Chromosome-specific maturation of the epigenome in the male germline of *D. melanogaster*

James Anderson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Steve Henikoff, Chair

David Shechner

Barbara Wakimoto

Program authorized to offer degree:

Molecular and Cellular Biology

University of Washington

Abstract

Chromosome-specific maturation of the epigenome in the male germline of *Drosophila melanogaster*

James Anderson

Chair of the Supervisory Committee:

Steve Henikoff

Department of Genome Sciences

The male germline of *Drosophila melanogaster* is a powerful system for investigating chromatin regulation in a differentiating tissue. Testis-specific genes utilize a distinct system of gene activation, involving tissue-specific rather than general transcription factors, and a unique promoter architecture. Furthermore, the sex chromosomes are regulated quite distinctly in the pre-meiotic male germline, with diminishing transcription of X-linked genes and hyperactivation of the spermatogenesis genes on the heterochromatic Y chromosome. It has been unclear which chromatin components are associated with the unique chromosome-specific trends of expression in the male germline. This dissertation advances our understanding of how chromatin relates to gene expression in the differentiating male germline, particularly regarding: unique patterns of histone 3 lysine 4 dimethylation (H3K4me2) and elongating RNA Polymerase II (RNA PolII ser2phos) at promoters and gene bodies of active spermatogenesis genes; loss of dosage compensation from the X chromosome; the apparent lack of silencing of the X chromosome; and the broad enrichment of ubiquitylated histone H2A (uH2A) across the hyperactivated Y chromosome in spermatocytes. These findings resolve long-open questions about the X chromosome in the male germline, by concluding that the X is neither dosage compensated nor silenced. Intriguingly, the finding of a uH2A body on the Y chromosome in spermatocytes provides new evidence that uH2A plays more versatile roles in chromatin regulation aside from just Polycomb repression. More broadly, this dissertation contributes—from the perspective of chromatin—to recent efforts such as the Fly Cell Atlas to understand what drives gene expression and cell type differences across differentiation and cell types of the fly.

Acknowledgements

Thanks to everyone in the Henikoff Lab, especially Steve and Kami for many years of great mentorship. Thanks as well to my committee members: Dr. Cole Trapnell, Dr. Barbara Wakimoto, Dr. Celeste Berg, and Dr. Dave Shechner.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Chapter 1

Introduction

1.1 Gene Activation

The expression of an organism's genes must be coordinated precisely over the course of development to ensure fitness and prevent disease. This is no small feat; the genomes of many organisms comprise a vast quantity of genetic information, from the fruit fly *Drosophila melanogaster* (~13,000-15,000 protein-coding genes) (Adams *et al.*, 2000) to *Homo sapiens* (~20,000 - 25,00 protein-coding genes) (Piovesan *et al.*, 2019). Coordination of gene expression depends on regions of the genome called non-coding elements: sequences of DNA which do not code for proteins, but nevertheless serve a crucial role as regulators of expression (Nelson & Wardle, 2013).

A major non-coding element in all eukaryotic genomes is the promoter. The promoter is the region of DNA proximal to the transcription start site (TSS) of a gene, often ~10 – 100bp long, which facilitates gene activation (Sloutskin *et al.*, 2021). Promoters are landing sites for proteins called transcription factors (TFs), which recognize and bind DNA sequence motifs (often 6 – 8bp long) like a biochemical lock-and-key. Transcription factors are the largest class of proteins encoded in the genome, comprising thousands of unique proteins. Many forms of cancer and heritable disease are driven by transcription factors which have received mutations causing them to recognize inappropriate DNA motifs, thus binding inappropriate genomic loci and driving the misexpression of genes (Barrera *et al.*, 2016). Other mutations sometimes ablate the ability of a TF to bind DNA at all.

A set of six TFs are called the general transcription factors (GTFs) because they are highly conserved across eukaryotes and archaea, and they are essential for the activation of most genes. The six GTFs are TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Thomas & Chiang, 2006). The precise role each of these GTFs in initiating transcription of a gene is distinct—TFIID, for example, contains a subunit called TATA-binding protein (TBP), which recognizes a sequence of DNA in some promoters called the TATA box. The GTFs associate with a large complex called Mediator, which mediates the recruitment of RNA Polymerase II (RNA PolII), the enzyme which transcribes genes into messenger RNA (mRNA) (Lu & Fuller, 2015). This large complex of GTFs, Mediator, and often dozens of other DNA-associated proteins is called the preinitiation complex (PIC) (Thomas & Chiang, 2006).

Intriguingly, there are particular tissue-specific exceptions to the essentiality of GTFs in gene activation. As will be discussed in Chapter 1.6, the male germline of *D. melanogaster* utilizes its own set of testis-specific transcription factors, mostly dispensing with GTFs for the activation of testis-specific genes (White-Cooper *et al.*, 2010). This feature is specific just to the testis gene program; other tissue-specific genes in other tissues of the fly utilize the traditional system of GTFs for gene activation. The testis-specific TFs are highly specialized, becoming active only in the differentiating male germline, and binding only their exclusive testis-specific promoters. This highly specialized scheme for gene regulation in the testis has been the subject of intense research for several decades within the of community *Drosophila*

genetics researchers (Laktionov *et al.*, 2018; Lu *et al.*, 2020). Chapter 2 will discuss further how my thesis work has expanded our understanding of the testis-specific transcription program.

After TF-mediated recruitment of RNA PolII to a promoter and formation of the PIC, the next step in gene expression is transcription initiation. RNA PolII catalyzes the first phosphodiester bond in its transcript and begins processing along the gene body at the TSS, polymerizing a progressively longer messenger RNA molecule (mRNA). However, this initiated form of RNA PolII typically pauses proximal to the promoter, within ~100bp. This initiated form of RNA PolII is characterized by phosphorylation of serine 5 in the C-terminal domain of the enzyme (RNA PolII ser5) (Buratowski, 2009). Promoter-proximal pausing of RNA PolII serves an important role in gene regulation; it is an important checkpoint to ensure proper establishment of chromatin structure prior to gene activation, and to enable rapid, robust response to gene-activating cues at the right moment (Core & Adelman, 2019). Chapters 2 and 3 will discuss how my thesis work has revealed a possible model whereby the pre-meiotic spermatocyte cells of the *Drosophila* male germline utilize a unique approach for gene activation which dispenses with the promoter-proximal pausing checkpoint which is typical of the activation of other tissue-specific genes; this would highlight yet another way in which the testis-specific transcription program is distinct from other tissue-specific genes.

Once this pause-checkpoint has been passed and RNA PolII has received the necessary cues (such as removal of the negative elongation factor NELF), RNA PolII enters its elongating phase, processing forward into the gene body. This elongating phase of RNA PolII is characterized by phosphorylation of serine 2 in the C-terminal domain of the enzyme (RNA PolII ser2) (Buratowski, 2009).

Gene activation can be enhanced by non-coding elements called enhancers. Enhancers are short sequences of DNA containing motifs which can be bound by tissue-specific transcription factors, chromatin modifiers, or architectural proteins. TF-bound enhancers become active and promote the expression of the genes with which they are associated (Panigrahi & O'Malley, 2021). The tissue-specific nature of enhancers is important to their function; a neuron-specific enhancer, for example, might become active and enhance the expression of its associated neuron-specific gene only in the brain, and not in the intestine. Thus, the activation of a gene can be modulated according to the appropriate tissue context, via the combinatorial effect of multiple enhancers exerting their tissue-specific effects.

The means by which enhancers boost transcription vary by context. One mechanism may be engagement with the Mediator complex to promote RNA PolII recruitment to the target gene. A key feature differentiating enhancers from promoters is the genomic distance across which they exert their effect. Enhancers often act on genes which are >10 kb away—sometimes >100kb (Panigrahi & O'Malley, 2021). The ability of enhancers to act across such wide distances depends on higher-order genomic structure in 3D space. For example, an enhancer which is far from a gene by linear distance can be brought proximal to its target by looping of DNA (Popay & Dixon, 2022).

This has been a brief overview of the mechanisms of gene activation—but what genomic features confer cell type-specific gene expression? Why does a spermatocyte, for example, access a distinct gene program from a white blood cell or a neuron or a stem cell or a cancerous cell? The key is chromatin.

1.2 Chromatin

Chromatin is the complex of interacting DNA, RNA, and protein that makes up eukaryotic genomes (Fyodorov *et al.*, 2018; Luger *et al.*, 2012). The fundamental unit of chromatin is the nucleosome. Nucleosomes are wrapped with genomic DNA analogously to thread around a spool, with each nucleosome wrapping ~147bp of DNA on average. Because genes may be many thousands of nucleotides long, a single gene body typically comprises many nucleosomes, looking like beads on a string. Nucleosomes play an important role both as packagers of DNA, and as recipients of marks associated with the active or repressed status of genes. The subunits of nucleosomes which receive these marks are the histones.

Histones are the protein subunits of nucleosomes. A canonical nucleosome is made up of 4 distinct histones, each represented twice: histone 2A (H2A), histone 2B (H2B), histone 3 (H3), and histone 4 (H4). Histones have N-terminal tails which can receive a variety of post-translational modifications (histone PTMs), *e.g.* methylation, acetylation, and ubiquitinylation (Kouzarides, 2007). The role of histone PTMs in gene regulation is one of the central topics in epigenetics research. Next, I will introduce some of the histone PTMs which are especially pertinent to this dissertation.

A classic example of a histone PTM associated with active chromatin is histone 3 lysine 4 trimethylation (H3K4me3) (Bernstein et al., 2005; Hyun et al., 2017). H3K4me3 is associated with active promoter regions, and is written by the Trithorax group (TrxG) in *D. melanogaster*, or by mixed lineage leukemia (MLL) in *H. sapiens*. There are 3 potential methylation states of H3K4— H3K4me1, H3K4me2, and H3K4me3—representing one, two, or three methyl groups covalently bonded to lysine respectively. These methyl groups are deposited sequentially onto H3K4, and each methylation state is associated with distinct states of active chromatin; whereas H3K4me3 is associated with active promoters, H3K4me1 is associated with active enhancers (Bernstein et al., 2005; Hyun et al., 2017). H3K4me2 is intermediate, enriched both at active promoters and sometimes also active enhancers.

A variety of other histone PTMs are associated with distinct chromatin states. For example, histone 3 lysine 27 trimethylation (H3K27me3) is a repressive mark associated with inactivated genomic regions called facultative heterochromatin (Blackledge & Klose, 2021; Hyun *et al.*, 2017). H3K27me3 is associated with Polycomb repression, called such because it is written by the Polycomb group (PcG) complex Polycomb Repressive Complex 2 (PRC2). PcG proteins were among the earliest-identified chromatin modifiers, and the name “Polycomb” specifically originated from the phenotype of one particular PcG-mutant factor causing flies to develop an excess of the anatomical feature called sex combs (Schuettengruber et al., 2017). Other mutations in key PcG components can induce a variety of development defects in fruit flies (and other species, including humans), referred to as homeotic transformations: inappropriate body patterning, sometimes resulting in the addition, loss, or replacement of body parts at incorrect anatomical regions, such as the development of an additional pair of wings, or an extra thorax (Schuettengruber *et al.*, 2017). H3K27me3 is typically enriched at genomic regions (such as the HOX genes (Hubert & Wellik, 2023)) which must be repressed according to

the anatomical context. The genomic enrichment of H3K27me3 can form more broad domains which may cover >10,000 bps, sometimes >100,000 bps.

Ubiquitinylated histone H2A (uH2A) at lysine 119 (in vertebrates) or lysine 118 (in flies) is another example of a Polycomb-mediated histone PTM. uH2A is deposited by the *sex combs extra* (*Sce*) subunit of Polycomb Repressive Complex 1 (PRC1 (Baarends *et al.*, 1999; Blackledge & Klose, 2021). PRC1 and PRC2 collaborate to mediate Polycomb repression: PRC2 deposits H3K27me3 at regions which are the target of repression, and then the Pc subunit of PRC1, having a chromodomain which binds H3K27me3, recruits PRC1 to the locus. *Sce* then writes uH2A (Gorfinkiel *et al.*, 2004). Ultimately, this results in long-term repression of the target locus, which becomes physically compacted, segregated away from active regions of the genome, and marked with additional repressive factors (Aloia *et al.*, 2013; Blackledge *et al.*, 2020).

uH2A has classically been considered a repressive mark of Polycomb silencing. However, recent research has uncovered nuances to the role of uH2A in chromatin regulation (Bonnet *et al.*, 2022; Loubiere *et al.*, 2020). Increasing evidence suggests that uH2A may be a more versatile histone PTM than was previously thought, playing important roles in gene activation, higher-order chromatin looping, and 3D chromatin structure, in addition to its association with Polycomb repression. In Chapter 2, I will discuss how my thesis work has also highlighted uH2A as a versatile histone PTM.

A few additional examples further illustrate the diversity of histone PTMs. Histone 3 lysine 9 dimethylation (H3K9me2), and particularly trimethylation (H3K9me3), are marks of constitutive heterochromatin: regions of chromosomes such as the centromere and telomere which are cluttered with highly repetitive DNA and “selfish” DNA sequences such as transposable elements (Hyun *et al.*, 2017). H3K9 methylation, and other repressive factors such as heterochromatin protein 1 (HP1), maintains these constitutive heterochromatic regions in a repressed state.

Histone 4 lysine 16 acetylation (H4K16Ac) is typically enriched at active regions, especially the upregulated X chromosome in somatic tissues of *D. melanogaster* males (Conrad & Akhtar, 2012).

The examples discussed above offer just a small sampling of the vast array of histone PTMs involved in chromatin regulation, with a particular focus just on the PTMs most pertinent to this dissertation. Hundreds of distinct histone PTMs have been identified and investigated for their role in development and disease. The collective ensemble of all histone PTMs has classically been referred to as “the histone code” (Strahl & Allis, 2000). The histone code hypothesis proposes that transcription and chromatin regulation depend in large part on histone PTMs, and that the language of histone PTMs can be deciphered to determine a gene’s expression status. Henikoff and Shilatifard wisely caution against oversimplistic interpretations of the histone code model: correlation of a histone PTM with specific chromatin states does not necessarily imply causation (Henikoff & Shilatifard, 2011).

1.3 Chromatin profiling

A seminal advance in our understanding of chromatin was made by Weintraub and Groudine in 1976, when they demonstrated that cell-specific active genes are especially sensitive to the DNA-cleaving enzyme DNase I (Weintraub & Groudine, 1976). This finding advanced our understanding of how nucleosome conformation relates to gene regulation. Since then, there has been a wave of new technical approaches for investigating the state of chromatin. These techniques for investigating chromatin are collectively called chromatin profiling methods.

By now, the toolbox of chromatin profiling methods is enormous, providing the ability to ask diverse questions about a cell's chromatin:

Which genomic regions are especially accessible or inaccessible (*e.g.* the methods: ATAC-seq (Buenrostro *et al.*, 2015) and DNase-seq (Song & Crawford, 2010))?

How can we infer the 3D structure of the genome via contact-frequencies between genomic loci (*e.g.* the methods: 3C, 4C, 5C, and Hi-C (Oluwadare *et al.*, 2019))?

What is the genome-wide enrichment of a given histone PTM, transcription factor, or other epitope of interest (*e.g.* the methods: ChIP-seq (Park, 2009), CUT&RUN (Skene & Henikoff, 2017), CUT&Tag (Kaya-Okur *et al.*, 2019))?

This last question deserves further discussion, as it is a primary focus of the Henikoff Lab's technology development, and it is central to this dissertation.

Until recently, the gold-standard chromatin profiling method for assaying the genome-wide landscape of an epitope of interest (*e.g.* a histone PTM or transcription factor) was a method called Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq) (Park, 2009). Briefly, in ChIP-seq: the sample is cross-linked to preserve interactions between the chromatin's protein and DNA; the chromatin is then fragmented into pieces, and incubated with antibody specific for an epitope of interest—typically a histone PTM, or a TF; the antibody-bound fragments are precipitated; the cross-links are reversed; undesirable protein and RNAs are degraded; and the enriched DNA fragments are PCR-amplified and sequenced. The mapped sequenced reads can be analyzed to determine the genome-wide enrichment of the epitope which was profiled.

ChIP-seq has major shortcomings (Kaya-Okur *et al.*, 2019; Park, 2009). First, the need for cross-linking can introduce artifacts or mask epitopes. Second, the signal:noise ratio is quite low, meaning that ChIP-seq requires a high amount of input material and needs very deep sequencing to yield high-quality data. Third, the workflow is more cumbersome than alternative approaches.

In recent years, the Henikoff Lab has developed two new methods which overcome the limitations of ChIP-seq for chromatin profiling: CUT&RUN (Cleavage Under Target and Release Using Nuclease) (Skene & Henikoff, 2017) and CUT&Tag (Cleave Under Target and Tagmentation) (Kaya-Okur *et al.*, 2019). While both methods are powerful approaches for chromatin profiling, here I will highlight just CUT&Tag, as it was the central chromatin profiling approach in my thesis work. CUT&Tag has a simple workflow: Antibody for an epitope of interest is incubated with the sample of cells or nuclei; a conjugate of protein

A (or protein AG) plus hyperactive Tn5 transposase (pA-Tn5) is incubated with the sample under high salt conditions; pA-Tn5 localizes to the antibody-bound epitope, as protein A is avid for antibody IgG domains; the transposase activity of Tn5 is activated by the addition of magnesium, inserting its DNA adapters into the genomic sequence proximal to the epitope of interest; the tagmented DNA is captured and PCR-amplified; and the enriched DNA is sequenced. CUT&Tag is easy to perform, yields sequencing data with high signal:noise, and can be performed in both low-input bulk samples and even single-cell experiments.

In addition to overcoming the limitations of ChIP-seq, CUT&Tag has inspired a variety of related methods within the "CUT&Tag family." Single-cell CUT&Tag (Wu *et al.*, 2021) has been used to identify distinct cell types from complex brain tumor populations, providing richer profiles at single-cell resolution than could have been achieved with earlier approaches and identifying mechanisms of cancer development. CUTAC (Cleavage Under Targeted Accessible Chromatin) (Henikoff *et al.*, 2020) identifies transcription-coupled accessibility sites with better technical quality than related methods such as ATAC-seq. CUT&Tag-2for1 (Janssens *et al.*, 2022) enables the profiling of both a repressive histone PTM (H3K27me3) and an active epitope (RNA PolII ser5phos) simultaneously in a single sample, where the resulting sequencing data is deconvolved *post hoc* to separate the landscapes of H3K27me3 and RNA PolII ser5phos individually. In Chapter 2, I will discuss my own innovation to CUT&Tag, where pure cells from a complex tissue can be isolated by flow cytometry and then profiled by CUT&Tag to characterize distinct cell types *vs* whole tissue.

