Title
The chromatin landscape of Drosophila: Heterochromatin differences between species, sexes, and ages

Author
Brown, Emily Jordan

Publication Date
2016

Peer reviewed|Thesis/dissertation

The chromatin landscape of Drosophila: Heterochromatin differences between species, sexes, and ages

By

Emily Jordan Brown

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Doris Bachtrog, Chair
Professor Michael Eisen
Professor Lior Pachter
Professor Gary Karpen

Fall 2016

Abstract

The chromatin landscape of Drosophila: Heterochromatin differences between species, sexes, and ages

By

Emily Jordan Brown

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Doris Bachtrog, Chair

Chromatin is composed of DNA and a variety of modified histones and non-histone proteins, and can be most broadly characterized by the gene-rich and repeat-poor euchromatin and the gene-poor and repeat-rich heterochromatin. Genome-wide profiling of chromatin components provides a method of generating a functional annotation of the underlying DNA sequences, as groups of correlated histone modifications are associated with both euchromatin and heterochromatin, as well as more specific functions such as active transcription or polycomb-mediated repression. Although there has been much progress towards understanding the general hallmarks of different chromatin functions, there have been many fewer efforts to characterize on a broad scale how different types of chromatin differ between species, sexes, or individuals of different ages.

In this dissertation, I investigate whether differences in sex chromosome content contributes to genome-wide differences in the chromatin landscape across species with different numbers of X chromosomes, between males and females, and between old and young individuals. Sex chromosomes have a unique chromatin structure compared to autosomes. The single male X chromosome recruits the dosage compensation complex and becomes hyper-acetylated, resulting in an approximate 2-fold increase of transcription. Females, however, have two copies of the X chromosome and do not recruit the dosage compensation complex to the X chromosome. In contrast to the hyper-transcription of the X chromosome in males, the male-limited Y chromosome is transcriptionally silenced via heterochromatin formation, as it is gene-poor and repeat-rich.

The Drosophila Y chromosome is known to harbor variation that effects position effect variegation (the ability of spreading heterochromatin to induce partial silencing of reporter genes in some cells, resulting in mosaic expression patterns). However, previous studies have not assayed the Y chromosome's effect on heterochromatin integrity genome-wide, nor have they directly assayed the role of the Y chromosome in generating differences in heterochromatin composition observed between males and females. Here, I use genome-wide profiles of

heterochromatic histone modifications in XO and XYY males, and XXY females, to assess the effect of the Y chromosome on genome-wide heterochromatin.

The chronic deterioration of chromatin structure has been implicated in aging, and an overall loss of heterochromatin has been observed in many old animals. Males and females differ in both their average lifespan as well as their total amount of heterochromatic sequences, due to the presence of the large heterochromatic Y chromosome in males. I compare lifespans, genome-wide heterochromatin profiles, and expression of repetitive elements during aging in males and females, as well as XO and XYY males, and XXY females, to interrogate whether the Y chromosome contributes to differences in lifespan and loss of chromatin organization between the sexes.

Acknowledgments

First I would like to thank my partner, Chris Ellison, for all of his support over the past 5 years. I am so fortunate to have a partner who is so completely empathetic to the challenges of academic science.

I would also like to thank my family for all of their words of encouragement, willingness to listen, and overall support.

To the many undergraduates who contributed to this work, I cannot thank you enough for your time and commitment. Hua Chen, Claire Liu, John Perez, Jana Guenther, and Jin Xing, thank you so much!

I want to thank my advisor, Doris Bachtrog, and the other members of my dissertation committee, Mike Eisen, Lior Pachter, and Gary Karpen, for years of advice and feedback on this work.

Finally, I want to thank the many current and former members of the Bachtrog lab, who taught me so much about bioinformatics and genomics, and who made the lab such a fun place to work.

Chapter 1

The chromatin landscape of *Drosophila*: comparison between species, sexes and chromosomes

Emily J. Brown & Doris Bachtrog
Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720

The chromatin landscape is key for gene regulation, but little is known about how it differs between the sexes or between species. Here, we study the sex-specific chromatin landscape of *Drosophila miranda*, a species with young sex chromosomes, and compare it to the model organism *D. melanogaster*. We analyze six histone modifications in male and female larvae of *D. miranda* (H3K4me1, H3K4me3, H3K36me3, H4K16ac, H3K27me3 and H3K9me2), and define seven biologically meaningful chromatin states that show different enrichment for transcribed and silent genes, repetitive elements, housekeeping and tissue-specific genes. The genome-wide distribution of both active and repressive chromatin states differs between males and females. In males, active chromatin is enriched on the X, relative to females, due to dosage compensation of the hemizygous X. Furthermore, a smaller fraction of the euchromatic portion of the genome is in a repressive chromatin state in males relative to females. However, sex-specific chromatin states appear not to explain sex-biased expression of genes. Overall, conservation of chromatin states between male and female *D. miranda* is comparable to conservation between *D. miranda* and *D. melanogaster*, which diverged over 30MY ago and lacks the secondary sex chromosomes of *D. miranda*. Active chromatin states are more highly conserved across species, while heterochromatin shows very low levels of conservation. Divergence in chromatin profiles contributes to expression divergence between species, with about 26% of genes in different chromatin states in the two species showing species-specific or species-biased expression, an enrichment of approximately 3-fold over null expectation. Our data are consistent with the hypothesis that heteromorphic sex chromosomes in males (that is, a hypertranscribed X and an inactivated Y) may contribute to globally redistribute active and repressive chromatin marks between chromosomes and sexes.

Introduction.

In the past several years, chromatin structure has been identified as a major component regulating gene expression (Schulze and Wallrath 2007). Chromatin is composed of DNA and a variety of modified histones and non-histone proteins, and genome-wide profiling of chromatin components has provided a rich functional annotation of the underlying DNA sequences (Filion et al. 2010; Kharchenko et al. 2011; The ENCODE Project Consortium 2012). Groups of correlated histone modifications (chromatin states) were found to be associated with specific biological functions, such as heterochromatic regions, active transcription or polycomb-mediated repression (Ernst et al. 2011; Riddle et al. 2011; Yin et al. 2011). Previous studies in *Drosophila melanogaster* have investigated chromatin states in various cell lines (Filion et al. 2010; Kharchenko et al. 2011) and mixed-sex adults (Yin et

al. 2011), and have greatly increased our understanding of the functional significance of chromatin marks. However, little is known about how chromatin structure varies across species or between sexes within a species. In particular, levels of gene expression vary considerably among species, and gene expression divergence has been implicated as an important factor driving adaptive divergence between species (Meiklejohn et al. 2003; McManus et al. 2010; Wittkopp and Kalay 2011), but it is generally unclear how expression divergence correlates with changes in chromatin structure (Cain et al. 2011).
Not only species, but also the two sexes within a species often vary considerably in which genes are expressed at what level (Khil et al. 2004; Zhang et al. 2007). Differential expression of genes between sexes (sex-biased gene expression) contributes to a variety of physiological, morphological and behavioral traits that differ between males and females (Parsch and Ellegren 2013). Recent genome-wide expression profiling studies have demonstrated that sex-related differences in gene expression are extensive across a range of taxa, including insects, nematodes, birds and mammals (Ellegren and Parsch 2007; Parsch and Ellegren 2013). How these differences in sex-biased gene expression are achieved on the cellular level and the involvement of chromatin structure differences are, for the most part, unknown, and no comprehensive analysis contrasting chromatin between males and females has yet been performed.
Finally, even chromosomes within an individual can show systematic differences in how genes are transcriptionally regulated. In particular, sex chromosomes often show unusual patterns of gene expression, with the Y being transcriptionally repressed in Drosophila, while the X is hypertranscribed in males. Both of these transcriptional modifications are mediated by changes to the chromatin landscape (Straub and Becker 2007; Girton and Johansen 2008; Gelbart and Kuroda 2009; Lemos et al. 2010). Sex chromosomes are derived from autosomes, but their dynamics are governed by unique evolutionary and functional pressures (Bachtrog 2006). The male-limited Y chromosome degenerates, that is, it looses most of its ancestral genes, accumulates repetitive junk DNA, and evolves a heterochromatic appearance (Charlesworth and Charlesworth 2000; Bachtrog 2013). Its former homolog, the X chromosome, acquires mechanisms to compensate for gene loss associated with Y degeneration, and evolves hypertranscription in Drosophila (i.e. dosage compensation; Gelbart and Kuroda 2009; Vicoso and Bachtrog 2009). Both heterochromatin formation and dosage compensation are accompanied by global changes in chromatin structure (Steinemann and Steinemann 2005; Straub and Becker 2007), but little is known about how these epigenetic modifications are acquired on a differentiating pair of proto-sex chromosomes (Zhou et al. 2013).

Here, we investigate the sex-specific chromatin landscape of *D. miranda*, a species that diverged from *D. melanogaster* about 30 MY ago and contains three sex chromosomes of different ages: XL, XR and the neo-X (**Figure 1**). XL is homologous to the X chromosome of *D. melanogaster* and has been segregating as a sex chromosome for over 60 MY (Vicoso and Bachtrog 2013). XR – which corresponds to chr 3L in *D. melanogaster* – became sex-linked about 15 MY ago in an ancestor of *D. miranda*, but has evolved most features characteristic of the ancestral X, including chromosome-wide dosage compensation and complete degeneration of its former homolog (Alekseyenko et al. 2013; Zhou et al. 2013). Finally, the neo-sex chromosomes of *D. miranda* (homologous to chr 2R of *D. melanogaster*) only formed about 1.5 MY ago, and harbor many characteristics that are intermediate between

ordinary autosomes and heteromorphic sex chromosomes. In particular, the neo-Y is undergoing genome-wide degeneration; over 1000 genes have acquired stop codons or frameshift mutations (Bachtrog 2005; Bachtrog et al. 2008; Zhou and Bachtrog 2012) and the neo-Y is evolving a heterochromatic configuration (Zhou et al. 2013). In response, the neo-X has begun to evolve partial dosage compensation (Alekseyenko et al. 2013). This species is therefore uniquely suited to study the evolutionary dynamics of chromatin associated with sex chromosome differentiation.

We set out to make comparisons of the chromatin landscape in Drosophila at three different levels: species, sexes and chromosomes. In particular, we first compare the chromatin landscape between autosomes and sex chromosomes of different ages, taking advantage of the unique karyotype of *D. miranda.* Second, we contrast the chromatin landscape of male and female *D. miranda*, to characterize sex-specific differences in chromatin. And finally, we compare chromatin states of *D. miranda* to the well-characterized chromatin landscape of *D. melanogaster* in order to identify evolutionary conservation and turn-over of chromatin marks within Drosophila.

Results & Discussion.

(1) The chromatin landscape of *Drosophila miranda & D. melanogaster*

To explore chromatin states in *Drosophila miranda*, we used ChIP-seq profiles from male and female 3rd instar larvae for six different histone modifications (Alekseyenko et al. 2013; Zhou et al. 2013): H3K36me3 (associated with transcription elongation); H4K16ac (associated with dosage compensation and transcribed regions); H3K4me3 (associated with active promoters and transcription start sites); H3K4me1 (associated with enhancers and introns); H3K27me3 (associated with polycomb-repressed regions) and H3K9me2 (associated with constitutive heterochromatin). These six histone modifications characterize a variety of chromatin states found in Drosophila, and provide a broad functional annotation of the *D. miranda* genome. Studies in *D. melanogaster* have assayed 18 histone modifications and applied a multivariate hidden Markov model that uses combinatorial patterns of chromatin marks to assign chromatin states to regions of the genome (Kharchenko et al. 2011). Nine major chromatin states (i.e. groups of correlated histone modifications) were identified in *D. melanogaster*. Using our subset of six histone modifications, we were able to define 7 biologically meaningful chromatin states in *D. melanogaster* using a multivariate hidden Markov model (Ernst and Kellis 2012). Comparison of our 7-state model to the 9-state model defined by Kharchenko et al. (2011) revealed that there is a generally good correspondence between the models in both their emissions parameters and their distribution throughout the genome (**Figure S1-S3**), even though the models were learned with different datasets from different tissues (SL2 cells versus 3rd instar larvae). This suggests that these chromatin states are robust and reflect true underlying biological combinations of histone modifications. Since we were able to directly compare the model learned from our subset of histone modifications to the model learned from a more complete panel of histone marks in *D. melanogaster*, we applied the same 7-state chromatin model to the *D. miranda* data (note that a chromatin state model recovered from *D. miranda* ChIP-seq profiles is very similar to the *D. melanogaster* 7-state model, **Figure S4**). Our map is derived from all cell types in the larvae of *D. melanogaster*

and *D. miranda*, weighted by their natural abundance, but many chromatin features are conserved in Drosophila cell lines, and across major developmental stages (Filion et al. 2010; Kharchenko et al. 2011; Yin et al. 2011). The 7 states show distinct biological enrichments: four active states (one each that corresponds to active promoters, transcription elongation, regulatory elements and introns, and spreading of the dosage compensation complex) as well as three silent or repressive states (one each for polycomb-mediated silencing, constitutive heterochromatin, and a background null state).

Previous studies have observed that there is a strong correlation between sequencing depth and amount of binding detected in ChIP-seq experiments (Kharchenko et al. 2008). In order to compare binding across multiple samples (male and female *D. miranda*, and *D. melanogaster*) with equal ability to detect peaks, we developed a normalization strategy that first ensured equal numbers of reads for each mark across all samples (**Figure S5**). Additionally, because we expect binding of certain marks (e.g. H4K16ac) to be highly enriched on sex chromosomes, we called binding events for autosomes independently of the sex chromosomes to detect lower-level binding events on the autosomes; this strategy is more conservative in estimating differences in histone enrichment profiles on autosomes between the sexes (**Figure S6**). To validate our normalization strategy, we performed ChIP-qPCR for H3K9me2 and H4K16ac, the two histone modifications that showed the greatest difference between the two sexes in absolute quantity, as detected by Western blot (**Figure S7**). Using a subset of targets from both autosomes and the X chromosomes, we confirmed that targets we called bound show a significant increase over the input compared to regions we called unbound for both marks (**Figure S8a,b**). Additionally, regions that were defined by ChIP-seq as being bound in both sexes showed a significantly different ratio of enrichment in females vs. males as compared to regions that were bound in only one sex (**Figure S8c,d**), suggesting that our normalization strategy captures real binding differences between the two sexes. This is further supported by replicate ChIP-seq experiments for H4K16ac and H3K9me3 in male and female *D. miranda* larvae (**Figure S9**), the two histone marks which show the most dramatic difference between sexes. The other four histone modifications are strongly correlated between male and female samples (**Figure S10**), and ChIP efficiencies are similar between sexes and species (**Figures S11, S12**).

Figure 2 gives an overview of the chromatin states identified in *D. miranda* and their genomic distribution. States 1 and 2 are enriched for histone marks of active transcription, characterized by the transcriptional elongation signature H3K36me3 and enriched at exons and 5' and 3' flanking regions of genes. State 1 shows strong enrichment within 5' UTRs, consistent with its enrichment for H3K4me3, a mark associated with active transcription start sites (TSS) and promoters. State 2, enriched primarily for the transcription elongation mark H3K36me3, is highly enriched within 3' UTRs. State 3, which is associated with H3K4me1 but not the other active histone modifications, is enriched in introns and 5' UTRs, consistent with the association of H3K4me1 with enhancers. State 4 is distinguished by high enrichment of H4K16ac only, and enrichment of this histone mark on the Drosophila male X chromosome is considered a signature of dosage compensation (Straub and Becker 2007; Gelbart and Kuroda 2009). States 5 and 6 are repressive states found in silenced genes, intergenic regions, and introns. State 5 (H3K27me3) corresponds to regions

of polycomb-mediated repression while state 6 (H3K9me2) represents constitutive heterochromatin. State 7 corresponds to silent domains that are not enriched for any of the histone modifications assayed, and is mainly found in intergenic regions and introns (Kharchenko et al. 2011). States 1-4 represent "active" chromatin states, while states 5-7 represent "repressed" chromatin states. Most transcribed genes are in active chromatin states (most frequently states 1 and 2), while most silent genes are in repressed chromatin states (**Figure 2c**). Among expressed genes, most housekeeping genes (as defined by their breadth of expression using the tissue-specificity index tau; Larracuente et al. 2008; **Figure S13**) are in states 1 or 2 (the H3K36me3 associated states), while most tissue-specific genes are characterized by repressive or silent chromatin states, even though they have detectable expression in 3rd instar larvae (**Figure 2c**). Thus, while states associated with H3K36me3 (states 1 and 2) are characteristic of genes that are broadly expressed, genes with more narrow expression patterns, even if transcribed, are not enriched for these chromatin states. Lower power to detect H3K36me3 binding of tissue-specific genes expressed only in a subset of cells in larvae could contribute to this effect, but we also see a deficiency in states associated with H3K36me3 in tissue-specific but expressed genes in SL2 cells (**Figure S14**). This corroborates a recent study in *D. melanogaster* that found that chromatin of housekeeping genes is enriched for the transcriptional elongation mark H3K36me3, but tissue-specific genes have a distinct chromatin structure that does not show enrichment for this histone modification (Filion et al. 2010).

States 1 and 2, which are both associated with H3K36me3, have the highest gene content of any of the states, with approximately 36% of state 1 (35% in females, 36% in males) and 48% of state 2 (48% in females, 48% in males) comprised of coding regions (**Figure S15a**). State 4, which is characterized by high levels of H4K16ac, also has a strong enrichment for coding regions, especially in females, although females have a smaller fraction of their genome in this state. State 6 (constitutive heterochromatin) has the highest transposable element content, with approximately 20% of state 6 (19% in females, 20% in males) composed of transposable elements (**Figure S15b**). Median expression levels of genes in different states also vary, with genes in the four active states having higher expression than genes in the three repressive states (**Figure S16**). Tissue-specificity, as measured by the tissue-specificity index tau, also differs between genes across states; genes in active chromatin states tend to be broadly expressed (indicative of housekeeping genes) and genes in repressive chromatin states show more tissue-specific expression (**Figure S17**).

(2) The sex- and chromosome-specific chromatin landscape of *D. miranda*

To explore the sex-specific chromatin landscape in *D. miranda*, we compared chromatin states between male and female larvae. In principle, differences in the chromatin structure between males and females could result from various sources: the hemizygous X chromosome is dosage compensated in males, which is accomplished by changes in the chromatin structure (Straub and Becker 2007; Gelbart and Kuroda 2009); males contain a large heterochromatic Y chromosome which could alter the stoichiometric balance of heterochromatin/euchromatin between the sexes (Weiler and Wakimoto 1995; Deng et al. 2009; Lemos et al. 2010; Zhou et al. 2012); and many genes show sex-biased expression which could be associated with sex-specific chromatin modifications**. Figure 3** shows a genome-wide karyotype view of the chromatin domains derived from female and male *D.*

miranda larvae. Several prominent chromatin organization features that differ between sexes are apparent, most noticeably the relative enrichment of active chromatin states (in particular, state 4) on the male X chromosomes, a signature of dosage compensation, but also more repressive chromatin states (both states 5 and 6) in the assembled, euchromatic part of the genome in females.

