre1, RpL35, Rab18, CG4119* | *Gr5a, CG42449, PGAP1* |
| *X178* | 5B8 | 5755296 | intergenic | *lin-52* | *lin-52, CG15772, CG4078, CG3125* | *CG15771, MAPk-Ak2, CG42699* |
| *X186* | 5C7 | 5901698 | intron | *Act5C* | - | *CG12236, CG4020* |
| *X226* | 5C7 | 5901776 | intron | *Act5C* | - | *CG12236, CG4020* |
| *X98* | 5D1 | 5965760 | intergenic | - | *fs(1)M3, CG6067* | *Grip* |
| *X180* | 5D1 | 5965823 | intergenic | - | *fs(1)M3, CG6067* | *Grip* |
| *X221* | 5D1 | 5988734 | intergenic | - | - | *CG5966, CG4766* |
| *X68* | 5D3 | 6069294 | 3’ UTR | *Tsp5D* | *CG5928, CG43115, CG5921* | - |
| *X151* | 5F2 | 6339095 | intron | *CG3842* | *CG42240, CG15894* | *CG3847, ND-B16.6* |
| *X34* | 6B1 | 6524533 | 5’ UTR | *CG3918* | *CG42340* | *CG3342, Spat, dx, RpL7A* |
| *X207* | 6E2 | 6868129 | intron | *CG43736* | *CG14434* | - |
| *X21* | 6E4 | 6968661 | 3’ UTR, 5’ UTR | *CG14431, Setd3* | - | *CG4586* |
| *X2* | 6E4 | 6998549 | intergenic | - | *Inx7, ogre* | *Inx2* |
| *X27* | 6E4 | 6998869 | 5’ UTR/intron | *Inx2* | *Inx7* | - |

**B. Supplemental Tables**

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X195* | 7D1 | 7967248 | intron | *sws, sn* | - | - |
| *X26* | 7D5 | 8057219 | intron | *fs(1)h* | - | *mys* |
| *X182* | 7D17 | 8176963 | 5' UTR/intron | *slpr* | *CG2278* | *sdt* |
| *X197* | 7D18 | 8193003 | intron | *sdt* | - | - |
| *X222* | 7F4 | 8561564 | intron | *Trf4-1* | *CG12112, IntS4, Nrg* | *CG12111, spirit* |
| *X11* | 8B6 | 8894224 | intron | *Moe* | - | *CG1885, Rbm13, e(r)* |
| *X94* | 8C4 | 9042594 | intron | *Nost* | - | - |
| *X202* | 8C4 | 9047984 | intron | *Nost* | - | *CG15365* |
| *X212* | 8C9 | 9103706 | 5' UTR | *CG7766* | - | *Bx42, Arfrp1, AP-1gamma* |
| *X174* | 8C13 | 9134501 | 5' UTR | *fend* | - | *CG7065* |
| *X148* | 8C13 | 9134915 | 5' UTR | *fend* | - | *CG7065* |
| *X191* | 8C13 | 9135026 | 5' UTR | *fend* | - | *CG7065* |
| *X196* | 8C13 | 9135037 | 5' UTR | *fend* | - | *CG7065* |
| *X220* | 8F5 | 9651090 | intron | *Ptpmeg2* | - | *CG3106* |
| *X65* | 9B1 | 10070875 | intergenic | - | *CG15312, Gr9a* | *ZAP3* |
| *X1* | 9B6 | 10318800 | intron | *alpha-Man-I* | - | *CG2909* |
| *X209* | 9B7 | 10325568 | intron | *alpha-Man-I* | - | *CG2909, Gip, CG15306, CG43740* |
| *X231* | 9B12 | 10366287 | intergenic | - | *l(1)G0289* | *CG32679, CG17841, CG32686, CG2898* |
| *X35* | 9E1 | 10730636 | intron | *stx* | - | - |

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X214* | 10C1 | 11493193 | 5' UTR | *Drak* | *CG33235, Spase25, Uba5, Cyp4g15* | - |
| *X23* | 10C5 | 11558520 | 5' UTR/intron | *CG1572* | *Drak* | *PGRP-SA, RpII215* |
| *X179* | 10D6 | 11713588 | 5' UTR | *Amun* | *FucT6, CG2444, PhKgamma, mir-9370* | *prtp, Tango10* |
| *X59* | 10D6 | 11714665 | 5' UTR | *Amun* | *mir-9370, FucT6, CG2444* | *CG15221, Tango10, prtp* |
| *X106* | 10E1 | 11729178 | intron | *CG15221, inaF-A, inaF-B, inaF-C* | *inaF-D, Tango10* | - |
| *X215* | 10E1 | 11729218 | intron | *CG15221, inaF-A, inaF-B, inaF-C* | *inaF-D, Tango10* | - |
| *X193* | 11A11 | 12003975 | intergenic | - | *fw* | *CG1806, CG1492, regucalcin* |
| *X185* | 11A11 | 12463520 | intergenic | - | *Rab40, CG15731, CG15927, PKCd* | *CG42258* |
| *X170* | 11B1 | 12575595 | intron | *Pits* | *hwt* | *LIMK1* |
| *X22* | 12C1 | 13762705 | intergenic | - | *Yp3, Rtc1, CG32625* | *rdgB* |
| *X223* | 12C1 | 13762705 | intergenic | - | *Yp3, Rtc1, CG32625* | *rdgB* |
| *X25* | 12E5 | 14222240 | intron | *l(1)G0007* | - | *CG11674, mRpL38* |

**B. Supplemental Tables**

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X238* | 12E5 | 14222330 | intron | *l(1)G0007* | - | *CG11674, mRpL38, CG42271* |
| *X5* | 12F4 | 14826069 | 5' UTR | *rut* | - | *CG14408, CG14411* |
| *X145* | 12F4 | 14826071 | 5' UTR | *rut* | - | *CG14408, CG14411* |
| *X225* | 12F4 | 14826071 | 5' UTR | *rut* | - | *CG14408, CG14411* |
| *X9* | 13A9 | 15075124 | intron | *Lsd-2* | *dob, opm* | *CG33177, CG33178, CG9065* |
| *X169* | 13A10 | 15089555 | intergenic | - | *CG5599, CG15027* | *Rab3-GEF* |
| *X205* | 13E10 | 15788882 | intergenic | - | *mRpS30, Paf-AHalpha, Efhc1.1, CG43673* | *Rhp, CG8974, CG32581* |
| *X239* | 14A6 | 16005774 | 5' UTR | *vap* | *CG8939, exd* | *CG12698, Muc14A* |
| *X245* | 14A7 | 16068555 | intron | *dpr18* | *CG12395* | *Nipsnap, Tob* |
| *X177* | 14B7 | 16278337 | 5' UTR | *eas* | *Pros28.1, Mfe2, kat80* | *CG32576, UQCR-14* |
| *X183* | 14B7 | 16278373 | 5' UTR | *eas* | *Pros28.1, Mfe2, kat80* | *CG32576, UQCR-14, caz* |
| *X200* | 14D14 | 16332768 | intron | *CG9921* | *CG9919, CalpC, CG32579, CG9917* | *Dsp1* |
| *X173* | 15A1 | 16670810 | intron | *CG13012* | *r, CG15865* | *CG13010, sing, Axs* |
| *X181* | 16A1 | 17296493 | intron | *Fim, CG5445* | - | - |

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X62* | 16B9 | 17625176 | 3' UTR/intron | *stas* | - | *CG8326, corolla, ND-24, CG8289* |
| *X64* | 16B10 | 17632078 | intergenic | - | *corolla, CG8326, stas* | *ND-24, CG8289, CG5800, RhoGAPp190* |
| *X105* | 16C1 | 17698803 | intergenic | - | *chas* | *CG8188, CG5884, par-6* |
| *X247* | 16C5 | 17738674 | 5' UTR/intron | *raskol* | - | - |
| *X233* | 17C1 | 18505312 | UTR'5 | *Wnt5* | *Xrcc2, Pgant7* | *Ggt-1, CG6470* |
| *X116* | 17C3 | 18534603 | intron | *Bx* | *CG15047, CG15042* | - |
| *X244* | 18A4 | 19154997 | 5' UTR/intron | *RhoGAP18B* | - | - |
| *X228* | 18A8 | 19180243 | UTR'5 | *CG7453* | *CG7556, CG33939, Inx5* | *CG33253, Mec2* |
| *X216* | 18B7 | 19262659 | intron | *Vav* | *rictor* | *CG8010* |
| *X201* | 18C1 | 19298145 | 5' UTR | *CG8034* | - | *CG8051* |
| *X218* | 18C1 | 19308076 | intergenic | *intergenic* | *CG8051, CG8034* | *out* |
| *X131* | 18C8 | 19494492 | 3' UTR | *kish* | *CG12204, pcm, Naa15-16* | *Pfrx, CG14200* |
| *X198* | 18C8 | 19498182 | intron | *Pfrx, CG14200* | *CG14199, CG12204, ND-18, pcm* | - |
| *X189* | 18C8 | 19498190 | intron | *Pfrx, CG14200* | *CG14199, CG12204, ND-18, pcm* | - |
| *X227* | 18D13 | 19663160 | 5' UTR/intron | *CG14223* | *Elys* | *Ssu72, Zw, et* |

**B. Supplemental Tables**

Table B2 continued

| Line ID | Cyt. band[1] | Coordinate[2] | Location[3] | Affected gene[4] | Proximal gene[5] | Distal gene[5] |
|---|---|---|---|---|---|---|
| *X118* | 18D13 | 19665841 | 3' UTR | *Ssu72* | *CG14223, Elys* | *Zw, et* |
| *X236* | 18E1 | 19693482 | 5' UTR | *Mer* | *CG14227, Ubqn, dome* | *CG14229, Cdc42, CG14231, CG14232* |
| *X171* | 19A2 | 19887117 | exon, intron | *amn, Hers* | - | - |
| *X28* | 18F4 | 19887117 | exon, intron | *amn, Hers* | - | - |
| *X224* | 19A2 | 19887128 | exon, intron | *amn, Hers* | - | - |
| *X104* | 18F4 | 19887264 | exon, intron | *amn, Hers* | - | - |
| *X234* | 19A2 | 19887348 | exon, intron | *amn, Hers* | - | - |
| *X219* | 19D2 | 20489108 | intron | *RhoGAP19D* | *CG15461* | - |
| *X190* | 19F1 | 21117218 | 5' UTR | *CG15445* | *unc, CG11566, stg1* | *CG34120* |
| *X24* | 19F4 | 21319242 | 5' UTR | *SLIRP1, anox* | *Rpt6, CG1801* | *Dd, CG1486* |

[1]Cytological band; [2]Genomic position of the insertion; [3]Type of genomic location (Slash indicates the location of insertion within alternatively transcribed units of the same gene); [4]Gene overlapping with the insertion; [5]Gene located within 10 kb of the insertion.

**Table B3.** Reporter gene expression in different tissues for the whole body dissection.

| Line ID | Female[1] | | | Male[1] | | |
|---|---|---|---|---|---|---|
| | Carcass | Head | Gonad | Carcass | Head | Gonad |
| *X83* | 13.548 | 26.072 | 1.622 | 16.573 | 14.360 | 1.053 |
| *X210* | 12.557 | 10.289 | 1.184 | 13.701 | 7.357 | 0.409 |
| *X194* | 8.019 | 11.118 | 0.436 | 8.702 | 8.489 | 0.537 |
| *X232* | 8.226 | 10.377 | 1.033 | 13.476 | 8.423 | 0.836 |
| *X82* | 6.401 | 7.041 | 0.670 | 5.442 | 5.710 | 0.292 |
| *X54* | 12.349 | 19.318 | 1.623 | 12.958 | 18.082 | 0.538 |
| *X29* | 14.665 | 15.941 | 1.425 | 11.087 | 10.117 | 0.465 |
| *X213* | 7.852 | 10.452 | 0.620 | 11.572 | 7.512 | 0.589 |
| *X8* | 11.795 | 16.853 | 1.217 | 13.807 | 11.304 | 0.930 |
| *X157* | 15.335 | 13.843 | 0.974 | 12.530 | 10.167 | 0.377 |
| *X188* | 6.871 | 10.881 | 0.728 | 9.478 | 6.120 | 0.505 |
| *X53* | 11.059 | 10.722 | 0.996 | 12.465 | 13.368 | 0.547 |
| *X192* | 8.814 | 22.298 | 0.718 | 16.661 | 18.080 | 0.600 |
| *X20* | 9.521 | 16.855 | 0.710 | 14.387 | 8.786 | 0.495 |
| *X45* | 14.816 | 24.137 | 1.053 | 23.907 | 15.802 | 0.446 |
| *X15* | 13.268 | 17.269 | 1.385 | 15.124 | 11.681 | 0.939 |
| *X60* | 12.174 | 20.844 | 1.466 | 18.448 | 17.700 | 0.669 |
| *X187* | 10.528 | 13.996 | 1.118 | 9.966 | 6.908 | 0.360 |
| *X199* | 13.562 | 23.778 | 1.581 | 15.735 | 17.169 | 1.342 |
| *X178* | 7.019 | 11.390 | 0.812 | 10.131 | 6.078 | 0.502 |
| *X186* | 10.317 | 12.635 | 1.163 | 17.651 | 9.497 | 1.078 |
| *X226* | 10.941 | 16.051 | 1.502 | 22.984 | 8.365 | 0.935 |
| *X98* | 11.986 | 17.710 | 0.593 | 12.371 | 19.173 | 0.420 |
| *X180* | 6.410 | 8.396 | 0.528 | 6.950 | 7.627 | 0.442 |
| *X221* | 12.516 | 17.409 | 0.978 | 14.692 | 14.320 | 0.905 |
| *X68* | 15.405 | 19.158 | 1.055 | 13.730 | 12.204 | 0.350 |
| *X151* | 15.678 | 21.916 | 3.167 | 29.189 | 15.546 | 1.222 |
| *X34* | 13.249 | 15.160 | 0.675 | 10.129 | 6.870 | 0.364 |
| *X207* | 8.388 | 9.629 | 1.060 | 11.803 | 5.871 | 0.573 |
| *X21* | 10.308 | 19.562 | 1.035 | 12.443 | 11.069 | 0.719 |
| *X2* | 16.445 | 19.152 | 1.512 | 20.315 | 15.028 | 1.051 |
| *X27* | 11.393 | 14.340 | 0.943 | 19.133 | 8.903 | 0.639 |