1.4 *D. melanogaster* as a model organism

For over one hundred years, the fruit fly *Drosophila melanogaster* has been a powerful model organism for genetics and epigenetics research. Thomas Hunt Morgan laid the foundation of fruit fly genetics research with the establishment of his fly room at Columbia University in the 1910s. Morgan and his colleagues used their flies to discover some of the fundamental principles of genetics, such as the chromosomal theory of inheritance, the chromosomal basis of sex determination, chromosomal crossing-over during meiosis, and gene linkage (Green, 2010; Morgan, 1910).

Since the seminal work from Morgan's fly room, we have developed a much deeper understanding of *D. melanogaster*'s genome, anatomy, and life cycle. The entire genome is ~180 Mb, with 4 chromosomes: X/Y, 2, 3, and 4. The autosomal chromosomes 2 and 3 are typically referred to more specifically according to the particular chromosomal arm: 2L or 2R, 3L or 3R.

It takes roughly 10 days from fertilization for *D. melanogaster* to develop into an adult: 1 day as an embryo; ~4 days as a larva; ~5 days as a pupa; and then "eclosion" on day ~10—the process of emerging from the pupal shell as a young adult fly.

D. melanogaster's short generation time and large brood size make it especially amenable for genetics studies. A prime example of the utility of *D. melanogaster* as a genetic tool is a technique called the enhancer trap, an approach for identifying promoters or enhancers that act in specific cell types of interest (O'Kane & Gehring, 1987). Typically, a reporter gene is integrated randomly into the genomes of

many flies across a generation, and the flies are screened to identify if the reporter is being expressed in the cell type of interest.

Identification of cell type-specific enhancers/promoters gives researchers the opportunity to manipulate a fly's genome in powerful ways. For example, the promoter of the gene *bag-of-marbles* (*bam*) is highly specific to the early, pre-meiotic germline cell type called spermatogonia (McKearin & Ohlstein, 1995). Thus, geneticists can easily design a cross to generate flies with the *bam* promoter acting like a cell type-specific toggle driving expression of a transgene of interest just in the early male germline. This germline-driven transgene could encode a variety of constructs, *e.g.*: a fluorescent reporter to label early germline cells; an RNAi or CRISPR construct to knockdown/knockout a target just in the germline; a mutant version of a transcription factor whose in vivo germline-specific activity is being investigated. This description of the *bam* promoter is just one example; the opportunities for leveraging enhancers/promoters for cell-specific transgene studies are endless, and they are especially amenable in *Drosophila*. Although *D. melanogaster* is certainly not the only model organism in which enhancers are utilized for cell-specific studies, the short generation time and large brood sizes of flies make these enhancer-driven studies especially powerful for genetics research.

1.5 The testis of *D. melanogaster*

The testis of *D. melanogaster* is a useful system for studying chromatin in a differentiating tissue. The spiral-shaped testis is simple to dissect and contains the entire lineage of the male germline, from gonial stem cells all the way through fully differentiated mature spermatozoa. Also conveniently, differentiation proceeds spatially along the spiral, from the apical to distal ends, meaning that immunohistochemistry can often clearly distinguish all cell types of the differentiating lineage at once (Figure 2.1). These convenient facets of the testis, combined with the powerful tools for genetic manipulation in *D. melanogaster*, make the fly testis a powerful system for investigating questions about chromatin during differentiation.

The gonial stem cells (GSCs) reside at the apex of the testis, anchored to a cluster of somatic cells called the hub (Demarco *et al.*, 2014). With each division, a GSC produces one GSC daughter which remains at the hub to maintain the pool of stem cells, and another daughter called a gonialblast. The gonialblast migrates distally and proceeds through four rounds of mitosis, as spermatogonia. A key factor in the transition to the next stage is *bag-of-marbles* (*bam*), mentioned briefly in Chapter 1.4. (White-Cooper *et al.*, 2010). Without *bam*, the germline arrests in this spermatogonial stage and accumulates cysts of early germline cells.

Bam mutants are among a category of useful fly genotypes called stage-arrest mutants. Stage-arrest mutants bear mutations which cause the developing germline to halt at a certain point, accumulating only the germline cell types up until that development point. Stage-arrest mutants are useful systems for studying the male germline at distinct points in differentiation. In Chapter 2, I will compare/contrast CUT&Tag profiles of *bam* and wild-type testis to reveal the accumulation of active chromatin at testis-specific genes in the later germline.

After four rounds of mitosis and a final S phase, spermatogonia transition to the pre-meiotic stage called primary spermatocytes. These primary spermatocytes grow ~3x in volume over a period of 3 days as they prepare for meiosis, upregulating a dramatic transcriptional program of spermatogenesis genes. A stage-arrest mutant called *always early* (*aly*) halts in this early spermatocyte stage (White-Cooper *et al.*, 2010); I will discuss my CUT&Tag profiling of *aly* testis in Chapter 2.

Spermatocytes undergo two rounds of meiosis. Following meiosis I and II, the haploid daughters are called round spermatids. These round spermatids begin the process of spermiogenesis, the morphological change into mature spermatozoa. A variety of changes happen during this time: The cells elongate, developing long sperm tails and need-shaped heads; the mitochondria fuse into an organelle called the nebenkern; histones are replaced with sperm-specific proteins called protamines which enable the genome to be packed very tightly in the sperm nucleus; and transcription of most genes in the genome has stopped, with the exception of a subset of genes called the cup-and-comet genes (White-Cooper *et al.*, 2010).

While the focus of this dissertation is primarily on chromatin, here I will briefly point to other biological topics (with references, to some of the pertinent studies and reviews) which have been investigated using the *Drosophila* male germline as a model system: the process of dedifferentiation, whereby early germline cells reacquire a stem cell identity to maintain a depleted pool of stem cells (Eun *et al.*, 2017); the intricate signaling pathways between germline cells, somatic hub cells, and the somatic cyst cells (de Cuevas & Matunis, 2011); the temporal transcription of genes across differentiation and the identification of an intriguing subset of genes which are transcribed post-meiotically while most other genes are not transcribed (Raz *et al.*, 2023); and the formation of sperm-specific structures and organelles such as the nebenkern, acrosome, or the axoneme in the sperm tail (Fabian & Brill, 2012).

1.6 The spermatogenesis transcription program

The spermatogenesis transcription program is highly specialized compared to other tissue-specific gene programs. As was mentioned briefly in Chapter 1.1, the regulation of testis-specific genes dispenses with the GTFs which are essential for the activation of almost all other tissue-specific and housekeeping genes. Instead, the male germline swaps-in its own proprietary set of testis-specific TFs (Laktionov *et al.*, 2018; White-Cooper *et al.*, 2010). Two of these testis-specific TF complexes are: (1) testis-specific TATA-binding protein-associated factor (tTAF), and (2) testis-specific meiotic arrest complex (tMAC). The interplay of tTAF and tMAC in activating testis-specific genes is a topic of ongoing research, but it is clear that the activation of most testis-specific genes is reliant on one or both of these complexes in an interdependent manner. In fact, many of the stage-arrest mutations mentioned earlier in Chapter 1.5 are null mutations in components of either tTAF or tMAC, disrupting the proper activation of the spermatogenesis transcription program and causing the germline to arrest. *Aly*, for example, is a component of tMAC (White-Cooper *et al.*, 2010). Aside from the unique regulation by tTAFs and tMAC, testis-specific genes are unique in generally having short promoter sequences.

1.7 Sex chromosomes in the testis of *D. melanogaster*

The X and Y chromosomes are regulated quite distinctly in the *Drosophila* testis.

The X chromosome in the male germline loses the effects of dosage compensation (Rastelli & Kuroda, 1998), a process which occurs in all the other somatic cell types of the fly. Dosage compensation broadly upregulates expression of all X-linked genes approximately 2x, to compensate for the sex-dependent difference in X chromosome dosage between males (XY) and females (XX) (Conrad & Akhtar, 2012). Other species have evolved different approaches for solving the X-dosage difference between XY males and XX females; mammals such as *H. sapiens* use an alternative system, where one of the two X chromosomes in each of a female's nuclei is inactivated.

The chromatin components mediating dosage compensation in *D. melanogaster* are well understood. The complex Male-specific lethal (MSL) deposits the active mark histone 4 lysine 16 acetylation (H4K16Ac) across the entire X chromosome, promoting a ~2x increase in transcription of X-linked genes in males. This canonical dosage compensation occurs in all somatic cells of the fly, but does not occur in the male germline. For example, cytological studies have described that the MSL complex is absent from the male germline (Rastelli & Kuroda, 1998). Transcriptomics studies have reported a diminishing trend of expression of X-linked genes in progressively later germline cell types (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2021), consistent with the expectation that loss of dosage compensation should return X-expression to approximately autosomal levels in the male germline.

It has been unclear whether the loss of dosage compensation in the male germline of *D. melanogaster* is also accompanied by the gain of repressive chromatin on the X chromosome. This would be analogous to meiotic sex chromosome inactivation (MSCI), a process in therian males (such as humans and mice) whereby the unsynapsed chromatin of chromosomes X and Y in the male germline becomes inactivated (Turner, 2015). In MSCI, chromosomes X and Y are decorated with repressive histone PTMs such as uH2A and H3K9me2, and are physically compacted and segregated from the rest of the nucleus in a structure called the sex body.

However, the fundamental differences between chromatin regulation in flies and mammals (*e.g.* distinct mechanisms of dosage compensation, and the lack of meiotic cross-over of chromosomes in the male germline of *D. melanogaster* (McKee & Handel, 1993), *etc*.) suggests that MSCI may not occur in flies—it is possible that the X chromosome simply loses dosage compensation, without any concomitant gain of repressive chromatin. In Chapter 2, I will discuss how my thesis work demonstrates that the X chromosome in the male germline of *D. melanogaster* does not gain the histone PTMs associated with MSCI in other organisms. My work supports the model that the X chromosome loses dosage compensation, but does not gain repressive chromatin.

The 40 Mb Y chromosome in *D. melanogaster* is almost entirely heterochromatic and gene-poor (Chang & Larracuente, 2019). Due to its density of repetitive satellite DNA and heterochromatin, sequencing the Y chromosome has been historically challenging, but modern genome assemblies have brought us closer to a complete build of the chromosome (Chang & Larracuente, 2019). The few genes which are on the Y chromosome are highly specific for expression in the testis, such: *Mst77Y*, encoding a chromatin

component specific to sperm; the piRNA-producing Supressor of Stellate (Su:Ste) locus; and the male fertility factors *kl-2*, *kl-3*, and *kl-5* (Fingerhut *et al.*, 2019). Although silent in other cell types of the fly, these Y-linked genes become dramatically activated in primary spermatocytes. Cytologically, the Y chromosome expands into long, diffuse, hypertranscribed loops within the interior of the spermatocyte nucleus. In Chapter 2, I will summarize the trends of chromatin components which accumulate on this highly active Y chromosome in the male germline. I will discuss the intriguing finding that uH2A—classically associated with polycomb repression—broadly coats the active Y chromosome, providing new evidence to the emerging picture of uH2A as a more versatile histone PTM than was previously thought.

1.8 Modern genomics on the male germline of *D. melanogaster*

Over just the past several years, single-cell RNA-seq studies have revealed intriguing new features of transcription during fly spermatogenesis. Lawlor *et al.* observed a burst of expression of transposable elements (TEs) in early spermatocytes, suggesting that the hyperactivity of transcription in spermatocytes provides an exploitable window of opportunity for selfish TEs (Lawlor *et al.*, 2021). Mahadevaraju *et al.* applied scRNA-seq to characterize the chromosome-specific dynamics of transcription in progressively later germline cell types (Mahadevaraju *et al.*, 2021). Their findings are consistent with our expectations that the loss of dosage compensation means that X-linked gene expression in spermatocytes is less than X-linked gene expression in somatic tissues, where dosage compensation is upregulating the X.

One of the crowning achievements of modern genomics research in *D. melanogaster* is the Fly Cell Atlas (FCA), a collaborative effort to perform single-cell RNA-seq on all cell types of the fly, producing an "atlas" of transcription in all unique cell types (Li *et al.*, 2022). This was an enormous consortium involving dozens of academic research labs and participation from Genentech, the Chan Zuckerberg Biohub, and the National Institute of Health. Ultimately, 250 distinct single-cell clusters were identified across all tissues of *D. melanogaster*, and the single-cell datasets have been made publicly available as a resource for future genetics research on *Drosophila*, such as this dissertation.

The Fly Cell Atlas is an incredibly valuable resource for single-cell transcriptomics data on the fly testis. The data covers >40,000 single nuclei across all cell types of the testis. This FCA data is already being mined by fly researchers to reveal new facets of transcription in all tissues of the fly, including the testis. Raz *et al.* mined this FCA data to identify novel marker genes for germline and somatic cell types of the testis, including rare cell types which have been challenging to profile in past studies (Raz *et al.*, 2023). They also clarify the temporal dynamics of transcription before/after meiosis, reporting a subset of genes expressed post-meiotically, as well as describing the perdurance of transcripts which are transcribed pre-meiotically but stored for their role in spermiogenesis post-meiotically.

In my own thesis work, this FCA data was a valuable resource enabling multi-modal genomics analysis between transcriptomics and chromatin-profiling data. As I discuss in Chapter 2, I leveraged the single-nucleus FCA data to annotate cell-specific gene categories for the major cell types of the testis. I then described how my CUT&Tag profiling for active epitopes such as H3K4me2 and RNA PolII ser2phos

reveals unique aspects of the activation of spermatogenesis genes compared to other tissue-specific genes.

Chapter 2

Chromosome-specific maturation of the epigenome in the male germline of *D. melanogaster*

Modified from an article currently in review at *eLife*:

James Anderson, Steve Henikoff, and Kami Ahmad.

Abstract

Spermatogenesis in the *Drosophila* male germline proceeds through a specialized transcriptional program controlled both by germline-specific transcription factors and by testis-specific versions of core transcriptional machinery. This program includes the activation of genes on the heterochromatic *Y* chromosome, and reduced transcription from the *X* chromosome. The expression from these sex chromosomes is regulated has not been defined. To resolve this, we profiled active chromatin features in the testes from wildtype and stage arrest mutants and integrate this with single-cell gene expression data from the Fly Cell Atlas. Our data assign the timing of promoter activation for genes with germline-enriched expression throughout spermatogenesis, and general alterations of promoter regulation in germline cells. By profiling both active RNA polymerase II and histone modifications in isolated spermatocytes, we detail widespread patterns associated with regulation of the sex chromosomes. Our results demonstrate that the *X* chromosome is not enriched for silencing histone modifications, implying that X chromosome inactivation does not occur in *Drosophila*. Instead, a lack of dosage compensation in spermatocytes accounts for the reduced expression from this chromosome. Finally, profiling uncovers dramatic ubiquitinylation of histone H2A and lysine-16 acetylation of histone H4 across the *Y* chromosome in spermatocytes that may contribute to the activation of this heterochromatic chromosome.

Introduction

The germline in animals is responsible for producing the specialized gametes that transmit genetic information to the next generation, and gene expression in these cells is tightly controlled to ensure proper cell differentiation and genome stability. In many cases gene regulation is distinctive from that in somatic cells (Freiman, 2009). A dramatic example of this is the deployment of testis-specific variants of core transcriptional machinery in the *Drosophila* male germline that are used to activate and regulate gene promoters during sperm development (Hiller *et al.*, 2004). These variants enable the activation of testis-specific promoters, but the effects they have on transcription and chromatin are less characterized.

Gene regulation in *Drosophila* spermatogenesis also involves widespread changes on entire chromosomes. The sex chromosomes each have distinct chromosomal features which uniquely impact their expression in the germline: the single *X* chromosome of males is thought to be upregulated in early germline cells, but then suffers a broad reduction in expression during spermatogenesis (Mahadevaraju *et al.*, 2021; Witt *et al.*, 2021). In contrast, the *Y* chromosome is largely heterochromatic and silenced in somatic cells but becomes highly active in spermatocytes, extruding long, diffuse loops of actively transcribing genes within the nucleus (Fingerhut *et al.*, 2019). How these chromosome-wide changes are orchestrated remains unknown.

Here, we apply CUT&Tag chromatin profiling (Kaya-Okur *et al.*, 2019) on the adult testis of *Drosophila* to characterize the chromatin features of the spermatogenic transcription program. By integrating chromatin profiles with published single-cell transcriptional data for the *Drosophila* testis, we track active chromatin features of germline-specific gene promoters, detailing their timing of activation. Further, we describe the enrichment of RNA polymerase II and select histone modifications across the sex chromosomes in isolated spermatocytes. These profiles show that the *X* chromosome does not accumulate repressive chromatin marks, supporting a model where reduced expression of this chromosome is due to a lack of chromosomal dosage compensation in this cell type. Surprisingly, we find that high levels of mono-ubiquitinylated histone H2A accumulate across the *Y* chromosome in spermatocytes, implicating this otherwise repressive histone modification in the activation of heterochromatic regions in the male germline.