The hyper-transcribed X in males. Males and females differ in their number of X chromosomes, and the X chromosome in male *Drosophila* is hypertranscribed (i.e. dosage compensated). Dosage compensation is acquired through histone modifications (Gelbart and Kuroda 2009), and the X chromosome is expected to show a different chromatin profile in males vs. females. Indeed, we see a larger fraction of the X chromosome (both XL and XR) to be in state 4 (H4K16ac associated state) in males, relative to females (20.9% vs. 4.9% in total, p<2.2e-16; 43.5% vs. 16.6% for CDS, p<2.2e-16, Fisher's exact test), and the male X is also slightly enriched for state 1 (all active histone modifications) relative to females (13.6% vs. 12.5% in females total, p=2.1e-7; 32% vs. 30% for CDS, p=0.008, Fisher's exact test). On the other hand, very few regions in state 2 (H3K36me3 binding only) are found on the X of males (0.4% vs. 4.7% in females total, p<2.2e-16; 0.3% vs. 13.7% for CDS, p<2.2e-16, Fisher's exact test), since most genes in state 2 on the female X are also associated with H4K16ac in males (states 1 and 4, see **Figure 3b,c**). This is consistent with the known mechanism of dosage compensation in Drosophila, whereby the MSL complex targets actively transcribed genes along the X chromosome by recognizing H3K36me3 and then inducing H4K16 acetylation (Larschan et al. 2007). Only the neo-X in males has a significant fraction of genes classified as state 2 (3.5% of CDS) but is also clearly enriched for genes in state 4 relative to females (42.8% in males vs. 14.5% in females within CDS, p<2.2e-16, Fisher's exact test); this is consistent with partial dosage compensation on this chromosome (Alekseyenko et al. 2013). The fraction of the genome associated with H4K16ac (states 1 and 4) differs between males and females (24.8 % in males, vs. 16.9% in females; see **Figure 2b**); however, this modification is significantly less abundant on autosomes in males relative to females (11% vs. 14.3%, p<2.2e-16, Fisher's exact test) but enriched along the male X chromosomes (34.5% on the male XL and XR vs. 17.4% on the female XL and XR; p<2.2e-16, Fisher's exact test; **Figure 3b**). A similar sex-biased distribution of H4K16ac is also observed in *D. melanogaster* (i.e. an excess on the male X, but an enrichment for H4K16ac on female relative to male autosomes; **Figure S18b,c**). All these differences in chromatin states between the sexes are exaggerated in transcribed genes on the X, consistent with the dosage compensation complex specifically targeting expressed genes (**Figure 3c**; Alekseyenko et al. 2006). In females, the dosage compensation complex does not assemble, and the chromatin landscape is more similar between the X and autosomes (**Figure 3b, c**). However, there is a significant enrichment for states associated with H4K16ac (states 1 and 4) on XL and XR in females relative to autosomes (17.4% vs. 14.3%; p<2.2e-16, Fisher's exact test). This may indicate that sequence features of the X that enable hypertranscription in males, to a small extent, manifest themselves in females (Zhang and Oliver 2010).

The distribution of active chromatin states across the *D. miranda* genome differs for different sequence features and among chromosomes (**Figure 4**). State 1 is enriched at genes, particularly at their 5' end, in both sexes. State 2 is also enriched within genes but

with a 3' bias, consistent with its enrichment for the transcription elongation mark H3K36me3 (**Figure 4b**). State 3 is enriched in the 5' region of genes, but underrepresented in gene bodies. State 4, which is highly enriched on male X chromosomes, shows enrichment within genes and at their 3' ends. This is consistent with the observed 3' bias of MSL-binding along genes (Alekseyenko et al. 2008), and the MSL-complex inducing H4K16ac at male X-linked genes. All of these states are enriched at transcribed genes, but not silent ones (**Figure 4c**).

In general, the distribution of active chromatin states along genes is similar across chromosomes in females, i.e. X chromosomes and autosomes show similar chromatin state profiles. In males, the X and the autosomes differ from each other, and neither matches the chromatin profiles of females (**Figure 4b,c**). Specifically, the distribution of chromatin state 2 in females resembles that of male autosomes but is basically absent on the male X, while the distribution of chromatin state 4 in females resembles that of the male X chromosomes (although enriched at a lower level) but is largely absent from male autosomes. Thus, chromatin-mediated dosage compensation of the male X chromosome appears to generally redistribute activating chromatin marks genome-wide, even on autosomes. In particular, our data suggest that the H4K16ac modification may be sequestered preferentially to X-linked genes in males and diluted away from potential autosomal targets.

Heterochromatin marks in males vs. females. Epigenetic sex differences have been observed in *D. melanogaster*. Depending on different Y-chromosome backgrounds, males differ in their propensity to silence a heterochromatin-sensitive reporter gene (Lemos et al. 2010; Zhou et al. 2012), and RNAi knockdown of the heterochromatin protein HP1 preferentially reduces male viability (Liu et al. 2005). Position effect variegation (PEV), the partial silencing of reporter genes in some cells that normally express a gene resulting in mosaic expression patterns, is often used as an indicator of the local heterochromatic environment, and frequently shows dose-dependent effects (Girton and Johansen 2008). This dosage sensitivity demonstrates the importance of a stoichiometric balance among protein components in the formation of heterochromatin. Males contain a Y chromosome that is highly heterochromatic and may shift this balance by acting as a sink for heterochromatin proteins (Yasuhara and Wakimoto 2008; Lemos et al. 2010; Zhou et al. 2012). This could mean that the genome-wide heterochromatin/ euchromatin balance differs between the sexes. In particular, if the Y chromosome sequesters proteins associated with heterochromatin formation, females might have higher levels of heterochromatin-like features in the rest of their genome as compared to males. Females generally show a higher degree of silencing in assays for PEV, suggesting that normally euchromatic regions are more prone to acquire a heterochromatic structure in females (Wallrath and Elgin 1995; Girton and Johansen 2008). Indeed, a significantly larger fraction of the assembled, mostly euchromatic part of the genome is in chromatin state 5 or 6 (heterochromatin or polycomb repression) in females relative to males (**Figure 2a & 4a**, $p<2.2e-16$, Fisher's exact test). Overall, 10.2% of the genome is associated with the H3K9me2 state (state 6) in females, and only 5.8% in males, and 15.3% is associated with the H3K27me3 state (state 5) in females, and 10.7% in males. Females thus have 1.8x as much heterochromatin overall than males and 1.4x as much polycomb overall in the

euchromatic portion of the genome. A similar excess of H3K9me2 is seen in the assembled portion of the genome in *D. melanogaster* females relative to males (**Figure S18a)**. Heterochromatin is recruited to regions with high repeat densities, resulting in a correlation between repetitive DNA sequence and the propensity of a genomic region to adopt a heterochromatic appearance (Dorer and Henikoff 1994; Pimpinelli et al. 1995; Lippman et al. 2004; Sentmanat and Elgin 2012). Indeed, we find that genomic regions that are heterochromatic in both males and females have highly increased transposable element densities compared to genomic regions that are not associated with H3K9me2 in either sex (transposable element densities of 21% vs. 6%, p<2.2e-16, Fisher's exact test; **Figure S19**). Interestingly, the genomic regions that are heterochromatic (state 6) only in males (approximately 1.6Mb) show transposable element densities slightly lower but similar to those regions that are heterochromatic in both sexes (transposable element densities of 16% vs. 21%), while genomic regions that are heterochromatic in females only (approximately 9.6Mb) display intermediate levels of repeat density (13.8%). This pattern supports the idea that transposable elements trigger heterochromatin formation and that the euchromatin/ heterochromatin balance differs between the sexes (Yasuhara and Wakimoto 2008; Lemos et al. 2010; Zhou et al. 2012). Regions of intermediate repeat content may not be able to initiate or propagate heterochromatin formation in males, as the repeat-rich Y may sequester the structural components required for heterochromatin formation, but heterochromatin may form in these regions of increased repeat-density in females. Thus the global landscape of repressive chromatin differs between the sexes and manifests itself mainly in regions of intermediate repeat content that can trigger heterochromatin formation or spreading more easily in females than in males.

Sex-biased expression vs. chromatin states. Many genes are expressed differently between the sexes, and we were interested in whether genes with sex-biased expression show sex-specific differences in chromatin. That is, genes that show sex-biased expression may be in an active chromatin state in the sex in which they are expressed more highly, and a repressive state in the opposite sex. To evaluate if differences in chromatin state contribute to sex-biased expression patterns, we (1) identified regions of the genome that are in a different chromatin state in males and females, and compared their expression between the sexes; (2) identified genes that are differentially expressed between male and female larvae (genes with sex-biased or sex-specific expression) and compared their chromatin profiles.

We classified genes according to their chromatin state as being in an active chromatin environment in both sexes (i.e. states 1-3; note that we did not include genes belonging to state 4 since they are expressed at an intermediate level, see **Figure S16**), being in a silent or repressive state in both sexes (state 5-7), and being active in one sex and silent or repressed in the other (**Figure 5a**). As expected, we find genes that are in an active chromatin environment in both sexes are expressed at a higher level than those that are in a silent/repressive state (**Figure 5a**, p<2.2e-16 for both female and male active vs. repressed genes, Wilcoxon test). However, levels of gene expression in males and females are similar (and not significantly different) at genes that are in an active chromatin environment in one sex, and a repressive state in the other (**Figure 5a**). We estimate that only 9% of genes that are in an active chromatin state in one sex but a repressed chromatin

state in the other show the expected bias in expression, which is not statistically different from the overall frequency of sex-biased genes (p=0.18, Fisher's exact test). Thus, sex-specific differences in the chromatin state of genes do not manifest themselves as sex-specific expression patterns.

Using gene expression data from male and female larvae, we identified 679 male-specific and 874 male-biased genes, and 95 female-specific and 362 female-biased genes, and characterized their chromatin state in both sexes. We find that genes that are expressed at a similar level in both sexes are highly enriched for active chromatin marks (**Figure 5b**). In contrast, genes with sex-biased or sex-specific expression show no marked differences in their histone profiles between sexes, and are in fact not enriched for activating histone marks at all, in either sex (**Figure 5b, Figure S20**). We find that only 1.6% of genes that have sex-specific or sex-biased expression have the expected sex-specific chromatin state, which is actually less than expected from the overall frequency of sex-specific active or repressed chromatin (2.7%; p=0.003, Fisher's exact test). Instead, genes with either male- or female- biased expression are most often in the background state 7, suggesting that they are not targeted by any of the histone modifications surveyed here, despite being highly transcribed. This could reflect a general difference in the chromatin signature of genes with housekeeping function versus genes with a more restricted function (such as sex-specific/sex-biased or tissue-specific/tissue-biased genes). Indeed, comparing tissue-specific genes to genes that are broadly expressed (as defined by the tissue-specificity index tau), we find a similar difference in their chromatin profile to that of unbiased and sex-biased genes (**Figure 2c, Figure S20**). Additionally, sex-specific genes (and also sex-biased genes, to a lesser extent) show higher tissue-specificity than genes that show similar expression in the two sexes (**Figure S21**). Thus, tissue-specificity of sex-biased genes appears to dominate their chromatin landscape, and not their sex-biased expression patterns.

To conclude, while there are more than a thousand genes that show sex-specific or sex-biased expression, these expression differences do not manifest themselves at the chromatin level between sexes, at least for the histone marks studied here. Instead, genes with sex-limited or sex-biased expression have chromatin profiles that differ from broadly transcribed housekeeping genes and resemble that of other tissue-specific genes.

(3) Comparisons of the chromatin landscape of *D. miranda* vs. *D. melanogaster*

To investigate the evolutionary dynamics of chromatin states, we compared various aspects of the chromatin landscape between *D. miranda* and *D. melanogaster*. Genome-wide, the chromatin landscape looks largely similar between the two species (**Figure S22**).

We categorized genes over 1kb in both species by the chromatin state at their transcription start site (see Methods). Genes without an ortholog in the other species were found more often to be in a repressive chromatin state (states 5-7) than genes shared between species (50% of *D. melanogaster* genes with no ortholog and 47% of *D. miranda* genes with no ortholog vs. 29% of orthologs in *D. melanogaster* and 28% of orthologs in *D. miranda*; **Figure 6a**). We classified the subset of genes that have orthologs in both species (7159 genes) as either conserving their chromatin state (a gene is in the same chromatin state in

D. melanogaster and both male and female *D. miranda*) or not conserving their chromatin state (a gene is in a different chromatin state in *D. melanogaster* than it is in both male and female *D. miranda*). We found that genes whose chromatin state is conserved across the two species have many characteristics of housekeeping genes (**Figure 6**). Genes with a conserved chromatin state are enriched for active chromatin states, especially state 1, while genes that do not have a conserved chromatin state are depleted for active chromatin states as compared to the set of all orthologous genes examined ($p<2.2e-16$ for all three samples, Fisher's exact test, **Figure 6b**). Genes with a conserved chromatin state have higher expression levels (**Figure 6c**) and also have broader expression across tissues (as measured by the tissue-specificity index tau, **Figure 6d**) than genes whose chromatin state is not conserved. Finally, genes with a conserved chromatin state across species have lower K_a/K_s values than the set of all orthologous genes, while genes whose chromatin state is not conserved have higher K_a/K_s values (**Figure 6e**). Thus, there is an association between sequence divergence and chromatin turnover across species; however, it remains to be determined whether chromatin turnover results directly from DNA divergence, or whether DNA and chromatin divergence both reflect lower functional constraint at fast-evolving genes. To conclude, genes that have conserved their chromatin states across species tend to be highly and widely expressed, with low rates of protein evolution. All these patterns hold true if the sexed *D. miranda* data are computationally merged to resemble the unsexed *D. melanogaster* ChIP-seq data (**Figure S23**).

We then looked at the number of genes that conserve their chromatin state between sexes vs. between species (**Figure S24**). Genes in all 7 states show a higher level of conservation across all three samples than what would be expected by chance based on a permutation test (see Methods). Globally, we find that genes in state 1 are most conserved across species and sexes: of all genes present in state 1 in either data set (3905 total), 76% are in state 1 in all three samples, representing an additional 1973 genes shared among all samples than expected by chance, and only 6.5% are restricted to a single sample (vs. 19% expected to be restricted to a single sample, $p<0.001$, permutation test). The least conserved state within orthologous genes, on the other hand, is the heterochromatin-associated state 6: only 4% of the 368 genes assigned to this state in either sample are shared among all three samples, representing only 13 additional shared genes over random expectation, and 60% are restricted to a single sample (vs. 62% expected by chance, $p=0.087$, permutation test). The second least conserved state is state 4 (H4K16ac only), with 4% of the 798 genes classified in this state shared among all samples (only 24 additional genes over neutral expectation) and 57% unique to a sample (vs. 56% expected, $p=0.91$, permutation test). For many states, the fraction of genes in the same state between *D. melanogaster* and *D. miranda* is similar to that between male and female *D. miranda* (state 1: 84% vs. 85%; state 2: 31% vs. 36%, state 3: 43% vs. 48%; state 4: 19% vs. 29%, state 5: 35% vs. 53%, state 6: 12% vs. 32%, state 7: 48% vs. 50%), i.e. there roughly is similar sharing in chromatin states between species as there is between sexes. The main exception here are genes in the polycomb and constitutive heterochromatin state (state 5 and 6) where male and female within *D. miranda* are considerably more similar than they are to *D. melanogaster*. This could reflect lineage-specific divergence in underlying developmental pathways causing lineage-specific silencing of polycomb target genes and evolution of lineage-specific heterochromatin, but could also result from differing ChIP

efficiencies in the two species, especially for H3K27me3 (**Figure S12).** State 4, representing the H4K16ac only state, also shows more similarity in males and females of *D. miranda* than it does across species, likely reflecting the fact that two chromosomal arms are sex-linked in *D. miranda* but autosomal in *D. melanogaster*. In general, however, we find that males and females from the same species and with a nearly identical genome sequence are almost as similar to each other at the chromatin level as they are to a species from which they diverged over 30MY ago and which shows high levels of sequence divergence (Richards et al. 2005).

In both species, housekeeping genes are mainly in an active chromatin state, while tissue-specific genes show a relative enrichment for repressive chromatin states, especially the background state 7 (**Figure S25**). However, both housekeeping and tissue-specific genes show higher levels of conservation of chromatin state across species than expected by chance. Housekeeping genes conserve their mainly active chromatin state in both species (**Figure S26, S27**); about 73% of housekeeping genes have the same state in *D. melanogaster* and *D. miranda* (vs. 68% expected by chance, $p<0.001$, permutation test), while about 33% of tissue-specific genes have the same chromatin state in both species (vs. 20% expected by chance, $p<0.001$, permutation test). We also find that there is considerable variation in conservation of chromatin states across chromosomes and sexes (**Figure S28**). Autosomes show similar levels of chromatin conservation for males and females, males show a higher level of conservation on the chromosomal arm that is X-linked in both species (X in *D. melanogaster*, XL in *D. miranda*), while females share the same chromatin state as *D. melanogaster* more often than males on XR and the neo-X (which are autosomal in *D. melanogaster*), reflecting the recruitment of histone modifications associates with dosage compensation to these chromosomes in male but not female *D. miranda*.

To evaluate the contribution of chromatin state to expression divergence, we classified genes according to their chromatin state in male and female *D. miranda* and in *D. melanogaster*, and then compared their expression patterns. As expected, we find that genes located in an active chromatin environment in both *D. miranda* and *D. melanogaster* are expressed at a higher level in both species than those that are in a silent/repressive state (**Figure 7**, $p<2.2e-16$, Wilcoxon test). Further, we also find that genes that are in different chromatin states in the two species are expressed more highly in the species where they reside in active chromatin, although the difference is not always statistically significant (**Figure 7a,b**). We used gene expression data from *D. melanogaster* mixed-sex larvae, and male and female *D. miranda* larvae to identify 2524 genes that are expressed at a similar level in both species, 751 genes that are silent in both species, 378 genes that are expressed exclusively in *D. melanogaster*, and 199 *D. miranda*-specific genes (**Figure 7c**), and characterized which histone marks they are associated with in either species. We find that genes that are expressed at a similar level in both species are highly enriched for active chromatin marks (**Figure 7c**), while non-transcribed genes are enriched for the background chromatin state (and also polycomb in *D. miranda;* see **Figure S29** for genes with species-biased expression levels). Genes with species-specific expression patterns are most often found in the silent background state, which mimics the chromatin profile of tissue-specific genes, and indeed, we find that genes that are expressed in both species are

generally more broadly expressed, and genes that are expressed only in one species or silent in both tend to be expressed in fewer tissues (**Figure S30, S31**). Nevertheless, we find that genes with *D. melanogaster*-specific expression tend to be more often in an active chromatin state (states 1-3) in *D. melanogaster* relative to *D. miranda* (41% are active in *D. melanogaster*, versus 30% in *D. miranda* females and 28% in *D. miranda* males, p<0.003, Fisher's exact test), and genes expressed specifically in *D. miranda* likewise tend to more often be in active chromatin in *D. miranda* than *D. melanogaster* (although the difference is small and not statistically significant; 36% are active in *D. melanogaster*, versus 38% in *D. miranda* females and males, p=0.2; Fisher's exact test, **Figure 7c**). Again, all these patterns remain if the sexed *D. miranda* data are computationally merged to resemble the unsexed *D. melanogaster* ChIP-seq data (**Figure S32**). We quantified the contribution of chromatin state turnover to lineage-specific gene expression as the fraction of genes with lineage-specific expression that show the expected difference in chromatin state (e.g. *D.* melanogaster-specific genes that are in an active state in *D. melanogaster* and a repressed state in *D. miranda*). We estimate that almost 25% of species-specific expression is accompanied by the expected change in chromatin state (p<2.2e-16, Fisher's exact test). Additionally, when we interrogate the fraction of genes that are in an active state in one species but a repressed state in the other, we find that about 26% of these genes show the expected direction of expression bias or specificity, as compared to approximately 8% of genes that would be expected to show these expression patterns due to chance (p<2.2e-16, Fisher's exact test). Thus, while sex-specific expression patterns appear not to manifest themselves at the chromatin level, we find that differences in the chromatin landscape between species are associated with lineage-specific expression patterns for a fraction of orthologous genes.