**B. Supplemental Tables**

Table B3 continued

| Line ID | Female[1] | | | Male[1] | | |
|---|---|---|---|---|---|---|
| | Carcass | Head | Gonad | Carcass | Head | Gonad |
| *X195* | 8.682 | 12.282 | 0.948 | 13.858 | 9.938 | 0.684 |
| *X26* | 11.127 | 17.954 | 1.011 | 26.225 | 16.545 | 1.041 |
| *X182* | 11.083 | 16.867 | 0.930 | 11.676 | 8.718 | 0.667 |
| *X197* | 14.849 | 17.591 | 1.850 | 10.517 | 10.228 | 0.637 |
| *X222* | 12.009 | 14.927 | 0.918 | 13.559 | 11.635 | 0.842 |
| *X11* | 14.062 | 22.917 | 2.531 | 21.450 | 17.817 | 0.910 |
| *X94* | 27.246 | 36.462 | 2.152 | 20.346 | 14.804 | 1.372 |
| *X202* | 15.886 | 23.917 | 3.829 | 22.008 | 14.136 | 0.959 |
| *X212* | 9.102 | 15.755 | 1.046 | 8.020 | 7.682 | 0.455 |
| *X174* | 11.707 | 17.004 | 1.084 | 13.158 | 12.285 | 0.522 |
| *X148* | 9.693 | 11.196 | 1.035 | 9.029 | 6.018 | 0.249 |
| *X191* | 10.498 | 24.585 | 1.109 | 9.784 | 17.607 | 0.820 |
| *X196* | 8.014 | 14.784 | 0.694 | 8.769 | 13.365 | 0.639 |
| *X220* | 20.338 | 33.626 | 2.569 | 31.056 | 25.886 | 1.858 |
| *X65* | 9.797 | 13.047 | 1.389 | 8.212 | 7.492 | 0.491 |
| *X1* | 11.408 | 19.247 | 1.139 | 17.544 | 11.899 | 0.996 |
| *X209* | 13.872 | 18.586 | 1.314 | 16.518 | 9.397 | 1.043 |
| *X231* | 7.276 | 7.339 | 0.848 | 8.047 | 6.329 | 0.326 |
| *X35* | 14.300 | 24.413 | 1.215 | 20.365 | 19.898 | 1.075 |
| *X214* | 14.497 | 22.215 | 1.102 | 19.734 | 16.042 | 1.064 |
| *X23* | 15.518 | 42.130 | 1.284 | 20.826 | 29.921 | 0.660 |
| *X179* | 22.353 | 31.968 | 2.398 | 15.769 | 15.735 | 0.827 |
| *X59* | 9.093 | 13.688 | 1.194 | 13.345 | 10.999 | 0.754 |
| *X106* | 23.854 | 39.778 | 4.843 | 31.716 | 25.794 | 1.110 |
| *X215* | 6.878 | 9.844 | 1.184 | 8.582 | 9.685 | 0.230 |
| *X193* | 8.297 | 18.015 | 0.677 | 12.783 | 13.651 | 1.210 |
| *X185* | 13.741 | 17.704 | 1.666 | 17.914 | 9.522 | 0.907 |
| *X170* | 5.616 | 14.010 | 0.436 | 11.639 | 9.788 | 0.298 |
| *X22* | 14.685 | 24.234 | 1.972 | 18.903 | 15.583 | 1.269 |
| *X223* | 13.303 | 58.961 | 1.817 | 13.178 | 30.213 | 0.464 |
| *X25* | 8.542 | 10.875 | 0.927 | 10.838 | 9.279 | 0.490 |
| *X238* | 10.316 | 9.519 | 0.753 | 9.225 | 6.187 | 0.471 |
| *X5* | 12.232 | 19.076 | 0.978 | 17.838 | 13.805 | 0.638 |

Table B3 continued

| Line ID | Female[1] | | | Male[1] | | |
|---|---|---|---|---|---|---|
| | Carcass | Head | Gonad | Carcass | Head | Gonad |
| *X145* | 12.716 | 22.604 | 1.458 | 27.439 | 15.704 | 0.697 |
| *X225* | 13.253 | 31.653 | 0.818 | 10.859 | 19.726 | 1.277 |
| *X9* | 16.860 | 19.489 | 1.501 | 29.358 | 21.485 | 0.733 |
| *X169* | 9.404 | 14.021 | 0.766 | 11.294 | 12.034 | 0.220 |
| *X205* | 34.054 | 47.229 | 3.190 | 31.081 | 58.236 | 2.330 |
| *X239* | 7.972 | 10.576 | 1.135 | 7.510 | 6.825 | 0.812 |
| *X245* | 5.187 | 5.359 | 0.692 | 5.934 | 4.472 | 0.183 |
| *X177* | 11.538 | 13.363 | 0.951 | 24.424 | 15.134 | 0.586 |
| *X183* | 16.696 | 24.551 | 1.196 | 18.946 | 14.266 | 1.206 |
| *X200* | 11.671 | 16.971 | 1.189 | 14.871 | 11.920 | 0.787 |
| *X173* | 8.565 | 18.646 | 0.588 | 11.963 | 10.961 | 0.460 |
| *X181* | 7.625 | 8.974 | 0.883 | 9.787 | 5.439 | 0.338 |
| *X62* | 21.722 | 23.947 | 2.101 | 22.217 | 9.186 | 2.647 |
| *X64* | 11.221 | 23.422 | 1.631 | 9.634 | 11.525 | 0.681 |
| *X105* | 9.883 | 11.193 | 1.317 | 9.399 | 8.838 | 0.572 |
| *X247* | 14.598 | 17.338 | 1.341 | 15.471 | 10.168 | 0.584 |
| *X233* | 6.744 | 7.220 | 0.658 | 6.379 | 4.329 | 0.319 |
| *X116* | 7.050 | 13.548 | 0.805 | 7.927 | 8.453 | 0.231 |
| *X244* | 12.262 | 12.436 | 1.933 | 13.135 | 12.641 | 0.871 |
| *X228* | 6.406 | 12.244 | 0.592 | 6.806 | 6.327 | 0.339 |
| *X216* | 12.643 | 20.502 | 0.944 | 18.972 | 18.698 | 1.063 |
| *X201* | 5.466 | 8.673 | 0.478 | 8.163 | 4.953 | 0.379 |
| *X218* | 10.508 | 17.409 | 0.718 | 9.095 | 13.025 | 0.407 |
| *X131* | 14.704 | 33.321 | 1.950 | 19.445 | 16.252 | 0.680 |
| *X198* | 13.321 | 22.288 | 1.974 | 17.599 | 16.260 | 0.740 |
| *X189* | 8.042 | 16.061 | 0.964 | 12.989 | 12.096 | 0.591 |
| *X227* | 10.944 | 19.243 | 0.822 | 9.461 | 10.494 | 0.541 |
| *X118* | 14.615 | 26.841 | 0.927 | 16.208 | 20.507 | 0.967 |
| *X236* | 9.137 | 15.012 | 1.400 | 10.488 | 11.007 | 0.666 |
| *X171* | 7.263 | 12.099 | 0.812 | 8.767 | 6.566 | 0.530 |
| *X28* | 10.726 | 19.336 | 1.021 | 11.657 | 16.205 | 0.765 |
| *X224* | 7.245 | 6.795 | 1.366 | 4.647 | 4.072 | 0.272 |
| *X104* | 7.915 | 12.034 | 0.650 | 12.991 | 8.451 | 0.562 |

**B. Supplemental Tables**

Table B3 continued

| Line ID | Female[1] | | | Male[1] | | |
|---|---|---|---|---|---|---|
| | Carcass | Head | Gonad | Carcass | Head | Gonad |
| *X234* | 6.870 | 12.717 | 0.860 | 8.236 | 7.969 | 0.377 |
| *X219* | 16.686 | 22.836 | 1.798 | 20.564 | 23.294 | 1.832 |
| *X190* | 6.315 | 13.226 | 0.610 | 6.046 | 7.359 | 0.735 |
| *X24* | 16.003 | 26.035 | 1.674 | 12.600 | 19.839 | 0.732 |

[1]Normalized β-galactosidase activity (units/mg).

**Table B4.** Relationship between the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites for the whole body dissection with 50kb cutoff for the maximum distance to the nearest DCC binding sites.

| Expression[2] | Tissue[3] | DCC binding site[4] | Spearman's correlation[5] | | Linear regression[5] | |
|---|---|---|---|---|---|---|
| | | | *Rho* | *P*-value[1] | $R^2$ | *P*-value[1] |
| Male/Female | Carcass | HAS | -0.243 | **0.033** | 0.062 | **0.030** |
| | | MLE | -0.270 | **0.016** | 0.082 | **0.011** |
| | | MSL2 | -0.227 | **0.048** | 0.060 | **0.035** |
| | | MSL3 | -0.205 | 0.051 | 0.063 | **0.022** |
| | Head | HAS | -0.268 | **0.021** | 0.045 | 0.056 |
| | | MLE | -0.174 | 0.084 | 0.024 | 0.112 |
| | | MSL2 | -0.382 | **0.002** | 0.097 | **0.010** |
| | | MSL3 | -0.193 | 0.062 | 0.018 | 0.145 |
| | Gonad | HAS | 0.028 | 0.582 | 4.9E-03 | 0.700 |
| | | MLE | 0.077 | 0.728 | 7.7E-03 | 0.755 |
| | | MSL2 | -0.018 | 0.448 | 1.1E-04 | 0.530 |
| | | MSL3 | 0.017 | 0.553 | 1.7E-03 | 0.628 |
| Female | Carcass | HAS | -0.079 | 0.278 | 0.013 | 0.200 |
| | | MLE | -0.098 | 0.220 | 0.018 | 0.148 |
| | | MSL2 | -4.2E-03 | 0.488 | 2.2E-03 | 0.366 |
| | | MSL3 | -0.171 | 0.087 | 0.040 | 0.056 |
| | Head | HAS | 0.059 | 0.671 | 5.7E-04 | 0.571 |
| | | MLE | 7.3E-04 | 0.502 | 9.1E-03 | 0.226 |
| | | MSL2 | 0.071 | 0.698 | 7.6E-03 | 0.737 |
| | | MSL3 | -0.062 | 0.311 | 0.030 | 0.083 |

Table B4 continued

| Expression[2] | Tissue[3] | DCC binding site[4] | Spearman's correlation[5] | | Linear regression[5] | |
|---|---|---|---|---|---|---|
| | | | *Rho* | *P*-value[1] | $R^2$ | *P*-value[1] |
| Female | Gonad | HAS | -0.024 | 0.428 | 0.014 | 0.187 |
| | | MLE | -0.090 | 0.240 | 0.024 | 0.110 |
| | | MSL2 | 0.005 | 0.516 | 5.2E-03 | 0.301 |
| | | MSL3 | -0.010 | 0.467 | 0.013 | 0.183 |
| Male | Carcass | HAS | -0.196 | 0.070 | 0.046 | 0.052 |
| | | MLE | -0.224 | **0.037** | 0.051 | **0.036** |
| | | MSL2 | -0.127 | 0.178 | 0.026 | 0.119 |
| | | MSL3 | -0.273 | **0.014** | 0.077 | **0.013** |
| | Head | HAS | -0.113 | 0.198 | 0.003 | 0.342 |
| | | MLE | -0.135 | 0.143 | 0.017 | 0.154 |
| | | MSL2 | -0.126 | 0.178 | 6.7E-03 | 0.276 |
| | | MSL3 | -0.182 | 0.073 | 0.045 | **0.044** |
| | Gonad | HAS | -0.070 | 0.299 | 1.7E-03 | 0.378 |
| | | MLE | -0.071 | 0.288 | 2.9E-03 | 0.336 |
| | | MSL2 | -0.052 | 0.354 | 1.7E-03 | 0.382 |
| | | MSL3 | -0.042 | 0.368 | 2.4E-03 | 0.350 |

[1]Cells with bold font indicate significant correlations ($P<0.05$); [2]Male-to-female reporter gene expression ratio, expression in females and males; [3]Target tissue; [4]The nearest HAS and the components of the DCC; [5]Spearman's correlation and linear regression for the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites.

## B. Supplemental Tables

**Table B5.** Relationship between the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites for the whole body dissection, for the whole body dissection without 50kb cutoff for the maximum distance to the nearest DCC binding sites.

| Expression[2] | Tissue[3] | DCC binding site[4] | Spearman's correlation[5] | | Linear regression[5] | |
|---|---|---|---|---|---|---|
| | | | *Rho* | *P*-value[1] | $R^2$ | *P*-value[1] |
| Male/Female | Carcass | HAS | 0.113 | 0.847 | 0.027 | 0.929 |
| | | MLE | -0.007 | 0.476 | 8.1E-03 | 0.790 |
| | | MSL2 | 0.107 | 0.832 | 0.022 | 0.908 |
| | | MSL3 | 0.074 | 0.747 | 0.027 | 0.933 |
| | Head | HAS | -0.052 | 0.320 | 4.4E-05 | 0.476 |
| | | MLE | 0.053 | 0.682 | 0.022 | 0.907 |
| | | MSL2 | -0.081 | 0.233 | 5.3E-06 | 0.508 |
| | | MSL3 | 0.016 | 0.556 | 0.012 | 0.835 |
| | Gonad | HAS | -0.021 | 0.426 | 8.6E-04 | 0.604 |
| | | MLE | 7.1E-03 | 0.526 | 1.2E-05 | 0.513 |
| | | MSL2 | -0.022 | 0.421 | 3.8E-03 | 0.711 |
| | | MSL3 | -0.010 | 0.464 | 3.3E-04 | 0.435 |
| Female | Carcass | HAS | -0.054 | 0.315 | 0.010 | 0.184 |
| | | MLE | -0.066 | 0.276 | 0.018 | 0.115 |
| | | MSL2 | -0.023 | 0.419 | 0.006 | 0.237 |
| | | MSL3 | -0.143 | 0.098 | 0.037 | **0.042** |
| | Head | HAS | 0.041 | 0.645 | 2.1E-03 | 0.341 |
| | | MLE | -0.004 | 0.486 | 0.016 | 0.125 |
| | | MSL2 | 0.030 | 0.608 | 2.2E-03 | 0.338 |
| | | MSL3 | -0.069 | 0.269 | 0.036 | **0.043** |
| | Gonad | HAS | -0.054 | 0.313 | 5.6E-03 | 0.250 |
| | | MLE | -0.145 | 0.095 | 0.025 | 0.079 |
| | | MSL2 | -0.043 | 0.348 | 4.7E-03 | 0.268 |
| | | MSL3 | -0.130 | 0.120 | 0.027 | 0.068 |
| Male | Carcass | HAS | -0.034 | 0.380 | 9.4E-04 | 0.608 |
| | | MLE | -0.116 | 0.148 | 3.7E-04 | 0.431 |
| | | MSL2 | -0.010 | 0.465 | 1.4E-03 | 0.630 |
| | | MSL3 | -0.134 | 0.114 | 0.001 | 0.388 |
| | Head | HAS | 4.1E-03 | 0.515 | 1.3E-03 | 0.372 |
| | | MLE | -0.010 | 0.466 | 8.5E-04 | 0.397 |
| | | MSL2 | -0.015 | 0.447 | 0.002 | 0.359 |
| | | MSL3 | -0.072 | 0.259 | 0.011 | 0.175 |

Table B5 continued

| Expression[2] | Tissue[3] | DCC binding site[4] | Spearman's correlation[5] | | Linear regression[5] | |
|---|---|---|---|---|---|---|
| | | | *Rho* | *P*-value[1] | $R^2$ | *P*-value[1] |
| Male | Gonad | HAS | -0.042 | 0.352 | 4.9E-04 | 0.421 |
| | | MLE | -0.103 | 0.176 | 0.006 | 0.241 |
| | | MSL2 | -0.035 | 0.377 | 4.2E-07 | 0.502 |
| | | MSL3 | -0.090 | 0.209 | 0.013 | 0.157 |

[1]Cells with bold font indicate significant correlations ($P<0.05$); [2]Male-to-female reporter gene expression ratio, expression in females and males; [3]Target tissue; [4]The nearest HAS and the components of the DCC; [5]Spearman's correlation and linear regression for the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites.