Results

Profiling active promoters in the Drosophila testes

To profile chromatin features of gene activity during spermatogenesis, we performed CUT&Tag profiling for histone H3 lysine-4 dimethylation (H3K4me2), which marks active promoters and enhancers (Bernstein, Kamal, Lindblad-Toh, Bekiranov, Bailey, Huebert, McMahon, Karlsson, Kulbokas, Gingeras, Schreiber, Lander, et al., 2005). The adult testis contains all developmental stages of spermatogenesis, as cells in the germline continually proliferate and differentiate (**Figure 2.1A**) (White-Cooper & Bausek, 2010). Germline stem cells at the apical tip of the testis asymmetrically divide to birth spermatogonia; these undergo four rapid mitotic divisions and then complete one final S phase before entering an extended G2 phase as primary spermatocytes before meiosis and differentiation into sperm. The transcriptional program of premeiotic and post-meiotic stages has been detailed most extensively by single-cell RNA-seq profiling (Witt *et al.*, 2019) (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2019; Witt *et al.*, 2021). To assess the developmental timing of active chromatin features, we profiled the H3K4me2 modification in wildtype adult testes and in two stage-arrest mutants. The germline of *bag-of-marbles* (*bam*) mutant males arrests in the spermatogonial stage, so their testes are full of early germline cells (Chen *et al.*, 2011). The germline of *always early* (*aly*) mutants arrests in the early primary spermatocyte stage, thus enriching for this cell type (Laktionov *et al.*, 2018). Between these three genotypes, only wildtype testes contain late spermatocytes and post-meiotic stages. Thus, profiling testes from these flies distinguishes when in development active chromatin features appear. We dissected testes from one-day-old adult wildtype, *bam*, and *aly* males, sequenced three replicate

libraries for each genotype, and mapped reads to a repeat-masked version of the *Drosophila* dm6 genome assembly (see Methods). To aid distinguishing germline from somatic cell type chromatin features, we also profiled wing imaginal discs from larvae. Sequencing from replicates was pooled, providing 4.4M-19.2M reads/genotype, and publicly available coverage tracks are posted (https://genome.ucsc.edu/s/jamesanderson12358/analysis230508___UCSC_session_germline_MS).

Inspection of these tracks reveals active chromatin features that track with gene expression timing during spermatogenesis. For example, the promoter for the germline stem cell marker *nanos* (*nos*) is marked with the H3K4me2 modification in *bam* testes, with less signal in *aly* or in wildtype testes (**Figure 2.1B**). In contrast, the promoter of the spermatocyte-expressed gene *loopin-1* lacks the H3K4me2 modification in *bam* testes but is heavily marked in *aly* and in wildtype testes (**Figure 2.1C**). Similarly, the promoter of the meiotically-expressed genes *Mst36Fa* and *Mst36Fb* (Di Cara *et al.*, 2006) are marked with the H3K4me2 modification in wildtype testes where the later stages of spermatogenesis are present (**Figure 2.1D**). In contrast, no signal is present at the *nos*, *loopin-1*, *Mst36Fa* or *Mst36Fb* genes in wing discs (**Figure 2.1B-D**). Finally, profiles for all three testes samples show the H3K4me2 mark at the promoters of the homeotic genes *abdominal-A* and *Abdominal-B* that are expressed in the somatic cells of the testis (**Figure 2.1E**), since somatic cells are present in all three samples.

To visualize chromatin features across active promoters during spermatogenesis, we categorized genes by their timing and level of mRNA expression in the male germline in single-nucleus RNA-seq profiling (Raz *et al.*, 2023). We focused on five categories of genes with germline-enriched expression in germline stages (**Supplementary Figure 2.1, Supplementary Table 2, 3**), including 845 genes that are predominantly expressed in spermatogonia, 1,510 expressed in early spermatocytes, 1,524 expressed in mid-spermatocytes, 2,052 expressed in late spermatocytes, and 475 genes expressed in spermatids. We then displayed the summed H3K4me2 signal spanning -200 – +500 bp around each of the promoters for these genes (**Figure 2.2A**). In the whole testis, germline stages compose a small proportion of all cells, and so genes specifically expressed in these stages have low signal for both mRNA and the H3K4me2 modification (**Supplementary Table 2**). Nevertheless, the overall tendency is that promoters of genes expressed in spermatogonia are marked with the H3K4me2 modification in *bam* mutant testes, which are enriched for early stages compared to *aly* mutant and wildtype testes (**Figure 2.2A**). For example, the promoters of early germline markers *nos*, *vasa* (*vas*), and *zero population growth* (*zpg*) are each heavily marked with H3K4me2 in *bam* mutant samples, reflecting their activity (**Figure 2.2A,B**). Similarly, the promoters for many genes primarily expressed in early spermatocytes and mid-stage spermatocytes are often most heavily marked with H3K4me2 in *bam* mutant testes, implying that these genes first become active in spermatogonia and accumulate mRNA in spermatocyte stages. Finally, gene promoters for mRNA that accumulate in late spermatocytes and in differentiating spermatids are predominantly marked with the H3K4me2 modification in wildtype testes, as this is the only sample that contains these stages of spermatogenesis (**Figure 2.2A**). This includes the activation of promoters for the *Y* chromosome-linked fertility factor genes (*kl-2*, *kl-3*, and *kl-5*) which are primarily expressed in late spermatocytes (**Figure 2.2A,B**). However, the activation timing of these genes appear to differ, as the promoters of *kl-2* and *kl-3* accumulate the H3K4me2 modification earlier than does the *kl-5* promoter (**Figure 2.2B**). The most dramatic instance of precocious activation of a *Y*-linked promoter is the *FDY* gene which becomes active in early spermatocytes, matching its early production of mRNA (**Figure 2.2A,B**).

There are three exceptions to the overall trend of correspondence between promoter activation and mRNA accumulation. First, active promoters with very low levels of H3K4me2 are most heavily marked in wildtype testes, regardless of when the genes are expressed. These might be active genes where the histone modification accumulates during the extended growth phase of spermatocytes. Additional examples are the *kumgang* (*kmg*) and *cookie monster* (*comr*) gene promoters which produce mRNA in mid-stage spermatocytes, but are most heavily marked with the H3K4me2 modification in wildtype testes. Second, some promoters acquire the H3K4me2 modification well before mRNA accumulates. A small number of genes are expressed in the post-meiotic stages of spermatogenesis, including the "cup" genes (Barreau *et al.*, 2008). Many of these promoters are most heavily marked with the H3K4me2 modification in wildtype testes, but the promoter for *ryder cup* (*r-cup*) is already marked in *aly* mutant testes, implying that it is already active in pre-meiotic stages (**Figure 2.2B**). Third, we cannot measure H3K4me2 modification at the *bam* promoter in *bam* mutant testes or at the *aly* promoter in *aly* mutant testes (**Figure 2.2B**). As these stage-arrest mutations have pleiotropic effects on gene expression (Barreau *et al.*, 2008) some discrepancies between promoter marking in mutants and transcript accumulation in wildtype testes may be due to aberrant transcriptional regulation.

Notably, the importance of the H3K4me2 modification appears to diminish as spermatogenesis proceeds, as the average signal of this mark around promoters for germline-enriched genes in wildtype testes drops in the later stages of spermatogenesis (**Figure 2.2C**). While the promoters of early germline-expressed genes and somatically-expressed genes have comparable levels of the H3K4me2 modification centered around their TSS, the promoters of late spermatocyte and spermatid genes display very little marking, and this low level extends into gene bodies. This suggests that the activities of H3K4-modifying enzymes are reduced in these later stages.

Profiling FACS-isolated primary spermatocytes

Germline cells undergo massive nuclear expansion and extensive transcriptional activation between mitotic spermatogonia and meiotic division, in part directed by germline-specific variants of general transcription factors (Lim *et al.*, 2012). To specifically profile chromatin features in spermatocytes, we used a spermatocyte-enriched GFP marker to isolate these cells by fluorescence-activated cell sorting (FACS). The *hephaestus* (*heph*) gene encodes an RNA-binding protein that is broadly expressed, but in spermatocytes Heph binds the abundant nuclear transcripts from the *Y* chromosome fertility genes (**Figure 2.3A**) (Fingerhut *et al.*, 2019). We performed FACS on 40 dissociated testes from males carrying a *heph-GFP* transgene, recording the forward scatter (FSC) and GFP signal of each event (**Figure 2.3B**). FACS profiles from *heph-GFP* samples display a large proportion of events with high GFP signal, which are absent in profiles of wild-type testes. These GFP-labeled spermatocytes have distinct sizes, consistent with the progressive growth of spermatocytes as they approach meiosis (White-Cooper, 2010): in a typical FACS experiment, ~5-10% of GFP-positive events have moderate GFP signal and moderate size (Gate 1), ~50% of events have very high GFP signal and moderate size (Gate 2), and ~5-10% have high GFP signal and large size (Gate 3). The cells of Gate 3 may represent the latest stage of spermatocytes when heph-GFP signal decreases just before meiosis. Because Gate 2 contained the most GFP-positive events, we focused further analysis on these spermatocytes. We used ~3,000 isolated spermatocytes for each profiling experiment, and since the

resulting libraries were comparatively small with high duplication rates, we pooled unique reads from multiple replicates to provide 200,000 – 900,000 unique reads for each profile (**Supplementary Table 1**).

We first profiled the distribution of the elongating form of RNA Polymerase II, marked with phosphorylation at Serine-2 (RNAPIIS2p) of the C-terminal tail of the largest subunit of the complex. Inspection of genome landscapes demonstrates the high quality of these profiles. For example, the meiotic beta-tubulin variant gene *betaTub85D* is broadly coated with RNAPIS2p in isolated spermatocytes, while signal is absent across this gene in somatic cells (**Figure 2.3C**).

Elongating RNAPII is also detectable at many genes that accumulate transcripts in late spermatocytes. For example, 14 genes encoding protamines that package the genome in sperm (Chang *et al.*, 2023) are heavily coated with RNAPII in isolated spermatocytes (**Figure 2.3E**). Similarly, elongating RNAPII is detectable at genes normally thought to be expressed in post-meiotic cells, such as *heineken-cup* (*h-cup*), implying that RNAPII is engaged at some genes well before their transcripts are detected.

While these results confirm the cell-type identity of the FACS-isolated cells, we noted that the distribution of RNAPIIS2p across genes such as *betaTub85D* differs from the typical pattern across active genes in somatic cells. Serine-2-phosphorylation of RNAPII is associated with transcriptional elongation, but in *Drosophila* and in mammalian somatic cells it shows a prominent peak at the 5’ end of active genes (Ahmad & Henikoff, 2021; Kaya-Okur *et al.*, 2019). The broad distribution of elongating RNAPII across active genes is typical in spermatocytes. Genes with germline-enriched expression show a broad distribution of elongating RNAPII downstream of their promoters, while somatically expressed genes show a prominent 5’ peak (**Figure 2.3F**).

To clearly compare RNAPII distributions between cell-types, we examined long genes that are commonly expressed in both spermatocytes and in wing imaginal discs. One example is the *shuttle craft* (*stc*) gene, which is almost 5 kb and is highly expressed. Strikingly, RNAPIIS2p signal at *stc* shows a prominent peak near the promoter in somatic wing imaginal disc cells but is broadly distributed across the gene in spermatocytes (**Figure 2.3D**). More generally, distribution of RNAPIIS2p is broadly distributed across all active genes in spermatocytes, in contrast to the proximally peaked distribution in somatic cells (**Supplementary Figure 2.2**)

The change from peaked to broad distributions of RNAPIIS2p is mirrored in the distributions of the H3K4me2 modification, as this histone modification shows an atypical broad and low distribution across the active *betaTub85D* and *stc* genes in spermatocytes (**Figure 2.3C,D**). As this histone modification occurs co-transcriptionally, its change in distribution is likely the result of the altered distribution of RNAPII across these genes.

The X chromosome is not dosage compensated in spermatocytes

In somatic cells of *Drosophila* males, the expression of genes on the single *X* chromosome is approximately doubled to equalize expression to autosomal genes. Canonical dosage compensation is accomplished by the Male Specific-Lethal (MSL) RNA-protein complex, which coats the *X* chromosome, catalyzes acetylation of histone H4 at lysine 16, and increases RNAPII density (Akhtar & Becker, 2000). However, in germline cells in the testis cytology detects no enrichment of the H4K16Ac modification, suggesting that *X* chromosome

dosage compensation does not occur in this cell type (Rastelli & Kuroda, 1998). Transcriptomic profiling showed that multiple components of the dosage compensation machinery are not expressed in the male germline (Witt *et al.*, 2021). In spite of this, single-cell RNA-seq studies have suggested that a perhaps non-canonical form of *X* chromosome dosage compensation occurs in early spermatogonial stages, but is not detectable by spermatocyte stages (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2019; Witt *et al.*, 2021).

The distribution of RNAPII in isolated spermatocytes is consistent with the lack of dosage compensation by this stage of spermatogenesis. Plotting the distribution of RNAPIIS2p across *Drosophila* chromosomes in wing imaginal discs shows a substantial enrichment across the *X* chromosome, resulting from dosage compensation in these somatic cells (**Figure 2.4A**). This enrichment of RNAPIIS2p across the *X* chromosome is lost in spermatocytes. To quantify chromosomal changes, we summarized RNAPIIS2p signal for the autosomal *2nd* and *3rd* chromosomes, the quasi-heterochromatic *4th* chromosome, and the sex chromosomes (**Figure 2.4B**). As the sex chromosomes are hemizygous, we doubled counts for genes on these chromosomes to calculate polymerase densities per gene copy, and then scaled gene scores to the median value of gene scores on the *2nd* and *3rd* autosomal chromosomes. As expected, the median expression of *X*-linked genes in wing imaginal disc cells is close to twice that of the major autosomes, showing they are dosage compensated. In contrast, median expression from *X*-linked genes is equal to that of the major autosomes in spermatocytes (**Figure 2.4B**).

To assess the chromosomal distribution of the H4K16Ac modification, we profiled it in wing imaginal discs and in isolated spermatocytes. This acetylation is widespread across the genome, consistent with its association with transcriptional activity, and in wing imaginal disc cells it is noticeably enriched across the dosage-compensated *X* chromosome (**Figure 2.4C**). In stark contrast, the *X* chromosome is depleted for H4K16ac in spermatocytes. Thus, chromatin profiling for both elongating RNAPII and the H4K16ac modification demonstrate there is no dosage compensation of the *X* chromosome in spermatocytes.

RNAPII density on the quasi-heterochromatic *4th* chromosome is reduced in spermatocytes, consistent with decreased transcript production from this chromosome in RNA-seq studies (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2019; Witt *et al.*, 2021). The *4th* chromosome is an evolutionary derivative of the *X* chromosome and it has been speculated that it may be subject to similar chromosomal regulation as the X (Larsson & Meller, 2006); however, only 10 genes are expressed from this small chromosome in spermatocytes, and this limits any inference about down-regulation of this chromosome. In contrast, the specific activation and accumulation of RNAPIIS2p on *Y* chromosome genes in spermatocytes is dramatic (**Figure 2.4A,B**). Likewise, the *Y* chromosome becomes conspicuously enriched for the H4K16ac modification in spermatocytes (**Figure 2.4C**), suggesting this modification is involved in gene activation from this chromosome.

Profiling silencing chromatin marks in spermatocytes

The sex chromosomes of male therian mammals form a cytological sex body in pre-meiotic cells (Solari, 1974). This body is a manifestation of meiotic sex chromosome inactivation (MSCI), where repressive histone

modifications silence unpaired chromosomes (Turner, 2015). MSCI has been suggested to occur in the *Drosophila* male germline to explain the mysterious dominant male sterility of many *X*-to-autosome translocations (Lifschytz & Lindsley, 1972). However, transcriptional profiling of the *Drosophila* testis has not observed silencing of the *X* chromosome (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2019; Witt *et al.*, 2021). We therefore profiled silencing histone modifications in isolated spermatocytes to determine if molecular marks of MSCI are enriched on the *Drosophila X* chromosome. Methylation of histone H3 at lysine-9 (H3K9me) is generally associated with heterochromatic silencing and marks the precociously silenced *X* chromosome in male mouse spermatogenesis (Ernst *et al.*, 2019; Khalil *et al.*, 2004). In wing imaginal disc cells dimethylation of H3K9 (H3K9me2) is enriched in silenced pericentromeric regions of all chromosomes, as well as throughout the heterochromatic *Y* chromosome and the quasi-heterochromatic *4th* chromosome (**Figure 2.5A**), consistent with the silencing of repetitive sequence regions in somatic cells. However, the genome in spermatocytes gains the H3K9me2 modification throughout chromosome arms, including those of the major autosomes and the *X* chromosome. In contrast the H3K9me2 modification is reduced across the *Y* chromosome, consistent with the activation of *Y*-linked genes in this cell type (**Figure 2.5A**). Although the H3K9me2 mark is reduced across this chromosome and across pericentromeric regions, substantial chromosomal methylation remains in spermatocytes.

A major system of chromatin repression uses trimethylation of histone H3 at lysine-27 (H3K27me3) to direct developmental gene silencing (Grossniklaus & Paro, 2014). Although this mark is not associated with MSCI in mammals (Mu *et al.*, 2014), we profiled it in *Drosophila* spermatocytes. There is little change in the chromosomal distribution of the H3K27me3 modification between wing imaginal disc cells and spermatocytes (**Figure 2.5B**). There is a slight apparent reduction of this modification across the *X* chromosome and a slight gain across the *Y* chromosome. Overall, the constancy of chromosomal patterns of the H3K27me3 modification is consistent with the inactivation of the histone methyltransferase *Enhancer of zeste* (*E(z)*) in spermatocytes (Chen *et al.*, 2011).

Mono-ubiquitinylation of histone H2A marks the active Y *chromosome*

An additional histone modification associated with precocious silencing of the *X* chromosome in male mammals is the mono-ubiquitinylation of histone H2A at lysine-119 (uH2A) (Baarends *et al.*, 1999). This modification is conserved at the homologous lysine-118 position of *Drosophila* histone H2A, and is linked to Polycomb-mediated silencing across eukaryotes (Barbour *et al.*, 2020). We therefore profiled the distribution of the uH2A modification in wing imaginal discs and in isolated spermatocytes to determine if this modification marked the *X* chromosome. The uH2A mark is broadly enriched throughout the arms of autosomes in both cell types, but shows no enrichment across the *X* chromosome in spermatocytes (**Figure 2.6A**). Thus, this chromatin marker of mammalian MSCI is also absent from the *Drosophila X* chromosome. However, the active *Y* chromosome is strikingly enriched for the uH2A modification in spermatocytes, with additional moderate enrichment in the repetitive pericentromeric regions of all chromosomes (**Figure 2.6A**).