Conclusions

Using six histone modifications, we have identified different active and repressive chromatin domains in males and females of *D. miranda*. Genome wide, the conservation of chromatin state of genes between male and female *D. miranda* is about 74%, and among genes with an ortholog in *D. melanogaster*, males and females share the same chromatin state in 75% of genes. These differences are not simply driven by dosage compensation of the hemizygous male X chromosomes alone, where widespread differences in chromatin structure between males and females are expected; nearly 12% of autosomal genes also have a different chromatin state in male and female *D. miranda*. Nevertheless, our data are compatible with the hypothesis that most differences in the chromatin landscape between sexes are a direct or indirect consequence of sex chromosomes. In particular, both the dose (for the X) and the presence (for the Y) of sex chromosomes differ between males and females. The dosage compensation machinery in male Drosophila alters the chromatin structure by inducing H4K16 acetylation at X-linked genes in males (Gelbart and Kuroda 2009), and we see a marked depletion of H4K16ac on male autosomes, relative to the X, or to H4K16ac levels on female autosomes. Thus, chromatin reorganizing due to dosage compensation in males may have a genome-wide effect by redistributing active chromatin marks. The other sex chromosome, the Y, might also function as a sink for chromatin marks, in this case, repressive chromatin (Yasuhara and Wakimoto 2008; Lemos et al. 2010; Zhou et al. 2012). In particular, we see less repressive chromatin in the assembled, mostly euchromatic portion of the genome in males, and we find a higher density of transposable

elements in male heterochromatin than in heterochromatic regions found only in females. These findings are consistent with a model in which the heterochromatic Y chromosome induces a global re-organization of repressive marks in males compared to females (Yasuhara and Wakimoto 2008; Lemos et al. 2010; Zhou et al. 2012), and dilutes these marks away from genomic regions with less heterochromatic characteristics (i.e. slightly lower repeat densities) in males. Our results suggest that differences in chromatin structure between males and females extend genome-wide and are not limited to the sex chromosomes, although the heterochromatic Y and the dosage compensated X may be ultimately responsible for global changes in chromatin structure (Gelbart and Kuroda 2009; Yasuhara and Wakimoto 2008; Lemos et al. 2010; Zhou et al. 2012). This view of the sex-specific chromatin landscape being largely a consequence of a stoichiometric redistribution of active and repressive chromatin marks is in line with our observation that sex-specific chromatin states do not explain sex-specific expression patterns. Instead, most genes with sex-specific or sex-biased gene expression are not targeted by any of the histone modifications assayed in either sex, and their chromatin profile largely resembles that of tissue-specific genes. Previous studies have suggested that many sex-biased genes are tissue specific, and that tissue-specific genes have a distinct chromatin structure that is characterized by a lack of H3K36me3 binding despite showing appreciable expression levels (Filion et al. 2010; Assis et al. 2012). Thus, while most sex-biased genes are unbound by the six histone marks that we assayed, profiling of additional chromatin marks may allow a better understanding of the relationship between sex-specific chromatin structure and sex-biased gene expression.

Unlike for sex-specific chromatin, we find that lineage-specific differences in the chromatin landscape do contribute to expression divergence between species. We find that approximately 62% of orthologous genes have the same chromatin state in *D. melanogaster* and *D. miranda*. The level of conservation of chromatin states between these two species is similar in magnitude to that observed between males and females of *D. miranda* (75% of orthologous genes are in the same chromatin state between sexes), despite the genome of male and female *D. miranda* being nearly identical, while *D. miranda* and *D. melanogaster* diverged over 30MY ago and are highly divergent at the DNA sequence level (Richards et al. 2005). Studies in primates have also indicated that chromatin states may be relatively stable over evolutionary time (Bradley et al. 2010; Cain et al. 2011). The high level of chromatin conservation between Drosophila species might be especially surprising given that two chromosome arms are sex-linked in *D. miranda* but autosomal in *D. melanogaster*, and genes on these chromosomes show comparatively poor conservation of chromatin state. Thus, species that share the same sex chromosome karyotype might show even higher levels of conservation of chromatin states. Future studies contrasting individuals within a species, and species at different divergence levels and with different sex chromosome karyotypes will help to shed light on the evolution of chromatin structure.

Experimental procedures

***ChIP-seq data*.** We analyzed ChIP-seq data from 6 histone marks (H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9me2, and H4K16ac) and an input control for unsexed 3rd instar larvae of *D. melanogaster* and sexed male and female 3rd instar larvae separately for *D.*

miranda. Data for *D. melanogaster* were obtained from modENCODE, and data for H3K27me3, H3K36me3, H3K9me2, and H4K16ac for *D. miranda* male and female larvae were obtained from (Alekseyenko et al. 2013; Zhou et al. 2013). To obtain H3K4me1 and H3K4me3 ChIP-seq data, and replicate data for H4K16ac and H3K9m2, we sexed 3rd instar larvae of *D. miranda*, and followed the protocol of (Alekseyenko et al. 2013; Zhou et al. 2013) to isolate chromatin and perform ChIP-seq. Briefly, chromatin was isolated from 0.5g sexed 3rd instar larvae, cross-linked using formaldehyde, and sheared by sonication. Chromatin pull-down was performed with Dynabeads Protein G (Life Technologies, 10003D) following overnight incubation with antibody against either H3K4me1 (5µl per immunoprecipitation, Abcam ab8895), H3K4me3 (5µl per immunoprecipitation, Abcam ab8580), H4K16ac (5µl per immunoprecipitation, Millipore 07-329), or H3K9me2 (5µl per immunoprecipitation, Abcam ab1220). Immunoprecipitated DNA was then purified and libraries were prepared according to the standard Illumina HiSeq protocol. Paired-end 100bp DNA sequencing was performed on the Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley.

***ChIP-seq data analysis*.** ChIP-seq reads from the 6 histone marks and an input control for unsexed *D. melanogaster* larvae were mapped to the *D. melanogaster* release 5 assembly using bowtie2 (Langmead and Salzberg 2012). First, we called regions of significant enrichment using ChromHMM ("BinarizeBed" function) for the entire genome, and then for only the autosomes; the enrichment profiles from all chromosomes were then combined (Ernst and Kellis 2012). We then used ChromHMM to learn models with different numbers of chromatin state by learning a multivariate hidden Markov model. Since we used the 9 state model from (Kharchenko et al. 2011) to validate our chromatin models, we set 9 as an upper bound on the number of chromatin states we could reliably learn from our subset of histone modifications. However, we did not achieve good concordance with our model and the published 9 state model until we reduced the number of states to 7, at which point we saw good agreement between the two models, i.e. the states from our 7 state model mostly correspond to either a single or two combined states from the 9 state model (**Figure S1-S3**).

ChIP-seq reads from the 6 histone marks and an input control for male and female *D. miranda* larvae were first clipped to remove any adapter sequences. Then, the unmapped reads were downsampled in the sex with more initial reads to match the number of reads in the other sex. Reads were then mapped to the assembled *D. miranda* genome using bowtie2 (Langmead and Salzberg 2012). To remove any reads that mapped to the neo-Y chromosome (which shows high levels of sequence similarity to the neo-X), any read (or its mate) that overlapped a diagnostic neo-Y SNP was removed from further analysis (Zhou et al. 2013). These filtered reads were then used to call regions of significant enrichment using ChromHMM ("BinarizeBed" function) first for the entire genome, and then for only the autosomes; the enrichments from all chromosomes were then combined and the chromatin model learned from *D. melanogaster* was used to assign each genomic region to one of the seven chromatin states (Ernst and Kellis 2012).

We evaluated our strategy of calling chromatin states for autosomes and sex chromosomes separately by comparing the differences in chromatin state between male and female *D.*

miranda both when we binarized sex chromosomes and autosomes separately ("bin separately") and when we binarized the entire genome simultaneously ("bin together"). We found that binning separately was the more conservative strategy in evaluating the sex-specific chromatin landscape in that the magnitude of difference between males and females was lower for all states on the autosomes than when the data were binned together (**Figure S5**). The differences between the sexes were slightly higher on the neoX for states 4 and 5 when the data were binned separately, but significantly lower for states 2 and 6, also consistent with binning separately being a more conservative way to estimate differences in the chromatin landscape between the sexes.

ChIP efficiency was evaluated by calculating the fraction of all mappable reads that mapped within each of the 7 chromatin states. We then calculated the enrichment for each chromatin state by normalizing the fraction of reads that mapped to each state by the genome-wide proportion of each state. This allowed us to compare the ChIP efficiency of each mark across all three samples, since each chromatin annotation was generated from the same 7 state model learned from *D. melanogaster* (**Figure S11, S12**).

To evaluate whether comparing mixed-sex *D. melanogaster* data to sexed *D. miranda* data influenced our conclusions on the turnover of chromatin across species, we simulated mixed-sex *D. miranda* data by merging the reads from the downsampled (i.e. equal number) male and female samples for all 6 histone marks and the input controls. We then called chromatin states for autosomes and sex chromosomes separately as described above, and used the 7 state chromatin model learned from the *D. melanogaster* data to define the mixed-sex chromatin landscape. We then classified genes by chromatin state as described below, and re-analyzed various aspects (chromatin state, gene expression, tissue specificity, K_a/K_s) of genes that conserved their chromatin state across the two species (**Figure S23**). We also looked at the correlation between species-specific expression and species-specific chromatin (**Figure S32**).

Western blotting and quantification. Sexed 3rd instar larvae were lysed and homogenized in Triton Extraction Buffer (PBS pH 7.4 with 0.5% Triton X100, 2mM phenylmethylsulfonyl fluoride, and 0.02% sodium azide) and histones were acid-extracted in 0.2N HCl overnight. Acid-extracted histones were then run on a 4-12% gradient Bis-Tris gel (Invitrogen) and blotted onto a nitrocellulose membrane. Antibodies for H3K9me2 (Abcam 1220), H3K27me3 (Abcam 6002), H3K36me3 (Abcam 9050), H4K16ac (Millipore 07-329) and actin (Abcam 1801) were all incubated at a dilution of 1:1000 in Hikari Signal Enhancer (Nacalai). Westerns blots were imaged and protein levels quantified using the Licor Odyssey software.

ChIP-qPCR. Approximately 0.5g of sexed 3rd instar larvae were flash-frozen in liquid nitrogen and chromatin was prepared by cross-linking with formaldehyde followed by shearing by sonication. Chromatin was immunoprecipitated using 5uL anti-H3K9me2 (Abcam 1220) or 5uL anti-H4K16ac (Millipore 07-329) followed by pull-down with Dynabeads® Protein G (Invitrogen 10003D). Crosslinks were reversed by incubating overnight at 65°C, and samples were extracted with phenol-chloroform. Genomic targets for qPCR were selected from the repeat-masked genome; we assayed 8 regions not bound

by any of the four histone marks assayed, 8 regions bound by H3K9me2 in both sexes, 4 regions bound by H3K9me2 in females but not males, 8 regions bound by H4K16ac in both sexes, 4 regions bound by H4K16ac in females but not males, and 4 regions bound by H4K16ac in males but not females. qPCR was performed using the KAPA SYBR FAST qPCR master mix (KAPA Biosystems 4600) and the data was analyzed using the Applied Biosystems StepOnePlus Real-Time PCR System (Invitrogen). ChIP enrichment was calculated as percent input by first calculating the normalized C_t of the ChIP as $\Delta C_{t\,norm} = C_{t\,ChIP} - (C_{t\,input} - \log_2(\text{input dilution factor}))$ where the input dilution factor was 2.5 in this experiment. The percent input could then be calculated as $2^{-\Delta Ct\,norm}$. Each PCR product was visualized on an agarose gel to confirm that there was a single amplicon and no primer-dimer.

Categorizing genes by chromatin state. Similar to the procedure used in (Kharchenko et al. 2011), we categorized genes by chromatin state if they were longer than 1kb, the resolution at which our chromatin model was learned. We then categorized genes as being in the chromatin state of their transcription start site.

To evaluate whether the overlap of genes in each state was significantly enriched or depleted between all three samples (*D. melanogaster*, and male and female *D. miranda*), we performed a permutation test. For each permutation, we shuffled the chromatin state of the genes for all three samples, keeping the same total number of genes in each state as in our observed data. For each of 1000 permutations, we calculated the proportion of genes in each state that were shared across all three samples, or that were restricted to a single sample, and compared the observed proportion to those generated by permutations to generate p-values. To generate expected values, we calculated the average proportion of genes in a given state that were shared across all three samples or restricted to a single sample across all permutations.

Replicate ChIP-seq datasets. We used two independent methods to calculate the level of correlation between male and female ChIP experiments. First we called peaks for each mark in each sample individually using MACS (Zhang et al. 2008). We then defined the top 40% of peaks for each histone modification for each sex based on average enrichment over the input, and calculated the fraction of peaks that were also called as peaks in the opposite sex (**Figure S10a**). Histone marks with at least 80% of the top 40% of their peaks present in the opposite sex were considered adequately replicated, as described in (Kharchenko et al. 2011). As an independent measure of replication, we also calculated the mean enrichment over the input in 10kb intervals across the genome for each histone modification for both males and females. We then calculated the Spearman correlation between male and female experiments for each histone modification (**Figure S10b**).

Since all histone modifications except H3K9me2 and H4K16ac showed very high levels of correlation between sexes (**Figure S10**), and since we found marked differences between males and females in chromatin states involving H3K9me2 and H4K16ac binding, we produced replicate ChIP-seq datasets for male and female *D. miranda* 3rd instar larvae for these two marks. The data were normalized and mapped as for the other datasets, as described above. We then substituted the published H3K9me2 and H4K16ac datasets with

our replicate datasets and re-ran ChromHMM as described, still with all 6 histone modifications but with different H3K9me2 and H4K16ac datasets. We found that the sex-biased distribution of these marks was consistent across both datasets (**Figure S9**).

Gene expression and tissue-specificity. RNA-seq reads from sexed 3rd instar *D. miranda* larvae were mapped to the assembled genome using TopHat (Trapnell et al. 2009). To be conservative in making comparisons in expression level between males and females, we equalized the number of mapped reads by subsampling the reads of the sex with higher coverage (in this case, males). We then used Cufflinks to estimate transcript abundances, measured as FPKMs. For *D. melanogaster*, reads generated from mixed-sexed 3rd instar larvae were mapped to the *D. melanogaster* release 5 assembly using TopHat, and Cufflinks (Trapnell et al. 2012) was used to estimate transcript abundances. The cutoff for expressed genes was set at FPKM=1 based on the transcription level of intronic regions. Sex-specific genes were defined as those genes expressed in one sex (FPKM>1) but not expressed in the other sex (FPKM<1); sex-biased genes were defined as genes expressed in both sexes (FPKM>1), but where the expression level in one sex was at least twice as high as the expression level in the other sex.

To calculate the tissue-specificity index tau, we used published expression data from seven tissues (female body, female head, ovary, male body, male head, testis, and accessory gland) from *D. melanogaster* and *D. pseudoobscura*, a close relative of *D. miranda* (Assis et al. 2012). Tau is calculated for each gene by $\tau = \frac{\sum_{i=1}^{N} 1 - \frac{logE_i}{logE_{max}}}{N-1}$ where N is the number of tissues, E_i is the expression in tissue *i*, and E_{max} is the maximum expression of the gene in all tissues. Tau is only calculated for genes that are expressed in at least one tissue, so any gene that is not expressed in any of the seven tissues assayed was not included in analyses of tissue-specificity.

Data access

The ChIP-seq data generated are deposited to Short Read Archive under the accession number SRP040696.

References

Alekseyenko A, Larschan E, Lai W, Park P, Kuroda M. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev* **20**(7): 848 - 857.

Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila. *Genes Dev* **27**(8): 853-858.

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**(4): 599 - 609.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. *Genome Biol Evol* **4**(11): 1189-1200.
Bachtrog D. 2005. Sex chromosome evolution: Molecular aspects of Y-degeneration in Drosophila. *Genome Research* **15**(10): 1393 - 1401.
Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev* **16**(6): 578 - 585.
Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**(2): 113-124.
Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol* 9(2): R30.
Bradley RK, Li XY, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB. 2010. Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. *PLoS Biol* **8**(3): e1000343.
Cain CE, Blekhman R, Marioni JC, Gilad Y. 2011. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187**(4): 1225-1234.
Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci* **355**(1403): 1563-1572.
Cherbas L, Willingham A, Zhang D, Yang L, Zou Y, Eads BD, Carlson JW, Landolin JM, Kapranov P, Dumais J et al. 2011. The transcriptional diversity of 25 Drosophila cell lines. *Genome Res* **21**(2): 301-314.
Deng X, Koya SK, Kong Y, Meller VH. 2009. Coordinated regulation of heterochromatic genes in Drosophila melanogaster males. *Genetics* **182**(2): 481-491.
Dorer DR, Henikoff S. 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**(7): 993-1002.
The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**(7414): 57-74.
Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* **8**(9): 689 - 698.
Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**(3): 215-216.
Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**(7345): 43-49.
Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**(2): 212-224.
Gelbart M, Kuroda M. 2009. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* **136**(9): 1399-1410.
Girton JR, Johansen KM. 2008. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. *Adv Genet* **61**: 1-43.
Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**(7339): 480-485.

Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**(12): 1351-1359.
Khil P, Smirnova N, Romanienko P, Camerini-Otero R. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet* **36**(6): 642 - 646.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**(4): 357-359.
Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in Drosophila. *Trends Genet* **24**(3): 114-123.
Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ, Kuroda MI. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**(1): 121 - 133.
Lemos B, Branco AT, Hartl DL. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **107**(36): 15826-15831.
Lippman Z, Gendrel A, Black M, Vaughn M, Dedhia N, McCombie W, Lavine K, Mittal V, May B, Kasschau K et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**(6998): 471 - 476.
Liu LP, Ni JQ, Shi YD, Oakeley EJ, Sun FL. 2005. Sex-specific role of Drosophila melanogaster HP1 in regulating chromatin structure and gene transcription. *Nat Genet* **37**(12): 1361 - 1366.
McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res* **20**(6): 816-825.
Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in Drosophila. *Proc Natl Acad Sci U S A* **100**(17): 9894 - 9899.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**(2): 83-87.
Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C, Gatti M. 1995. Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. *Proc Natl Acad Sci U S A* **92**(9): 3804-3808.
Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP et al. 2005. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. *Genome Res* **15**(1): 1-18.
Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res* **21**(2): 147-163.
Schulze SR, Wallrath LL. 2007. Gene regulation by chromatin structure: paradigms established in Drosophila melanogaster. *Annu Rev Entomol* **52**: 171-192.

Sentmanat MF, Elgin SC. 2012. Ectopic assembly of heterochromatin in Drosophila melanogaster triggered by transposable elements. *Proc Natl Acad Sci U S A* **109**(35): 14104-14109.
Steinemann S, Steinemann M. 2005. Y chromosomes: born to be destroyed. *Bioessays* **27**(10): 1076 - 1083.
Straub T, Becker PB. 2007. Dosage compensation: the beginning and end of generalization. *Nat Rev Genet* **8**(1): 47 - 57.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**(3): 562-578.
Vicoso B, Bachtrog D. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation. *Chromosome Research* **17**(5): 585-602.
Vicoso B, Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in Drosophila. *Nature*.
Wallrath LL, Elgin SC. 1995. Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes Dev* **9**(10): 1263-1277.
Weiler KS, Wakimoto BT. 1995. Heterochromatin and gene expression in Drosophila. *Annu. Rev. Gene.* **29** 577–605.
Wittkopp PJ, Kalay G. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**(1): 59-69.
Yasuhara JC, Wakimoto BT. 2008. Molecular landscape of modified histones in Drosophila heterochromatic genes and euchromatin-heterochromatin transition zones. *PLoS Genet* **4**(1): e16.
Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2011. A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster. *PLoS Genet* **7**(12): e1002380.
Zhang Y, Oliver B. 2010. An evolutionary consequence of dosage compensation on Drosophila melanogaster female X-chromatin structure? *BMC Genomics* **11**: 6.
Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. *Nature* **450**(7167): 233 - 237.
Zhou J, Sackton TB, Martinsen L, Lemos B, Eickbush TH, Hartl DL. 2012. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. *Proc Natl Acad Sci U S A* **109**(25): 9941-9946.
Zhou Q, Bachtrog D. 2012. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. *Science* 337(6092): 341-345.
Zhou Q, Ellison CE, Kaiser VB, Alekseyenko AA, Gorchakov AA, Bachtrog D. 2013. The epigenome of evolving Drosophila neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol* **11**(11): e1001711.

Figure Legends

Figure 1. Karyotypes of *D. melanogaster* and *D. miranda*. The euchromatic karyotypes of *D. melanogaster* and *D. miranda*, with chromosome arms colored by orthology. Chromosome arms that are autosomal in both species are in gray and black, the shared X chromosome is in red, and the chromosome arms that are sex-linked in *D. miranda* but not *D. melanogaster* are in pink and purple.

Figure 2. The chromatin landscape of *Drosophila miranda* and *D. melanogaster.* a. A model of prevalent chromatin states found in *D. melanogaster* and genomic coverage of chromatin states in *D. melanogaster* and male and female *D. miranda*. b. Coverage of chromatin states in different categories of the genome. 5' and 3' flanking regions refer to ±1kb up- and downstream of the coding sequence (CDS). c. Coverage of chromatin states in expressed and silent genes, and expressed genes categorized by their breath of expression (housekeeping genes vs. tissue specific genes as measured by the tissue-specificity index tau). d. Example of chromatin profiles across the *D. miranda* genome.

Figure 3. The sex-specific chromatin landscape in *D. miranda*, and transitions in chromatin states between sexes. a. A genome-wide karyotype view of the chromatin domains derived from female larvae (top) and male larvae (bottom) smoothed across 50kb windows. b. Transitions of chromatin states across chromosomes between the sexes. The solid background color indicates the fraction of a particular state in a given sex, and the crosshatch color indicates transitions to a given state in the other sex; solid regions indicate regions in the same chromatin state in both sexes. c. Same as b, but for genomic regions overlapping CDS of actively transcribed genes (FPKM >1).

Figure 4. Distribution of chromatin states in *D. miranda* across genomic features and genes bodies. a. Distribution of different chromatin states across chromosomes and different functional categories. b. Enrichment of chromatin states across all genes (size-normalized, indicated by gray box), including 1kb 5' and 3' of CDS. c. Frequency of each chromatin state across size-normalized genes (including 1kb 5' and 3' of CDS) categorized by RNA-seq as expressed or silent (FPKM > 1 or FPKM < 1).

Figure 5. Sex-specific chromatin states and gene expression in *D. miranda*. a. Female (pink) and male (blue) expression of genes (measured as log2 FPKM) categorized in active chromatin in both sexes (State 1-3), active in one sex but not the other, or repressive chromatin in both sexes (States 5-7). b. Chromatin states of genes in females (top panel) and males (bottom panel) that are expressed similarly in both sexes or expressed in a sex-specific manner. Expressed unbiased genes are those expressed (FPKM>1, cutoff based on intergenic expression levels) in both sexes, with the ratio of expression between the two sexes less than 2. Sex-specific genes are those genes that are expressed in one sex (FPKM>1) but not expressed in the other (FPKM<1).

Figure 6. Conservation of chromatin states of genes in *D. miranda* and *D. melanogaster*. a. Chromatin state of orthologous and nonorthologous genes in *D. miranda* females and males and *D. melanogaster*. Orthologous genes defined by FlyBase. b.

Chromatin states of orthologous genes whose state is conserved across species and sexes (left), and those whose state is not conserved (right). c. Expression levels (log2(FPKM)) of genes (*D. melanogaster* in white, *D. miranda* males in blue, *D. miranda* females in pink) of orthologous genes with conserved and not conserved chromatin state. d. Tissue-specificity (measured by the tissue-specificity index tau) of genes with conserved and not conserved chromatin state in *D. melanogaster* (white) and *D. miranda* (grey). e. Rates of protein evolution (Ka/Ks) of genes with conserved and not conserved chromatin state between *D. melanogaster* and *D. miranda*.

Figure 7. Species-specific chromatin states and gene expression. a. *D. melanogaster* (white) and *D. miranda* female (pink) expression of genes categorized in active chromatin in both species (states 1-3), active in one species but not the other, or in repressive chromatin in both species (states 5-7). Genes in active chromatin in both species are significantly more highly expressed than genes that are in an active state in only one species ($p<2.6e-6$ for all comparisons, Wilcoxon test). b. Same as a. but comparing *D. melanogaster* to *D. miranda* male (blue). Genes in active chromatin in both species are significantly more highly expressed than genes in an active state in only one species ($p<3e-4$ for all comparisons, Wilcoxon test). c. Chromatin states of genes expressed similarly in both species, genes expressed in one species but not the other, and genes not expressed in either species. Only genes with unbiased expression between male and female *D. miranda* were considered.

Supplementary Figures

Figure S1. Comparison of 7 state model generated from 3rd instar *D. melanogaster* larvae to 9 state model generated from S2 cells (Kharchenko et al. 2011). a. Regions in different states from 7 state model partitioned by state from 9 state model. b. Regions in different states from 9 state model partitioned by state from 7 state model. In general, the dominant chromatin mark(s) of a state from the two models overlap. c. A model of prevalent chromatin states found in *D. melanogaster* and genomic coverage of chromatin states in *D. melanogaster* (the 9 state model from Kharchenko et al. (2011) is shown for comparison).

Figure S2. Comparison of 8 state model generated from 3rd instar *D. melanogaster* larvae to 9 state model generated from S2 cells (Kharchenko et al. 2011). a. Regions in different states from 8 state model partitioned by state from 9 state model. b. Regions in different states from 9 state model partitioned by state from 8 state model. c. A model of prevalent chromatin states found in *D. melanogaster* and genomic coverage of chromatin states in *D. melanogaster* for the 8 state model. The 8 state model splits state 1 from the 9 state model (that also correspond to a single state in the 7 state model) into two separate states (state 1 and 4 of the 8 state model).

Figure S3. Overlap of states from 6 histone marks with the published 9 state model from Kharchenko et al. (2011). a. Comparison of the 7 state model with the 9 state model. a. Comparison of the 8 state model with the 9 state model.

Figure S4. A 7-state model generated from *D. miranda* 3rd instar larvae ChIP-seq profiles.

Figure S5. Overview of normalization procedure to correct for unequal number of reads between samples and biased signal from different chromosomes.

Figure S6. Effect of normalization procedure to correct for biased signal from different chromosomes. Binarizing the X chromosomes and autosomes separately is generally more conservative in estimating differences in chromatin state between the sexes on the autosomes. Note that XL and XR do not change because they are binarized the same way in both methods.

Figure S7. Western blots and quantification of histone modifications in male (blue) and female (red) *D. miranda*. Quantification was calculated relative to actin signal for all four histone marks assayed (averaged across four replicates). For each histone modification, a 2-fold serial dilution was loaded to ensure quantification within the dynamic range; example blots for each mark are shown.

Figure S8. ChIP-qPCR of H3K9me2 and H4K16ac targets, and relative enrichment in male and female *D. miranda*, as described in Experimental Procedures. Significance was assayed using the Wilcoxon rank-sum test.

Figure S9. Direction and magnitude of sex-biased distribution of chromatin states annotated using published (v1) or replicate H3K9me2 and H4K16ac data for male and female *D. miranda* 3rd instar larvae; published data from the remaining 4 histone modifications were used in both the v1 and replicate annotations. The ratio of genomic regions in each state in females vs. males was calculated for each chromatin annotation.

Figure S10. Correlation between male and female samples for each histone modification. a. Fraction of the top 40% of female (red) or male (blue) peaks that are also called as peaks in the opposite sex. b. Correlation between enrichment intensities of 10kb windows for male and female samples for each histone modification. Spearman's correlation coefficients are also shown.

Figure S11. Immunoprecipitation efficiency using published (v1) or replicate H3K9me2 and H4K16ac data for male and female *D. miranda* 3rd instar larvae. Efficiency was calculated as the fraction of all mappable reads that mapped within each chromatin state, normalized by the genome-wide frequency of each chromatin state.

Figure S12. Immunoprecipitation efficiency in *D. melanogaster*. Efficiency was calculated as described for Figure S11 except for mixed-sex *D. melanogaster* 3rd instar larvae (modENCODE data).

Figure S13. Distribution of tau values for *D. melanogaster* (black) and *D. pseudoobscura* (red, the sister species of *D. miranda*, which is used as a proxy to estimate tissue-specificity of *D. miranda* genes). Genes with tau values below 0.4 are considered housekeeping genes, and genes with tau values greater than 0.6 are considered tissue-specific. Genes with intermediate tau values are not considered in analyses of the effect of tissue specificity.

Figure S14. Chromatin states of housekeeping vs. tissue-specific genes in SL2 cells. Using the published 9-state chromatin model for S2 cells (Kharchenko et al. 2011), we defined genes over 1kb by the chromatin state at their transcription start site. We then categorized genes expressed in S2 cells (FPKM>4, data from Cherbas et al. (2011)) as housekeeping (tau<0.4) or tissue-specific (tau>0.6) and looked at their chromatin states. We find an excess of genes in an active state (1-3) in housekeeping genes compared to tissue-specific genes ($p<2.2e-16$, Fisher's exact test).

Figure S15. Coding sequence (CDS) and transposable element (TE) content of chromatin states. a. Proportion of each state that overlaps CDS in female and male *D. miranda* and *D. melanogaster*. b. Proportion of each state that overlaps TEs in female and male *D. miranda* and *D. melanogaster* (TE annotation derived from RepeatModeler).

Figure S16. Expression values, measured by log2(FPKM) of genes by chromatin state for (a) female larvae *D. miranda*, (b) male larvae *D. miranda*, and (c) mixed-sex larvae *D. melanogaster*. Significance values from Wilcoxon test.

Figure S17. Tissue specificity, measured by tau, of genes by chromatin state for (a) female *D. miranda*, (b) male *D. miranda*, and (c) *D. melanogaster*. Significance values from Wilcoxon test.

Figure S18. Sex-specific histone modification in *D. melanogaster*. a. We compared regions of H3K9me2 enrichment based on ChIP-chip data for male and female 3rd instar *D. melanogaster* larvae from modENCODE (accession numbers 3694 and 3772) and found that more of the assembled, euchromatic portion of the genome is enriched for H3K9me2 in females relative to males. b. We compared regions of H4K16ac enrichment on chromosomes X and 2L based on ChIP-chip data for male and female 3rd instar *D. melanogaster* larvae data from Gelbart and Kuroda (2009); (GEO accession number GSE14884) and found that while the X chromosome is highly male-biased in H4K16ac binding, chromosomes 2L is highly female-biased in H4K16ac binding. c. Normalized intensity values for H4K16ac ChIP-chip show higher values for males on the X, but higher values for females on chromosomes 2L. We defined bound regions as those with a normalized intensity greater than 1.

Figure S19. Transposable elements in heterochromatic regions. The genome was characterized for female and male *D. miranda* as either not heterochromatic (i.e. not in state 6), heterochromatic in one sex but not the other, or heterochromatic in both sexes. We then looked at the overlap between these regions and transposable elements (annotated from RepeatModeler).

Figure S20. Chromatin state of *D. miranda* genes in different expression categories. a. Sex-biased genes are defined by expression in both sexes with the biased sex having at least twice the expression level of the other sex. b. Based on tau values from *D. pseudoobscura*, housekeeping genes have tau<0.4, tissue-specific genes have tau>0.6, and head-specific genes have tau>0.6 with maximum expression observed in male or female head.

Housekeeping, tissue-specific, and head-specific genes had to be expressed (FPKM>1) in at least one sex of *D. miranda* larvae to be included in the analysis. Gene numbers for each category are given in parentheses.

Figure S21. Tau values by expression category in *D. miranda*. Expressed unbiased genes are those expressed (FPKM>1, cutoff based on intergenic expression levels) in both sexes, with the ratio of expression between the two sexes less than 2. Sex-specific genes are those genes that are expressed in one sex (FPKM>1) but not expressed in the other (FPKM<1). Sex-biased genes are defined by expression in both sexes with the biased sex having at least twice the expression level of the other sex. For each category, tau values from *D. pseudoobscura* were analyzed. Statistical significance was assessed with Wilcoxon tests.

Figure S22. Chromosome-wide distribution of chromatin states. The fraction of each chromosome in each chromatin state for (a) *D. miranda* females, (b) *D. miranda* males, and (c) *D. melanogaster*.

Figure S23. Same as Figure 6 but *D. miranda* male and female data were computationally merged.

Figure S24. Conservation of chromatin states of orthologous genes in *D. miranda* and *D. melanogaster*. Overlap of orthologous genes by chromatin state (*D. melanogaster* top in gray, *D. miranda* males lower left in blue, *D. miranda* females lower left in pink).

Figure S25. Chromatin states of genes by tissue specificity in *D. melanogaster* and *D. miranda*, as calculated by tau. Housekeeping genes are those with tau values less than 0.4; tissue-specific genes are those with tau values greater than 0.6.

Figure S26. Conservation of chromatin state of orthologous housekeeping genes in *D. miranda* and *D. melanogaster*. Transitions of chromatin states between *D. melanogaster* and *D. miranda* females (top) and males (bottom). The solid background color indicates the number of genes in a given state in *D. melanogaster*, and the crosshatch color indicates transitions to a given state in *D. miranda*; solid regions indicate genes that are shared in the two species in a given state. Only genes with tau<0.4 in both species are considered.

Figure S27. Conservation of chromatin state of orthologous tissue-specific genes in *D. miranda* and *D. melanogaster*. Transitions of chromatin states between *D. melanogaster* and *D. miranda* females (top) and males (bottom). The solid background color indicates the number of genes in a given state in *D. melanogaster*, and the crosshatch color indicates transitions to a given state in *D. miranda*; solid regions indicate genes that are shared in the two species in a given state. Only genes with tau>0.6 in both species are considered.

Figure S28. Fraction of orthologous genes with a conserved chromatin state between *D. melanogaster* and *D. miranda* females (pink) and males (blue) by chromosome.

Figure S29. Chromatin states of genes with species-biased expression. *D. miranda* biased genes are expressed in both species, and both male and female *D. miranda* have at least

twice the expression level of *D. melanogaster*. *D. melanogaster* biased genes are expressed in both species, and *D. melanogaster* expression values are at least twice as high as those of both male and female *D. miranda*. Only genes that are expressed in both male and female *D. miranda* are considered.

Figure S30. Tissue-specificity of genes in different chromatin categories. a. Tissue specificity, measured by tau and normalized to have the same median between species, in *D. melanogaster* (white) and *D. pseudoobscura* (gray) for genes categorized as active (states 1-3) in both *D. melanogaster* and *D. miranda* females, active in only *D. melanogaster* or *D. miranda* females, or repressed (states 5-7) in both *D. melanogaster* and *D. miranda* females. For each species, all differences in tau across chromatin categories are highly significant ($p<3e-7$, Wilcoxon test). b. Same as (a) but chromatin states compared between *D. melanogaster* and *D. miranda* males. For each species, all differences in tau across chromatin categories are highly significant ($p<2e-10$, Wilcoxon test).

Figure S31. Tissue-specificity of genes by expression pattern in *D. melanogaster* and *D. miranda*. Unbiased genes are expressed in both species (FPKM>1) and the ratio of expression between the two species is less than 2. Species-specific genes are those expressed in one species but not the other. Species-biased genes are those expressed in both species, but *D. melanogaster* have at least twice the expression level of both male and female *D. miranda* (*D. melanogaster* biased), or both male and female *D. miranda* have at least twice the expression level of *D. melanogaster* (*D. miranda* biased). Silent genes are those not expressed in either species. Only genes that are expressed in both male and female *D. miranda* are considered.

Figure S32. Same as Figure 7 but *D. miranda* male and female data were computationally merged.