**Table B6.** Relation between $Log_2$(Male/Female expression) of the reporter gene, $Log_2$(Male/Female expression) of endogenous genes and the distance to the nearest DCC binding sites for the groups of homozygous and heterozygous females.

| Dependent variable[1] | Source of variance[3] | Df (effect, error) | *F*-value | *P*-value[5] |
|---|---|---|---|---|
| $Log_2$(M/F) of the reporter gene in carcass | 1: $Log_2$(M/F) of endogenous genes in carcass | 1, 38 | 1.553 | 0.22 |
| | 2: Distance to HAS[2] | 1, 38 | 1.219 | 0.138 |
| | 1 × 2 interaction | 1, 38 | 1.0E-04 | 0.993 |
| | 1: $Log_2$(M/F) of endogenous genes in carcass | 1, 41 | 1.672 | 0.203 |
| | 2: Distance to MLE binding sites | 1, 41 | 3.917 | 0.027 |
| | 1 × 2 interaction | 1, 41 | 0.013 | 0.909 |
| | 1: $Log_2$(M/F) of endogenous genes in carcass | 1, 37 | 1.822 | 0.185 |
| | 2: Distance to MSL2 binding sites[2] | 1, 37 | 3.148 | **0.042** |
| | 1 × 2 interaction | 1, 37 | 0.003 | 0.955 |
| | 1: $Log_2$(M/F) of endogenous genes in carcass | 1, 43 | 0.404 | 0.528 |
| | 2: Distance to MSL3 binding sites[2] | 1, 43 | 1.233 | 1.137 |
| | 1 × 2 interaction | 1, 43 | 0.013 | 0.91 |

**B. Supplemental Tables**

Table B6 continued

| Dependent variable[1] | Source of variance[3] | Df (effect, error) | F-value | P-value[5] |
|---|---|---|---|---|
| $Log_2$(M/F) of the reporter gene in head | 1: $Log_2$(M/F) of endogenous genes in head | 1, 39 | 0.191 | 0.664 |
| | 2: Distance to HAS[2] | 1, 39 | 3.624 | **0.032** |
| | 1 × 2 interaction | 1, 39 | 1.1E-04 | 0.974 |
| | 1: $Log_2$(M/F) of endogenous genes in head | 1, 43 | 0.623 | 0.434 |
| | 2: Distance to MLE binding sites[2] | 1, 43 | 1.697 | 0.1 |
| | 1 × 2 interaction | 1, 43 | 0.51 | 0.479 |
| | 1: $Log_2$(M/F) of endogenous genes in head | 1, 38 | 0.632 | 0.432 |
| | 2: Distance to MSL2 binding sites[2] | 1, 38 | 7.295 | **0.005** |
| | 1 × 2 interaction | 1, 38 | 0.432 | 0.515 |
| | 1: $Log_2$(M/F) of endogenous genes in head | 1, 44 | 0.487 | 0.489 |
| | 2: Distance to MSL3 binding sites[2] | 1, 44 | 1.378 | 0.123 |
| | 1 × 2 interaction | 1, 44 | 0.071 | 0.792 |
| $Log_2$(M/F) of the reporter gene in gonad | 1: $Log_2$(M/F) of endogenous genes in gonad | 1, 39 | 0.453 | 0.505 |
| | 2: Distance to HAS[2] | 1, 39 | 3.066 | 0.456 |
| | 1 × 2 interaction | 1, 39 | 0.608 | 0.44 |
| | 1: $Log_2$(M/F) of endogenous genes in gonad | 1, 42 | 0.465 | 0.499 |
| | 2: Distance to MLE binding sites[2] | 1, 42 | 1.388 | 0.377 |
| | 1 × 2 interaction | 1, 42 | 0.317 | 0.577 |
| | 1: $Log_2$(M/F) of endogenous genes in gonad | 1, 38 | 0.529 | 0.500 |
| | 2: Distance to MSL2 binding sites[2] | 1, 38 | 1.494 | 0.385 |
| | 1 × 2 interaction | 1, 38 | 0.529 | 0.472 |
| | 1: $Log_2$(M/F) of endogenous genes in gonad | 1, 44 | 1.118 | 0.296 |
| | 2: Distance to MSL3 binding sites[2] | 1, 44 | 0.931 | 0.33 |
| | 1 × 2 interaction | 1, 44 | 0.165 | 0.687 |
| $Log_2$(M/F) of the reporter gene in carcass | 1: Distance to HAS (bp) | 1, 26 | 4.591 | **0.021** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 34.04 | **3.8E-06** |
| | 1 × 2 interaction | 1, 26 | 0.011 | 0.918 |
| | 1: Distance to MLE (bp) | 1, 26 | 3.888 | **0.03** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 33.381 | **4.4E-06** |
| | 1 × 2 interaction | 1, 26 | 0.121 | 0.731 |

Table B6 continued

| Dependent variable[1] | Source of variance[3] | Df (effect, error) | F-value | P-value[5] |
|---|---|---|---|---|
| $Log_2$(M/F) of the reporter gene in carcass | 1: Distance to MSL2 (bp) | 1, 24 | 3.478 | **0.037** |
| | 2: Dosage of the reporter gene[4] | 1, 24 | 30.199 | **1.2E-05** |
| | 1 × 2 interaction | 1, 24 | 0.016 | 0.9 |
| | 1: Distance to MSL3 (bp) | 1, 26 | 1.508 | 0.115 |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 30.794 | **7.9E-06** |
| | 1 × 2 interaction | 1, 26 | 0.175 | 0.679 |
| $Log_2$(M/F) of the reporter gene in head | 1: Distance to HAS (bp) | 1, 26 | 15.561 | **2.7E-04** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 52.423 | **1.1E-07** |
| | 1 × 2 interaction | 1, 26 | 0.287 | 0.597 |
| | 1: Distance to MLE (bp) | 1, 26 | 19.112 | **9.0E-05** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 56.721 | **5.4E-08** |
| | 1 × 2 interaction | 1, 26 | 0.167 | 0.686 |
| | 1: Distance to MSL2 (bp) | 1, 24 | 15.536 | **3.1E-04** |
| | 2: Dosage of the reporter gene[4] | 1, 24 | 44.161 | **7.1E-07** |
| | 1 × 2 interaction | 1, 24 | 0.042 | 0.839 |
| | 1: Distance to MSL3 (bp) | 1, 26 | 9.181 | **2.7E-03** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 44.228 | **4.7E-07** |
| | 1 × 2 interaction | 1, 26 | 0.125 | 0.727 |
| $Log_2$(M/F) of the reporter gene in gonad | 1: Distance to HAS (bp) | 1, 26 | 3.012 | **0.047** |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 16.862 | **3.5E-04** |
| | 1 × 2 interaction | 1, 26 | 0.001 | 0.975 |
| | 1: Distance to MLE (bp) | 1, 26 | 2.095 | 0.08 |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 16.444 | **4.1E-04** |
| | 1 × 2 interaction | 1, 26 | 0.200 | 0.659 |
| | 1: Distance to MSL2 (bp) | 1, 24 | 1.777 | 0.098 |
| | Dosage of the reporter gene[4] | 1, 24 | 18.450 | **2.5E-04** |
| | 1 × 2 interaction | 1, 24 | 0.014 | 0.908 |
| | 1: Distance to MSL3 (bp) | 1, 26 | 1.381 | 0.125 |
| | 2: Dosage of the reporter gene[4] | 1, 26 | 15.953 | **4.8E-04** |
| | 1 × 2 interaction | 1, 26 | 0.067 | 0.797 |
| $Log_2$(M/F) of the reporter gene in carcass | 1: $Log_2$(M/F) of endogenous genes in carcass | 1, 20 | 4.674 | **0.043** |
| | 2: Dosage of the reporter gene[4] | 1, 20 | 28.534 | **3.2E-05** |
| | 1 × 2 interaction | 1, 20 | 1.605 | 0.220 |

**B. Supplemental Tables**

Table B6 continued

| Dependent variable[1] | Source of variance[3] | Df (effect, error) | F-value | P-value[5] |
|---|---|---|---|---|
| $Log_2$(M/F) of the reporter gene in head | 1: $Log_2$(M/F) of endogenous genes in head | 1, 18 | 2.130 | 0.162 |
| | 2: Dosage of the reporter gene[4] | 1, 18 | 19.501 | **3.3E-04** |
| | 1 × 2 interaction | 1, 18 | 0.064 | 0.803 |
| $Log_2$(M/F) of the reporter gene in gonad | 1: $Log_2$(M/F) of endogenous genes in gonad | 1, 20 | 0.406 | 0.531 |
| | 2: Dosage of the reporter gene[4] | 1, 20 | 16.815 | **1.0E-03** |
| | 1 × 2 interaction | 1, 20 | 0.682 | 0.419 |

[1]Male-to-female reporter gene expression ratio measures as $Log_2$(Male/female β-galactosidase activity) [2]The groups of distances to the nearest DCC binding sites: "close" (0–25 kb), "distant" (25–50 kb); [3]"1" – Covariate, "2" – Categorical explanatory variable. [4] The groups of females with homozygous and heterozygous reporter genes; [5] Cells with bold font indicate significant correlations ($P$<0.05).

**Table B7.** Relationship between the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest G4s.

| Expression[1] | Tissue[2] | Spearman's correlation[3] | | Linear regression[3] | |
|---|---|---|---|---|---|
| | | *Rho* | *P*-value[4] | $R^2$ | *P*-value[4] |
| Male/Female | Carcass | -0.014 | 0.903 | 0.001 | 0.746 |
| | Head | 0.117 | 0.290 | 0.001 | 0.765 |
| | Gonads | 0.103 | 0.356 | 0.012 | 0.330 |
| Male | Carcass | 0.168 | 0.128 | 0.021 | 0.196 |
| | Head | 0.245 | 0.025 | 0.032 | 0.105 |
| | Gonads | 0.147 | 0.185 | 0.069 | 0.016 |
| Female | Carcass | 0.172 | 0.121 | 0.032 | 0.106 |
| | Head | 0.218 | 0.047 | 0.051 | 0.040 |
| | Gonads | 0.177 | 0.109 | 0.011 | 0.339 |

[1]Male-to-female reporter gene expression ratio, expression in females and males; [2]Target tissue; [3]Spearman's correlation and linear regression for the expression of the *CMV-lacZ* reporter genes in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites. [4] Cells with bold font indicate significant correlations ($P$<0.05).

**Table B8.** Reporter gene expression for the whole body dissection in heterozygous females, homozygous females and males.

| Line ID | Heterozigos female[1] | | | Homozygous female[1] | | | Male[1] | | |
|---|---|---|---|---|---|---|---|---|---|
| | Carcass | Head | Gonad | Carcass | Head | Gonad | Carcass | Head | Gonad |
| *X232* | 8.999 | 10.951 | 1.251 | 16.234 | 18.995 | 1.693 | 18.969 | 11.006 | 1.335 |
| *X29* | 15.640 | 14.356 | 1.106 | 22.345 | 29.703 | 3.242 | 10.254 | 9.149 | 0.488 |
| *X8* | 10.673 | 12.956 | 1.216 | 23.054 | 30.138 | 2.722 | 14.578 | 10.053 | 0.703 |
| *X15* | 12.464 | 16.140 | 1.406 | 21.650 | 23.487 | 4.057 | 17.177 | 9.244 | 1.240 |
| *X221* | 12.963 | 13.814 | 1.274 | 22.000 | 20.089 | 3.321 | 15.159 | 9.769 | 1.071 |
| *X182* | 13.452 | 16.409 | 1.093 | 23.900 | 31.571 | 2.048 | 11.093 | 8.984 | 0.961 |
| *X212* | 12.064 | 22.174 | 1.161 | 14.004 | 17.235 | 1.586 | 9.464 | 9.396 | 0.606 |
| *X196* | 8.784 | 14.434 | 0.861 | 30.587 | 22.882 | 4.331 | 8.969 | 10.943 | 0.653 |
| *X65* | 10.107 | 15.215 | 1.812 | 10.461 | 14.004 | 0.666 | 6.890 | 6.089 | 0.396 |
| *X1* | 11.824 | 21.209 | 1.145 | 26.294 | 36.187 | 3.722 | 16.534 | 9.484 | 1.135 |
| *X35* | 14.969 | 12.000 | 1.381 | 28.057 | 27.928 | 3.746 | 15.902 | 14.244 | 1.255 |
| *X239* | 8.990 | 12.506 | 1.309 | 24.133 | 24.484 | 3.958 | 8.615 | 8.727 | 1.255 |
| *X183* | 15.629 | 23.983 | 1.285 | 33.964 | 35.996 | 2.938 | 20.846 | 14.035 | 1.413 |
| *X200* | 15.542 | 20.317 | 1.597 | 17.732 | 23.914 | 2.026 | 18.352 | 12.175 | 0.989 |
| *X244* | 14.542 | 12.268 | 2.773 | 16.219 | 18.708 | 1.868 | 13.813 | 11.981 | 0.816 |

[1]Normalized β-galactosidase activity (units/mg).