We confirmed the chromosomal enrichment of the uH2A modification by immunostaining spermatocytes (**Figure 2.6B**). The subnuclear pattern of a Polycomb-GFP (PcGFP) fusion protein distinguishes early from late spermatocytes (El-Sharnouby *et al.*, 2013). Using this marker, we see that uH2A is largely absent from the

nucleus of early spermatocytes. In mid-stage spermatocytes a stringy wedge of uH2A staining appears in the interchromosomal space between the chromatin bodies and expands to one or two wedges in late spermatocytes (**Figure 2.6B**). The timing of appearance and position of these stained bodies resemble that of the chromosome loops that unfold from *Y*-linked genes as they are expressed (Bonaccorsi *et al.*, 1988). We therefore engineered *XO* male flies lacking a *Y* chromosome and immunostained their spermatocytes. The uH2A modification is present in these cells, but with a distinctly different focal appearance, suggesting that the ubiquitinylated histone aggregates in spermatocytes without a *Y* chromosome (**Figure 2.6B**). To confirm the timing of the appearance of the uH2A body, we immunostained germline nuclei from *bam* and *aly* stage-arrest mutants. The spermatogonial nuclei from *bam* mutant testes lack uH2A staining, while early spermatocyte nuclei from *aly* mutant testes have only a small uH2A body that always abuts against the nucleolus (**Figure 2.6B**). Thus, the uH2A body accumulates during the early spermatocyte stage and expands as spermatocytes develop. The timing and position of the uH2A body is consistent with the idea that it contains the *Y* chromosome, which is transcriptionally active in these stages.

To further characterize the relationship between transcriptional activity of the *Y* chromosome and the uH2A modification, we plotted signal for the uH2A modification and for RNAPIIS2p at *Y*-linked gene promoters (**Figure 2.6C**). In wing imaginal discs these promoters have little RNAPIIS2p or uH2A modification, but there is strong enrichment for both features at these promoters in spermatocytes, consistent with transcriptional activity of the *Y* chromosome. Since the uH2A modification also becomes enriched in pericentromeric regions, we compared the enrichment of the uH2A modification and RNAPIIS2p at repetitive transposons that constitute a large fraction of these regions. A number of transposons gain RNAPIIS2p signal specifically in spermatocytes (**Figure 2.6D**; **Supplementary Table 4**; **Supplementary Figure 2.3**), consistent with their transcriptional activation (Lawlor *et al.*, 2021). These activated transposons also gain the uH2A modification. These correspondences suggest that the uH2A modification contributes to transcriptional activation of heterochromatic regions in spermatocytes.

Discussion

Germline cells use distinctive variations on transcriptional gene regulation, and studies of *Drosophila* spermatogenesis have detailed many specialized alterations of core general transcription factors that direct expression programs as differentiation proceeds (Hiller *et al.*, 2004). However, the chromatin features of gene regulation in spermatogenesis have been less thoroughly characterized, in part because of the complexity of the tissue and limiting numbers of germline cells. We have addressed this by performing efficient CUT&Tag chromatin profiling for both active and repressive chromatin marks in the *Drosophila* testis and in isolated spermatocytes. These profiles reveal several notable features of the epigenome in the differentiating germline. First, integration of chromatin marks with published gene expression data details a general correspondence between gene promoter activation and mRNA production as expected, but for a fraction of genes their promoters activate earlier than expected. Second, genes that are activated late in spermatogenesis tend to have very reduced active chromatin marks at their promoters. Third, while many active genes in somatic cells accumulate RNA polymerase II near their gene starts, such accumulation is absent in spermatocytes. Fourth, integration of chromatin profiling for multiple chromatin marks and profiling of RNA polymerase II demonstrate that the single *X* chromosome is neither dosage compensated

nor globally inactivated in spermatocytes. Finally, histone H2A mono-ubiquitinylation appears to have a specialized role in activation of the heterochromatic *Y* chromosome.

Quirks of gene expression in the male germline

The uniform distribution of RNAPII throughout active genes in male germline cells is strikingly different from the typical accumulation of RNAPII near promoters in somatic cells. Promoter-proximal accumulation results from dynamic pausing of RNAPII before conversion into the productive elongating isoform, and is a major control point for transcriptional regulation in somatic cells (Muniz *et al.*, 2021). Thus, the uniformity of RNAPII across expressed genes in spermatocytes suggests that pausing does not occur, necessitating gene regulation solely by transcription factor and RNAPII recruitment. This may allow for a simpler promoter architecture, and indeed spermatogenic gene promoters are distinctively small (White-Cooper, 2010). Further, a lack of pausing likely affects chromatin features of promoters. Paused RNAPII binds enzymes that progressively methylate the lysine-4 residue of histone H3, and so active promoters in somatic cells are typically marked with both H3K4-dimethylation and -trimethylation (Bernstein et al., 2005). A consequence of no pausing in spermatocytes would be reduced methylation of active promoters (which we observe is most severe in late spermatocyte and post-meiotic stages), and thereby reduced reliance of these marks for promoting transcription.

The X *chromosome is neither dosage-compensated nor inactivated in spermatocytes*

While protein and lncRNA components of the somatic dosage compensation machinery are not produced in the male germline, transcriptional profiling described up-regulation of the single *X* chromosome in early germline stages (Mahadevaraju *et al.*, 2021; Raz *et al.*, 2023; Witt *et al.*, 2021). The mechanism for this non-canonical dosage compensation remains unknown (Witt *et al.*, 2021), but by the spermatocyte stage *X*-linked genes are no longer up-regulated. Our profiling of RNAPII confirms that there is no up-regulation of this chromosome in spermatocytes. There has been substantial investigation into whether the *X* is inactivated in spermatocytes, inspired by the regulation of sex chromosomes in mammals (Turner, 2015; Vibranovski, 2014). In mammalian male germlines the *X* and *Y* chromosomes undergo Meiotic Sex Chromosome Inactivation (MSCI), where these chromosomes are precociously silenced just before meiosis (Turner, 2015). A variety of chromatin features accumulate across the sex chromosomes at this time, including the enrichment of both H3K9me2 and uH2A histone modifications. However, we find that neither of these histone modifications is enriched on the *X* chromosome in the *Drosophila* male germline. This, combined with the equivalent amounts of RNA polymerase II on *X*-linked genes and autosomal ones implies that there is no *X* chromosome inactivation in *Drosophila*.

Why do flies differ from mammals? Mammalian MSCI is considered to be an elaborated response of germline cells to the detection of the unpaired sex chromosomes (Huynh & Lee, 2005). However, meiosis in *Drosophila* males is unusual in that all the chromosomes do not synapse nor recombine (McKee & Handel, 1993). Logically, the evolutionary loss of synapsis must have required the concomitant loss of an unpaired chromosome response in *Drosophila* males. Indeed, many of the proteins recruited to the mammalian sex body are normally involved in meiotic recombination but have been repurposed for unpaired chromosome inactivation (Abe *et al.*, 2022), supporting the idea that MSCI is mechanistically linked to synapsis. Alternatively, mammalian MSCI may be a variant of the X-inactivation system that operates in females for

dosage compensation (Huynh & Lee, 2005) but since *Drosophila* solved the dosage compensation problem without inactivation, this precluded the evolution of MSCI. These possibilities are not mutually exclusive.

Activation of the Y *chromosome*

The *Y* chromosome in *Drosophila* is unique in that it is almost entirely composed of repetitive sequence, but also carries some unique genes required for male fertility (Chang & Larracuente, 2019). Thus, in somatic cell types this a heterochromatic chromosome, but is heavily transcribed in spermatocytes (Bonaccorsi *et al.*, 1988). This activated chromosome accumulates multiple histone modifications (Hennig & Weyrich, 2013), some of which we have profiled here. Surprisingly, one of the histone modifications that coats the activated *Y* chromosome is mono-ubiquitinylation of histone H2A modification. This modification is typically associated with Polycomb-mediated gene silencing in somatic cells (Aloia *et al.*, 2013), where it is catalyzed by the Sce/RING1B enzyme subunit of the Polycomb Repressive Complex 1 (PRC1) (Gorfinkiel *et al.*, 2004; Wang *et al.*, 2004). However, the Polycomb subunit of PRC1 does not co-localize with the uH2A-coated *Y* chromosome in spermatocytes, suggesting that it is not catalyzed by Sce here. Indeed, uH2A also coats the *X* and *Y* chromosomes in mammalian spermatocytes, where this modification is catalyzed by a distinct enzyme, the UBR2 E3 ubiquitin ligase (An *et al.*, 2010). The *Drosophila* genome encodes a homolog of this protein family, as well as many other ubiquitin ligases, some of which target the H2A histone (Tasaki *et al.*, 2005). We do not know which enzyme is responsible for the uH2A modification in the *Drosophila* male germline, although some candidates have male sterile phenotypes when mutated (Rathke *et al.*, 2007). While it is surprising to find a histone modification conventionally associated with silencing enriched on an activated chromosome (and indeed the reason we profiled it was as a putative MSCI marker), the roles of uH2A in gene regulation are diverse even in somatic cells. The uH2A modification is associated with both silenced and some active genes in developing eye tissue (Loubiere *et al.*, 2020), and counters chromatin compaction in the early embryo (Bonnet *et al.*, 2022). In addition to the *Y* chromosome, a number of heterochromatic transposons in the *Drosophila* male germline are also activated and accumulate the uH2A modification. Thus, it is conceivable that this modification – perhaps in combination with other marks – works generally to activate extremely heterochromatic regions in spermatocytes.

Methods

Fly strains

All crosses were performed at 25℃. All mutations and chromosomal rearrangements used here are described in Flybase (http://www.flybase.org). The w^{1118} strain was used as a wildtype control. The *heph-GFP* males used for profiling have the genotype *y w/Y ; P[PTT-GC]heph*^{CC00664}*/TM3, Ser Sb*. The *bam* mutant males have the genotype *w/Y ; e bam*^{D86}*/Df(3R)FDD-0089346*. The *aly* mutant males have the genotype *P[ry11]ry2, mwh aly*^{1} *ry*^{506} *e/Df(3L)BSC428*. Additional genotypes used for cytological characterization were *y w P[bam-GAL4:VP16,w*$^{+}$*]1/Y ; P[UAS-RFP,w*$^{+}$*]2/2 ; P[PTT-GC]heph*^{CC00664}*/3* and w^{1118}*/Y ; P[Pc-eGFP,w*$^{+}$*]3*.

Antibodies

The following antibodies were used: Epicypher 13-0027 anti-H3-K4-dimethyl, Cell Signalling Technology E1Z3G anti-RNAPII-Serine-2-phosphorylation, Cell Signalling Technology 8240 anti-H2A-K119-ubiquitinylation, EMD Millipore 05-1249 anti-H3-K9-dimethyl, Cell Signalling Technology C36B11 anti-H3-K27-trimethyl, Abcam ab109463 anti-H4-K16-acetyl, and Abcam ab5821 anti-fibrillarin.

Imaging whole testes

Testes from one-day old adult males were dissected and fixed in 4% formaldehyde/PBS with 0.1% Triton-X100 (PBST) for 10 minutes, stained with 0.5 µg/mL DAPI/PBS, and mounted in 80% glycerol on slides. Testes were imaged by epifluorescence on an EVOS FL Auto 2 inverted microscope (Thermo Fisher Scientific) with a 10X objective.

Imaging spermatocytes

Testes from third-instar male larvae were prepared as described [Bonaccorsi *et al*, 2000]. Briefly, one testis was dissected in a drop of PBS on a Histobond glass slide (VWR 16004-406), squashed gently with a RainX (ITW Global Brands) coated coverslip, then flash-frozen in liquid nitrogen. After popping off the coverslip, the sample was fixed with 4% formaldehyde/PBST for 5 minutes, and incubated with 0.3% sodium deoxycholate/PBST twice for 20 minutes each. Samples were incubated with primary antiserum in PBST supplemented with 0.1% bovine serum albumin (BSA) at 4˚overnight, and finally with fluorescently-labeled secondary antibodies (1:200 dilution, Jackson ImmunoResearch). Slides were stained with 0.5 µg/mL DAPI/PBS, mounted in 80% glycerol, and imaged by epifluorescence on an EVOS FL Auto 2 inverted microscope (Thermo Fisher Scientific) with a 40X objective. Pseudo-colored images were adjusted and composited in Adobe Photoshop and Adobe Illustrator.

Whole-mount CUT&Tag

To perform CUT&Tag for whole tissues (“whole-mount CUT&Tag”), we dissected 10 testes from one-day-old adults or 10 imaginal wing discs from 3rd instar larvae in PBS buffer supplemented with cOmplete protease inhibitor (Roche 11697498001). Dissected tissues were permeabilized with 0.1% Triton/PBS for 30 minutes at room temperature, and then manually transferred into the following CUT&Tag solutions sequentially between wells of a glass dissection plate: primary antibody solution (diluted in Wash+ buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 2 mM EDTA, 1% BSA, with cOmplete protease inhibitor)) overnight at 4˚, secondary antibody solution (in Wash+ buffer) for 1 hour at room temperature, and then incubated with loaded protein-A-Tn5 (in 300Wash+ buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine with cOmplete protease inhibitor) for 1 hour. After one wash with 300Wash+ buffer, samples were incubated in 300Wash+ buffer supplemented with 10 mM MgCl2 for 1 hour at 37˚to tagment chromatin. Tissues were then dissociated with collagenase (2 mg/mL, Sigma C9407) in HEPESCA (50mM HEPES buffer pH 7.5, 360 µM $CaCl_2$) solution at 37˚for 1 hour. We then added SDS to 0.16%, protease K to 0.3 mg/mL, and EDTA to 16 mM and incubated at 58˚for 1 hour, and DNA was purified by phenol:chloroform extraction and ethanol precipitation. Libraries were prepared as described (Kaya-Okur *et al.*, 2019) (https://www.protocols.io/view/bench-top-cut-amp-tag-kqdg34qdpl25/v3), with 14 cycles of PCR. Libraries were sequenced in PE50 mode on the Illumina NextSeq 2000 platform at the Fred Hutchinson Cancer Center Genomics Shared Resource.

FACS-CUT&Tag of spermatocytes

40 testes were dissected from one-day old adult *heph-GFP* males and digested in 200 µL of 2 mg/mL collagenase (Sigma C9407) in HEPESCA solution at 37˚for one hour. The sample was then repeatedly pipetted with a P200 pipette tip to dissociate the tissue, then passed through a 35 µM filter with 5 mL collection tube (Corning 352235) on ice. The filter was washed with PBS to bring the total volume of collected filtrate to 1 mL. A Sony MA900 Multi-Application Cell Sorter with a 100 µM nozzle, flow pressure of 2, GFP laser settings of 32% and FSC=1 was used for isolating cells. Isolated cells were collected in 1 mL of PBS in 5 mL tubes. Benchtop CUT&Tag was performed on these samples as described (Kaya-Okur *et al.*, 2019), and sequenced in PE50 paired-end mode.

Genome mapping

To streamline analysis of repetitive transposons in the fly genome, we used a modified version of the release r6.30 *D. melanogaster* genome for mapping where repetitive sequences are masked out of the genome (http://hgdownload.cse.ucsc.edu/goldenPath/dm6/bigZips/dm6.fa.masked.gz) and with consensus sequences for 128 transposon sequences (https://github.com/bergmanlab/*Drosophila*-transposons/blob/master/misc/D_mel_transposon_sequence_set.fa) appended [Ashburner *et al*, 2021]. Paired-end reads were mapped to this assembly using Bowtie2 (using parameters, *e.g.*: --end-to-end --very-sensitive --no-mixed --no-discordant -q --phred33 -I 10 -X 700).

Mapped reads from whole-tissue replicates were merged using samtools-merge and converted to coverage tracks using bedtools-genomecov with options -scale -fs. For profiling FACS-isolated spermatocytes, duplicate reads were removed from each library using Picard-remove duplicates, and then replicates were merged using samtools-merge and converted to coverage tracks using bedtools-genomecov with options -scale -fs. These tracks are hosted at UCSC (https://genome.ucsc.edu/s/jamesanderson12358/analysis230508___UCSC_session_germline_MS) for visualization, and selected regions were exported as PDF files.

Processing of FCA testis snRNA-seq data

We downloaded snRNA-seq data generated by the Fly Cell Atlas project (Li *et al.*, 2022) as a Seurat object linked in supplementary data of (Raz *et al.*, 2023), summarizing gene expression data of single nuclei from dissociated *Drosophila* adult testes. We used the Seurat function AverageExpression() to get the average expression of all genes in each of 40 UMAP groups which represent distinct cell types of the testis. This produced a 40 groups x 15,833 genes table.

The 18 germline groups and 22 somatic groups assigned in (Raz *et al.*, 2023) are:

1. Spermatogonium

2. spermatogonium-spermatocyte transition

3. mid-late proliferating spermatogonia

4. spermatocyte 0

5. spermatocyte 1
6. spermatocyte 2
7. spermatocyte 3
8. spermatocyte 4
9. spermatocyte 5
10. spermatocyte 6
11. spermatocyte 7a
12. maturing primary spermatocyte
13. spermatocyte
14. late primary spermatocyte
15. early elongation stage spermatid
16. early-mid elongation-stage spermatid
17. mid-late elongation-stage spermatid
18. spermatid
19. hub
20. cyst stem cell
21. early cyst cell 1
22. early cyst cell 2
23. cyst cell intermediate
24. spermatocyte cyst cell branch a
25. spermatocyte cyst cell branch b
26. cyst cell branch a
27. cyst cell branch b
28. male gonad associated epithelium
29. seminal vesicle
30. adult tracheocyte
31. muscle cell
32. testis epithelium

33. hemocyte
34. hcc
35. tcc
36. pigment cell
37. adult fat body
38. secretory cell of the male reproductive tract
39. adult neuron
40. “Unannotated”

Promoter and gene scoring tables

We compiled a list of genes with male-germline-enriched expression as follows. We compiled a list of unique protein-coding mRNAs and lncRNA genes from the *Drosophila* dmel_r6.31 genome assembly (http://ftp.flybase.net/releases/FB2019_06/dmel_r6.31/gtf/), and matched FCA expression data for 40 testis cell types [Raz 2023] to each gene with a lookup table. There were 1,062 genes that are not represented in the expression dataset; these genes are listed with #N/A values for gene expression.

For each gene, we calculated its average expression in the 18 germline groups (gexp) and its average expression in 21 somatic groups (sexp). The 40th ‘unannotated’ group was not considered for gexp or sexp values. We then used k-means clustering (k=10) to group genes by cell-type expression within the testis (**Supplementary Figure 2.1**). The k-means groups 1-5 were associated with gene expression in germline clusters 1-18. The remaining groups 6-10 were associated with gene expression in somatic clusters 19-40, and we collapsed these into one somatic group called “all somatic categories” (k-cluster group 11). We added these k-cluster annotations to each gene.