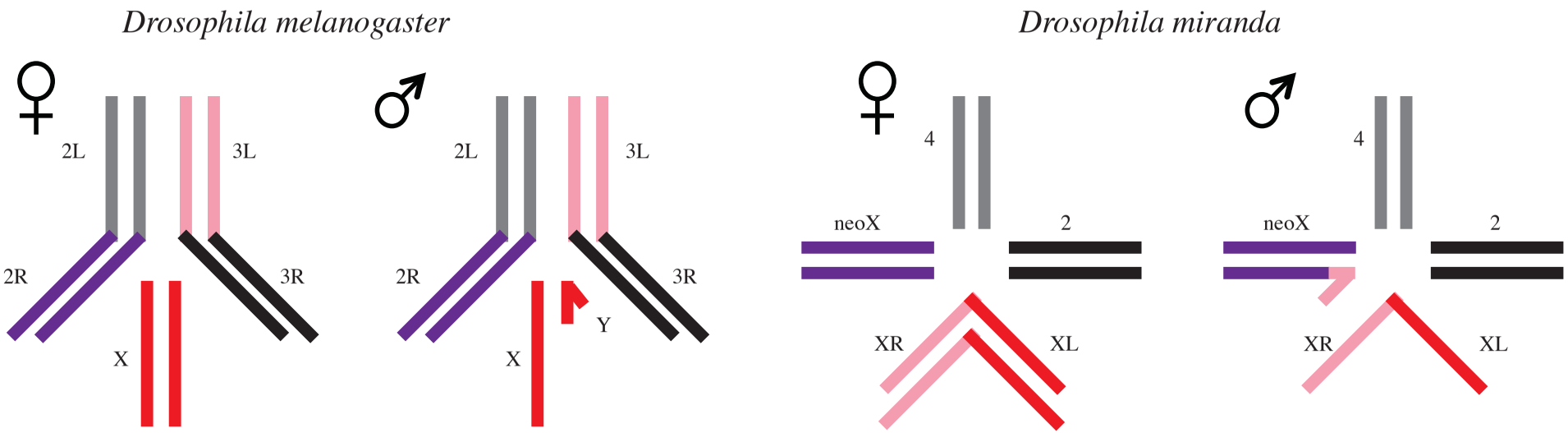

Drosophila melanogaster
2L
3L
2R
3R
X
2L
3L
2R
3R
X
Y
Drosophila miranda
4
neoX
2
XR
XL
4
neoX
2
XR
XL

Figure 1

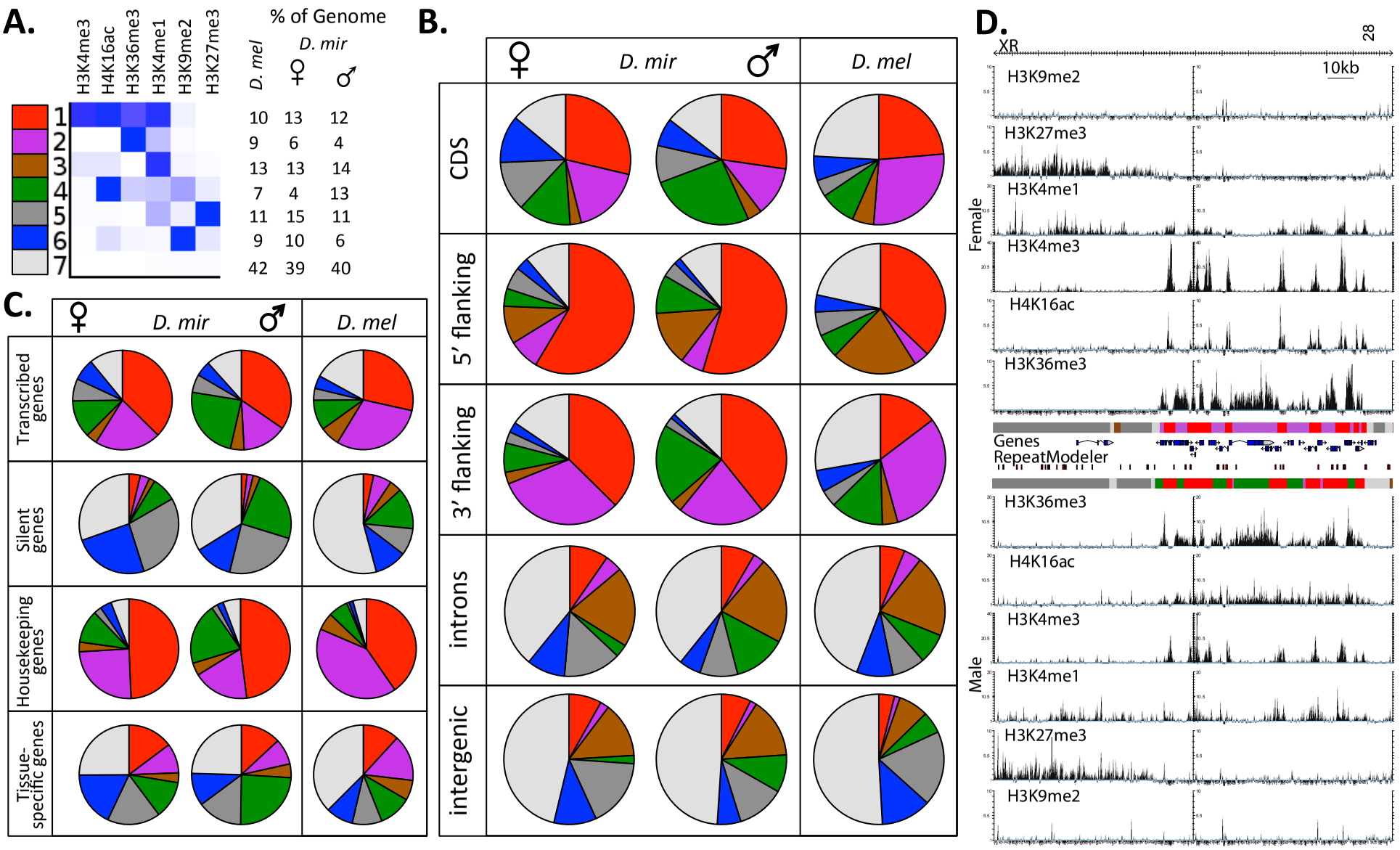

A.
H3K4me3
H4K16ac
H3K36me3
H3K4me1
H3K9me2
H3K27me3
1
2
3
4
5
6
7
% of Genome
D. mel
D. mir
♀
♂
10 13 12
9 6 4
13 13 14
7 4 13
11 15 11
9 10 6
42 39 40
B.
♀
D. mir
♂
D. mel
CDS
5' flanking
3' flanking
introns
intergenic
C.
♀
D. mir
♂
D. mel
Transcribed genes
Silent genes
Housekeeping genes
Tissue-specific genes
D.
XR
28
10kb
H3K9me2
H3K27me3
H3K4me1
H3K4me3
H4K16ac
H3K36me3
Female
Genes
RepeatModeler
H3K36me3
H4K16ac
H3K4me3
H3K4me1
H3K27me3
H3K9me2
Male

Figure 2

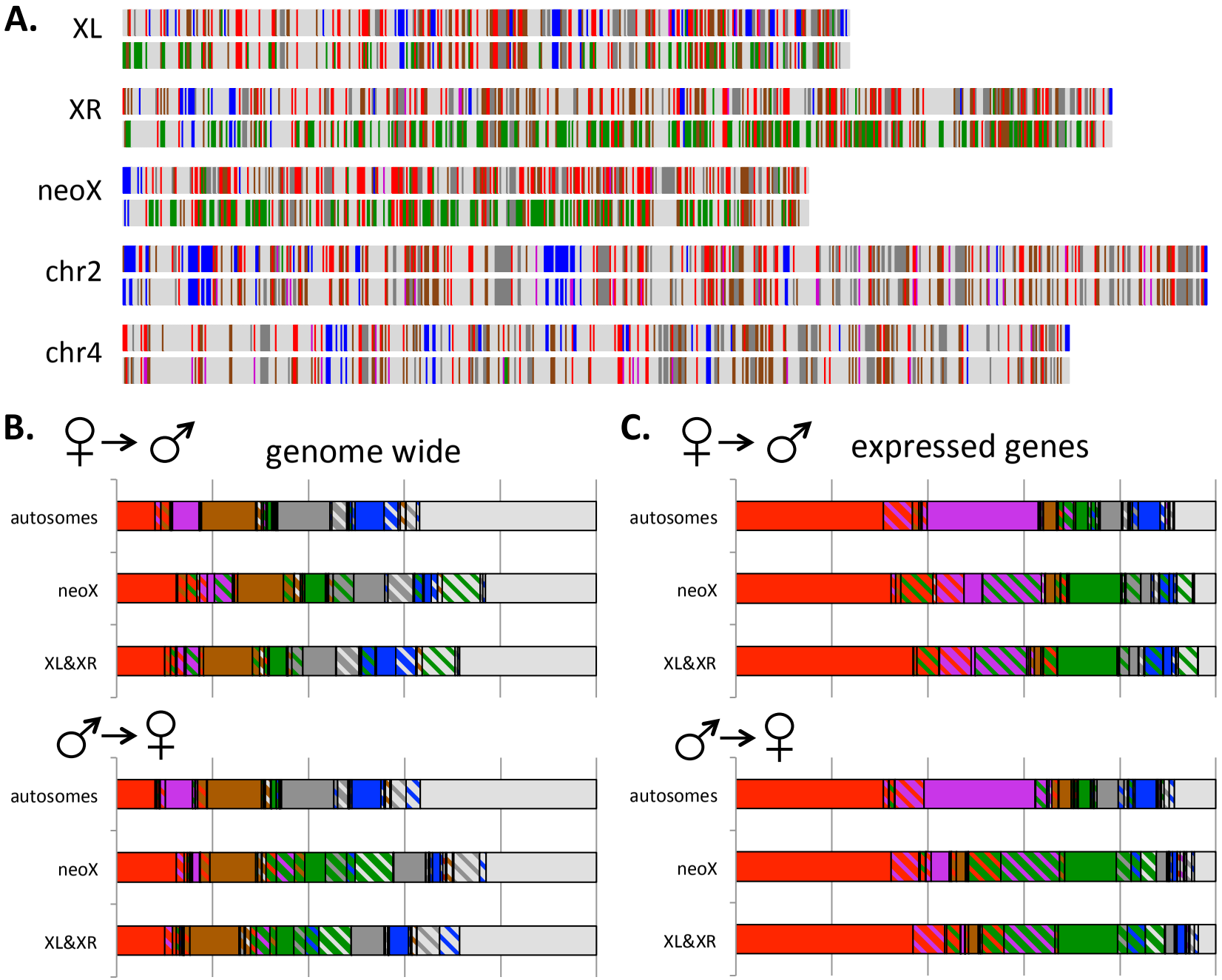

A.
XL
XR
neoX
chr2
chr4
B.
genome wide
autosomes
neoX
XL&XR
C.
expressed genes
autosomes
neoX
XL&XR

Figure 3

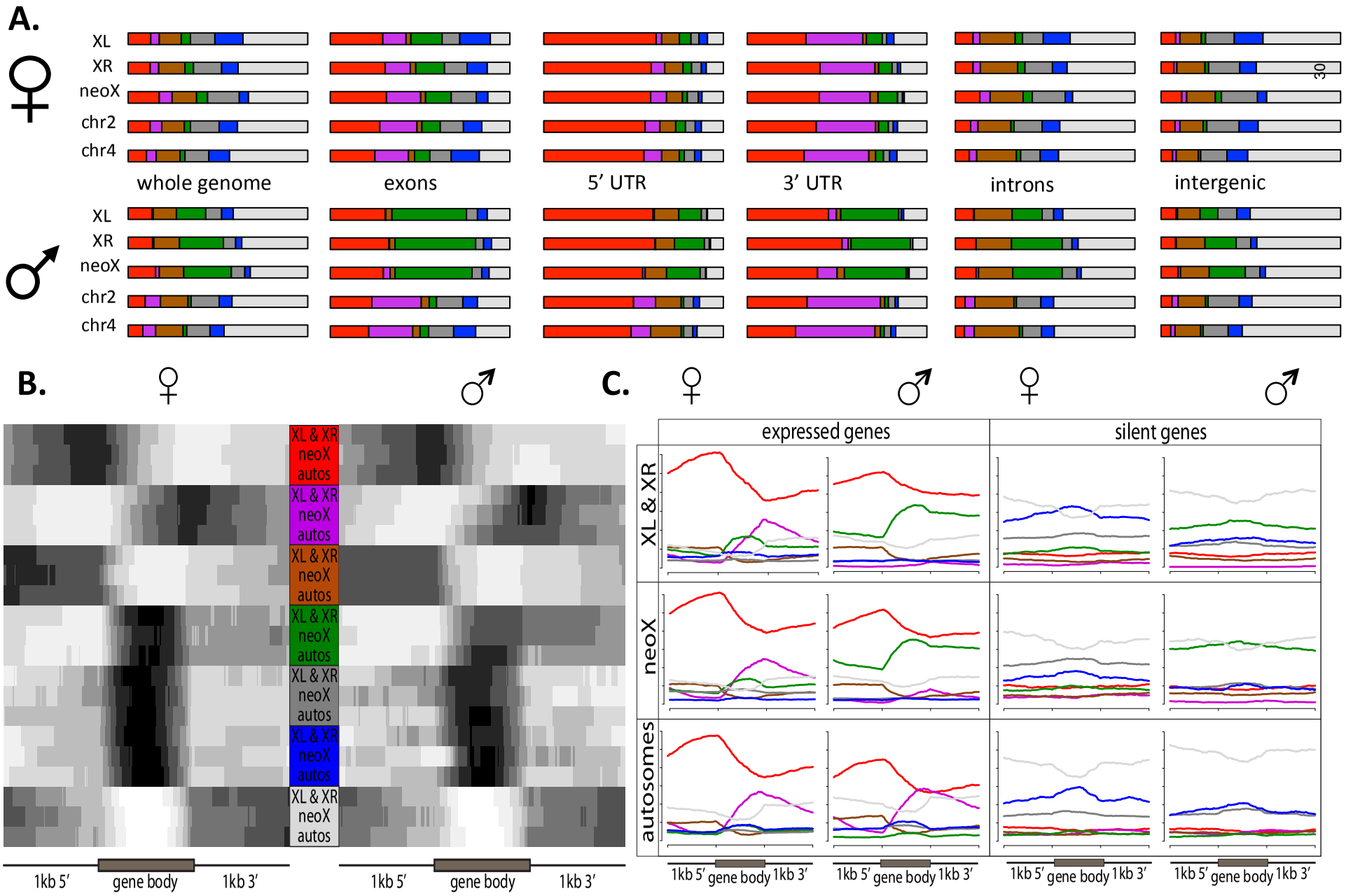
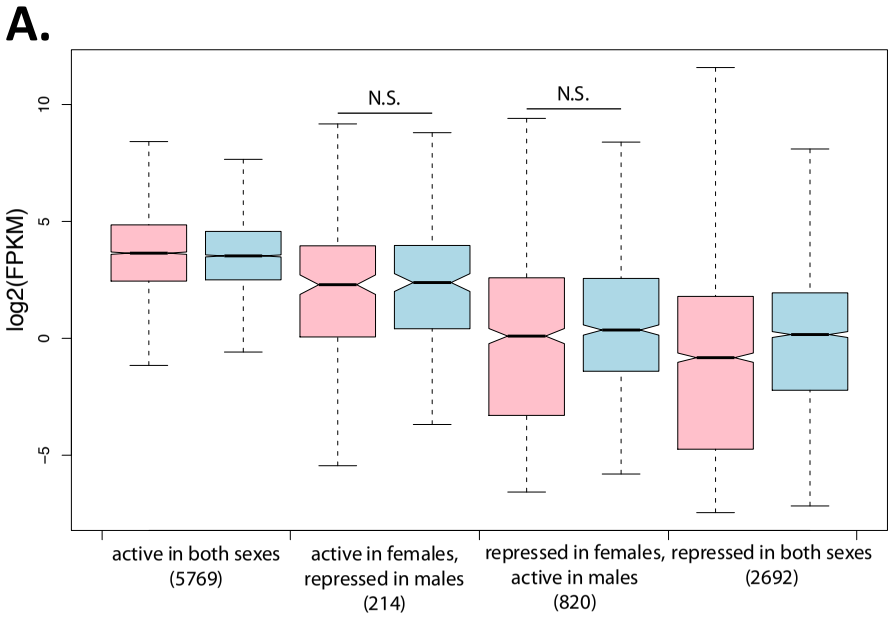

A.
XL
XR
neoX
chr2
chr4
whole genome
exons
5′ UTR
3′ UTR
introns
intergenic
30
B.
C.
XL & XR neoX autos
1kb 5′
gene body
1kb 3′
expressed genes
silent genes
XL & XR
neoX
autosomes

Figure 4


A.
log2(FPKM)
10
5
0
−5
N.S.
N.S.
active in both sexes
(5769)
active in females,
repressed in males
(214)
repressed in females,
active in males
(820)
repressed in both sexes
(2692)