**B. Supplemental Tables**

**Table B9.** Reporter gene expression in different tissues for the head dissection.

| Line ID | Female | | Male | |
|---|---|---|---|---|
| | Head case | Brain | Head case | Brain |
| *X232* | 27.594 | 0.614 | 21.395 | 0.716 |
| *X82* | 25.823 | 0.455 | 21.880 | 0.662 |
| *X29* | 25.404 | 0.823 | 15.578 | 0.736 |
| *X8* | 36.944 | 2.116 | 21.554 | 1.086 |
| *X157* | 22.478 | 1.340 | 10.959 | 1.309 |
| *X15* | 42.155 | 2.555 | 21.460 | 1.012 |
| *X60* | 19.534 | 0.909 | 16.254 | 0.306 |
| *X199* | 60.344 | 2.547 | 31.332 | 2.098 |
| *X221* | 29.433 | 2.039 | 22.554 | 1.144 |
| *X34* | 26.615 | 0.274 | 13.738 | 0.304 |
| *X195* | 27.753 | 3.358 | 19.379 | 0.388 |
| *X26* | 21.032 | 0.380 | 14.824 | 0.687 |
| *X182* | 33.021 | 1.129 | 18.793 | 1.093 |
| *X197* | 27.985 | 0.341 | 19.317 | 0.269 |
| *X212* | 24.887 | 0.808 | 17.577 | 0.906 |
| *X196* | 30.722 | 0.745 | 17.914 | 0.846 |
| *X65* | 19.907 | 1.018 | 13.677 | 0.680 |
| *X1* | 27.012 | 0.632 | 19.699 | 0.659 |
| *X209* | 36.939 | 1.370 | 25.684 | 1.045 |
| *X231* | 28.442 | 1.142 | 18.340 | 0.538 |
| *X35* | 38.211 | 1.105 | 22.902 | 1.164 |
| *X59* | 27.657 | 0.883 | 14.906 | 0.443 |
| *X215* | 55.342 | 3.642 | 37.457 | 2.663 |
| *X223* | 29.129 | 0.719 | 17.197 | 0.578 |
| *X239* | 32.084 | 0.560 | 18.374 | 0.724 |
| *X183* | 58.028 | 2.627 | 31.904 | 2.335 |
| *X200* | 36.063 | 0.705 | 17.154 | 1.051 |
| *X62* | 28.805 | 0.853 | 20.369 | 0.459 |
| *X247* | 40.349 | 0.896 | 32.320 | 1.362 |
| *X244* | 23.227 | 0.789 | 17.752 | 0.337 |
| *X234* | 25.239 | 1.218 | 17.454 | 0.650 |
| *X24* | 35.780 | 1.439 | 22.138 | 1.797 |

[1]Normalized β-galactosidase activity (units/mg).

**Table B10.** Relationship between the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites for the head dissection.

| Expression[1] | Tissue[2] | DCC binding site[3] | Spearman's correlation[4] | | Linear regression[4] | |
|---|---|---|---|---|---|---|
| | | | *Rho* | *P*-value | $R^2$ | *P*-value |
| Male/Female | Brain | HAS | 0.066 | 0.720 | 0.004 | 0.744 |
| | | MLE | 0.068 | 0.711 | 0.012 | 0.545 |
| | | MSL2 | 0.021 | 0.910 | 0.002 | 0.832 |
| | | MSL3 | 0.079 | 0.666 | 0.020 | 0.441 |
| | Head case | HAS | 0.004 | 0.983 | 0.010 | 0.583 |
| | | MLE | 0.082 | 0.652 | 0.019 | 0.448 |
| | | MSL2 | 0.039 | 0.831 | 0.021 | 0.434 |
| | | MSL3 | -0.046 | 0.801 | 0.010 | 0.594 |
| Male | Brain | HAS | -0.125 | 0.495 | 0.060 | 0.175 |
| | | MLE | -0.091 | 0.618 | 0.035 | 0.308 |
| | | MSL2 | -0.168 | 0.357 | 0.069 | 0.145 |
| | | MSL3 | -0.088 | 0.629 | 0.035 | 0.306 |
| | Head case | HAS | -0.257 | 0.156 | 0.055 | 0.196 |
| | | MLE | -0.067 | 0.713 | 0.017 | 0.473 |
| | | MSL2 | -0.209 | 0.249 | 0.051 | 0.213 |
| | | MSL3 | -0.192 | 0.290 | 0.027 | 0.373 |
| Female | Brain | HAS | -0.188 | 0.302 | 0.065 | 0.160 |
| | | MLE | -0.217 | 0.232 | 0.047 | 0.232 |
| | | MSL2 | -0.195 | 0.284 | 0.069 | 0.148 |
| | | MSL3 | -0.198 | 0.275 | 0.052 | 0.209 |
| | Head case | HAS | -0.243 | 0.179 | 0.068 | 0.150 |
| | | MLE | -0.150 | 0.411 | 0.043 | 0.256 |
| | | MSL2 | -0.271 | 0.133 | 0.087 | 0.101 |
| | | MSL3 | -0.133 | 0.465 | 0.040 | 0.272 |

[1]Male-to-female reporter gene expression ratio, expression in females and males; [2]Target tissue; [3]The nearest HAS and the components of the DCC; [4]Spearman's correlation and linear regression for the expression of the *CMV-lacZ* reporter gene in males, females, male-to-female expression ratio, and proximity to the nearest DCC binding sites.

## B. Supplemental Tables

**Table B11.** Genomic characteristics of mutations found by genome-wide analysis in *INXS 1/2*.

| Chr[1] | Coordinate[2] | Line ID | Ref[3] | Mut[4] | Location[5] | Affected gene[6] | Proximal gene[7] | Distal gene[7] |
|---|---|---|---|---|---|---|---|---|
| X | 1,697[8] | *Low1* | G | A | Coding region | *wol* reporter | - | - |
| X | 2,992[8] | *Low2* | G | A | Coding region | *wol* reporter | - | - |
| X | 2,640,476 | *INXS1* | C | T | Intron | *Sgg** | - | - |
| X | 5,326,178 | *INXS1* | C | T | Coding region | *Mcm3** | *CG3309* | - |
| X | 13,607,528 | *INXS2* | C | T | Intron | *CG11178* | - | - |
| X | 14,180,973 | *INXS2* | C | T | Intergenic | - | - | *CG11584* |
| X | 14,515,518 | *INXS2* | C | T | Intergenic | - | - | - |
| X | 15,283,862 | *INXS2* | C | T | Intergenic | - | - | - |
| X | 16,239,180 | *INXS1* | T | A | Intergenic | - | - | - |
| X | 16,596,787 | *INXS2* | G | A | Coding region | *CG9784* | - | - |
| X | 16,693,568 | *INXS2* | G | A | Intron | *Septin4* | - | - |
| X | 16,847,455 | *INXS1* | AAG CCC AAG TAC | - | Coding region | *CG13003* | - | - |
| X | 17,340,022 | *INXS2* | C | T | Intergenic | - | - | - |
| X | 17,566,412 | *INXS2* | T | A | Intergenic | - | *ppk23* | - |
| X | 17,566,675 | *INXS2* | C | T | 3' UTR | *ppk23* | - | - |
| X | 17,699,406 | *INXS2* | A | G | 3' UTR | *CG8188* | - | - |
| X | 17,845,247 | *INXS1* | G | A | 5' UTR | *Ucp4A** | - | *CG8142* |
| X | 18,390,699 | *INXS2* | A | T | 3' UTR | *CG32549* | - | - |
| X | 18,506,862 | *INXS2* | A | T | Intron | *Ggt-1* | - | - |
| X | 20,012,220 | *INXS2* | A | T | Intron | *Dop2R* | - | - |
| X | 20,589,173 | *INXS2* | C | T | Intergenic | | *RunxA* | - |
| X | 20,869,041 | *INXS2* | G | A | Coding region | *CG1314* | - | - |
| X | 20,991,513 | *INXS2* | C | T | Intergenic | - | - | - |
| X | 21,453,252 | *INXS2* | A | T | Intron | *DIP-beta* | - | - |
| X | 22,636,181 | *INXS2* | C | T | Intergenic | - | - | - |
| 3R | 4,427,171 | *INXS2* | G | A | Coding region | *CG31525** | - | - |
| 3R | 4,427,425 | *INXS1* | C | T | Coding region | *CG31525** | - | - |
| 2L | 19,102,646 | *INXS1* | A | T | Coding region, 3' UTR, intron | *CG17344, CG43731, Lim3* | - | - |
| 2R | 7,683,208 | *INXS1* | C | T | Coding region | *LRR** | - | - |
| 2L | 20,102,465 | *INXS1* | G | A | Coding region | *Kua** | *CG13970* | - |

Table B11 continued.

| Chr[1] | Coordinate[2] | Line ID | Ref[3] | Mut[4] | Location[5] | Affected gene[6] | Proximal gene[7] | Distal gene[7] |
|---|---|---|---|---|---|---|---|---|
| 3L | 19,954,508 | *INXS1* | G | A | Coding region | *CG42654** | *CG42655* | *CG42263* |
| 3R | 13,959,941 | *INXS1* | G | A | Coding region | *Ipp** | *CG9925* | - |

[1]Chromosomal location of the mutation; [2]Genomic position; [3]Nucleotide bases on the 5'->3' DNA strand in *D. melanogaster* reference genome. [4]Nucleotide bases on the 5'->3' DNA strand in INXS mutants. [5]Types of genomic location; [6]Gene carrying the mutation; [7]Gene located within 1.5 kb of the mutation site; [8]Location is based on the DNA sequence of the *P[wFl-ocnlacZ]* reporter gene construct *Genes used as genotypic markers for recombination sites.

**Table B12.** Male fertility test with RNAi targeting three candidate genes (*CG13003*, *CG1314*, *CG31525*).

| Target gene | Gal4 promoter[1] | UAS-RNAi[2] | | UAS control[3] | | *p*-value[4] |
|---|---|---|---|---|---|---|
| | | Fertile | Sterile | Fertile | Sterile | |
| *CG13003* | *Bam* | 4 | 22 | 24 | 0 | 1.7E-10 |
| | *TJ* | 26 | 3 | 23 | 2 | 0.572 |
| | *Nanos* | 24 | 4 | 22 | 1 | 0.201 |
| *CG1314* | *Bam* | 24 | 0 | 25 | 0 | 1 |
| | *TJ* | 24 | 0 | 24 | 3 | 1 |
| | *Nanos* | 25 | 0 | 25 | 0 | 1 |
| | *Act5C* | 18 | 7 | 22 | 3 | 0.146 |
| *CG31525* | *Bam* | 22 | 1 | 24 | 0 | 0.489 |
| | *TJ* | 25 | 0 | 23 | 2 | 1 |
| | *Nanos* | 25 | 0 | 22 | 1 | 1 |

[1]Gal4 promoters specific for various germline development stages (*Bam*, *Nanos*, *TJ*) and promoter, which drives ubiquitous *Gal4* expression (*Act5C*); [2]The number of fertile and sterile males in the offspring after crossing females carrying the Gal4 coding sequence and males with adjacent to UAS inverted repeat motive of a target gene; [3]The number of fertile and sterile males in the offspring after crossing females carrying the *Gal4* coding sequence and males of the line used for the creation of a specific RNAi transgenic line. [4]Proportion of sterile males was tested using Fisher's Exact Test.

# Appendix C

## Laboratory Protocols

**Protocol C1.** DNA extraction from fly tissues with MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA).

1. Collect 1 whole fly or 1 whole fly without abdomens into a microcentrifuge tube and freeze the samples in liquid nitrogen.
2. Dilute 1 µl of Proteinase K into 300 µl of Tissue and Cell Lysis Solution for each sample.
3. Grind frozen tissues and add 300µl of Tissue and Cell Lysis Solution containing the Proteinase K and mix thoroughly.
4. Incubate at 65ºC for 15 min at 300 rpm and vortex every 5 minutes.
5. Cool the samples to 37ºC and add 1µl RNAse A, then mix thoroughly.
6. Incubate at 37ºC for 30 min and then place samples on ice for 3 – 5 min.
7. Add 175 µl MPC Protein Precipitation reagent and vortex for 10 sec.
8. Place the samples on ice for 3 – 5 min.
9. Pellet the debris by centrifugation at 4ºC for 10 min at 13,200 rpm in a microcentrifuge.
10. Transfer the supernatant to a clean microcentrifuge tube and discard the pellet.
11. Add 500 µl of Isopropanol to the recovered supernatant. Invert the tube 30 – 40 times.
12. Pellet the DNA by centrifugation at 4ºC for 10 min at 13,200 rpm in a microcentrifuge.
13. Remove all of the Isopropanol without dislodging the DNA pellet.
14. Rinse with 400 µl of 70% Ethanol and centrifuge at 4ºC for 3 min at 13,200 rpm.
15. Remove all of the residual Ethanol and dry pellet for 15 min at room temperature.
16. Resuspend the DNA in 30 µl dd$H_2O$ (26 µl for sequencing) and store the samples at -20ºC.

## C. Laboratory Protocols

**Protocol C2.** Restriction digest for the inverse PCR.

1. Prepare the reaction solution in a microcentrifuge tube:
   - 13.5 µl $ddH_2O$
   - 2.5 µl CutSmart Buffer (1%)
   - 0.5 µl RNAse A
   - 1 µl NEB restriction enzyme (New England Biolabs, Ipswich, MA, USA)
   - 7.5 µl DNA
2. Incubation:
   1) 65ºC for 25 min
   2) Overnight at -20ºC
3. Test restriction products using a 1% agarose gel electrophoresis (35 min, 100V).

**Protocol C3.** Ligation of DNA fragments.

1. Prepare the reaction solution in a microcentrifuge tube:
   345.5 µl $ddH_2O$
   1 µl 10X T4 DNA Ligase Buffer
   1 µl T4 DNA Ligase
   1 µl Restriction enzyme (New England Biolabs, Ipswich, MA, USA)
   10 µl Digested DNA
2. Incubate at 16°C overnight.