We then selected 6,419 genes with $\log_2$Fold-change$\geq$1 average expression in germline groups than in testis somatic groups (termed genes with germline-enriched expression). This table is included as **Supplementary Table 3**. To assign alternative promoters to each gene, for each transcript in the .gtf file with orientation “+” we assigned the minimum coordinate as its TSS position, and for each transcript with orientation “-“ we assigned the maximum coordinate as its TSS position. We retained only one instance of duplicate TSSs for genes with TSS coordinates represented multiple times. In 144 instances two gene names share the same TSS coordinate, and we retained a TSS for each gene in the table. This table of 21,982 promoters is included as **Supplementary Table 2**.

For each gene, we determined the identity and distance to the nearest promoter of the next upstream gene using the bedtools/closest with parameters: -D a -fu , and for the next downstream gene with parameters: -D a -fd .

To summarize the enrichment of chromatin marks at promoters, we counted mapped reads in an interval from -200 – +500 bp around each TSS in merged profiling data by summing reads using deeptools/multiBamSummary with parameters: BED-file –BED

and scaled counts by the number of reads in each library (counts * 1,000,000) / (number of reads in library) to give Counts per Million (CPM). Profiling counts were transformed into z-scores for each promoter between *bam*, *aly*, and wildtype testis samples, and these values are appended to **Supplementary Table 2**.

To summarize the enrichment of chromatin marks across genes, we counted mapped reads from the start to the end of each gene in merged profiling data by summing reads using deeptools/multiBamSummary with parameters: BED-file –BED

and scaled counts by the number of reads in each library and the length of the gene (counts * 1,000,000) / (number of reads in library * gene length in kb) to give Counts per Kilobase per Million (CPKM), and these values are appended to **Supplementary Table 3**.

To summarize enrichment of chromatin marks across transposons, we counted mapped reads across consensus transposon sequences using deeptools/multiBamSummary with parameters: BED-file –BED

and scaled counts as CPKM. These values are provided in **Supplementary Table 4**.

Genomic display

For average plots of H3K4me2 signal around promoters, profiling coverage was summarized with deepTools/bamCoverage ±1 kb around annotated TSSs excluding regions with a second gene promoter in the display window with 10 bp binning, and plotted using plotHeatmap.

For average plots of H3K4me2 signal around promoters, profiling coverage was summarized with deepTools/bamCoverage ±1 kb around annotated TSSs excluding regions with a second gene promoter in the display window with 10 bp binning, and plotted using plotHeatmap.

For heatmapping of RNAPIIS2p signal at gene starts, we used deepTools/computeMatrix with parameters: -b 2000 -a 2000 -R and then we used deepTools/plotHeatmap with parameters: --colorMap viridis.

For visualizing the chromosomal distribution of CUT&Tag data as CIRCOS plots, we used the circlize package in R with default settings (circlize version 0.4.15). Genome coverage files for plotting were generated by deepTools/bamCoverage command, using parameters:

-bs 20000 --centerReads --effectiveGenomeSize 142573017 -of bedgraph

And plotted in consecutive 20 kb bins. The innermost 3 rings of each plot display genomic coverage, with color-coding set independently per ring.The color of bins in the outer ring correspond to the fold-change of signal in spermatocytes compared to wing imaginal discs calculated using deepTools/multiBigwigSummary command, with parameters:

bins -bs 20000 –outRawCounts

For boxplots, the enrichment score for each gene was scaled by dividing its read count by the median count in the "chromosome *2* & *3*" category and plotted, discarding genes with enrichment score = 0 in either dissociated testes, in spermatocytes, or in wing imaginal discs.

Figure 2.1

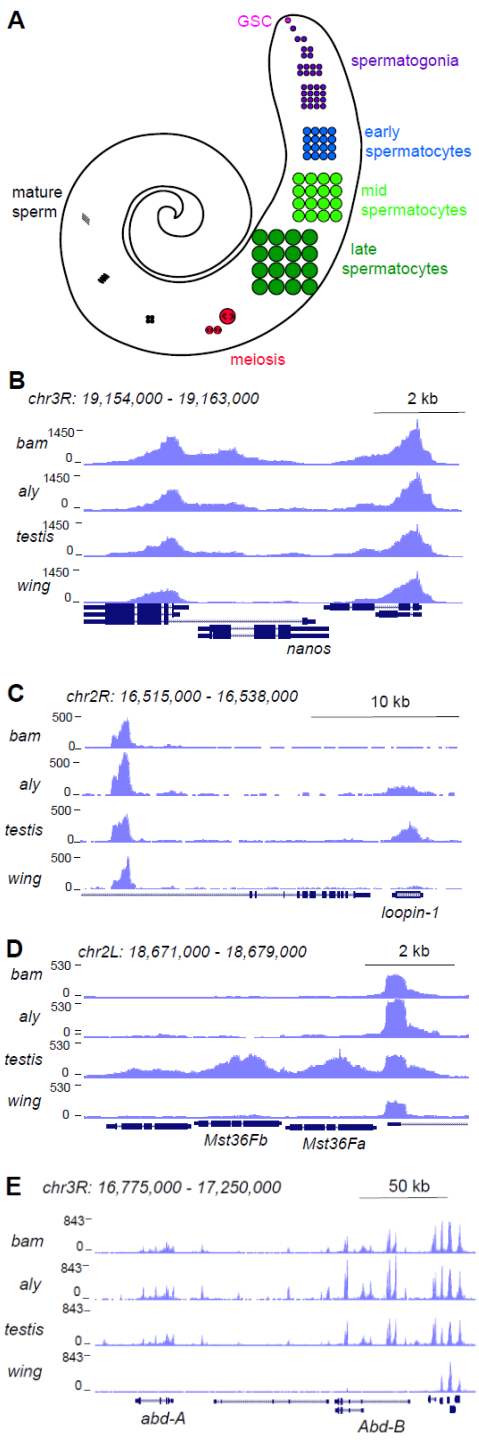


Figure 2.1. Profiling of the histone H3K4me2 modification in the *Drosophila* testis.
(**A**) Schematic of male germline stages in Drosophila. Germline stem cells (GSCs, fuschia) are located in the apical tip of the testis. After an asymmetric division, a progeny spermatogonium (purple) undergoes 4 rounds of mitotic divisions. After one last S phase, cells grow over ~3 days as spermatocytes (blue, light green, dark green) before meiosis (red). Post-meiotic differentiation produces mature sperm (black) with elongated nuclei. Somatic cell types of the testis are not shown.
(**B-E**) Distribution of the H3K4me2 modification in testes from *bam* mutants, from *aly* mutants, from wildtype animals, and from wing imaginal discs.
(**B**) H3K4me2 around the GSC-expressed *nanos* gene. low signal across *nanos* is highest in testes from *bam* mutants, and apparent in all three testes samples, while neighboring genes show peaks in all samples.
(**C**) H3K4me2 around the spermatocyte-expressed *loopin-1* gene. H3K4me2 signal appears in *aly* mutant samples (which contain early spermatocytes) and reach high levels in wildtype testes (which include later stages of spermatogenesis).
(**D**) H3K4me2 around the meiotically-expressed genes *Mst36Fa* and *Mst36Fb* genes. Signal across these genes only appears in wildtype testes.
(**E**) H3K4me2 around the *abd-A* and *Abd-B* genes, which are expressed in somatic cells of the testis.

Figure 2.2

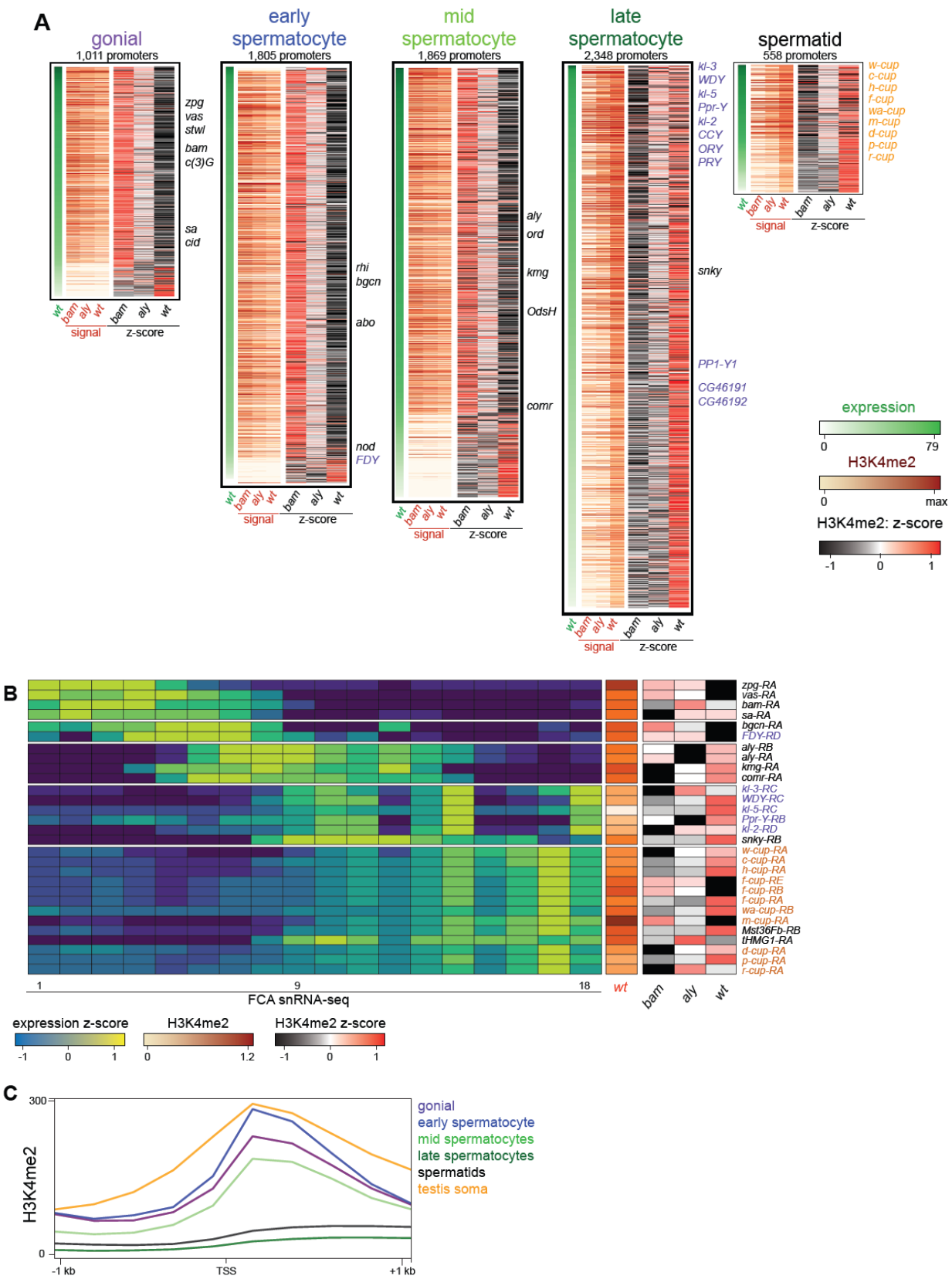

A
gonial
1,011 promoters
early spermatocyte
1,805 promoters
mid spermatocyte
1,869 promoters
late spermatocyte
2,348 promoters
spermatid
558 promoters
zpg
vas
stwl
bam
c(3)G
sa
cid
rhi
bgcn
abo
nod
FDY
aly
ord
kmg
OdsH
comr
kl-3
WDY
kl-5
Ppr-Y
kl-2
CCY
ORY
PRY
snky
PP1-Y1
CG46191
CG46192
w-cup
c-cup
h-cup
f-cup
wa-cup
m-cup
d-cup
p-cup
r-cup
wt
bam
aly
signal
z-score
expression
0
79
H3K4me2
0
max
H3K4me2: z-score
-1
0
1
B
zpg-RA
vas-RA
bam-RA
sa-RA
bgcn-RA
FDY-RD
aly-RB
aly-RA
kmg-RA
comr-RA
kl-3-RC
WDY-RC
kl-5-RC
Ppr-Y-RB
kl-2-RD
snky-RB
w-cup-RA
c-cup-RA
h-cup-RA
f-cup-RE
f-cup-RB
f-cup-RA
wa-cup-RB
m-cup-RA
Mst36Fb-RB
tHMG1-RA
d-cup-RA
p-cup-RA
r-cup-RA
1
9
18
FCA snRNA-seq
wt
bam
aly
wt
expression z-score
-1
0
1
H3K4me2
0
1.2
H3K4me2 z-score
-1
0
1
C
300
0
H3K4me2
-1 kb
TSS
+1 kb
gonial
early spermatocyte
mid spermatocytes
late spermatocytes
spermatids
testis soma

Figure 2.2. Changes in the histone H3K4me2 modification in germline-expressed genes. H3K4me2 signal around gene promoters (-200–+500 bp) with transcripts enriched in specific germline cell types. Transcript expression was derived from FCA snRNA-seq clustering [Raz *et al*, 2023].
(**A**) Expression (snRNA-seq) in wildtype testes and H3K4me2 enrichment in *bam* mutant, in *aly* mutant, and in wildtype testes in germline stages. The z-scores for H3K4me2 signal were calculated between the three genotypes. Notable germline-enriched genes are indicated, including those expressed in GSCs, in meiosis, or for cell division (black), linked to the *Y* chromosome (purple), or expressed post-meiotically (orange).
(**B**) Selected examples of promoters with germline-enriched expression. Expression z-scores from FCA snRNA-seq across 18 germline clusters [Raz 2023] and H3K4me2 enrichment in *bam* mutant, in *aly* mutant, and in wildtype testes.
(**C**) Distribution of H3K4me2 around promoters with germline-enriched expression. The testis somatic category comprises the top tercile of promoters with somatic cell-type expression in snRNA-seq data [Raz 2023]. Only promoters with no promoter of a second gene within 1 kb upstream are shown.

Figure 2.3

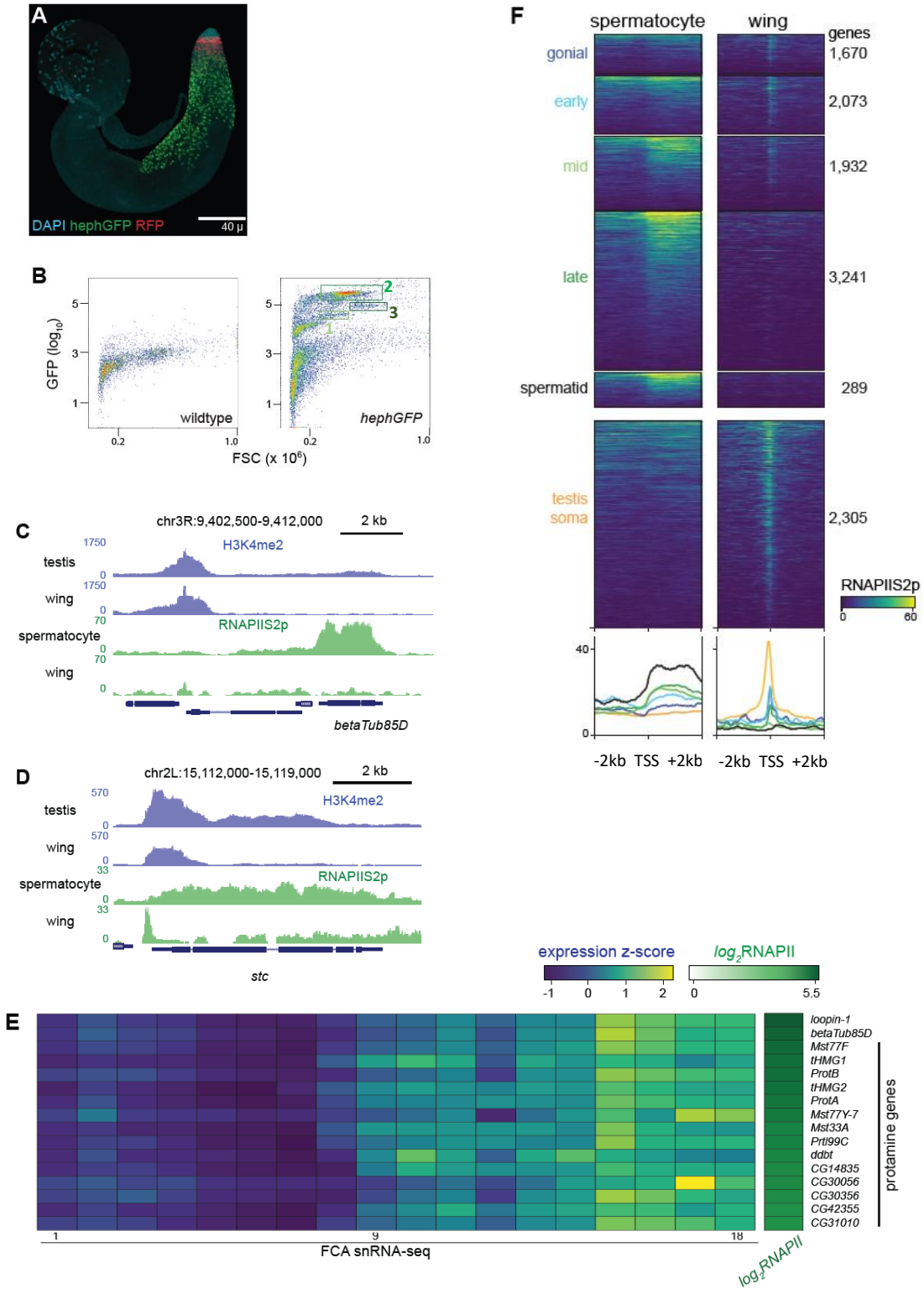

A
DAPI hephGFP RFP
40 µ
B
GFP (log10)
wildtype
hephGFP
FSC (x 10^6)
C
chr3R:9,402,500-9,412,000
2 kb
H3K4me2
testis
wing
RNAPIIS2p
spermatocyte
wing
betaTub85D
D
chr2L:15,112,000-15,119,000
2 kb
H3K4me2
testis
wing
RNAPIIS2p
spermatocyte
wing
stc
E
FCA snRNA-seq
loopin-1
betaTub85D
Mst77F
tHMG1
ProtB
tHMG2
ProtA
Mst77Y-7
Mst33A
Prtl99C
ddbt
CG14835
CG30056
CG30356
CG42355
CG31010
protamine genes
log2RNAPII
F
spermatocyte
wing
genes
gonial
1,670
early
2,073
mid
1,932
late
3,241
spermatid
289
testis soma
2,305
RNAPIIS2p
-2kb TSS +2kb
-2kb TSS +2kb
expression z-score
log2RNAPII

Figure 2.3. Profiling RNA Polymerase II in isolated spermatocytes.
(**A**) An adult testis carrying a *UASRFP* construct induced by a *bamGAL4* driver and a *hephGFP* construct. Gonial cells are labeled red while spermatocytes are labeled green with fluorescent proteins.
(B) FACS plots of recorded events from dissociated testes for forward scatter (FSC) and GFP signal. Boxes indicate events collected for chromatin profiling.
(C) Distribution of RNAPIIS2p at the spermatocyte-expressed *betaTub85D* gene in isolated spermatocytes and in wing imaginal discs.
(D) Distribution of RNAPIIS2p at the broadly-expressed *stc* gene in isolated spermatocytes and in wing imaginal discs. RNAPIIS2p is strongly localized at the *stc* promoter in wing imaginal discs, but more evenly distributed in spermatocytes.
(**E**) Selected examples of genes with late germline expression and for protamines. Expression z-scores from FCA snRNA-seq across 18 germline clusters [Raz *et al*, 2023] and RNAPIIS2p enrichment in isolated spermatocytes.
(**F**) Enrichment of RNAPIIS2p in isolated spermatocytes and in wing imaginal discs across genes with germline-enriched transcripts.