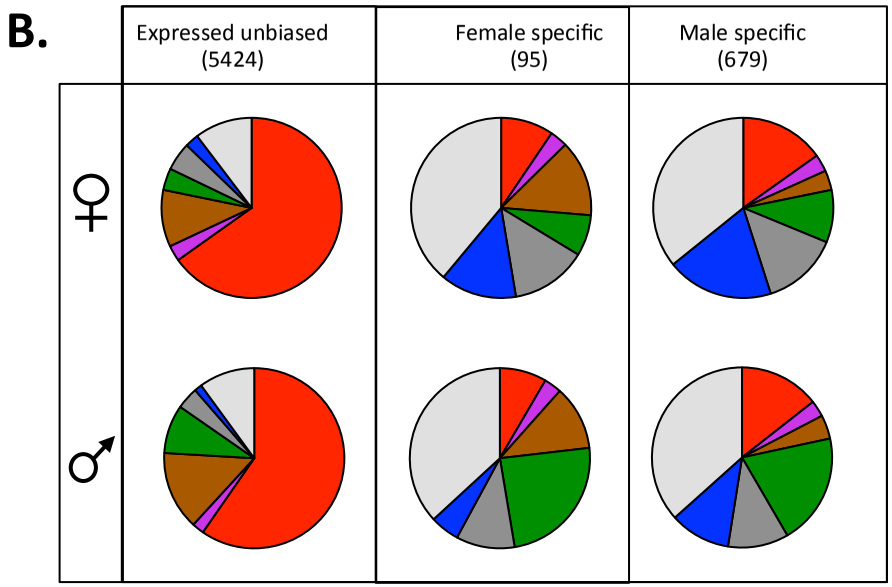

B.
Expressed unbiased
(5424)
Female specific
(95)
Male specific
(679)
♀
♂

Figure 5

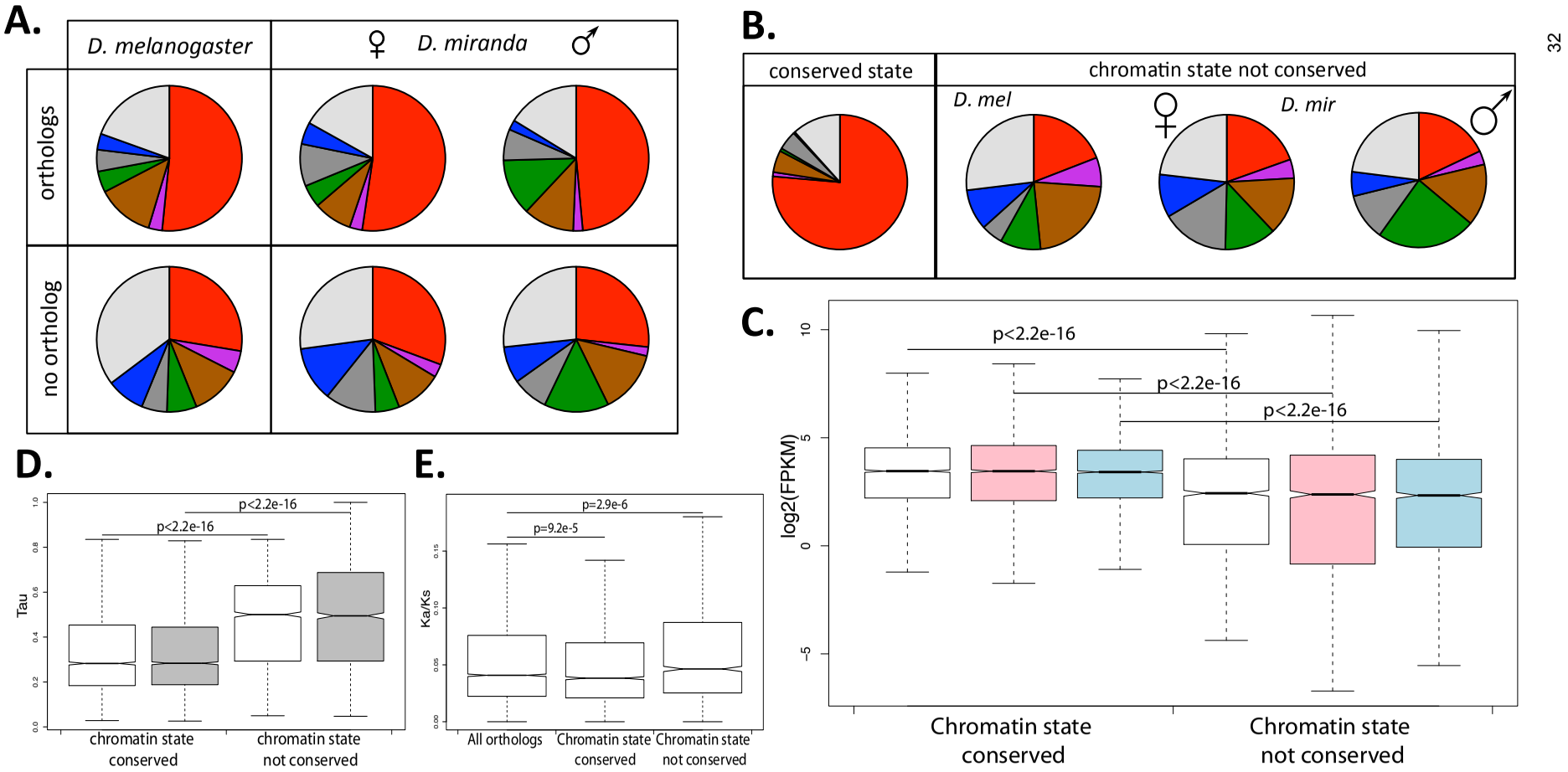
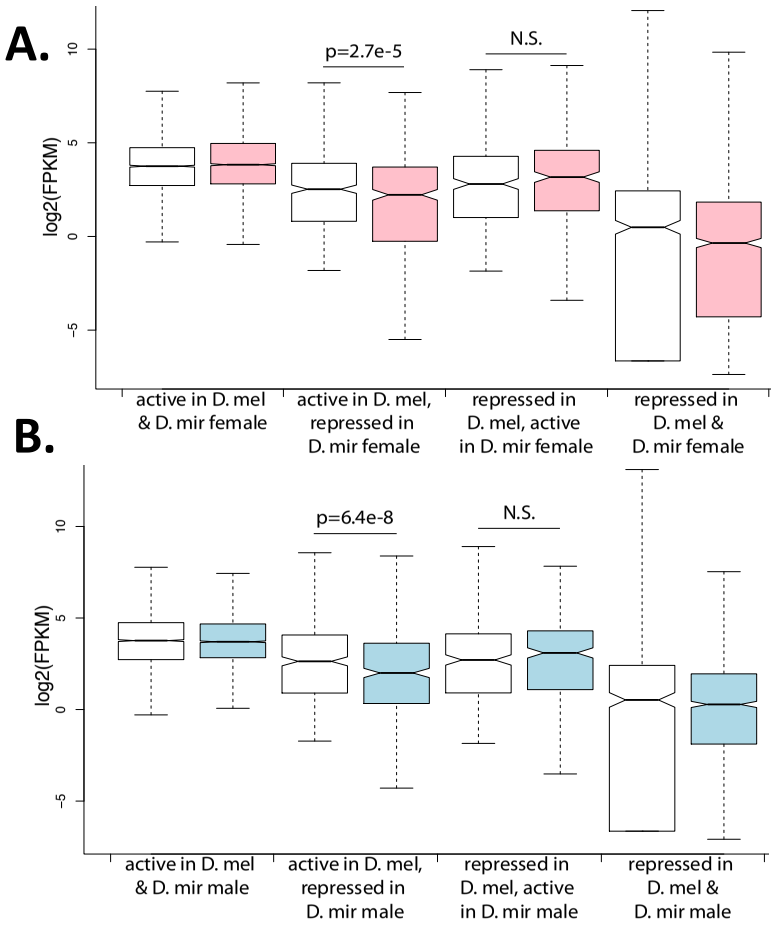

A.
D. melanogaster
♀ D. miranda ♂
orthologs
no ortholog
B.
conserved state
chromatin state not conserved
D. mel
D. mir
C.
p<2.2e-16
p<2.2e-16
p<2.2e-16
log2(FPKM)
Chromatin state conserved
Chromatin state not conserved
D.
p<2.2e-16
p<2.2e-16
Tau
chromatin state conserved
chromatin state not conserved
E.
p=2.9e-6
p=9.2e-5
Ka/Ks
All orthologs
Chromatin state conserved
Chromatin state not conserved

Figure 6


A.
log2(FPKM)
p=2.7e-5
N.S.
active in D. mel & D. mir female
active in D. mel, repressed in D. mir female
repressed in D. mel, active in D. mir female
repressed in D. mel & D. mir female
B.
log2(FPKM)
p=6.4e-8
N.S.
active in D. mel & D. mir male
active in D. mel, repressed in D. mir male
repressed in D. mel, active in D. mir male
repressed in D. mel & D. mir male

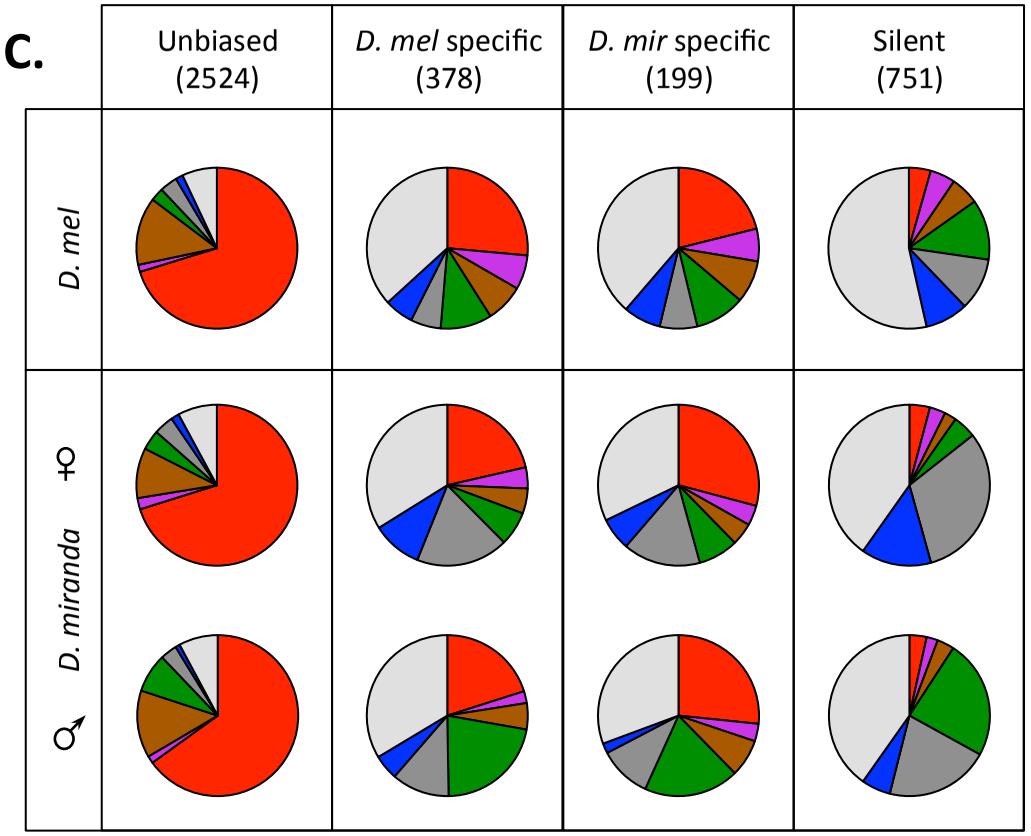
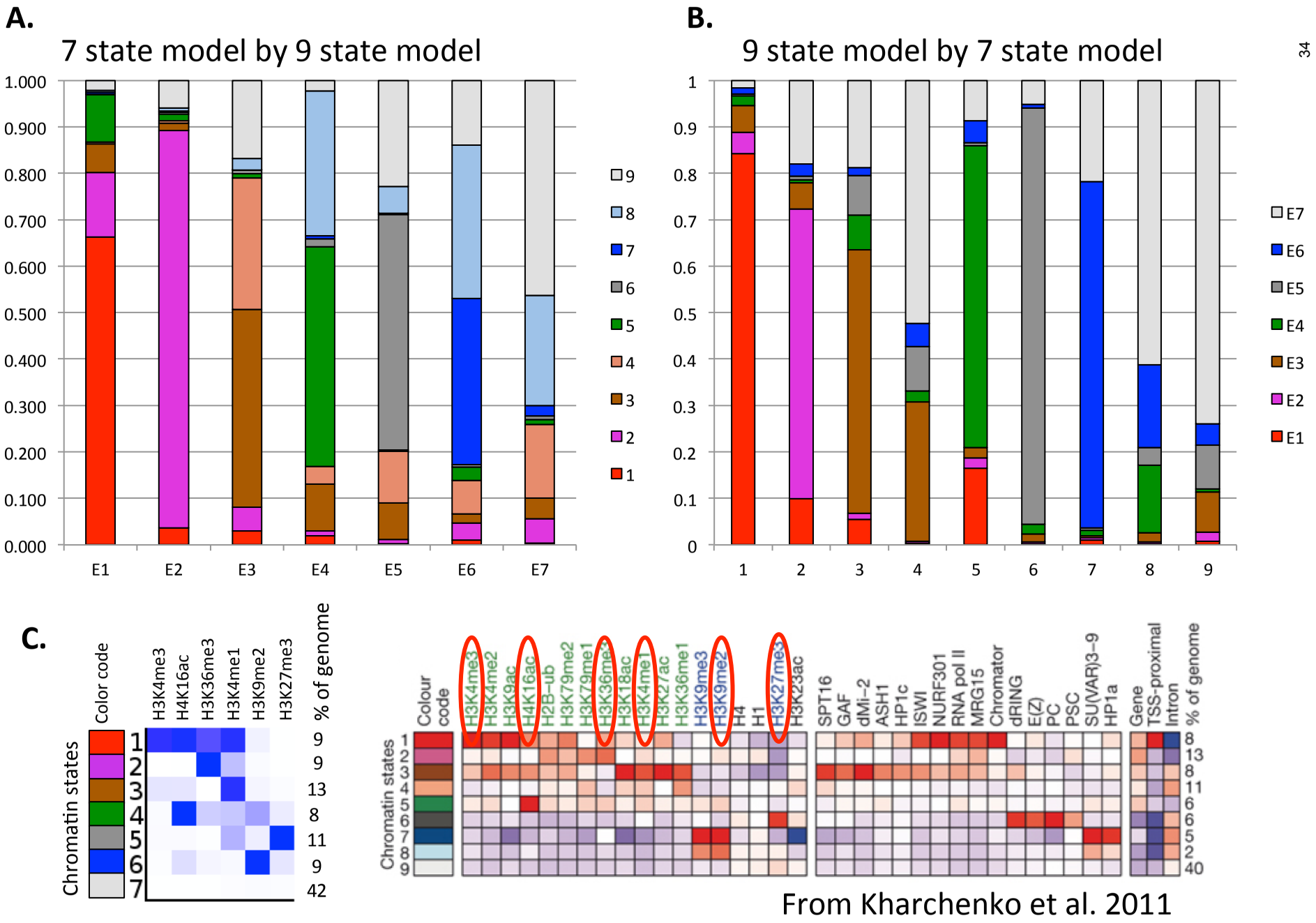
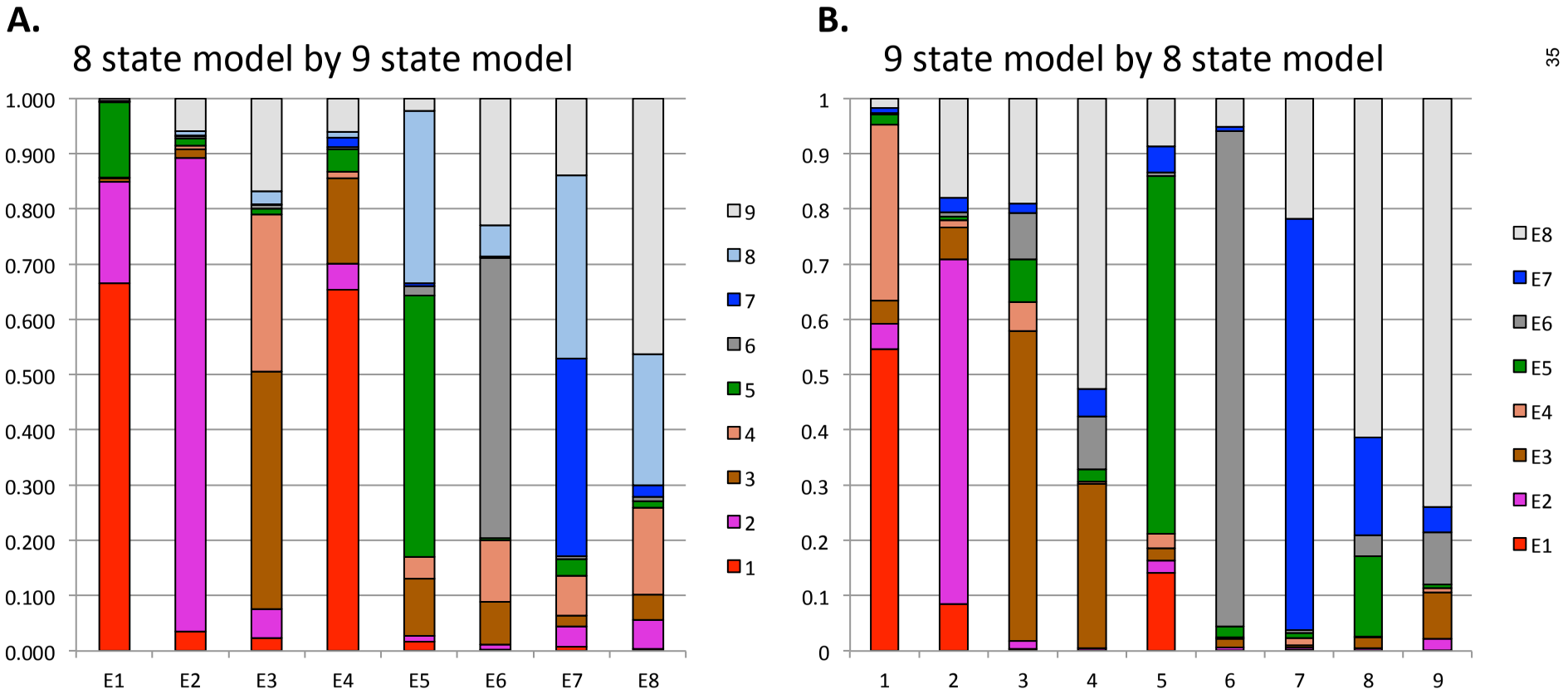

C.
Unbiased (2524)
D. mel specific (378)
D. mir specific (199)
Silent (751)
D. mel
D. miranda ♀
D. miranda ♂

Figure 7



From Kharchenko et al. 2011

Figure S1


A.
8 state model by 9 state model
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
E1
E2
E3
E4
E5
E6
E7
E8
9
8
7
6
5
4
3
2
1
B.
9 state model by 8 state model
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
E8
E7
E6
E5
E4
E3
E2
E1

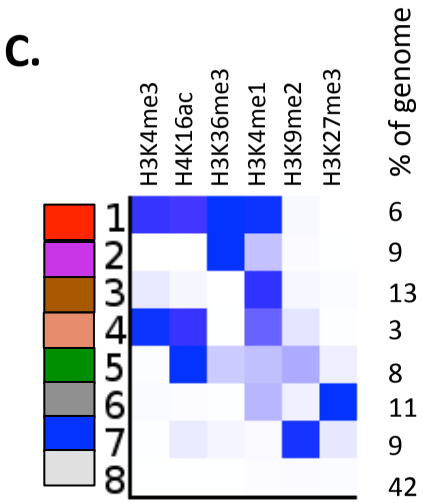
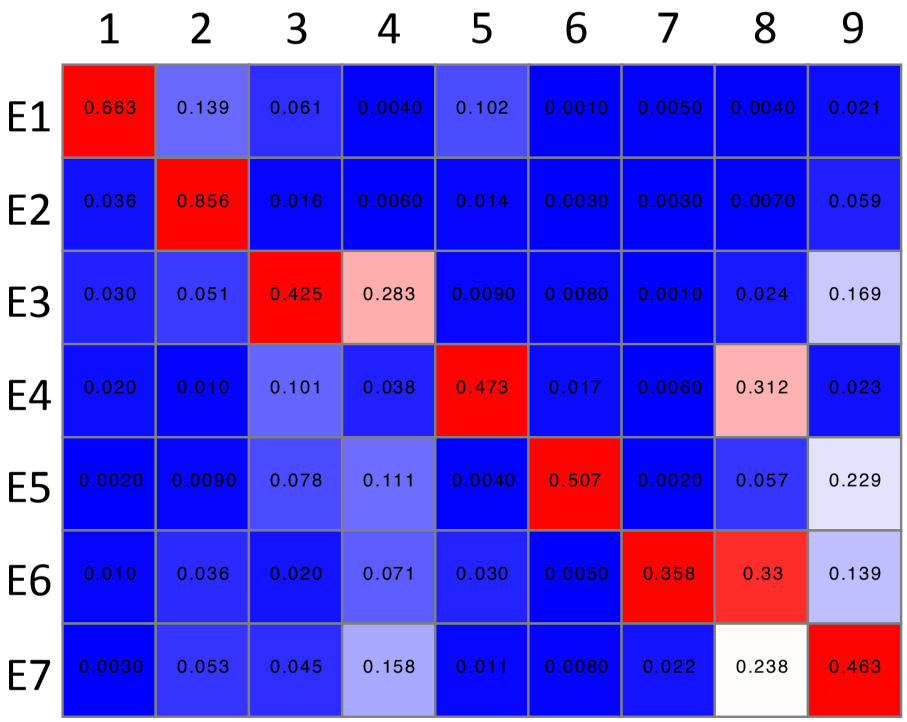

C.
H3K4me3
H4K16ac
H3K36me3
H3K4me1
H3K9me2
H3K27me3
% of genome
1
2
3
4
5
6
7
8
6
9
13
3
8
11
9
42

Figure S2

A.