DNA precipitation:

1. Add 40 µl of 3M sodium acetate to the ligated DNA.
2. Add 1ml 96% ethanol and mix by inverting the tube.
3. Incubate at -80°C overnight.
4. Pellet the DNA by centrifugation at 4°C for 30 min at 14,000 rpm.
5. Remove supernatant and wash pellet with 500 µl cold 70% ethanol.
6. Centrifuge at 4°C for 15 min at 14,000 rpm.
7. Remove all residual Ethanol with a pipette and air dry pellet for 15 min at room temperature.
8. Resuspend the RNA in 20 µL $ddH_2O$
9. Measure DNA concentration on Nano-Drop and store the samples at -80°C.

## C. Laboratory Protocols

**Protocol C4.** Inverse PCR of self-ligated DNA fragments.

1. Prepare the reaction solution in a 200 µl PCR tube:

   10.4 µl $ddH_2O$

   2 µl 10x PCR buffer (-$MgCl_2$)

   0.8 µl $MgCl_2$

   0.8 µl 10 mM dNTPs

   0.4 µl Forward Primer (10 µM)

   0.4 µl Reverse Primer (10 µM)

   0.2 µl LongAmp Taq DNA Polymerase (New England Biolabs, Ipswich, MA, USA)

   5 µl DNA template

2. Thermocycling program:
   1) 95°C for 5 min
   2) 95°C for 30 sec
   3) 55°C for 30 sec
   4) 72°C for 1.5 min
   5) Repeat steps 2 - 4 for 39 times
   6) 72°C 10 min
   7) Hold at 8°C

**Protocol C5.** Sequencing of PCR products with BigDye Terminator v1.1 procedure (Thermo Fisher Scientific, Waltham, MA, USA).

1. For PCR products cleaning by ExoSAP (Thermo Fisher Scientific, Waltham, MA, USA) prepare the reaction solution in a 200 µl PCR tube:
   3.5 µl $ddH_2O$
   0.5 µl 10x PCR buffer (-$MgCl_2$)
   1 µl ExoSAP
   20 µl PCR product
2. Thermocycling program:
   1) 3795ºC for 30 min
   2) 80ºC for 15 min
   3) Hold at 8ºC
3. Prepare the reaction solution in a 200 µl PCR tube:
   2 µl BigDye reaction mix
   1 µl 5x BigDye Sequencing buffer
   1 µl Primer
   3 µl $ddH_2O$
   3 µl DNA template
4. Thermocycling conditions:
   1) 95ºC for 1 min
   2) 96ºC for 10 sec
   3) 50ºC for 15 sec
   4) 60ºC for 4 min
   5) Repeat steps 2 - 4 for 35 times
   6) Hold at 8ºC

## C. Laboratory Protocols

**Protocol C6.** PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA)

1. Prepare the reaction solution in a 200 µl PCR tube:
   - 12.4 µl $ddH_2O$
   - 4 µl 5X Phusion HF Buffer
   - 0.4 µl dNTPs (10 µM)
   - 1 µl Forward Primer (10 µM)
   - 1 µl Reverse Primer (10 µM)
   - 0.2 µl Phusion DNA Polymerase
   - 0.5 -1 µl Template DNA
2. Thermocycling program:
   1) 98°C for 30 sec
   2) 98°C for 5 sec
   3) 60-70°C for 20 sec
   4) 72°C for 30 sec
   5) Repeat steps 2 - 4 for 30 times
   6) 72°C 7 min
   7) Hold at 8°C

**Protocol C7.** Restriction digest of PCR products.

1. Prepare the reaction solution in a microcentrifuge tube:
   3 µl $ddH_2O$
   2.5 µl CutSmart Buffer (1%)
   1 µl Restriction NEB enzymes
   15 µl DNA
2. Incubate overnight at 37ºC or as specified for each enzyme
3. Test restriction products using a 1% agarose gel electrophoresis (35 min, 100V).

## C. Laboratory Protocols

**Protocol C8.** RNA extraction from fly tissues with MasterPureTM DNA Purification Kit (Epicentre, Madison, WI, USA).

1. Clean pestles and all surfaces with RNAse AWAY (Thermo Fisher Scientific, U.S.A) and 70% Ethanol.
2. Collect 10 dissected testes or 7 fly abdomens per sample into a microcentrifuge tube and freeze the samples in liquid nitrogen.
3. Dilute 2.5 µl of Proteinase K into 300 µl of Tissue and Cell Lysis Solution for each sample.
4. Grind frozen tissues and add 300 µL of the proteinase K/TC lysis solution, keep on ice.
5. Vortex thoroughly and incubate at 65ºC for 15 min; vortex every 5 min.
6. Place the samples on ice for 3-5 min and then add 175 µl of MPC Protein Precipitation Reagent to the lysed sample and vortex for 10 sec.
7. Pellet the precipitate by centrifugation.
8. Transfer the supernatant to a clean microcentrifuge tube and discard the pellet.
9. Add 500 µl of Isopropanol to the supernatant and invert the tube 30 – 40 times.
10. Pellet the total nucleic acids by centrifugation at 4°C for 10 min at 14,000 rpm.
11. Remove all residual isopropanol and rinse the pellet once with 70% ethanol.
12. Centrifuge at 4ºC for 10 min at 14,000rpm and remove all ethanol with a pipette.
13. Prepare 200 µl of DNase I solution for each sample by diluting 10 µl of RNase-free DNase I up to 200 µl with 1x DNase Buffer.
14. Resuspend the total nucleic acid pellet in 200 µl of DNase I solution and incubate at 37ºC for 30 minutes.
15. Add 200 µl of 2x Tissue and Cell Lysis Solution and vortex for 5 sec.
16. Add 200 µl of MPC Protein Precipitation Reagent and vortex 10 sec.
17. Place the samples on ice for 3 – 5 min and then pellet the debris by centrifugation at 4ºC for 10 min at 14,000 rpm.
18. Transfer the supernatant into a clean microcentrifuge tube and discard the pellet.
19. Add 500 µl of Isopropanol to the supernatant and invert the tube 30 – 40 times.
20. Pellet the RNA by centrifugation at 4ºC for 10 min at 14,000 rpm.
21. Remove the Isopropanol without dislodging the RNA pellet.
22. Add 400 µl 70% Ethanol, invert tube several times, centrifuge at 4ºC for 10 min at 14,000 rpm.

23. Remove all residual Ethanol with a pipette and air dry pellet for 15 min at room temperature.
24. Resuspend the RNA in 20 µL dd$H_2O$ and add 1 µL RiboGuard RNase Inhibitor (Epicenter, USA). Store RNA samples at -80_C.

## C. Laboratory Protocols

**Protocol C9.** First-strand cDNA generation

1. Prepare the reaction solution in a 200 µl PCR tube:
   - n µl — 650 ng RNA template for samples with 10 testis [1.5 µg for 7 whole fly samples]
   - 1 µl — Random hexamer primer (New England Biolabs, Ipswich, MA, USA)
   - 1 µl — dNTPs (10 µM)
   - n µl — RNase-free $H_2O$ (to 13 µl total volume)
2. Incubate 65ºC for 5 minutes and place on ice (4ºC) for at least 1 minute.
3. Test restriction products using a 1% agarose gel electrophoresis (35 min, 100V).
4. Spin down and add 7µL of the following mix:
   - 4 µl — 5X First-strand buffer (Invitrogen Life Technologies)
   - 1 µl — DTT (0.1M)
   - 1 µl — SuperScript III Reverse Transcriptase (Invitrogen Life Technologies, Carlsbad, CA, USA)
   - 15 µl — RNase-free $H_2O$
5. Mix by pipetting and incubate: 50ºC for 60 min, 70ºC for 17 min.
6. Spin down and add 1 µL RNase H
7. Incubate: 37ºC for 20 min
8. Add 6µL of RNase-free $H_2O$ to have a final approximate concentration of 25 ng/µL.
9. Check cDNA quality by PCR (Protocol 7) and 1% agarose gel electrophoresis (35 min, 100V)

**Protocol C10.** qPCR with SYBR Green (Invitrogen Life Technologies)

1. Pool 2-4 µL of each cDNA sample into a single stock to have enough cDNA for 5 standard solutions in 2-3 technical replicates for one plate:
   1) 3 µL cDNA — 1: 1 dilution (75 ng)
   2) 2 µL cDNA + 1 µL $ddH_2O$ — 1: 1.5 dilution (50 ng)
   3) 1 µL cDNA + 2 µL $ddH_2O$ — 1: 3 dilution (25 ng)
   4) 3 µL cDNA + 27 µL $ddH_2O$ — 1: 10 dilution (7.5 ng)
   5) 3 µL cDNA of a 1:10 dilution + 12 µL $ddH_2O$ — 1: 50 dilution (1.5 ng)
2. Prepare the master mix for standards:
   5 µl SSO advanced SYBR green (BioRad, USA)
   1 µl Forward Primer (2 µM)
   1 µl Reverse Primer (2 µM)
   3 µl Template cDNA from stock dilutions
3. Prepare the master mix for test samples:
   5 µl SSO advanced SYBR green
   4 µl 5X Phusion HF Buffer
   1 µl Forward Primer (2 µM)
   1 µl Reverse Primer (2 µM)
   1 µl Template cDNA
   2 µl $ddH_2O$
4. Thermocycling program:
   1) 95°C for 30 sec
   2) 95°C for 10 sec
   3) 61-63°C for 30 sec
   4) Plate read
   5) Repeat steps 2 – 4 for 35-40 times
   6) 95°C 10 sec
   7) Melt curve with 65-95 °C, increment 0.5°C
   8) Plate read

## C. Laboratory Protocols

**Protocol C11.** β-galactosidase enzymatic activity assay for pooled samples

1. Collect tissues from 5 flies for the whole fly dissection [10 flies for the head dissection] in a microcentrifuge tube.
2. Homogenized tissue in 200 µl [135 µl] of lysis buffer (0.1 M Tris-HCl, 1 mM EDTA, 7 mM 2-mercaptoethanol; pH 7.5).
3. Incubate for 15 min on ice and centrifuge at 4ºC for 15 min at 12,500 rpm.
4. Transfer 130 µl of supernatant to a 200 µl PCR tube.
5. Transfer 50 µl of the protein extract into two wells of a 96-well spectrophotometer plate for two technical replicates. Store the rest of the protein extract in -20 ºC for the Lowry assay.
6. Add 53 µl of 2X assay buffer (200 mM sodium phosphate (pH 7.3), 2 mM $MgCl_2$, 100 mM 2-mercaptoethanol, 1.33 mg/ml o-nitrophenyl-β-D-galactopyranoside) to another 96-well plate.
7. Add 50 µl of the 2X assay buffer to the spectrophotometer plate with a multichannel pipette and immediately run measurement of absorbance with a spectrophotometer for a total of 58 min at 420 nm at 37°.

**Protocol C12.** β-galactosidase enzymatic activity assay for single male samples

1. Collect whole flies or fly abdomens in a microcentrifuge tube.
2. Homogenized tissue in 55 µl of lysis buffer (0.1 M Tris-HCl, 1 mM EDTA, 7 mM 2-mercaptoethanol; pH 7.5). Alternatively, freeze samples in liquid nitrogen, then grind frozen tissues and add 55 µl of the lysis buffer.
3. Incubate for 15 min on ice and centrifuge at 4ºC for 8 min at 12,500 rpm.
4. Transfer 50 µl of supernatant to a 96-well spectrophotometer plate.
5. Add 53 µl of the 2X assay buffer (200 mM sodium phosphate (pH 7.3), 2 mM $MgCl_2$, 100 mM 2-mercaptoethanol, 1.33 mg/ml o-nitrophenyl-β-D-galactopyranoside) to another 96-well plate.
6. Add 50 µl of the 2X assay buffer to the spectrophotometer plate with a multichannel pipette and immediately run measurement of absorbance with a spectrophotometer for a total of 58 min at 420 nm at 37°.

## C. Laboratory Protocols

**Protocol C13.** Total soluble protein concentration measuring with the Lowry assay

1. Dilute 10 µl (for the whole fly dissection) or 10 µl (for the head dissection) of the protein extract in 200 µl of $ddH_2O$ in a microcentrifuge tube.
2. Prepare standard BSA (Bovine Serum Albumin) solutions:
   1) 200 µL $ddH_2O$ 0 µL 0.5 mg/ml BSA
   2) 195 µL $ddH_2O$ 5 µL 0.5 mg/ml BSA
   3) 190 µL $ddH_2O$ 10 µL 0.5 mg/ml BSA
   4) 185 µL $ddH_2O$ 15 µL 0.5 mg/ml BSA
   5) 180 µL $ddH_2O$ 20 µL 0.5 mg/ml BSA
   6) 175 µL $ddH_2O$ 25 µL 0.5 mg/ml BSA
   7) 170 µL $ddH_2O$ 30 µL 0.5 mg/ml BSA
3. Add of 200 µl of CTC working solution (0.025% (wt/vol) copper sulfate, 0.025% (wt/vol) potassium tartrate, 2.5% (wt/vol) sodium carbonate, 0.2 N NaOH, 2.5% SDS) to test and standard tubes.
4. Incubate for 10 min at room temperature.
5. Add 20% of Folin and Ciocalteu's phenol reagent (Sigma–Aldrich, Steinheim, Germany) and incubate at room temperature for 30 min.

   Measure the absorbance at 340 nm was for two technical replicates with a spectrophotometer.