Figure 2.4

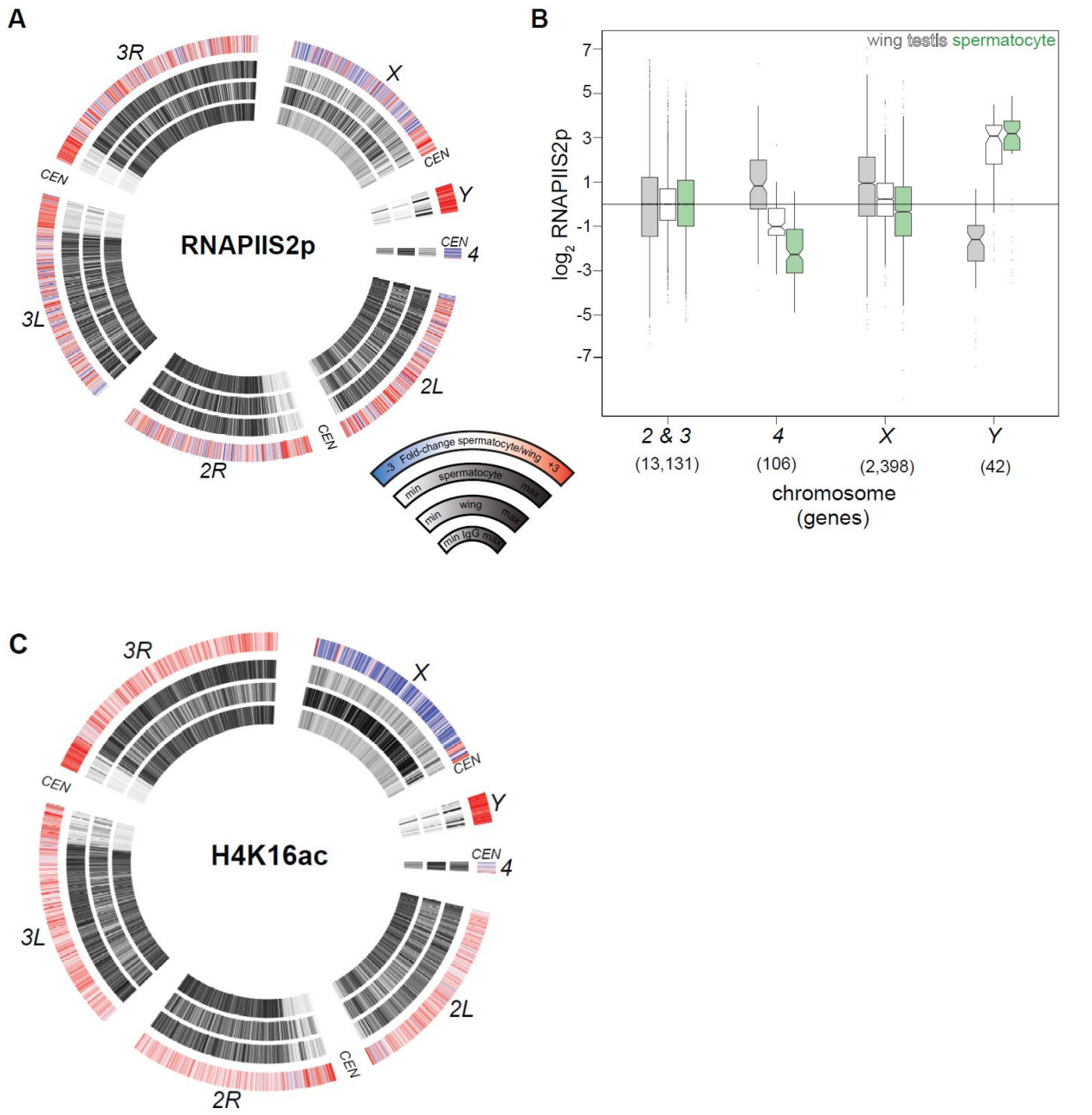


Figure 2.4. Chromosomal distribution of RNA polymerase II and H4K16ac in isolated spermatocytes.
(**A**) CIRCOS plot of RNAPIIS2p across Drosophila chromosomes. The signal (black) in IgG controls, in wing imaginal discs, and in isolated spermatocytes is shown in internal rings, and the log2 fold-change of signals between spermatocytes and wings is shown in the outer ring.
(**B**) Enrichment of RNAPIIs2p across gene bodies in wing imaginal discs, in whole dissociated testes, and in isolated spermatocytes separated by chromosomal location. Scores are scaled to the median score on the 2nd and 3rd chromosomes.
(**C**) CIRCOS plot of the dosage-compensation marker histone H4K16ac across Drosophila chromosomes. Signal (black) in IgG controls, in wing imaginal discs, and in isolated spermatocytes is shown in internal rings, and the log2 fold-change of signals between spermatocytes and wings is shown in the outer ring.

Figure 2.5

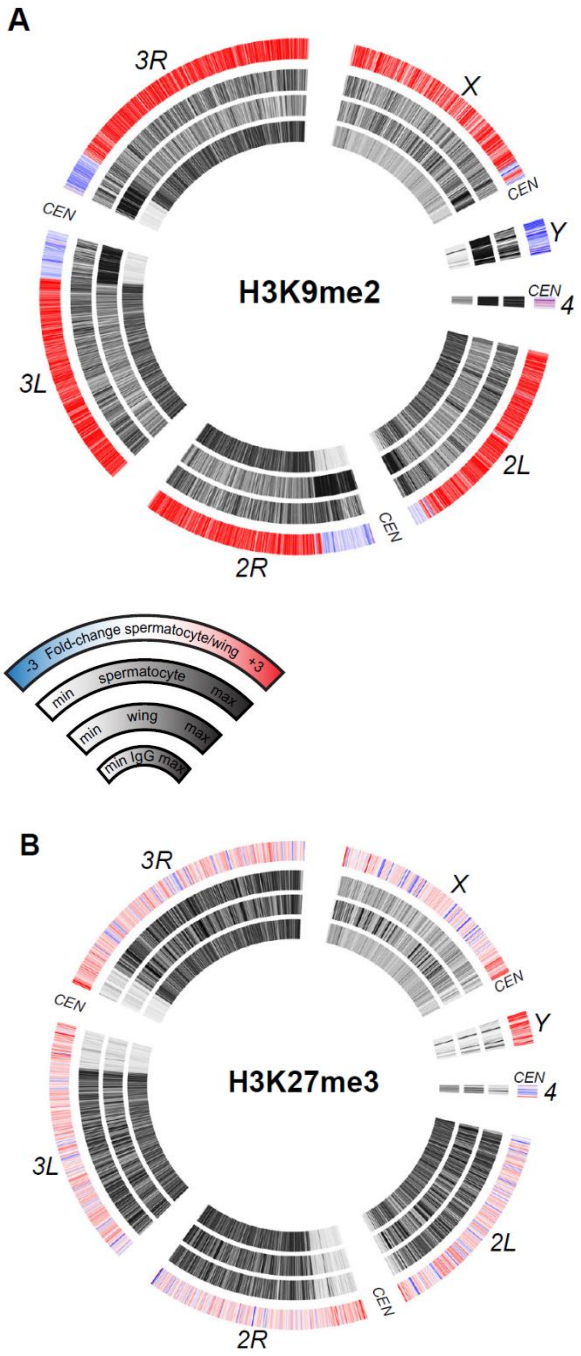


Figure 2.5. Chromosomal distribution of repressive histone modifications in isolated spermatocytes. CIRCOS plots across Drosophila chromosomes show signal (black) in IgG controls, in wing imaginal discs, and in isolated spermatocytes in internal rings, and the log2 fold-change of signals between spermatocytes and wings in the outer ring.
(**A**) Distribution of the heterochromatin-silencing marker H3K9me2.
(**B**) Distribution of the Polycomb-silencing marker H3K27me3.

Figure 2.6

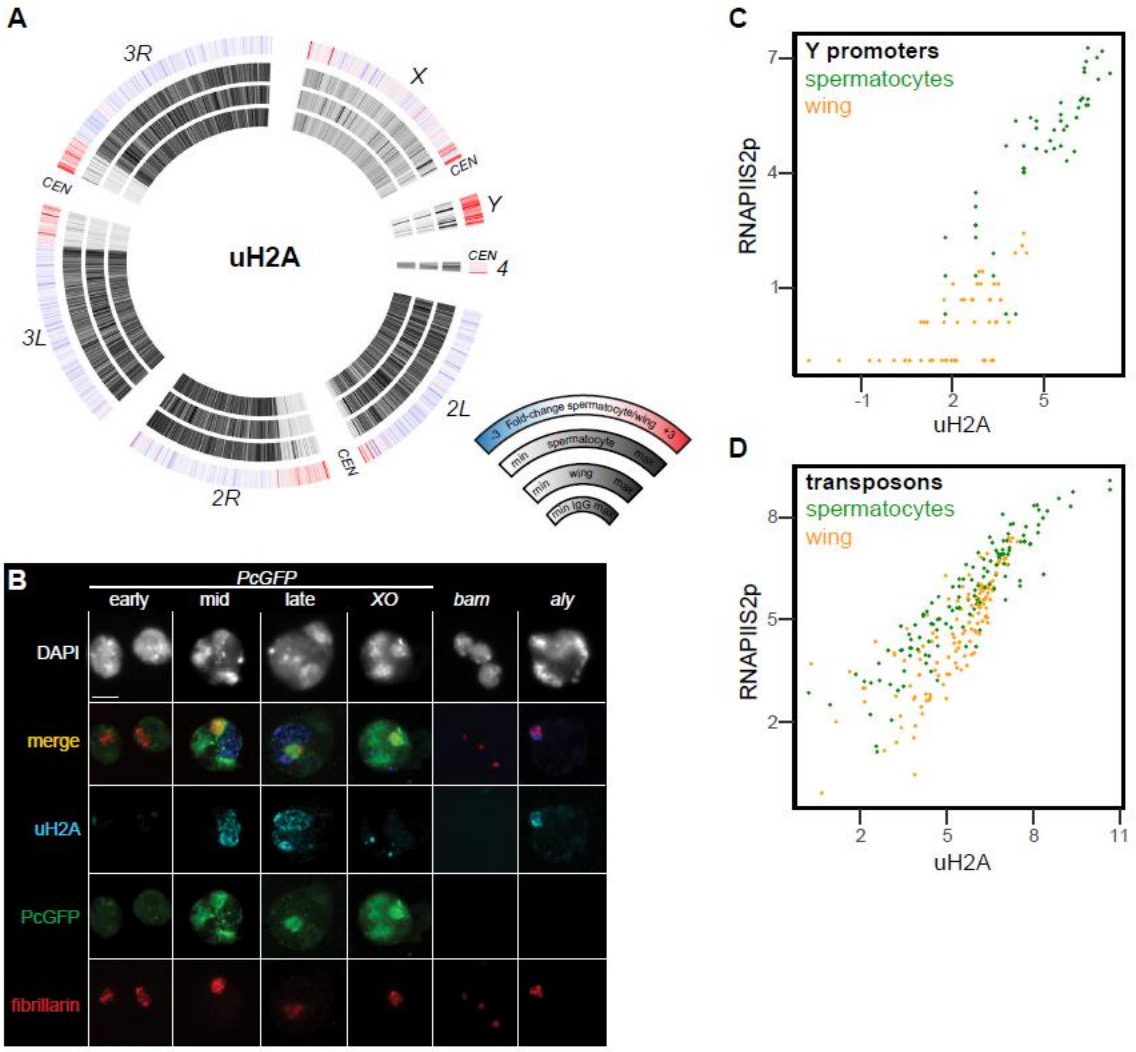
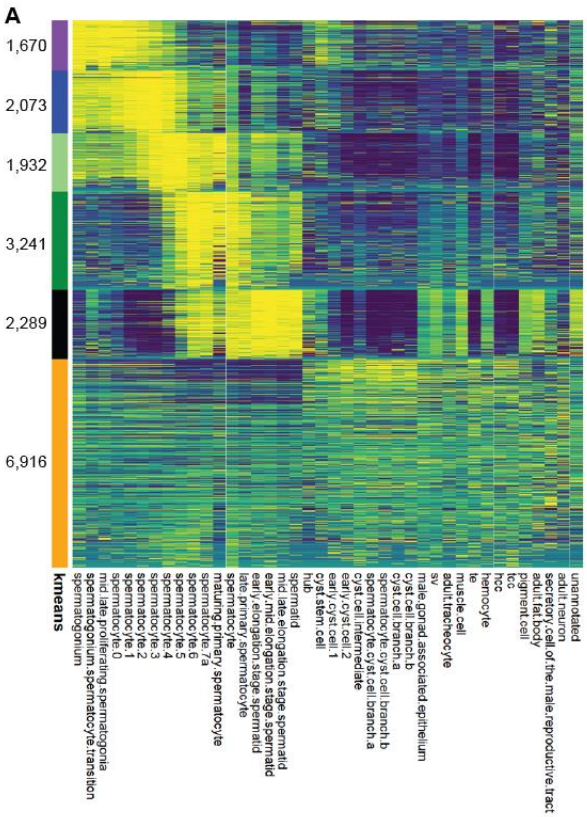
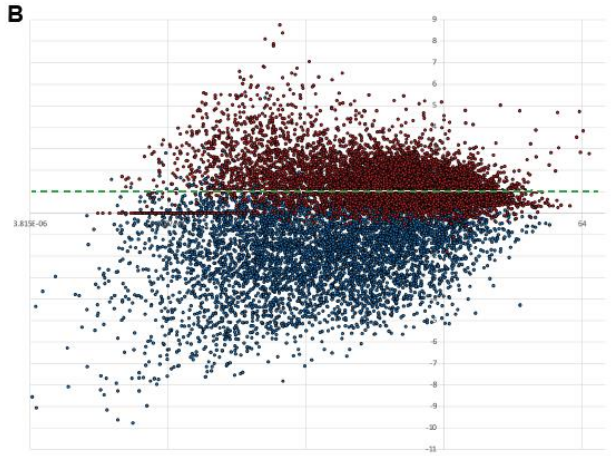


Figure 2.6. Chromosomal distribution of ubiquitinylated histone H2A in isolated spermatocytes.
(**A**) CIRCOS plots of uH2A across Drosophila chromosomes in IgG controls, in wing imaginal discs, and in isolated spermatocytes.
(**B**) Immunostaining of uHA (blue) and the nucleolar marker fibrillarin (red) on germline nuclei. Early-, mid-, and late-spermatocyte stage were identified by PcGFP (green) localization pattern in wildtype spermatocytes and in *X/O* spermatocytes. Testes from *bam* mutants contain gonial cells, while testes from *aly* mutants contain mostly early spermatocytes.
(**C**) Correspondence of uH2A and RNAPIIS2p signals around the promoters of genes on the *Y* chromosome in wing imaginal discs and in isolated spermatocytes.
(**D**) Correspondence of uH2A and RNAPIIS2p signals across transposon consensus sequences in wing imaginal discs and in isolated spermatocytes.

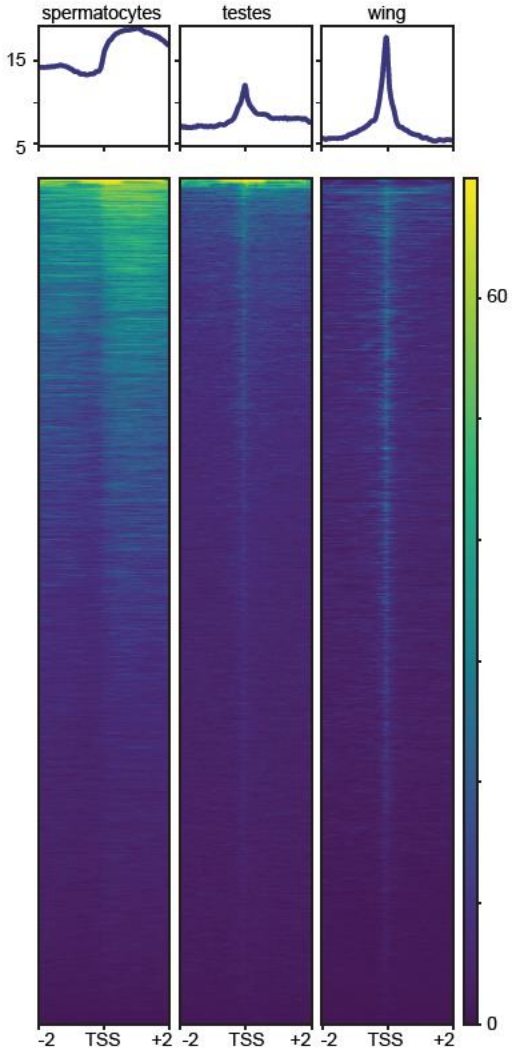
Supplementary Figure S2.1





Supplementary Figure S2.1. Genes with germline-enriched expression in testes.
(**A**) FCA clustering. Number of promoters in each cluster is marked on the left.
(**B**) MA plot and threshold for germline-enriched genes (snRNA-seq data showing gexp/sexp with ≥1 cut-off). Red marks genes in germline-expressed clusters 1-5, and blue marks genes in testes somatic clusters 6-10. The threshold for assigning genes with germline-enriched expression is in green

Supplementary Figure S2.2



Supplementary Figure S2.2. Distribution of elongating RNAPII in spermatocytes and in somatic cells. Heatmap of RNAPIIS2p signal around gene TSSs in spermatocytes, in dissociated testes, and in wing imaginal discs.