1 2 3 4 5 6 7 8 9
E1 0.663 0.139 0.061 0.0040 0.102 0.0010 0.0050 0.0040 0.021
E2 0.036 0.856 0.016 0.0060 0.014 0.0030 0.0030 0.0070 0.059
E3 0.030 0.051 0.425 0.283 0.0090 0.0080 0.0010 0.024 0.169
E4 0.020 0.010 0.101 0.038 0.473 0.017 0.0060 0.312 0.023
E5 0.0020 0.0090 0.078 0.111 0.0040 0.507 0.0020 0.057 0.229
E6 0.010 0.036 0.020 0.071 0.030 0.0050 0.358 0.33 0.139
E7 0.0030 0.053 0.045 0.158 0.011 0.0080 0.022 0.238 0.463

B.

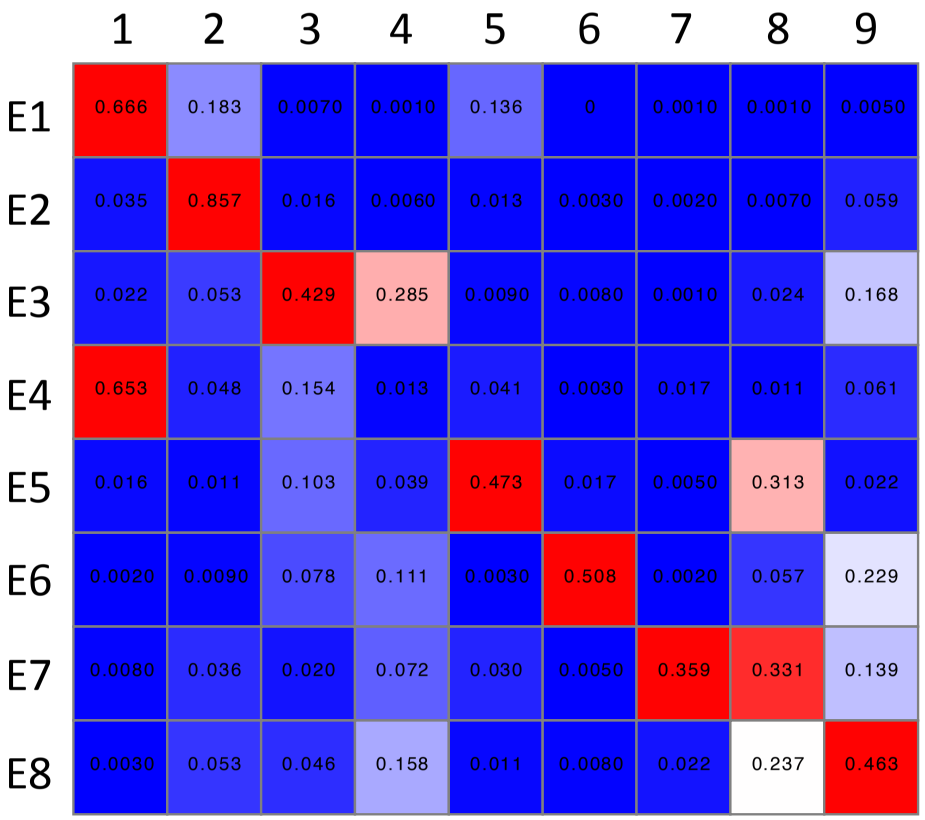
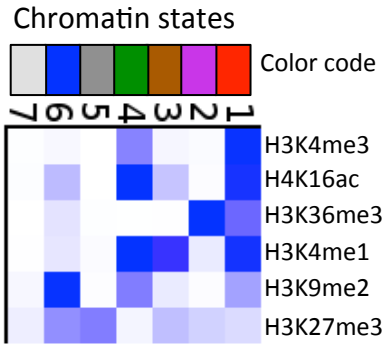

1 2 3 4 5 6 7 8 9
E1 0.666 0.183 0.0070 0.0010 0.136 0 0.0010 0.0010 0.0050
E2 0.035 0.857 0.016 0.0060 0.013 0.0030 0.0020 0.0070 0.059
E3 0.022 0.053 0.429 0.285 0.0090 0.0080 0.0010 0.024 0.168
E4 0.653 0.048 0.154 0.013 0.041 0.0030 0.017 0.011 0.061
E5 0.016 0.011 0.103 0.039 0.473 0.017 0.0050 0.313 0.022
E6 0.0020 0.0090 0.078 0.111 0.0030 0.508 0.0020 0.057 0.229
E7 0.0080 0.036 0.020 0.072 0.030 0.0050 0.359 0.331 0.139
E8 0.0030 0.053 0.046 0.158 0.011 0.0080 0.022 0.237 0.463

Figure S3

Figure S4


Chromatin states
Color code
1
2
3
4
5
6
7
H3K4me3
H4K16ac
H3K36me3
H3K4me1
H3K9me2
H3K27me3

Normalization procedure

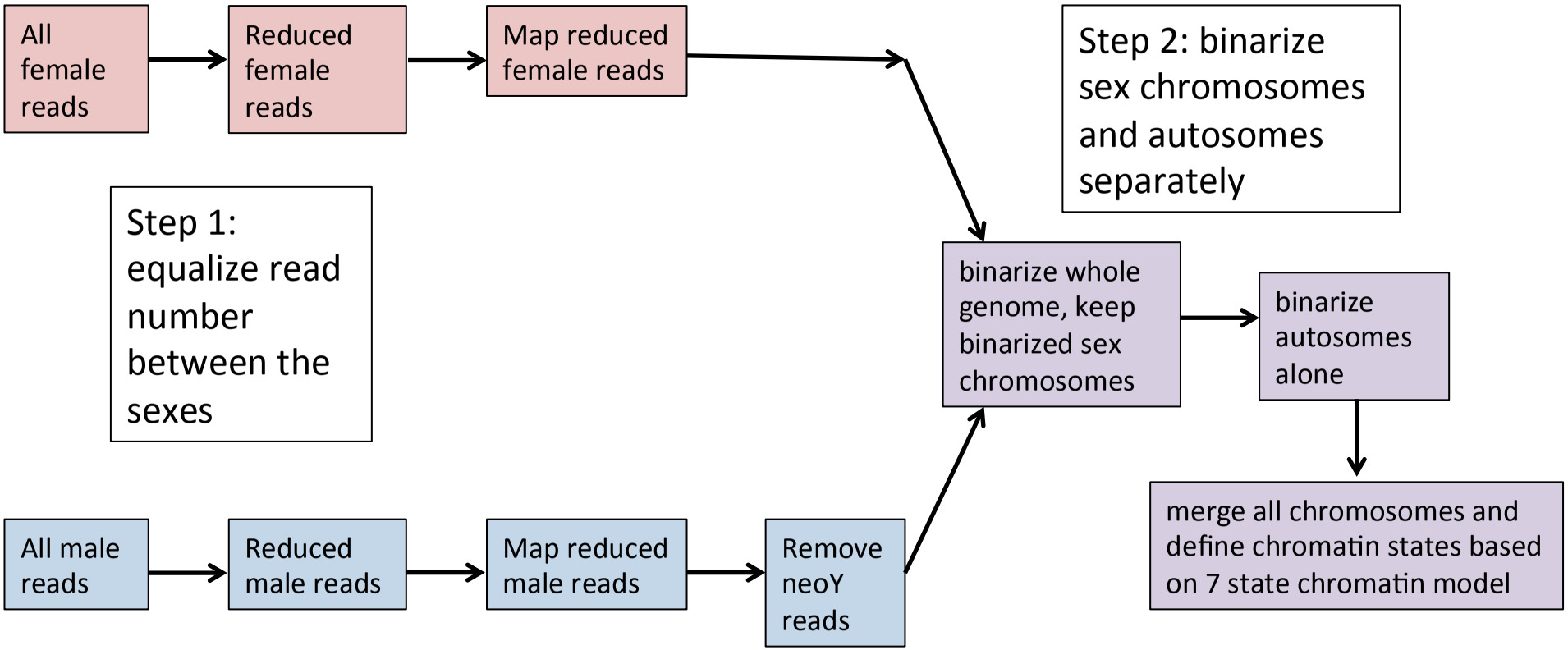


Figure S5

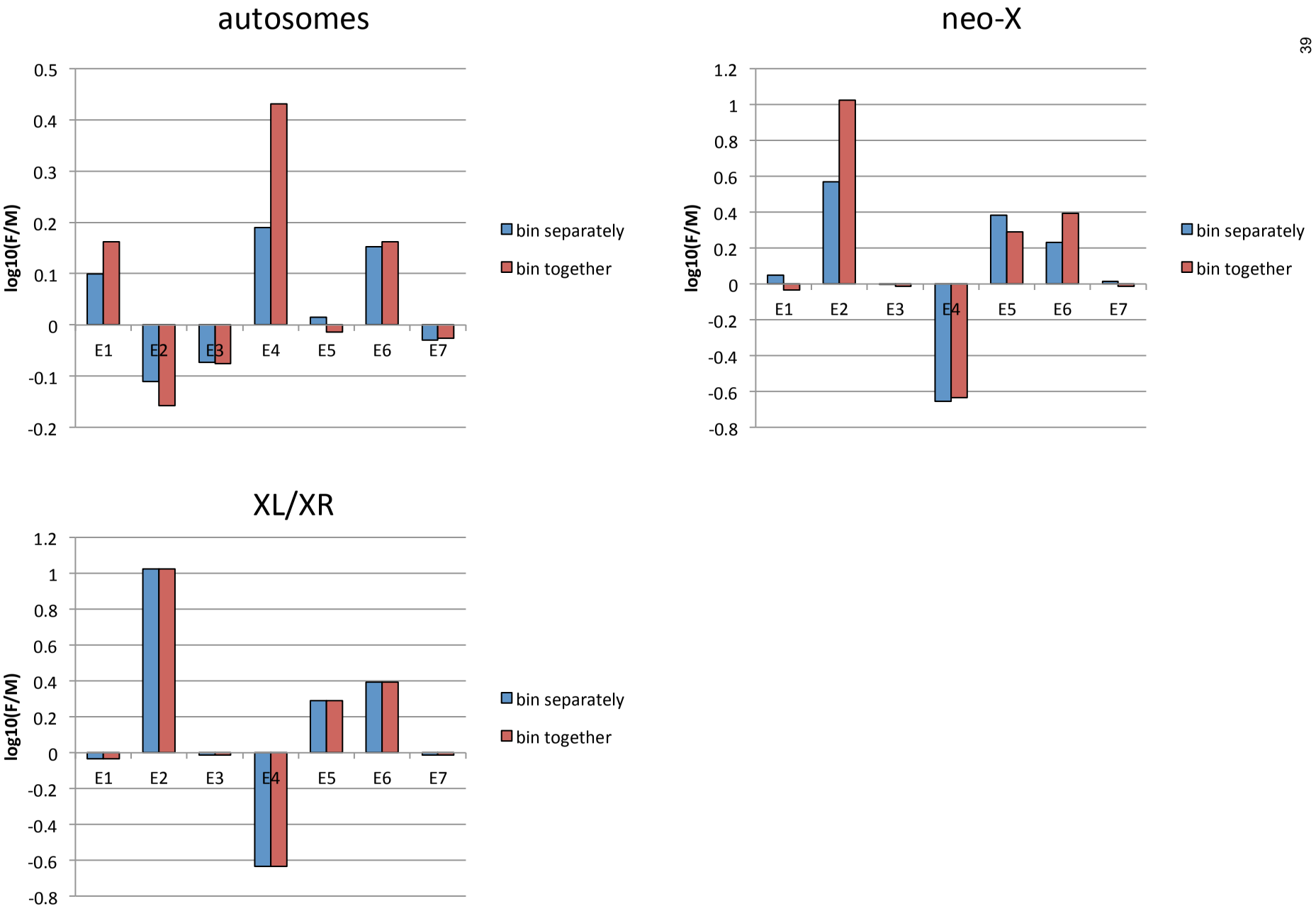
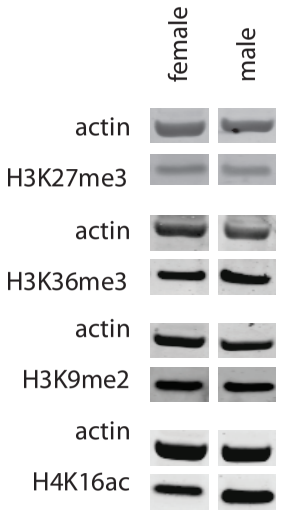

autosomes
log10(F/M)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
E1
E2
E3
E4
E5
E6
E7
bin separately
bin together
neo-X
log10(F/M)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
E1
E2
E3
E4
E5
E6
E7
bin separately
bin together
XL/XR
log10(F/M)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
E1
E2
E3
E4
E5
E6
E7
bin separately
bin together