# Bibliography

Albig C, Tikhonova E, Krause S, Maksimenko O, Regnard C and Becker PB (2019) Factor cooperation for chromosome discrimination in *Drosophila*. Nucleic Acids Res 47:1706–1724. doi: 10.1093/nar/gky1238

Alekseyenko AA, Ho JWK, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. (2012) Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. PLOS Genetics 8:e1002646. doi: 10.1371/journal.pgen.1002646

Alekseyenko AA, Larschan E, Lai WR, Park PJ and Kuroda MI (2006) High-resolution ChIP–chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. Genes Dev 20:848–857. doi: 10.1101/gad.1400206

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee O-K, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. (2008) A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. Cell 134:599–609. doi: 10.1016/j.cell.2008.06.033

Alexander ML (2015) Detection and separation of spontaneous and induced translocations in mature and immature germ cells of *Drosophila*. Am Nat. doi: 10.1086/282238

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi: 10.1016/S0022-2836(05)80360-2

**Bibliography**


Andrews J, Bouffard GG, Cheadle C, Lü J, Becker KG and Oliver B (2000) Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. Genome Res 10:2030–2043. doi: 10.1101/gr.159800

Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW and White KP (2002) Gene expression during the life cycle of *Drosophila melanogaster*. Science 297:2270–2275. doi: 10.1126/science.1072152

Argyridou E and Parsch J (2018) Regulation of the X chromosome in the germline and soma of *Drosophila melanogaster* males. Genes 9:242. doi: 10.3390/genes9050242

Arnqvist G and Rowe L (2005) Sexual Conflict. Princeton University Press

Ashburner M (2005) *Drosophila*, 2. ed. Cold Spring Harbor Laboratory Press

Assis R, Zhou Q and Bachtrog D (2012) Sex-biased transcriptome evolution in *Drosophila*. Genome Biol Evol 4:1189–1200. doi: 10.1093/gbe/evs093

Baarends WM, Wassenaar E, Laan R van der, Hoogerbrugge J, Sleddens-Linkels E, Hoeijmakers JHJ, Boer P de and Grootegoed JA (2005) Silencing of unpaired chromatin and histone H2A ubiquitination in mammalian meiosis. Mol Cell Biol 25:1041–1053. doi: 10.1128/MCB.25.3.1041-1053.2005

Bachtrog D (2013) Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. Nat Rev Genet 14:113–124. doi: 10.1038/nrg3366

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R et al. (2014) Sex determination: why so many ways of doing it? PLoS Biol 12:e1001899. doi: 10.1371/journal.pbio.1001899

Bachtrog D, Toda NRT and Lockton S (2010) Dosage compensation and demasculinization of X chromosomes in *Drosophila*. Curr Biol 20:1476–1481. doi: 10.1016/j.cub.2010.06.076

Bean CJ, Schaner CE and Kelly WG (2004) Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. Nature Genetics 36:100–105. doi: 10.1038/ng1283

Beckmann K, Grskovic M, Gebauer F and Hentze MW (2005) A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. Cell 122:529–540. doi: 10.1016/j.cell.2005.06.011

Bell G (1982) The masterpiece of nature: the genetics and evolution of sexuality. University of California Press, Berkeley and Los Angeles

Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, Evans-Holm M, Hiesinger PR, Schulze KL, Rubin GM et al. (2004) The *BDGP* gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. Genetics 167:761–781. doi: 10.1534/genetics.104.026427

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300. doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blumenstiel JP, Noll AC, Griffiths JA, Perera AG, Walton KN, Gilliland WD, Hawley RS and Staehling-Hampton K (2009) Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. Genetics 182:25–32. doi: 10.1534/genetics.109.101998

Bochman ML, Paeschke K and Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet 13:770–780. doi: 10.1038/nrg3296

Bouchon D, Rigaud T and Juchault P (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. Proc R Soc Lond B Biol Sci 265:1081–1090. doi: 10.1098/rspb.1998.0402

Brand AH and Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401–415.

Brenman JE, Gao F-B, Jan LY and Jan YN (2001) Sequoia, a tramtrack-related zinc finger protein, functions as a pan-neural regulator for dendrite and axon morphogenesis in *Drosophila*. Dev Cell 1:667–677. doi: 10.1016/S1534-5807(01)00072-7

Brown EJ, Nguyen AH and Bachtrog D (2020) The Y chromosome may contribute to sex-specific ageing in *Drosophila*. Nat Ecol Evol 4:853–862. doi: 10.1038/s41559-020-1179-5

Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM et al. (2014) Diversity and dynamics of the *Drosophila* transcriptome. Nature 512:393–399. doi: 10.1038/nature12962

**Bibliography**

Bull JJ (1983) Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.

Bulmer MG and Parker GA (2002) The evolution of anisogamy: a game-theoretic approach. Proc R Soc Lond B Biol Sci 269:2381–2388. doi: 10.1098/rspb.2002.2161

Burge S, Parkinson GN, Hazel P, Todd AK and Neidle S (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res 34:5402–5415. doi: 10.1093/nar/gkl655

Cabrero J, Teruel M, Carmona FD, Jiménez R and Camacho JPM (2007) Histone H3 lysine 9 acetylation pattern suggests that X and B chromosomes are silenced during entire male meiosis in a grasshopper. Cytogenet Genome Res 119:135–142. doi: 10.1159/000109630

Campos JL, Johnston KJA and Charlesworth B (2018) The effects of sex-biased gene expression and X-linkage on rates of sequence evolution in *Drosophila*. Mol Biol Evol 35:655–665. doi: 10.1093/molbev/msx317

Capel B (2017) Vertebrate sex determination: evolutionary plasticity of a fundamental switch. Nat Rev Genet 18:675–689. doi: 10.1038/nrg.2017.60

Carrel L and Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434:400–404. doi: 10.1038/nature03479

Carvalho AB, Koerich LB and Clark AG (2009) Origin and evolution of Y chromosomes: *Drosophila* tales. Trends Genet 25:270–277. doi: 10.1016/j.tig.2009.04.002

Carvalho AB, Vicoso B, Russo CAM, Swenor B and Clark AG (2015) Birth of a new gene on the Y chromosome of *Drosophila melanogaster*. PNAS 112:12450–12455. doi: 10.1073/pnas.1516543112

Castrillon DH, Gönczy P, Alexander S, Rawson R, Eberhart CG, Viswanathan S, DiNardo S and Wasserman SA (1993) Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single P element mutagenesis. Genetics 135:489–505.

Catalán A, Hutter S and Parsch J (2012) Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics 13:654. doi: 10.1186/1471-2164-13-654

Chang PL, Dunham JP, Nuzhdin SV and Arbeitman MN (2011) Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. BMC Genomics 12:364. doi: 10.1186/1471-2164-12-364

Charlesworth B, Campos JL and Jackson BC (2018) Faster-X evolution: theory and evidence from *Drosophila*. Mol Ecol 27:3753–3771. doi: 10.1111/mec.14534

Charlesworth B, Harvey PH, Charlesworth B and Charlesworth D (2000) The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563–1572. doi: 10.1098/rstb.2000.0717

Charlesworth D (2019) Young sex chromosomes in plants and animals. New Phytol 224:1095–1107. doi: https://doi.org/10.1111/nph.16002

Charlesworth D, Charlesworth B and Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118–128. doi: 10.1038/sj.hdy.6800697

Charnov EL and Bull J (1977) When is sex environmentally determined? Nature 266:828–830. doi: 10.1038/266828a0

Chen D, Ahlford A, Schnorrer F, Kalchhauser I, Fellner M, Viràgh E, Kiss I, Syvänen A-C and Dickson BJ (2008) High-resolution, high-throughput SNP mapping in *Drosophila melanogaster*. Nat Methods 5:323–329. doi: 10.1038/nmeth.1191

Chen W-F, Rety S, Guo H-L, Dai Y-X, Wu W-Q, Liu N-N, Auguin D, Liu Q-W, Hou X-M, Dou S-X et al. (2018) Molecular mechanistic insights into *Drosophila* DHX36-mediated G-quadruplex unfolding: a structure-based model. Structure 26:403-415.e4. doi: 10.1016/j.str.2018.01.008

Chippindale AK and Rice WR (2001) Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. PNAS 98:5677–5682. doi: 10.1073/pnas.101456898

Chlamydas S, Holz H, Samata M, Chelmicki T, Georgiev P, Pelechano V, Dündar F, Dasmeh P, Mittler G, Cadete FT et al. (2016) Functional interplay between MSL1 and CDK7 controls RNA polymerase II Ser5 phosphorylation. Nat Struct Mol Biol 23:580–589. doi: 10.1038/nsmb.3233

**Bibliography**

Coelho SM, Mignerot L and Cock JM (2019) Origin and evolution of sex-determination systems in the brown algae. New Phytol 222:1751–1756. doi: https://doi.org/10.1111/nph.15694

Connallon T and Knowles LL (2006) Evidence for overdominant selection maintaining X-linked fitness variation in *Drosophila melanogaster*. Evolution (N Y) 60:1445–1453. doi: https://doi.org/10.1111/j.0014-3820.2006.tb01223.x

Cooley L, Kelley R and Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements. Science (80- ) 239:1121–1128. doi: 10.1126/science.2830671

Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R and Becker PB (2006) Targeting determinants of dosage compensation in *Drosophila*. PLoS Genet 2:e5. doi: 10.1371/journal.pgen.0020005

Deal SL and Yamamoto S (2019) Unraveling novel mechanisms of neurodegeneration through a large-scale forward genetic screen in *Drosophila*. Front Genet. doi: 10.3389/fgene.2018.00700

Dean R and Mank JE (2016) Tissue specificity and sex-specific regulatory variation permit the evolution of sex-biased gene expression. Am Nat 188:E74–E84. doi: 10.1086/687526

Dean R, Perry JC, Pizzari T, Mank JE and Wigby S (2012) Experimental evolution of a novel sexually antagonistic allele. PLoS Genet 8:e1002917. doi: 10.1371/journal.pgen.1002917

Delph LF (2009) Sex allocation: evolution to and from dioecy. Curr Biol 19:R249–R251. doi: 10.1016/j.cub.2009.01.048

Demakova OV, Kotlikova IV, Gordadze PR, Alekseyenko AA, Kuroda MI and Zhimulev IF (2003) The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster*. Chromosoma 112:103–115. doi: 10.1007/s00412-003-0249-1

Demir E and Dickson BJ (2005) *fruitless* splicing specifies male courtship behavior in *Drosophila*. Cell 121:785–794. doi: 10.1016/j.cell.2005.04.027

Deng X, Koya SK, Kong Y and Meller VH (2009) Coordinated regulation of heterochromatic genes in *Drosophila melanogaster* males. Genetics 182:481–491.

Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 448:151–156. doi: 10.1038/nature05954

Duffy JB (2002) GAL4 system in drosophila: A fly geneticist's swiss army knife. Genesis 34:1–15. doi: https://doi.org/10.1002/gene.10150

Ellegren H, Hultin-Rosenberg L, Brunström B, Dencker L, Kultima K and Scholz B (2007) Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol 5:40. doi: 10.1186/1741-7007-5-40

Ellegren H and Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet 8:689–698. doi: 10.1038/nrg2167

Falconer DS and Mackay TFC (1996) Introduction to quantitative genetics, Subsequent edition. Benjamin-Cummings Pub Co, Harlow

Feltz CJ and Miller GE (1996) An asymptotic test for the equality of coefficients of variation from *k* populations. Stat Med 15:647–658. doi: 10.1002/(SICI)1097-0258(19960330)15:6<647::AID-SIM184>3.0.CO;2-P

Ferrari F, Plachetka A, Alekseyenko AA, Jung YL, Ozsolak F, Kharchenko PV, Park PJ and Kuroda MI (2013) "Jump start and gain" model for dosage compensation in *Drosophila* based on direct sequencing of nascent transcripts. Cell Rep 5:629–636. doi: 10.1016/j.celrep.2013.09.037

Figueiredo MLA, Kim M, Philip P, Allgardsson A, Stenberg P and Larsson J (2014) Non-coding *roX* RNAs prevent the binding of the MSL-complex to heterochromatic regions. PLoS Genet 10:e1004865. doi: 10.1371/journal.pgen.1004865

Franke A, Dernburg A, Bashaw GJ and Baker BS (1996) Evidence that MSL-mediated dosage compensation in *Drosophila* begins at blastoderm. Development 122:2751–2760.

Fry JD (2010) The genomic location of sexually antagonistic variation: some cautionary comments. Evolution (N Y) 64:1510–1516. doi: https://doi.org/10.1111/j.1558-5646.2009.00898.x

Gallach M and Betrán E (2016) Dosage compensation and the distribution of sex-biased gene expression in *Drosophila*: considerations and genomic constraints. J Mol Evol 82:199–206. doi: 10.1007/s00239-016-9735-y

**Bibliography**

Gao F-B, Brenman JE, Jan LY and Jan YN (1999) Genes regulating dendritic outgrowth, branching, and routing in *Drosophila*. Genes Dev 13:2549–2561.

Gebauer F, Merendino L, Hentze MW and Valcárcel J (1998) The *Drosophila* splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. RNA 4:142–150.

Geffroy B and Wedekind C (2020) Effects of global warming on sex ratios in fishes. J Fish Biol 97:596–606. doi: https://doi.org/10.1111/jfb.14429

Gerlach SU and Herranz H (2020) Genomic instability and cancer: lessons from *Drosophila*. Open Biol 10:200060. doi: 10.1098/rsob.200060

Gevedon O, Bolus H, Lye SH, Schmitz K, Fuentes-González J, Hatchell K, Bley L, Pienaar J, Loewen C and Chtarbanova S (2019) In vivo forward genetic screen to identify novel neuroprotective genes in *Drosophila melanogaster*. JoVE 59720. doi: 10.3791/59720

Gibson JR, Chippindale AK and Rice WR (2002) The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc R Soc Lond B Biol Sci 269:499–505. doi: 10.1098/rspb.2001.1863

Gilfillan GD, Straub T, Wit E de, Greil F, Lamm R, Steensel B van and Becker PB (2006) Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. Genes Dev 20:858–870. doi: 10.1101/gad.1399406

Gnad F and Parsch J (2006) Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. Bioinformatics 22:2577–2579. doi: 10.1093/bioinformatics/btl422

Grath S and Parsch J (2016) Sex-biased gene expression. Annu Rev Genet 50:29–44. doi: 10.1146/annurev-genet-120215-035429

Gu L, Reilly PF, Lewis JJ, Reed RD, Andolfatto P and Walters JR (2019) Dichotomy of dosage compensation along the neo Z chromosome of the monarch butterfly. Curr Biol 29:4071-4077.e3. doi: 10.1016/j.cub.2019.09.056

Gu L and Walters JR (2017) Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details. Genome Biol Evol 9:2461–2476. doi: 10.1093/gbe/evx154

Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, Malley JD, Eastman PS and Oliver B (2006) Global analysis of X-chromosome dosage compensation. J Biol 5:3. doi: 10.1186/jbiol30

Haelterman NA, Jiang L, Li Y, Bayat V, Sandoval H, Ugur B, Tan KL, Zhang K, Bei D, Xiong B et al. (2014) Large-scale identification of chemically induced mutations in *Drosophila melanogaster*. Genome Res 24:1707–1718. doi: 10.1101/gr.174615.114

Hafezi Y, Sruba SR, Tarrash SR, Wolfner MF and Clark AG (2020) Dissecting fertility functions of *Drosophila* Y chromosome genes with CRISPR. Genetics 214:977–990. doi: 10.1534/genetics.120.302672

Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A and Kadlec J (2012) Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. Mol Cell 48:587–600. doi: 10.1016/j.molcel.2012.09.014

Hamada FN, Park PJ, Gordadze PR and Kuroda MI (2005) Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. Genes Dev 19:2289–2294. doi: 10.1101/gad.1343705

Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M et al. (2016) G-quadruplex structures mark human regulatory chromatin. Nat Genet 48:1267–1272. doi: 10.1038/ng.3662

Hense W, Baines JF and Parsch J (2007) X chromosome inactivation during *Drosophila* spermatogenesis. PLoS Biol 5:e273. doi: 10.1371/journal.pbio.0050273

Holleley CE, O'Meally D, Sarre SD, Marshall Graves JA, Ezaz T, Matsubara K, Azad B, Zhang X and Georges A (2015) Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. Nature 523:79–82. doi: 10.1038/nature14574

Hörandl E (2009) A combinational theory for maintenance of sex. Heredity (Edinb) 103:445–457. doi: 10.1038/hdy.2009.85

Hoyle HD, Hutchens JA, Turner FR and Raff EC (1995) Regulation of β-tubulin function and expression in *Drosophila* spermatogenesis. Dev Genet 16:148–170. doi: https://doi.org/10.1002/dvg.1020160208

Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM and Mackay TF (2016) Spontaneous mutations and the origin and maintenance of quantitative genetic variation. Elife 5:e14625. doi: 10.7554/eLife.14625

Huylmans AK and Parsch J (2015) Variation in the X:autosome distribution of male-biased genes among *Drosophila melanogaster* tissues and its relationship with dosage compensation. Genome Biol Evol 7:1960–1971. doi: 10.1093/gbe/evv117

Huylmans AK and Parsch J (2014) Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. G3 (Bethesda) 4:2307–2315. doi: 10.1534/g3.114.013417

Innocenti P and Morrow EH (2010) The sexually antagonistic genes of *Drosophila melanogaster*. PLoS Biol 8:e1000335. doi: 10.1371/journal.pbio.1000335

Itoh Y, Melamed E, Yang X, Kampf K, Wang S, Yehya N, Van Nas A, Replogle K, Band MR, Clayton DF et al. (2007) Dosage compensation is less effective in birds than in mammals. J Biol 6:2. doi: 10.1186/jbiol53

Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR and Oliver B (2011) Synthetic spike-in standards for RNA-seq experiments. Genome Res 21:1543–1551. doi: 10.1101/gr.121095.111

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G and Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nat Genet 29:389–395. doi: 10.1038/ng766

Kanippayoor RL, Alpern JHM and Moehring AJ (2020) A common suite of cellular abnormalities and spermatogenetic errors in sterile hybrid males in *Drosophila*. Proc R Soc Lond B Biol Sci 287:20192291. doi: 10.1098/rspb.2019.2291

Kavi HH, Fernandez H, Xie W and Birchler JA (2008) Genetics and biochemistry of RNAi in *Drosophila*. Curr Top Microbiol Immunol 320:37–75. doi: 10.1007/978-3-540-75157-1_3

Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL and Kuroda MI (1999) Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. Cell 98:513–522. doi: 10.1016/S0092-8674(00)81979-0

Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V and Kuroda MI (1995) Expression of *Msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. Cell 81:867–877. doi: 10.1016/0092-8674(95)90007-1

Kelley RL, Wang J, Bell L and Kuroda MI (1997) Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. Nature 387:195–199. doi: 10.1038/387195a0

Kemkemer C, Hense W and Parsch J (2011) Fine-scale analysis of X chromosome inactivation in the male germ line of *Drosophila melanogaster*. Mol Biol Evol 28:1561–1563. doi: 10.1093/molbev/msq355

Kennison JA (1983) Analysis of Y-linked mutations to male sterility in *Drosophila melanogaster*. Genetics 103:219–234.

Khodursky S, Svetec N, Durkin SM and Zhao L (2020) The evolution of sex-biased gene expression in the *Drosophila* brain. Genome Res 30:874–884. doi: 10.1101/gr.259069.119

Kim D, Blus BJ, Chandra V, Huang P, Rastinejad F and Khorasanizadeh S (2010) Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. Nat Struct Mol Biol 17:1027.

Kind J and Akhtar A (2007) Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. Genes Dev 21:2030–2040.

Koellner CM, Mensink KA and Highsmith WE (2020) Basic concepts in human molecular genetics. Essential Concepts in Molecular Pathology. Elsevier, pp 81–99

Koerich LB, Wang X, Clark AG and Carvalho AB (2008) Low conservation of gene content in the *Drosophila* Y chromosome. Nature 456:949–951. doi: 10.1038/nature07463

Kossack ME and Draper BW (2019) Genetic regulation of sex determination and maintenance in zebrafish (*Danio rerio*). Current Topics in Developmental Biology. Elsevier, pp 119–149

Kuroda MI, Hilfiker A and Lucchesi JC (2016) Dosage compensation in *Drosophila*—a model for the coordinate regulation of transcription. Genetics 204:435–450. doi: 10.1534/genetics.115.185108

Landeen EL, Muirhead CA, Wright L, Meiklejohn CD and Presgraves DC (2016) Sex chromosome-wide transcriptional suppression and compensatory *cis*-regulatory evolution mediate gene expression in the *Drosophila* male germline. PLoS Biol 14:e1002499. doi: 10.1371/journal.pbio.1002499

Lane N (2019) Why is life the way it is? Molecular Frontiers Journal 3:20–28.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B and Clark AG (2008) Evolution of protein-coding genes in *Drosophila*. Trends Genet 24:114–123. doi: 10.1016/j.tig.2007.12.001

Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ and Kuroda MI (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell 28:121–133. doi: 10.1016/j.molcel.2007.08.011

Larsson J, Svensson MJ, Stenberg P and Mäkitalo M (2004) Painting of fourth in genus *Drosophila* suggests autosome-specific gene regulation. PNAS 101:9728–9733. doi: 10.1073/pnas.0400978101

Laurie-Ahlberg CC and Stam LF (1987) Use of *P*-element-mediated transformation to identify the molecular basis of naturally occurring variants affecting *Adh* expression in *Drosophila melanogaster*. Genetics 115:129–140.

Leader DP, Krause SA, Pandit A, Davies SA and Dow JAT (2018) FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res 46:D809–D815. doi: 10.1093/nar/gkx976

Lemos B, Branco AT and Hartl DL (2010) Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. PNAS 107:15826–15831. doi: 10.1073/pnas.1010383107

Leonard JL (2018) The evolution of sexual systems in animals. In: Leonard JL (ed) Transitions Between Sexual Systems: Understanding the Mechanisms of, and Pathways Between, Dioecy, Hermaphroditism and Other Sexual Systems. Springer International Publishing, Cham, pp 1–58

Lewis EB and Bacher F (1968) Method of feeding ethyl methane sulfonate (EMS) to Drosophila males. Dros Inf Serv 43:193.

Li X-Y and Gui J-F (2018) Diverse and variable sex determination mechanisms in vertebrates. Sci China Life Sci 61:1503–1514. doi: 10.1007/s11427-018-9415-7

Lifschytz E and Lindsley DL (1972) The role of X-chromosome inactivation during spermatogenesis. PNAS 69:182–186. doi: 10.1073/pnas.69.1.182

Ligrone R (2019) Sexual reproduction. Biological innovations that built the world. Springer International Publishing, Cham, pp 233–249

Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25:402–408. doi: 10.1006/meth.2001.1262

Long M, Vibranovski MD and Zhang YE (2012) Evolutionary interactions between sex chromosomes and autosomes. In: Singh RS, Xu J and Kulathinal RJ (eds) Rapidly evolving genes and genetic systems. Oxford University Press, pp 101–114

Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA and Eisen MB (2011) Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. PLoS Biol 9:e1000590. doi: 10.1371/journal.pbio.1000590

Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8

Lowry OH (1951) Rosebrough NJ, Farr Al, and Randall RJ. Protein measurement with the Folin phenol reagent J Biol Chem 193:265–275.

Lundberg LE, Figueiredo MLA, Stenberg P and Larsson J (2012) Buffering and proteolysis are induced by segmental monosomy in *Drosophila melanogaster*. Nucleic Acids Res 40:5926–5937. doi: 10.1093/nar/gks245

Makowski MM, Gräwe C, Foster BM, Nguyen NV, Bartke T and Vermeulen M (2018) Global profiling of protein–DNA and protein–nucleosome binding affinities using quantitative mass spectrometry. Nat Commun 9:1653. doi: 10.1038/s41467-018-04084-0

Mank JE, Hultin-Rosenberg L, Zwahlen M and Ellegren H (2008) Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. Am Nat 171:35–43. doi: 10.1086/523954

Mank JE, Nam K, Brunström B and Ellegren H (2010) Ontogenetic complexity of sexual dimorphism and sex-specific selection. Mol Biol Evol 27:1570–1578. doi: 10.1093/molbev/msq042

Mankiewicz JL, Godwin J, Holler BL, Turner PM, Murashige R, Shamey R, Daniels HV and Borski RJ (2013) Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. Integr Comp Biol 53:755–765. doi: 10.1093/icb/ict093

Martin SG, Dobi KC and St Johnston D (2001) A rapid method to map mutations in *Drosophila*. Genome Biol 2:research0036.1. doi: 10.1186/gb-2001-2-9-research0036

McAnally AA and Yampolsky LY (2010) Widespread transcriptional autosomal dosage compensation in *Drosophila* correlates with gene expression level. Genome Biol Evol 2:44–52. doi: 10.1093/gbe/evp054

McKee BD and Handel MA (1993) Sex chromosomes, recombination, and chromatin conformation. Chromosoma 102:71–80. doi: 10.1007/BF00356023

McKee BD, Wilhelm K, Merrill C and Ren X (1998) Male sterility and meiotic drive associated with sex chromosome rearrangements in *Drosophila*: role of X-Y pairing. Genetics 149:143–155.

McNair A, Lokman PM, Closs GP and Nakagawa S (2015) Ecological and evolutionary applications for environmental sex reversal of fish. Q Rev Biol 90:23–44. doi: 10.1086/679762

Meiklejohn CD, Landeen EL, Cook JM, Kingan SB and Presgraves DC (2011) Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol 9:e1001126. doi: 10.1371/journal.pbio.1001126

Meiklejohn CD and Presgraves DC (2012) Little evidence for demasculinization of the *Drosophila* X chromosome among genes expressed in the male germline. Genome Biol Evol 4:1007–1016. doi: 10.1093/gbe/evs077

Meiklejohn CD and Tao Y (2010) Genetic conflict and sex chromosome evolution. Trends Ecol Evol 25:215–223. doi: 10.1016/j.tree.2009.10.005

Meisel RP and Connallon T (2013) The faster-X effect: integrating theory and data. Trends Genet 29:537–544. doi: 10.1016/j.tig.2013.05.009

Meisel RP, Malone JH and Clark AG (2012) Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res 22:1255–1265. doi: 10.1101/gr.132100.111

Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholz C, Kelley RL and Kuroda MI (2000) Ordered assembly of *roX* RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. Curr Biol 10:136–143.

Menon DU and Meller VH (2012) A role for siRNA in X-chromosome dosage compensation in *Drosophila melanogaster*. Genetics 191:1023–1028. doi: 10.1534/genetics.112.140236

Menon DU and Meller VH (2015) Identification of the *Drosophila* X chromosome: The long and short of it. RNA Biol 12:1088–1093. doi: 10.1080/15476286.2015.1086864

Mérel V, Boulesteix M, Fablet M and Vieira C (2020) Transposable elements in *Drosophila*. Mob DNA 11:23. doi: 10.1186/s13100-020-00213-z

Mikhaylova LM and Nurminsky DI (2011) Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila* X chromosome. BMC Biol 9:29. doi: 10.1186/1741-7007-9-29

Miller DE, Cook KR and Hawley RS (2019) The joy of balancers. PLoS Genet 15:e1008421. doi: 10.1371/journal.pgen.1008421

Miller DE, Cook KR, Kazemi NY, Smith CB, Cockrell AJ, Hawley RS and Bergman CM (2016) Rare recombination events generate sequence diversity among balancer chromosomes in *Drosophila melanogaster*. PNAS 113:E1352–E1361. doi: 10.1073/pnas.1601232113

Moore SA, Ferhatoglu Y, Jia Y, Al-Jiab RA and Scott MJ (2010) Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono or dimethyl lysine 20 on histone H4. J Biol Chem jbc-M110.

Muller HJ (1964) The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1:2–9. doi: 10.1016/0027-5107(64)90047-8

Namekawa SH and Lee JT (2009) XY and ZW: Is meiotic sex chromosome inactivation the rule in evolution? PLoS Genet 5:e1000493. doi: 10.1371/journal.pgen.1000493

**Bibliography**

Neukomm LJ, Burdett TC, Gonzalez MA, Züchner S and Freeman MR (2014) Rapid in vivo forward genetic approach for identifying axon death genes in *Drosophila*. PNAS 111:9965–9970. doi: 10.1073/pnas.1406230111

Nishihara S (2007) *Drosophila* development, RNAi, and glycobiology. In: Kamerling H (ed) Comprehensive glycoscience. Elsevier, Oxford, pp 49–79

Nissani M (1977) On the interpretation of mutagenically induced mosaicism in *Drosophila*. Genetics 86:779–787.