Supplementary Figure S2.3

			length		uH2A			H3K9me2			RNAPIIS2p		IgG	FC_RNAPIIS2p
name	type	family	(kb)	wing	testes	sp'cyte	wing	testes	sp'cyte	wing	testes	sp'cyte	sp'cyte	(sp'cytes/wing)
1731	LTR	Copia	4.648	6.6	6.1	8.0	10.1	9.2	9.6	5.9	8.5	8.3	7.1	2.5
Quasimodo	LTR	Gypsy	7.387	6.5	6.2	8.2	10.0	8.9	9.2	5.6	8.4	8.2	7.2	2.6
aurora-element	LTR	Pao	4.263	6.1	5.2	6.9	9.6	8.7	8.7	4.9	7.6	8.1	6.3	3.2
Idefix	LTR	Gypsy	7.411	6.2	5.7	7.2	9.9	8.8	8.9	5.0	7.7	7.8	6.8	2.8
GATE	LTR	Pao	8.507	6.4	5.7	8.2	9.9	8.9	9.4	5.3	8.1	7.8	6.8	2.4
Transpac	LTR	Gypsy	5.249	6.1	6.0	6.5	9.1	7.7	7.2	5.2	6.8	7.4	6.4	2.2
Circe	LTR	Composite	7.45	6.1	5.5	7.0	9.7	8.8	8.9	4.6	6.8	7.3	6.4	2.7
Stalker4	LTR	Gypsy	7.359	6.4	5.8	7.9	9.8	8.9	9.6	5.1	7.4	7.2	6.9	2.1
Stalker	LTR	Gypsy	7.256	6.3	5.8	7.9	9.8	8.9	9.6	5.1	7.4	7.2	6.9	2.1
micropia	LTR	Gypsy	5.461	6.0	5.5	7.0	9.8	8.8	8.9	4.6	7.1	6.9	6.3	2.3
Tabor	LTR	Gypsy	7.345	5.9	5.6	6.8	9.4	8.2	8.4	4.0	6.9	6.8	6.3	2.8
gypsy12	LTR	Gypsy	10.218	5.5	4.7	6.3	9.1	8.3	8.3	3.5	6.6	6.3	5.6	2.7
rooA	LTR	Pao	0.079	3.2	3.0	5.4	6.0	5.6	4.8	2.8	5.6	6.0	5.0	3.2
McClintock	LTR	Gypsy	6.45	5.6	4.9	6.3	9.4	8.4	8.6	3.9	6.7	5.9	6.1	2.0
accord2	LTR	Gypsy	7.65	5.3	4.4	6.0	9.3	8.3	8.4	3.7	6.4	5.9	5.7	2.1
Osvaldo	LTR	Gypsy	1.543	4.4	3.5	4.3	8.1	7.4	6.7	2.7	5.3	5.2	4.3	2.5
TART-B	LINE	telomeric	10.654	3.7	3.0	4.2	8.1	6.5	4.3	3.1	5.7	5.1	4.4	2.0
R1-2	LINE	R1	0.776	3.9	3.0	4.5	8.1	7.0	5.2	0.5	5.2	3.5	3.5	3.0

Supplementary Figure S2.3. Chromatin features across transposon consensus sequences in spermatocytes and in somatic cells.
Selected transposons with increased RNAPIIS2p signal in spermatocytes *versus* wing imaginal discs (log2FC≥2).

Chapter 3

Conclusion

My thesis work has expanded our understanding of the testis as a highly specialized tissue utilizing unique mechanisms of gene regulation. My findings about H3K4me2 smeariness and gene-body enrichment of elongating RNA Pol II adds a new layer to our understanding of the testis-specific chromatin regulation being quite distinct from other tissue-specific genes. I report new findings about the chromatin state of the X chromosome in spermatocytes, which nicely complement recent transcriptomics findings about X-linked gene expression in the male germline. Finally, I observed an intriguing uH2A body on the Y chromosome. This finding of a uH2A body on the active Y adds a new layer to our understanding of H2A ubiquitylation as a versatile histone PTM which plays important roles in chromatin regulation aside from just Polycomb repression. Here I will conclude my findings, and discuss how my thesis work has opened new avenues for future research, proposing some specific ideas for new experiments and studies moving forward.

3.1 Unique facets of gene regulation in the male germline

My thesis work has revealed new ways in which activation of the testis-specific gene program is distinct from other tissue-specific genes, namely: (1) The smeary enrichment of H3K4me2 into gene bodies of testis-specific genes—contrary to the promoter-proximal "peaky" pattern with which H3K4me2 is typically associated; and (2) the lack of a TSS-centered peak for RNA PolII in spermatocytes. These two features of testis-specific genes are likely related, as methylation of H3K4 occurs co-transcriptionally.

These findings are consistent with our understanding that gene regulation in the testis is more specialized than the regulation of other tissue-specific genes. In their recent paper investigating the activation of spermatogenesis genes, Lu *et al.* elegantly summarized this fundamental difference between testis-specific genes, and other tissue-specific genes:

> …a gene that is expressed in different specific places, such as Pitx1 in jaw, pituitary, and pelvis, would need to use different enhancer elements to specify activation in different regions, with the regulatory input from the different enhancers perhaps feeding into a common generic promoter. The situation and constraints may be quite different for terminal differentiation genes that are only expressed in a single tissue (Lu *et al.*, 2020).

The terminal differentiation genes to which Lu *et al.* refer are, in this case, the spermatogenesis genes. Since spermatogenesis genes are fated to be expressed only in the testis, it stands to reason that these genes can dispense with the more complex regulatory logic of tissue-specific enhancers which modulate gene expression according to the tissue context. Similarly, it is reasonable to think that promoter-proximal pausing of RNA PolII—which in the context of other tissue-specific genes acts as an important checkpoint prior to full gene activation—may be another layer of unnecessary complexity in the

relatively simpler scheme of spermatogenesis gene activation. This would facilitate the intensely transcriptionally active state of most testis-specific genes.

Future experiments could provide better support to my proposed model that spermatocytes forego promoter-proximal pausing as a checkpoint to gene activation. For example, I have profiled only RNA PolII ser2phos, *i.e.* the epitope associated with elongating polymerase. Additional experiments profiling total RNA PolII, or distinct subunits of the RNA PolII holoenzyme, or other phosphorylated residues in the C-terminal domain (CTD) of RNA PolII could provide more clarity to my model about promoter-proximal pausing during spermatocyte transcription.

In future studies, it would be valuable to characterize gene activation in more finely resolved cell populations from the testis by chromatin profiling. This could be achieved by a couple of different approaches: (1) single-cell CUT&Tag; or (2) FACS-CUT&Tag on flies with stage-specific enhancers driving fluorescent reporters in other testis cell types aside from just spermatocytes (*e.g.* gonial cells, cyst cells, or perhaps more finely resolved sub-populations of primary spermatocytes). This would enable us to address the question: When, specifically, in the germline do we observe the switch from promoter-proximal pausing of RNA PolII, to the gene body-biased pattern? Is this trend specific just to spermatocytes? Would we also observe this unique pattern of RNA PolII in sorted gonial cells?

Aside from addressing the specific question about RNA PolII, a single-cell CUT&Tag analysis on elongating RNA PolII and/or H3K4me2 would provide a nice complement to single-cell transcriptomics datasets such as the Fly Cell Atlas which I described in Chapters 1 and 2. This would enable us to correspond chromatin with transcription in discrete cell types of the differentiating male germline.

3.2 The X chromosome in the male germline of *D. melanogaster*

Early in my thesis work I had hypothesized that the X chromosome in the *Drosophila* male germline is silenced by repressive chromatin—but I eventually came to reject this hypothesis. Chapter 2 discusses my conclusion that the X chromosome is neither dosage compensated nor silenced in the male germline. My initial (now, rejected) hypothesis that the X is silenced was inspired by at least two ideas: (1) MSCI silences the sex chromosomes in heterogametic males of many other species, including most mammals; and (2) recent transcriptomics studies had reported that X-linked gene expression is diminished in the male germline compared to X-linked gene expression in somatic tissues.

However, these two considerations can be neatly and simply addressed, in respective order: (1) Chromatin regulation in the *Drosophila* male germline is quite distinct from regulation in the male germlines of many other species—*e.g.* there is no meiotic recombination in *Drosophila* males (McKee & Handel, 1993)—and thus it is reasonable to expect that a mechanism like MSCI has not evolved in *Drosophila*; and (2) the lack of dosage compensation is, in itself, sufficient to account for the diminished expression of chromosome X in the germline relative to X-linked expression in somatic tissues, without the need for any concomitant accumulation of repressive chromatin.

A few recent papers have speculated that the X chromosome is silenced in the male germline (Mahadevaraju *et al.*, 2021). Will my thesis work persuade these authors that the X chromosome is not actually silenced in spermatocytes? On the one hand, my profiling provides very clear evidence that the

three major repressive marks which I considered—H3K27me3, uH2A, and H3K9me2—are not enriched on the X chromosome in spermatocytes. However, this topic suggests some possibilities for future studies which might provide even further evidence demonstrating that the X chromosome truly is not silenced. A simple starting point could be to profile some additional repressive chromatin features such as H3K9me3 and HP1, to demonstrate (according to my expectations) that these marks—just like H3K27me3, uH2A, and H3K9me2—do not accumulate on the X chromosome.

Another interesting idea for a future study could be to drive transgene expression of the MSL complex in spermatocytes, since the endogenous MSL is not present or active in spermatocytes (Rastelli & Kuroda, 1998). In theory this transgene MSL expression could restore the expression of X-linked genes in spermatocytes to similar levels of X-expression in somatic tissues. If that were the case, it would support my conclusion that diminishing transcription of X-linked genes in the male germline can simply be attributed to the loss of dosage compensation. This would also support the idea that there are no major repressive chromatin features silencing the X chromosome in spermatocytes.

3.3 The uH2A body on the Y chromosome in spermatocytes

I have observed that the Y chromosome gains a prominent uH2A body in spermatocytes. Given that this uH2A accumulation corresponds with the activation of Y-linked transposons and Y-linked genes in spermatocytes, the simplest explanation is that uH2A facilitates activation of the Y chromosome in the male germline. These findings present several intriguing questions to address in future studies, which I will describe below.

Which chromatin modifier deposits uH2A on the Y chromosome in spermatocytes? My cytology has demonstrated that PRC1 (the Pc subunit, more specifically) does not localize to the Y chromosome in spermatocytes, so the writer is unlikely to be PRC1's ubiquitin ligase subunit *Sce*. Nonetheless, profiling and cytology of *Sce* in spermatocytes would be a useful experiment, to definitively demonstrate that Sce is not associated with the Y chromosome.

An intriguing, though possibly fraught, future study could be a large-scale screen to identify the E3 ubiquitin ligase which establishes the uH2A body in spermatocytes. There is a big caveat to this idea: In the *Drosophila* database FlyBase.org, querying the gene group "E3 ubiquitin ligase" returns 151 hits—a daunting list of potential chromatin modifiers associated with the uH2A body. To be sure, not all of these 151 candidate-uH2A-body-writers would need to be screened initially. For example, many of these ubiquitin ligases are associated with protein degradation pathways and do not have known histone-modifying activity. Conversely, some members such as Ubr1 (Tasaki *et al.*, 2005) have orthologs in mammals which are associated with ubiquitylation of histone H2A, making them high-priority candidates for screening. By carefully assessing what is known about these E3 ubiquitin ligases, a prioritized list of candidates could be assembled—possibly starting with an initial list of just ~10 genes.

Then, to assess whether the candidate is associated with the establishment of the uH2A body in spermatocytes, a stage-specific germline promoter could be used to drive RNA interference (RNAi) against the candidate uH2A writer. CRISPR-mediated gene deletion could be an alternative approach

instead of RNAi. The *bam* promoter (see Chapter 1.6) could be a good promoter to drive a perturbation such as this. Alternatively, the *chif* promoter drives transgene expression in spermatocytes (Bunt *et al.*, 2012), so *chif*-driven RNAi could also be another useful approach to knock-down the expression of chromatin modifiers in spermatocytes.

Following expression of the RNAi transgene against each candidate, the following steps would be taken to determine whether the candidate chromatin modifier is associated with the formation of a uH2A body in spermatocytes: assess the success of the knockdown of the candidate uH2A writer via cytology (and possibly also via quantitative RT-PCR on sorted spermatocytes), looking for diminished/ablated presence of the target of the knockdown; assess via cytology whether the perturbation has disrupted the uH2A body in spermatocytes; and assess whether the perturbation has affected male fertility by crossing the affected males to wild-type females. In addition to all of these steps, it would be interesting to perform CUT&Tag chromatin profiling for other chromatin features—perhaps the active mark H3K4me2—to assess whether ablation of the uH2A body has any unexpected second-order effects, *e.g.* disruption of H3K4me2 at active spermatogenesis genes due to the disruption of the uH2A landscape. A study such as this could be interesting not just for identifying the modifier which establishes the uH2A body, but also addressing the question: Is the uH2A body actually essential for successful differentiation of the male germline, and for fertility more generally?

In Chapter 2 I referred to studies from Bonnet *et al.* and Loubiere *et al.* which may point to an explanation for why uH2A is associated with the highly active Y chromosome. Bonnet *et al.* demonstrated that excess accumulation of uH2A causes increased DNA accessibility *in vivo*, and that uH2A interferes with nucleosomal stacking *in vitro*. Loubiere *et al.* described the association of uH2A with active chromatin and provided evidence that PRC1 is associated with chromatin looping and long-range contact between DNA elements. These studies are helpful for describing other contexts in which uH2A is associated with active/accessible chromatin, but there still looms the question: Why is uH2A associated just with the active Y chromosome, and not also associated with the many other autosomal genes which become highly upregulated in spermatocytes?

The answer may have to do with the unique nature of the Y chromosome, namely: it is heterochromatic; it is enriched with transposons and satellite repeats; and it contains several spermatogenesis genes with "intron gigantism," *i.e.* megabase-sized introns. Next, I will provide a brief primer on a study from Fingerhut *et al*. about intron gigantism, which raises some ideas for future studies to characterize the role of the uH2A body in spermatocytes.

To better understand intron gigantism, Fingerhut *et al.* applied FISH to visualize the expansion of transcripts from several of the megabase-long transcripts from the Y-linked genes *kl-3*, *kl-5*, and *ks-1*, using antisense probes against specific exons or introns of each gene (Fingerhut *et al.*, 2019). This allowed them to describe the temporal pattern of transcription of these gigantic genes and to identify spatially where the transcripts are localized within spermatocyte cells.

The method used by Fingerhut *et al.* to visualize expanding Y chromosome loops and transcripts in spermatocytes could be an interesting approach for characterizing the effects of losing the uH2A body, in the experiments I proposed earlier. If the previous experiments I proposed were successful in ablating

the uH2A body by knocking down the appropriate chromatin modifier, then FISH experiments (analogous to the Fingerhut *et al.* study) could be applied to assess whether loss of the uH2A body causes collapse of the Y chromosome loops or results in degradation of the megabase-sized transcripts coming from those gigantic genes.