Nüsslein-Volhard C and Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. Nature 287:795–801. doi: 10.1038/287795a0

Otto SP and Lenormand T (2002) Resolving the paradox of sex and recombination. Nat Rev Genet 3:252–261. doi: 10.1038/nrg761

Paris M, Villalta JE, Eisen MB and Lott SE (2015) Sex bias and maternal contribution to gene expression divergence in *Drosophila* blastoderm embryos. PLoS Genet 11:e1005592. doi: 10.1371/journal.pgen.1005592

Parisi M, Nuttall R, Edwards P, Minor J, Naiman D, Lü J, Doctolero M, Vainer M, Chan C, Malley J et al. (2004) A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. Genome Biol 5:R40. doi: 10.1186/gb-2004-5-6-r40

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S and Oliver B (2003) Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. Science 299:697–700. doi: 10.1126/science.1079190

Parsch J (2004) Functional analysis of *Drosophila melanogaster* gene regulatory sequences by transgene coplacement. Genetics 168:559–561. doi: 10.1534/genetics.104.028498

Parsch J and Ellegren H (2013) The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet 14:83–87. doi: 10.1038/nrg3376

Parsch J, Tanda S and Stephan W (1997) Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 94:928–933. doi: 10.1073/pnas.94.3.928

Pastink A, Heemskerk E, Nivard MJM, van Vliet CJ and Vogel EW (1991) Mutational specificity of ethyl methanesulfonate in excision-repair-proficient and -deficient strains of *Drosophila melanogaster*. Molec Gen Genet 229:213–218. doi: 10.1007/BF00272158

Patten MM (2018) Selfish X chromosomes and speciation. Mol Ecol 27:3772–3782. doi: https://doi.org/10.1111/mec.14471

Pennell MW, Mank JE and Peichel CL (2018) Transitions in sex determination and sex chromosomes across vertebrate species. Mol Ecol 27:3950–3963. doi: https://doi.org/10.1111/mec.14540

Perrin N (2012) What uses are mating types? The "Developmental switch" model. Evolution (N Y) 66:947–956. doi: https://doi.org/10.1111/j.1558-5646.2011.01562.x

Perry JC, Harrison PW and Mank JE (2014) The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. Mol Biol Evol 31:1206–1219. doi: 10.1093/molbev/msu072

Phadnis N, Baker EP, Cooper JC, Frizzell KA, Hsieh E, de la Cruz AFA, Shendure J, Kitzman JO and Malik HS (2015) An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. Science 350:1552–1555. doi: 10.1126/science.aac7504

Prayitno K, Schauer T, Regnard C and Becker PB (2019) Progressive dosage compensation during *Drosophila* embryogenesis is reflected by gene arrangement. EMBO Rep 20:e48138. doi: 10.15252/embr.201948138

Qian S-H, Chen L and Chen Z-X (2019) Enriched G-quadruplexes on the *Drosophila* male X chromosome function as insulators of dosage compensation complex. bioRxiv 656538. doi: 10.1101/656538

Qiao H-H, Wang F, Xu R-G, Sun J, Zhu R, Mao D, Ren X, Wang X, Jia Y, Peng P et al. (2018) An efficient and multiple target transgenic RNAi technique with low toxicity in *Drosophila*. Nat Commun 9:4160. doi: 10.1038/s41467-018-06537-y

Radzvilavicius AL, Hadjivasiliou Z, Pomiankowski A and Lane N (2016) Selection for mitochondrial quality drives evolution of the germline. PLoS Biol 14:e2000410. doi: 10.1371/journal.pbio.2000410

Raj A and van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. Cell 135:216–226. doi: 10.1016/j.cell.2008.09.050

Ramírez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung H-R, Lajoie BR, de Wit E, Zhan Y, de Laat W et al. (2015) High-affinity sites form an interaction network to

facilitate spreading of the MSL complex across the X chromosome in *Drosophila*. Mol Cell 60:146–162. doi: 10.1016/j.molcel.2015.08.024

Ranz JM, Castillo-Davis CI, Meiklejohn CD and Hartl DL (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742–1745. doi: 10.1126/science.1085881

Rastelli L and Kuroda MI (1998) An analysis of maleless and histone H4 acetylation in *Drosophila melanogaster* spermatogenesis. Mech Dev 71:107–117. doi: 10.1016/S0925-4773(98)00009-4

Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. Evolution (N Y) 38:735–742. doi: 10.1111/j.1558-5646.1984.tb00346.x

Rice WR and Chippindale AK (2001) Intersexual ontogenetic conflict. J Evolution Biol 14:685–693. doi: https://doi.org/10.1046/j.1420-9101.2001.00319.x

Riddle NC and Elgin SCR (2018) The *Drosophila* dot chromosome: Where genes flourish amidst repeats. Genetics 210:757–772. doi: 10.1534/genetics.118.301146

Robinett CC, Vaughan AG, Knapp J-M and Baker BS (2010) Sex and the single cell. II. there Is a time and place for sex. PLoS Biol 8:e1000365. doi: 10.1371/journal.pbio.1000365

Rørth P (1996) A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. PNAS 93:12418–12422. doi: 10.1073/pnas.93.22.12418

Rubin GM (2000) A brief history of *Drosophila's* contributions to genome research. Science 287:2216–2218. doi: 10.1126/science.287.5461.2216

Ruzicka F, Hill MS, Pennell TM, Flis I, Ingleby FC, Mott R, Fowler K, Morrow EH and Reuter M (2019) Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. PLoS Biol 17:e3000244. doi: 10.1371/journal.pbio.3000244

Sabath N, Itescu Y, Feldman A, Meiri S, Mayrose I and Valenzuela N (2016) Sex determination, longevity, and the birth and death of reptilian species. Ecol Evol 6:5207–5220. doi: 10.1002/ece3.2277

Salz H and Erickson JW (2010) Sex determination in *Drosophila*: the view from the top. Fly 4:60–70. doi: 10.4161/fly.4.1.11277

Samata M and Akhtar A (2018) Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. Annu Rev Biochem 87:323–350. doi: 10.1146/annurev-biochem-062917-011816

Schauer T, Ghavi-Helm Y, Sexton T, Albig C, Regnard C, Cavalli G, Furlong EE and Becker PB (2017) Chromosome topology guides the *Drosophila* dosage compensation complex for target gene activation. EMBO Rep 18:1854–1868. doi: 10.15252/embr.201744292

Schützenmeister A and Piepho H-P (2012) Residual analysis of linear mixed models using a simulation approach. Comput Stat Data Anal 56:1405–1416. doi: 10.1016/j.csda.2011.11.006

Sedlazeck FJ, Rescheneder P and von Haeseler A (2013) NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29:2790–2791. doi: 10.1093/bioinformatics/btt468

Serebriiskii IG and Golemis EA (2000) Uses of lacZ to study gene function: evaluation of β-galactosidase assays employed in the yeast two-hybrid system. Anal Biochem 285:1–15. doi: 10.1006/abio.2000.4672

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A and Cavalli G (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. Cell 148:458–472. doi: 10.1016/j.cell.2012.01.010

Siddiqui-Jain A, Grand CL, Bearss DJ and Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. PNAS 99:11593–11598. doi: 10.1073/pnas.182256799

Singh CM and Heberlein U (2000) Genetic control of acute ethanol-induced behaviors in *Drosophila*. Alcoholism Clin Exp Res 24:1127–1136. doi: 10.1111/j.1530-0277.2000.tb02075.x

Speijer D, Lukeš J and Eliáš M (2015) Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. PNAS 112:8827–8834. doi: 10.1073/pnas.1501725112

Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, Misra S and Rubin GM (1999) The berkeley *Drosophila* genome project gene disruption project: single *P*-element insertions mutating 25% of vital *Drosophila* genes. Genetics 153:135–177. doi: 10.1093/genetics/153.1.135

**Bibliography**

St Johnston D (2002) The art and design of genetic screens: *Drosophila melanogaster*. Nat Rev Genet 3:176–188. doi: 10.1038/nrg751

Stenberg P and Larsson J (2011) Buffering and the evolution of chromosome-wide gene regulation. Chromosoma 120:213–225.

Stenberg P, Lundberg LE, Johansson A-M, Rydén P, Svensson MJ and Larsson J (2009) Buffering of segmental and chromosomal aneuploidies in *Drosophila melanogaster*. PLoS Genet 5:e1000465. doi: 10.1371/journal.pgen.1000465

Straub T, Grimaud C, Gilfillan GD, Mitterweger A and Becker PB (2008) The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. PLoS Genet 4:e1000302. doi: 10.1371/journal.pgen.1000302

Straub T, Zabel A, Gilfillan GD, Feller C and Becker PB (2013) Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. Genome Res 23:473–485. doi: 10.1101/gr.146407.112

Suzuki MG (2018) Sex determination cascade in insects: a great treasure house of alternative splicing. In: Kobayashi K, Kitano T, Iwao Y and Kondo M (eds) Reproductive and developmental strategies. Springer Japan, Tokyo, pp 267–288

Tadros W and Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts. Development 136:3033–3042. doi: 10.1242/dev.033183

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V et al. (2019) FlyBase 2.0: the next generation. Nucleic Acids Res 47:D759–D765. doi: 10.1093/nar/gky1003

Turner JMA, Mahadevaiah SK, Fernandez-Capetillo O, Nussenzweig A, Xu X, Deng C-X and Burgoyne PS (2005) Silencing of unsynapsed meiotic chromosomes in the mouse. Nat Genet 37:41–47. doi: 10.1038/ng1484

Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A et al. (2016) Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res 26:70–84. doi: 10.1101/gr.196006.115

Urban JA, Doherty CA, Jordan WT, Bliss JE, Feng J, Soruco MM, Rieder LE, Tsiarli MA and Larschan EN (2017) The essential *Drosophila* CLAMP protein differentially regulates non-coding *roX* RNAs in male and females. Chromosome Res 25:101–113.

Vensko SP and Stone EA (2015) X-to-autosome expression and *msl-2* transcript abundance correlate among *Drosophila* melanogaster somatic tissues. PeerJ 3:e771. doi: 10.7717/peerj.771

Vensko SP and Stone EA (2014) No evidence for a global male-specific lethal complex-mediated dosage compensation contribution to the demasculinization of the *Drosophila melanogaster* X chromosome. PLOS ONE 9:e103659. doi: 10.1371/journal.pone.0103659

Vibranovski MD, Koerich LB and Carvalho AB (2008) Two new Y-linked genes in *Drosophila melanogaster*. Genetics 179:2325–2327. doi: 10.1534/genetics.108.086819

Vibranovski MD, Lopes HF, Karr TL and Long M (2009) Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet 5:e1000731. doi: 10.1371/journal.pgen.1000731

Villa R, Forné I, Müller M, Imhof A, Straub T and Becker PB (2012) MSL2 combines sensor and effector functions in homeostatic control of the *Drosophila* dosage compensation machinery. Mol Cell 48:647–654.

Villa R, Schauer T, Smialowski P, Straub T and Becker PB (2016) PionX sites mark the X chromosome for dosage compensation. Nature 537:244.

Wakimoto BT, Lindsley DL and Herrera C (2004) Toward a comprehensive genetic analysis of male fertility in *Drosophila melanogaster*. Genetics 167:207–216. doi: 10.1534/genetics.167.1.207

Waxman D and Peck JR (1998) Pleiotropy and the preservation of perfection. Science 279:1210–1213. doi: 10.1126/science.279.5354.1210

Weismann A, Poulton EB, Schönland S (Selmar), Shipley AE (Arthur E, Knight BCJG, University of London. Institute of Psychiatry Library former owner, Great Britain. Local Government Board former owner, Wellcome Library former owner, and King's College London (1889) Essays upon heredity and kindred biological problems. Oxford : at the Clarendon Press

White-Cooper H (2012) Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the *Drosophila* testis. Spermatogenesis 2:11–22. doi: 10.4161/spmg.19088

Witt E, Shao Z, Hu C, Krause HM and Zhao L (2021) Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in *Drosophila* testis. bioRxiv 2021.02.05.429952. doi: 10.1101/2021.02.05.429952

Wood AM, Van Bortle K, Ramos E, Takenaka N, Rohrbaugh M, Jones BC, Jones KC and Corces VG (2011) Regulation of chromatin organization and inducible gene expression by a *Drosophila* insulator. Mol Cell 44:29–38. doi: 10.1016/j.molcel.2011.07.035

Wu C-I and Yujun Xu E (2003) Sexual antagonism and X inactivation – the SAXI hypothesis. Trends Genet 19:243–247. doi: 10.1016/S0168-9525(03)00058-1

Wyman MJ, Agrawal AF and Rowe L (2010) Condition-dependence of the sexually dimorphic transcriptome in *Drosophila melanogaster*. Evolution (N Y) 64:1836–1848. doi: https://doi.org/10.1111/j.1558-5646.2009.00938.x

Yu J, Lan X, Chen X, Yu C, Xu Y, Liu Y, Xu L, Fan H-Y and Tong C (2016) Protein synthesis and degradation are essential to regulate germline stem cell homeostasis in *Drosophila* testes. Development 143:2930–2945. doi: 10.1242/dev.134247

Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, MacAlpine DM and Oliver B (2010a) Expression in aneuploid *Drosophila* S2 Cells. PLoS Biol 8:e1000320. doi: 10.1371/journal.pbio.1000320

Zhang YE, Vibranovski MD, Krinsky BH and Long M (2010b) Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. Genome Res 20:1526–1533. doi: 10.1101/gr.107334.110

# Acknowledgements

The last three and a half years have been the most amazing and eventful years of my life, not only because I have devoted them to this research work but also because of the people who have accompanied me. I would like to express many thanks to these people, without whom I could not have come all this way.

I express my sincere gratitude and appreciation to my main supervisor John Parsch, for providing me with an exciting research topic, for his help and valuable advice at all stages of my dissertation. I am grateful for the amount of work he spent in making me a scientist and for the positive attitude that helped me overcome my research challenges.

I express my deep gratitude to the members of the Thesis Advisory Committee, Tamara Mikeladze-Dvali and Jochen Wolf, for their highly qualified feedback and suggestions on my work, as well as their sincere personal support.

I can't imagine how challenging it would have been to start my work in a new country and in entirely new surroundings without the friendly and warm atmosphere in the Parsch group. Sonja, Amanda, Annabella and Vera, Selina were always around to discuss results and advice on work and many aspects of life in Germany. A special thanks to Eliza, who contributed to both projects and who was my mentor in the lab, and also to Timothy, with whom I have established the wonderful tradition of going outside for a coffee break, and who has not only helped me with some experiments but has given me the best psychological support. Hildegard was a constant help in the lab, thanking her for motivating me to learn German. I warmly remember the time I spent working in the microscopy room with Anna Lena, who inspired my interest in the German language and showed me some beautiful places in Bavaria. I would also like to thank Ingo and Derek, students of the Master's programme, for their participation in my research project.

**Acknowledgements**

What helped me a lot in my work was the friendly atmosphere in the Evolutionary Biology department and the regular cheerful meetings in the Beer garden with its staff. Thanks to everyone I worked with there and who was always interested in the progress of my work. A special thanks to Cynthia, who has often helped me organise my work mood and Michaela for her friendly support. My daily work in the office would not have been as pleasant without the company of Vera, whom I bored with endless conversations about politics and cycling together, and Ana, to whom I often turned for advice.

I am grateful to the Graduate School Life Science Munich, Francisca Mende and Nadine Hamze for organising the scientific seminars and communication between students. Thanks to all the doctoral students I met during my studies, who became friends and with whom I could share all the burdens of doctoral studies, Yen-Yu, Robert, Matteo, Deepak, Mariam, Rafael, Lingyun and Lisa.

Thanks to my friends and family, who, despite the long-distance, have always given me support during my work. Special thanks to my parents Aleksandr and Vera, for their love and for fostering my curiosity and love of knowledge. This work is dedicated to you.

I want to express my gratitude to a friend Anna who was there for ups and downs, and randomly provided valuable life words of advice.