References

Abe, H., Yeh, Y. H., Munakata, Y., Ishiguro, K. I., Andreassen, P. R., & Namekawa, S. H. (2022). Active DNA damage response signaling initiates and maintains meiotic sex chromosome inactivation. *Nat Commun*, *13*(1), 7212. https://doi.org/10.1038/s41467-022-34295-5 PMC9705562

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., . . . Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science*, *287*(5461), 2185-2195. https://doi.org/10.1126/science.287.5461.2185

Ahmad, K., & Henikoff, S. (2021). The H3.3K27M oncohistone antagonizes reprogramming in Drosophila. *PLoS Genet*, *17*(7), e1009225. https://doi.org/10.1371/journal.pgen.1009225 PMC8320987

Akhtar, A., & Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell*, *5*(2), 367-375. https://doi.org/10.1016/s1097-2765(00)80431-1

Aloia, L., Di Stefano, B., & Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development*, *140*(12), 2525-2534. https://doi.org/10.1242/dev.091553

An, J. Y., Kim, E. A., Jiang, Y., Zakrzewska, A., Kim, D. E., Lee, M. J., Mook-Jung, I., Zhang, Y., & Kwon, Y. T. (2010). UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. *Proc Natl Acad Sci U S A*, *107*(5), 1912-1917. https://doi.org/10.1073/pnas.0910267107 PMC2836623

Baarends, W. M., Hoogerbrugge, J. W., Roest, H. P., Ooms, M., Vreeburg, J., Hoeijmakers, J. H., & Grootegoed, J. A. (1999). Histone ubiquitination and chromatin remodeling in mouse spermatogenesis. *Dev Biol*, *207*(2), 322-333. https://doi.org/10.1006/dbio.1998.9155

Barbour, H., Daou, S., Hendzel, M., & Affar, E. B. (2020). Polycomb group-mediated histone H2A monoubiquitination in epigenome regulation and nuclear processes. *Nat Commun*, *11*(1), 5947. https://doi.org/10.1038/s41467-020-19722-9 PMC7683540

Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., & White-Cooper, H. (2008). Post-meiotic transcription in Drosophila testes. *Development*, *135*(11), 1897-1902. https://doi.org/10.1242/dev.021949

Barrera, L. A., Vedenko, A., Kurland, J. V., Rogers, J. M., Gisselbrecht, S. S., Rossin, E. J., Woodard, J., Mariani, L., Kock, K. H., Inukai, S., Siggers, T., Shokri, L., Gordân, R., Sahni, N., Cotsapas, C., Hao, T., Yi, S., Kellis, M., Daly, M. J., . . . Bulyk, M. L. (2016). Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science*, *351*(6280), 1450-1454. https://doi.org/10.1126/science.aad2257 PMC4825693

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L., & Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, *120*(2), 169-181. https://doi.org/10.1016/j.cell.2005.01.001

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L., Lander, E. S., Khalil, A. M., Boyar, F. Z., & Driscoll, D. J. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse

Dynamic histone modifications mark sex chromosome inactivation and reactivation during mammalian spermatogenesis. *Cell*, *120*(2), 169-181. https://doi.org/10.1128/mcb.25.16.7120-7136.2005

10.1016/j.cell.2005.01.001 1190250

Blackledge, N. P., Fursova, N. A., Kelley, J. R., Huseyin, M. K., Feldmann, A., & Klose, R. J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell*, *77*(4), 857-874.e859. https://doi.org/10.1016/j.molcel.2019.12.001 PMC7033600

Blackledge, N. P., & Klose, R. J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. *Nat Rev Mol Cell Biol*, *22*(12), 815-833. https://doi.org/10.1038/s41580-021-00398-y PMC7612013

Bonaccorsi, S., Pisano, C., Puoti, F., & Gatti, M. (1988). Y chromosome loops in Drosophila melanogaster. *Genetics*, *120*(4), 1015-1034. https://doi.org/10.1093/genetics/120.4.1015 PMC1203565

Bonnet, J., Boichenko, I., Kalb, R., Le Jeune, M., Maltseva, S., Pieropan, M., Finkl, K., Fierz, B., & Müller, J. (2022). PR-DUB preserves Polycomb repression by preventing excessive accumulation of H2Aub1, an antagonist of chromatin compaction. *Genes Dev*, *36*(19-20), 1046-1061. https://doi.org/10.1101/gad.350014.122 PMC9744231

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*, *109*, 21.29.21-21.29.29. https://doi.org/10.1002/0471142727.mb2129s109 PMC4374986

Bunt, S. M., Monk, A. C., Siddall, N. A., Johnston, N. L., & Hime, G. R. (2012). GAL4 enhancer traps that can be used to drive gene expression in developing Drosophila spermatocytes. *Genesis*, *50*(12), 914-920. https://doi.org/10.1002/dvg.22341

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell*, *36*(4), 541-546. https://doi.org/10.1016/j.molcel.2009.10.019 PMC3232742

Chang, C. H., & Larracuente, A. M. (2019). Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. *Genetics*, *211*(1), 333-348. https://doi.org/10.1534/genetics.118.301765 PMC6325706

Chang, C. H., Mejia Natividad, I., & Malik, H. S. (2023). Expansion and loss of sperm nuclear basic protein genes in Drosophila correspond with genetic conflicts between sex chromosomes. *Elife*, *12*. https://doi.org/10.7554/eLife.85249 PMC9917458

Chen, X., Lu, C., Morillo Prado, J. R., Eun, S. H., & Fuller, M. T. (2011). Sequential changes at differentiation gene promoters as they become active in a stem cell lineage. *Development*, *138*(12), 2441-2450. https://doi.org/10.1242/dev.056572 PMC3100706

Conrad, T., & Akhtar, A. (2012). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet*, *13*(2), 123-134. https://doi.org/10.1038/nrg3124

Core, L., & Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev*, *33*(15-16), 960-982. https://doi.org/10.1101/gad.325142.119 PMC6672056

de Cuevas, M., & Matunis, E. L. (2011). The stem cell niche: lessons from the Drosophila testis. *Development*, *138*(14), 2861-2869. https://doi.org/10.1242/dev.056242 PMC3119301

Demarco, R. S., Eikenes Å, H., Haglund, K., & Jones, D. L. (2014). Investigating spermatogenesis in Drosophila melanogaster. *Methods*, *68*(1), 218-227. https://doi.org/10.1016/j.ymeth.2014.04.020 PMC4128239

Di Cara, F., Morra, R., Cavaliere, D., Sorrentino, A., De Simone, A., Polito, C. L., & Digilio, A. F. (2006). Structure and expression of a novel gene family showing male germline specific expression in Drosophila melanogaster. *Insect Mol Biol*, *15*(6), 813-822. https://doi.org/10.1111/j.1365-2583.2006.00688.x

El-Sharnouby, S., Redhouse, J., & White, R. A. (2013). Genome-wide and cell-specific epigenetic analysis challenges the role of polycomb in Drosophila spermatogenesis. *PLoS Genet*, *9*(10), e1003842. https://doi.org/10.1371/journal.pgen.1003842 PMC3798269

Ernst, C., Eling, N., Martinez-Jimenez, C. P., Marioni, J. C., & Odom, D. T. (2019). Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nat Commun*, *10*(1), 1251. https://doi.org/10.1038/s41467-019-09182-1 PMC6424977

Eun, S. H., Feng, L., Cedeno-Rosario, L., Gan, Q., Wei, G., Cui, K., Zhao, K., & Chen, X. (2017). Polycomb Group Gene E(z) Is Required for Spermatogonial Dedifferentiation in Drosophila Adult Testis. *J Mol Biol*, *429*(13), 2030-2041. https://doi.org/10.1016/j.jmb.2017.04.012 PMC5516936

Fabian, L., & Brill, J. A. (2012). Drosophila spermiogenesis: Big things come from little packages. *Spermatogenesis*, *2*(3), 197-212. https://doi.org/10.4161/spmg.21798 PMC3469442

Fingerhut, J. M., Moran, J. V., & Yamashita, Y. M. (2019). Satellite DNA-containing gigantic introns in a unique gene expression program during Drosophila spermatogenesis. *PLoS Genet*, *15*(5), e1008028. https://doi.org/10.1371/journal.pgen.1008028 PMC6508621

Freiman, R. N. (2009). Specific variants of general transcription factors regulate germ cell development in diverse organisms. *Biochim Biophys Acta*, *1789*(3), 161-166. https://doi.org/10.1016/j.bbagrm.2009.01.005 PMC2686221

Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I., & Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol*, *19*(3), 192-206. https://doi.org/10.1038/nrm.2017.94 PMC5897046

Gorfinkiel, N., Fanti, L., Melgar, T., García, E., Pimpinelli, S., Guerrero, I., & Vidal, M. (2004). The Drosophila Polycomb group gene Sex combs extra encodes the ortholog of mammalian Ring1 proteins. *Mech Dev*, *121*(5), 449-462. https://doi.org/10.1016/j.mod.2004.03.019

Green, M. M. (2010). 2010: A century of Drosophila genetics through the prism of the white gene. *Genetics*, *184*(1), 3-7. https://doi.org/10.1534/genetics.109.110015 PMC2815926

Grossniklaus, U., & Paro, R. (2014). Transcriptional silencing by polycomb-group proteins. *Cold Spring Harb Perspect Biol*, *6*(11), a019331. https://doi.org/10.1101/cshperspect.a019331 PMC4413232

Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S., & Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *Elife*, *9*. https://doi.org/10.7554/eLife.63274 PMC7721439

Henikoff, S., & Shilatifard, A. (2011). Histone modification: cause or cog? *Trends Genet*, *27*(10), 389-396. https://doi.org/10.1016/j.tig.2011.06.006

Hennig, W., & Weyrich, A. (2013). Histone modifications in the male germ line of Drosophila. *BMC Dev Biol*, *13*, 7. https://doi.org/10.1186/1471-213x-13-7 PMC3602674

Hiller, M., Chen, X., Pringle, M. J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T. Y., Marino, S., & Fuller, M. T. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development*, *131*(21), 5297-5308. https://doi.org/10.1242/dev.01314

Hubert, K. A., & Wellik, D. M. (2023). Hox genes in development and beyond. *Development*, *150*(1). https://doi.org/10.1242/dev.192476 PMC10216783

Huynh, K. D., & Lee, J. T. (2005). X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. *Nat Rev Genet*, *6*(5), 410-418. https://doi.org/10.1038/nrg1604

Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Exp Mol Med*, *49*(4), e324. https://doi.org/10.1038/emm.2017.11 PMC6130214

Janssens, D. H., Otto, D. J., Meers, M. P., Setty, M., Ahmad, K., & Henikoff, S. (2022). CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells. *Genome Biol*, *23*(1), 81. https://doi.org/10.1186/s13059-022-02642-w PMC8928696

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*, *10*(1), 1930. https://doi.org/10.1038/s41467-019-09982-5 PMC6488672

Khalil, A. M., Boyar, F. Z., & Driscoll, D. J. (2004). Dynamic histone modifications mark sex chromosome inactivation and reactivation during mammalian spermatogenesis. *Proceedings of the National Academy of Sciences*, *101*(47), 16583-16587. https://doi.org/10.1073/pnas.0406325101

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693-705. https://doi.org/10.1016/j.cell.2007.02.005

Laktionov, P. P., Maksimov, D. A., Romanov, S. E., Antoshina, P. A., Posukh, O. V., White-Cooper, H., Koryakov, D. E., & Belyakin, S. N. (2018). Genome-wide analysis of gene regulation mechanisms during Drosophila spermatogenesis. *Epigenetics Chromatin*, *11*(1), 14. https://doi.org/10.1186/s13072-018-0183-3 PMC5879934

Larsson, J., & Meller, V. H. (2006). Dosage compensation, the origin and the afterlife of sex chromosomes. *Chromosome Res*, *14*(4), 417-431. https://doi.org/10.1007/s10577-006-1064-3

Lawlor, M. A., Cao, W., & Ellison, C. E. (2021). A transposon expression burst accompanies the activation of Y-chromosome fertility genes during Drosophila spermatogenesis. *Nat Commun*, *12*(1), 6854. https://doi.org/10.1038/s41467-021-27136-4 PMC8617248

Li, H., Janssens, J., De Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W., David, F. P. A., Brbić, M., Spanier, K., Leskovec, J., McLaughlin, C. N., Xie, Q., Jones, R. C., Brueckner, K., Shim, J., Tattikota, S. G., Schnorrer, F., Rust, K., . . . Zinzen, R. P. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science*, *375*(6584), eabk2432. https://doi.org/10.1126/science.abk2432 PMC8944923

Lifschytz, E., & Lindsley, D. L. (1972). The role of X-chromosome inactivation during spermatogenesis (Drosophila-allocycly-chromosome evolution-male sterility-dosage compensation). *Proc Natl Acad Sci U S A*, *69*(1), 182-186. https://doi.org/10.1073/pnas.69.1.182 PMC427571

Lim, C., Tarayrah, L., & Chen, X. (2012). Transcriptional regulation during Drosophila spermatogenesis. *Spermatogenesis*, *2*(3), 158-166. https://doi.org/10.4161/spmg.21775 PMC3469439

Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A. M., & Cavalli, G. (2020). Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv*, *6*(2), eaax4001. https://doi.org/10.1126/sciadv.aax4001 PMC6954061

Lu, C., & Fuller, M. T. (2015). Recruitment of Mediator Complex by Cell Type and Stage-Specific Factors Required for Tissue-Specific TAF Dependent Gene Activation in an Adult Stem Cell Lineage. *PLoS Genet*, *11*(12), e1005701. https://doi.org/10.1371/journal.pgen.1005701 PMC4666660

Lu, D., Sin, H. S., Lu, C., & Fuller, M. T. (2020). Developmental regulation of cell type-specific transcription by novel promoter-proximal sequence elements. *Genes Dev*, *34*(9-10), 663-677. https://doi.org/10.1101/gad.335331.119 PMC7197356

Luger, K., Dechassa, M. L., & Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol*, *13*(7), 436-447. https://doi.org/10.1038/nrm3382 PMC3408961

Mahadevaraju, S., Fear, J. M., Akeju, M., Galletta, B. J., Pinheiro, M., Avelino, C. C., Cabral-de-Mello, D. C., Conlon, K., Dell'Orso, S., Demere, Z., Mansuria, K., Mendonça, C. A., Palacios-Gimenez, O. M., Ross, E., Savery, M., Yu, K., Smith, H. E., Sartorelli, V., Yang, H., . . . Oliver, B. (2021). Dynamic sex chromosome expression in Drosophila male germ cells. *Nat Commun*, *12*(1), 892. https://doi.org/10.1038/s41467-021-20897-y PMC7873209

McKearin, D., & Ohlstein, B. (1995). A role for the Drosophila bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development*, *121*(9), 2937-2947. https://doi.org/10.1242/dev.121.9.2937

McKee, B. D., & Handel, M. A. (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma*, *102*(2), 71-80. https://doi.org/10.1007/bf00356023

Morgan, T. H. (1910). SEX LIMITED INHERITANCE IN DROSOPHILA. *Science*, *32*(812), 120-122. https://doi.org/10.1126/science.32.812.120

Mu, W., Starmer, J., Fedoriw, A. M., Yee, D., & Magnuson, T. (2014). Repression of the soma-specific transcriptome by Polycomb-repressive complex 2 promotes male germ cell development. *Genes Dev*, *28*(18), 2056-2069. https://doi.org/10.1101/gad.246124.114 PMC4173155

Muniz, L., Nicolas, E., & Trouche, D. (2021). RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *Embo j*, *40*(15), e105740. https://doi.org/10.15252/embj.2020105740 PMC8327950

Nelson, A. C., & Wardle, F. C. (2013). Conserved non-coding elements and cis regulation: actions speak louder than words. *Development*, *140*(7), 1385-1395. https://doi.org/10.1242/dev.084459

O'Kane, C. J., & Gehring, W. J. (1987). Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A*, *84*(24), 9123-9127. https://doi.org/10.1073/pnas.84.24.9123 PMC299704

Oluwadare, O., Highsmith, M., & Cheng, J. (2019). An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biol Proced Online*, *21*, 7. https://doi.org/10.1186/s12575-019-0094-0 PMC6482566

Panigrahi, A., & O'Malley, B. W. (2021). Mechanisms of enhancer action: the known and the unknown. *Genome Biol*, *22*(1), 108. https://doi.org/10.1186/s13059-021-02322-1 PMC8051032

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, *10*(10), 669-680. https://doi.org/10.1038/nrg2641 PMC3191340

Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Res Notes*, *12*(1), 106. https://doi.org/10.1186/s13104-019-4137-z PMC6391780

Popay, T. M., & Dixon, J. R. (2022). Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. *J Biol Chem*, *298*(8), 102117. https://doi.org/10.1016/j.jbc.2022.102117 PMC9283939

Rastelli, L., & Kuroda, M. I. (1998). An analysis of maleless and histone H4 acetylation in Drosophila melanogaster spermatogenesis. *Mech Dev*, *71*(1-2), 107-117. https://doi.org/10.1016/s0925-4773(98)00009-4

Rathke, C., Baarends, W. M., Jayaramaiah-Raja, S., Bartkuhn, M., Renkawitz, R., & Renkawitz-Pohl, R. (2007). Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in Drosophila. *J Cell Sci*, *120*(Pt 9), 1689-1700. https://doi.org/10.1242/jcs.004663

Raz, A. A., Vida, G. S., Stern, S. R., Mahadevaraju, S., Fingerhut, J. M., Viveiros, J. M., Pal, S., Grey, J. R., Grace, M. R., Berry, C. W., Li, H., Janssens, J., Saelens, W., Shao, Z., Hu, C., Yamashita, Y. M., Przytycka, T., Oliver, B., Brill, J. A., . . . Fuller, M. T. (2023). Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of Drosophila testes. *Elife*, *12*. https://doi.org/10.7554/eLife.82201 PMC9934865

Schuettengruber, B., Bourbon, H. M., Di Croce, L., & Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, *171*(1), 34-57. https://doi.org/10.1016/j.cell.2017.08.002

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, *6*. https://doi.org/10.7554/eLife.21856 PMC5310842

Sloutskin, A., Shir-Shapira, H., Freiman, R. N., & Juven-Gershon, T. (2021). The Core Promoter Is a Regulatory Hub for Developmental Gene Expression. *Front Cell Dev Biol*, *9*, 666508. https://doi.org/10.3389/fcell.2021.666508 PMC8461331

Solari, A. J. (1974). The behavior of the XY pair in mammals. *Int Rev Cytol*, *38*(0), 273-317. https://doi.org/10.1016/s0074-7696(08)60928-6

Song, L., & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, *2010*(2), pdb.prot5384. https://doi.org/10.1101/pdb.prot5384 PMC3627383

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41-45. https://doi.org/10.1038/47412

Tasaki, T., Mulder, L. C., Iwamatsu, A., Lee, M. J., Davydov, I. V., Varshavsky, A., Muesing, M., & Kwon, Y. T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol*, *25*(16), 7120-7136. https://doi.org/10.1128/mcb.25.16.7120-7136.2005 PMC1190250

Thomas, M. C., & Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol*, *41*(3), 105-178. https://doi.org/10.1080/10409230600648736

Turner, J. M. (2015). Meiotic Silencing in Mammals. *Annu Rev Genet*, *49*, 395-412. https://doi.org/10.1146/annurev-genet-112414-055145

Vibranovski, M. D. (2014). Meiotic sex chromosome inactivation in Drosophila. *J Genomics*, *2*, 104-117. https://doi.org/10.7150/jgen.8178 PMC4105432

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, *431*(7010), 873-878. https://doi.org/10.1038/nature02985

Weintraub, H., & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science*, *193*(4256), 848-856. https://doi.org/10.1126/science.948749

White-Cooper, H. (2010). Molecular mechanisms of gene regulation during Drosophila spermatogenesis. *Reproduction*, *139*(1), 11-21. https://doi.org/10.1530/rep-09-0083

White-Cooper, H., & Bausek, N. (2010). Evolution and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci*, *365*(1546), 1465-1480. https://doi.org/10.1098/rstb.2009.0323 PMC2871925

White-Cooper, H., D'Alessio, J. A., Wright, K. J., & Tjian, R. (2010). Molecular mechanisms of gene regulation during Drosophila spermatogenesis

Shifting players and paradigms in cell-specific transcription. *Reproduction*, *139*(1), 11-21. https://doi.org/10.1186/gb-2010-11-4-r42

10.1530/rep-09-0083 2884545

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. *Elife*, *8*. https://doi.org/10.7554/eLife.47138 PMC6697446

Witt, E., Shao, Z., Hu, C., Krause, H. M., & Zhao, L. (2021). Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in Drosophila testis. *PLoS Genet*, *17*(8), e1009728. https://doi.org/10.1371/journal.pgen.1009728 PMC8396764

Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D., Sarthy, J. F., Gottardo, R., Ahmad, K., Henikoff, S., & Patel, A. P. (2021). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol*, *39*(7), 819-824. https://doi.org/10.1038/s41587-021-00865-z PMC8277750