Figure S6


female
male
actin
H3K27me3
actin
H3K36me3
actin
H3K9me2
actin
H4K16ac

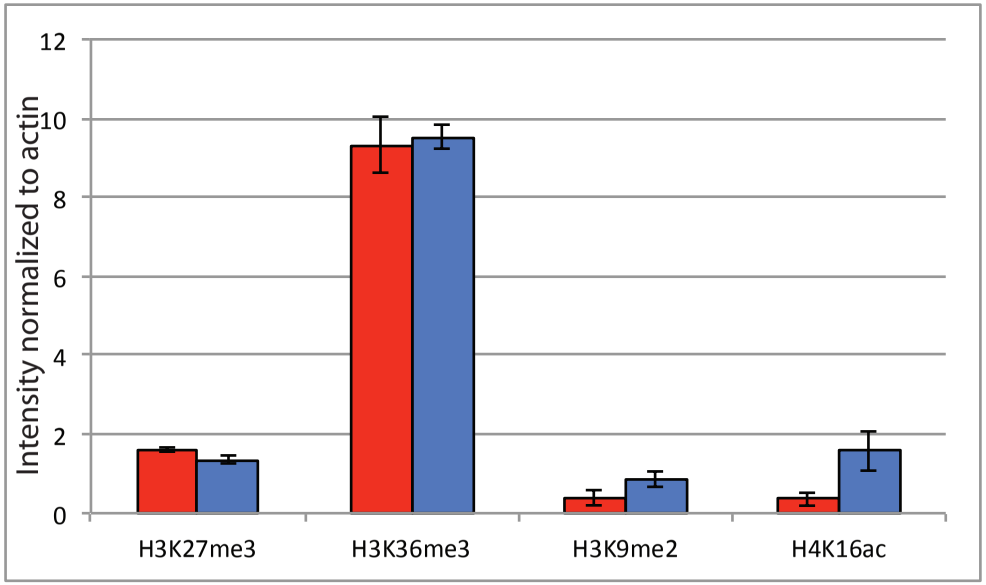

Intensity normalized to actin
12
10
8
6
4
2
0
H3K27me3
H3K36me3
H3K9me2
H4K16ac

Figure S7

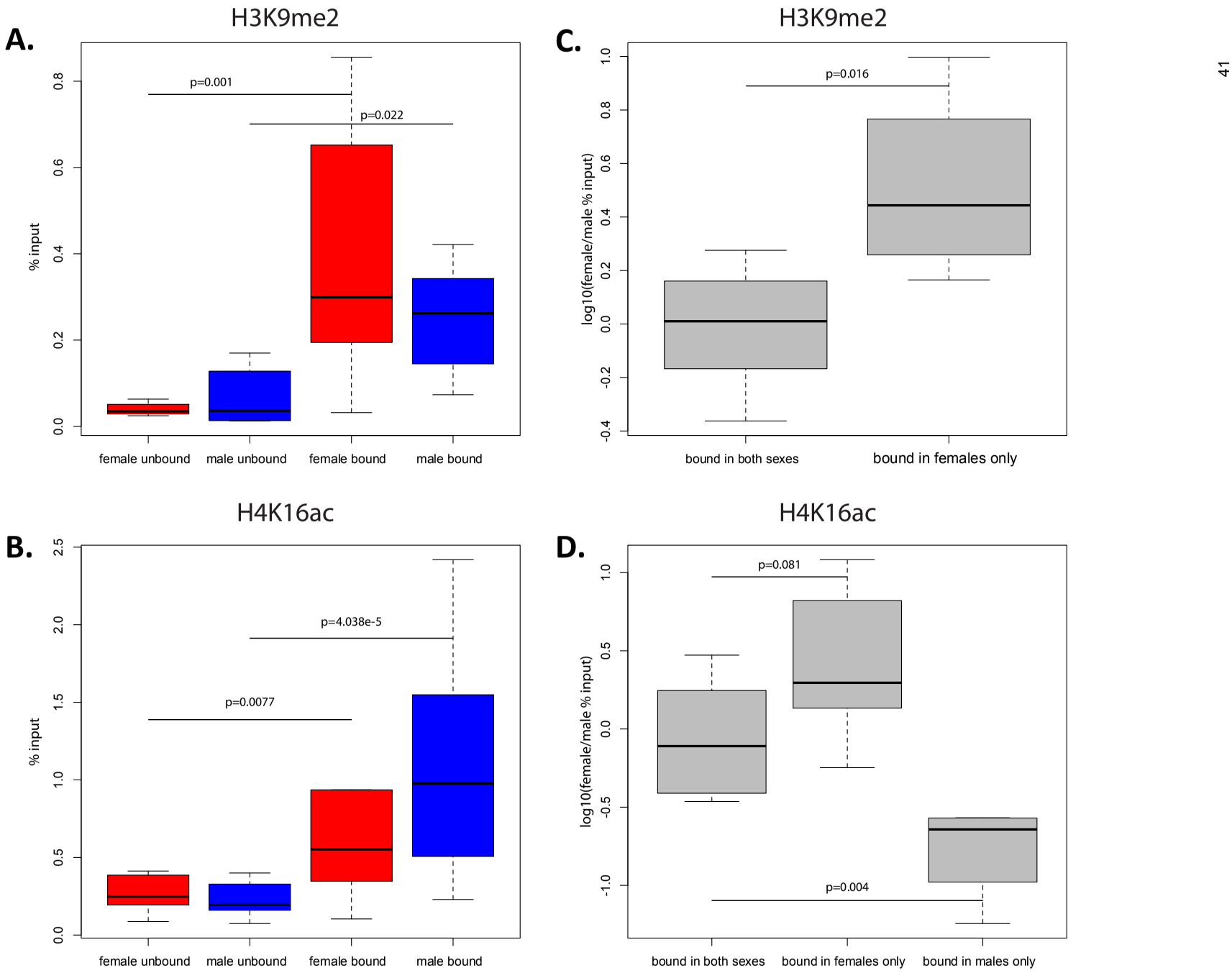
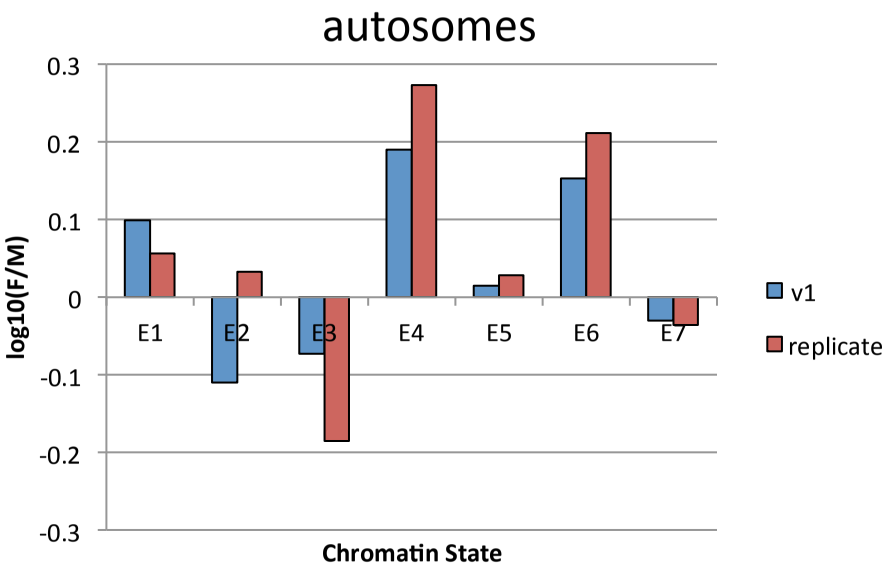
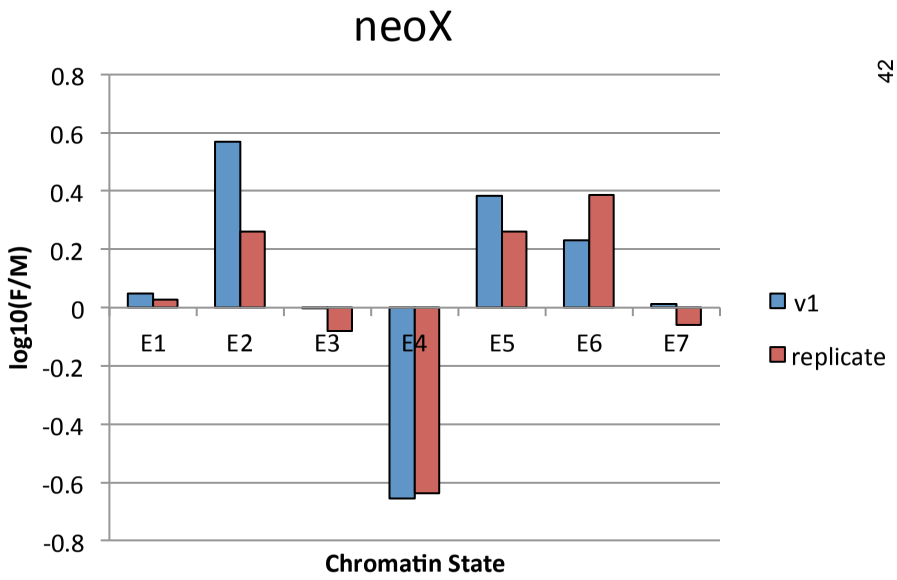

A.
H3K9me2
% input
p=0.001
p=0.022
female unbound
male unbound
female bound
male bound
C.
H3K9me2
log10(female/male % input)
p=0.016
bound in both sexes
bound in females only
B.
H4K16ac
% input
p=4.038e-5
p=0.0077
female unbound
male unbound
female bound
male bound
D.
H4K16ac
log10(female/male % input)
p=0.081
p=0.004
bound in both sexes
bound in females only
bound in males only

Figure S8


autosomes
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
log10(F/M)
E1
E2
E3
E4
E5
E6
E7
Chromatin State
v1
replicate


neoX
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
log10(F/M)
E1
E2
E3
E4
E5
E6
E7
Chromatin State
v1
replicate

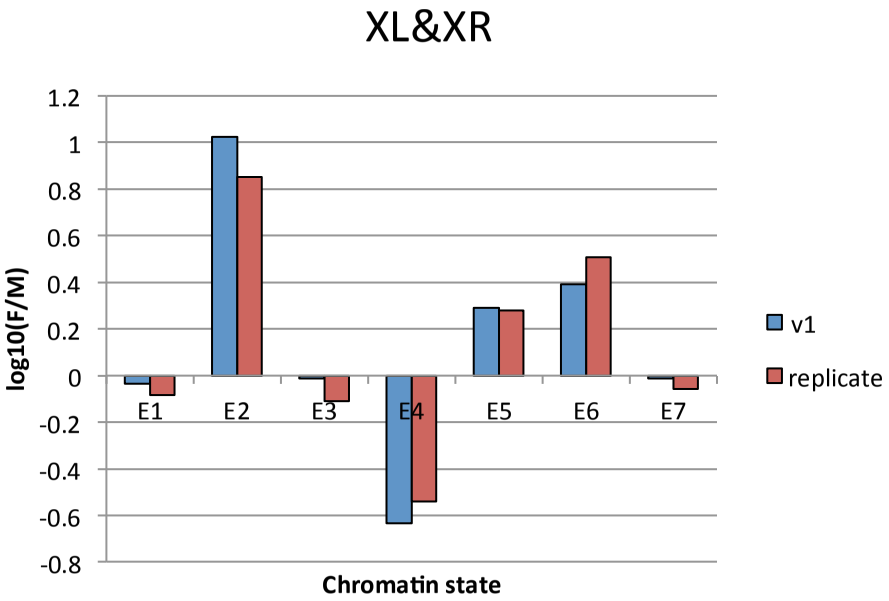

XL&XR
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
log10(F/M)
E1
E2
E3
E4
E5
E6
E7
Chromatin state
v1
replicate

Figure S9

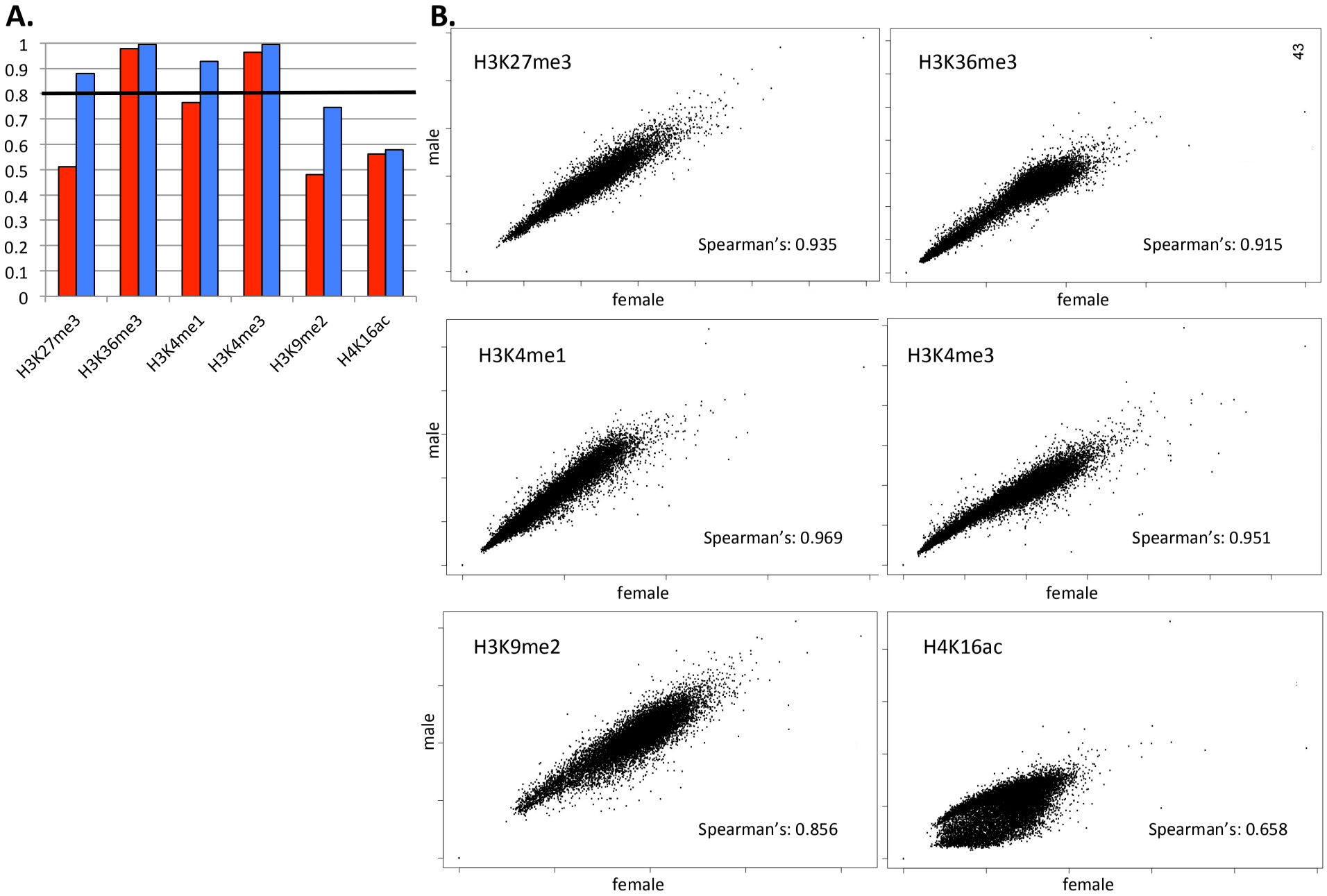
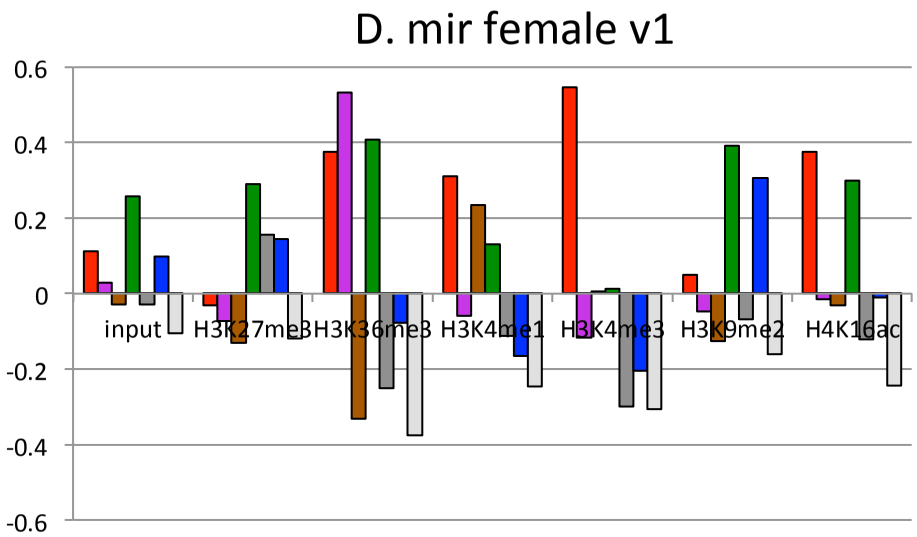
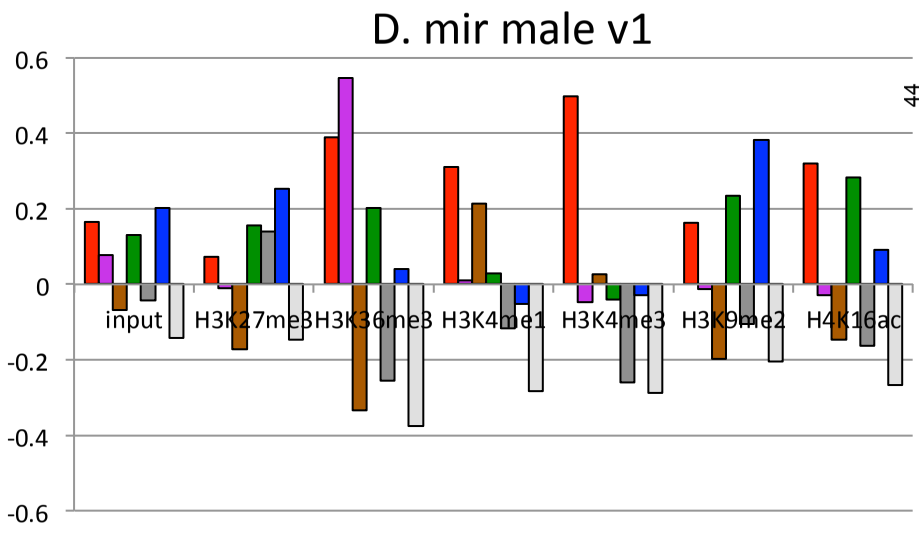
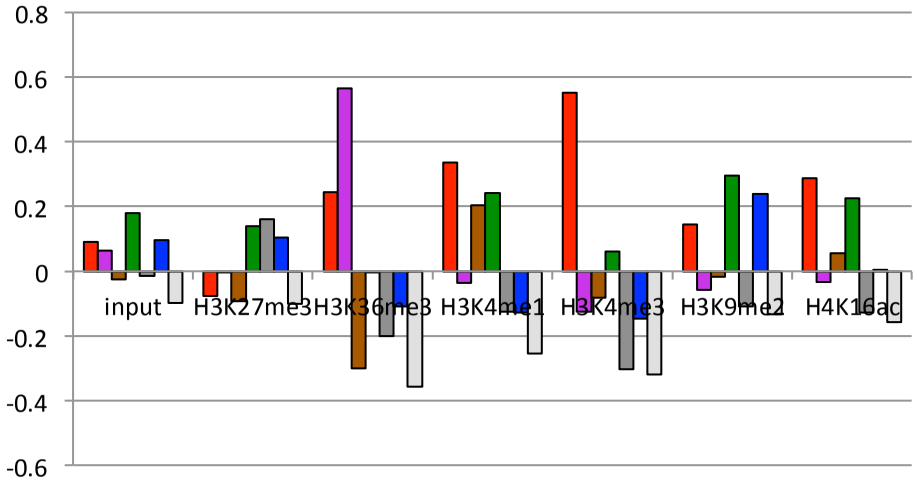

A.
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac
B.
H3K27me3
Spearman's: 0.935
H3K36me3
Spearman's: 0.915
H3K4me1
Spearman's: 0.969
H3K4me3
Spearman's: 0.951
H3K9me2
Spearman's: 0.856
H4K16ac
Spearman's: 0.658
male
female
43

Figure S10


D. mir female v1
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
input
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac


D. mir male v1
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
input
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac


D. mir female replicate
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
input
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac

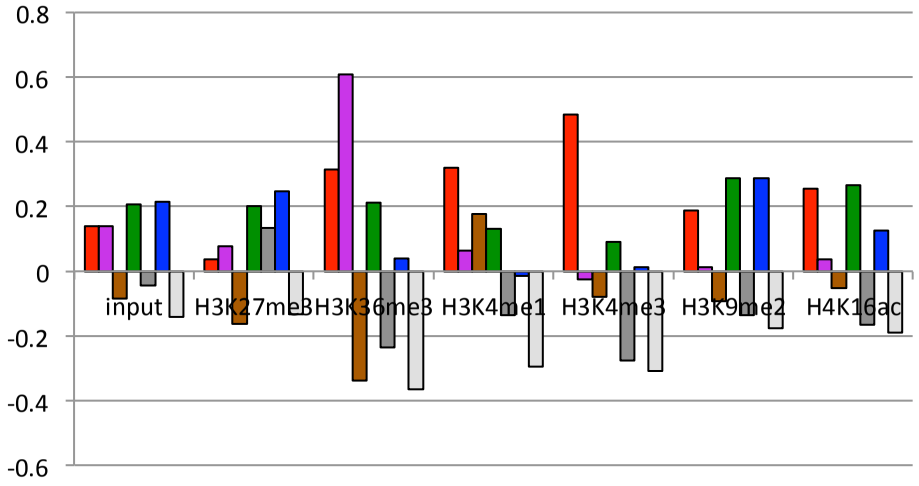

D. mir male replicate
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
input
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac

Figure S11

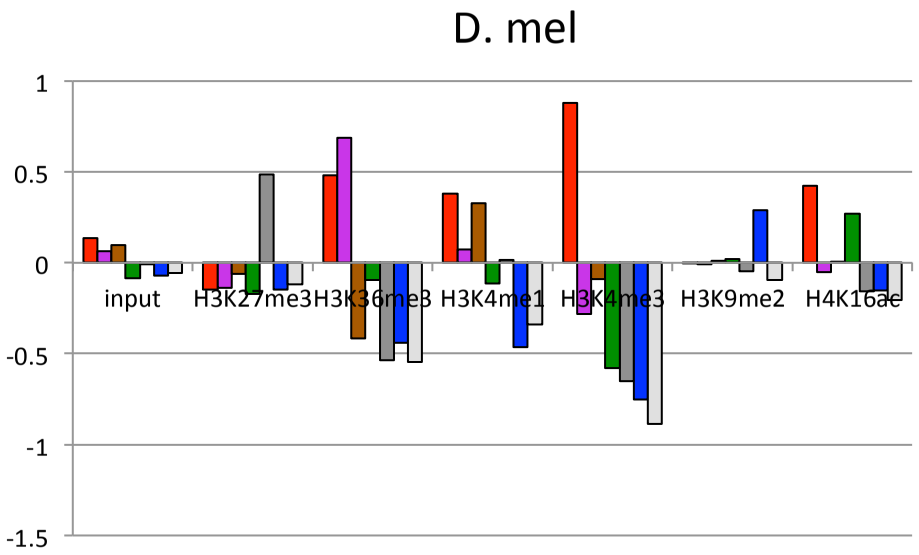

D. mel
1
0.5
0
-0.5
-1
-1.5
input
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me2
H4K16ac

Figure S12

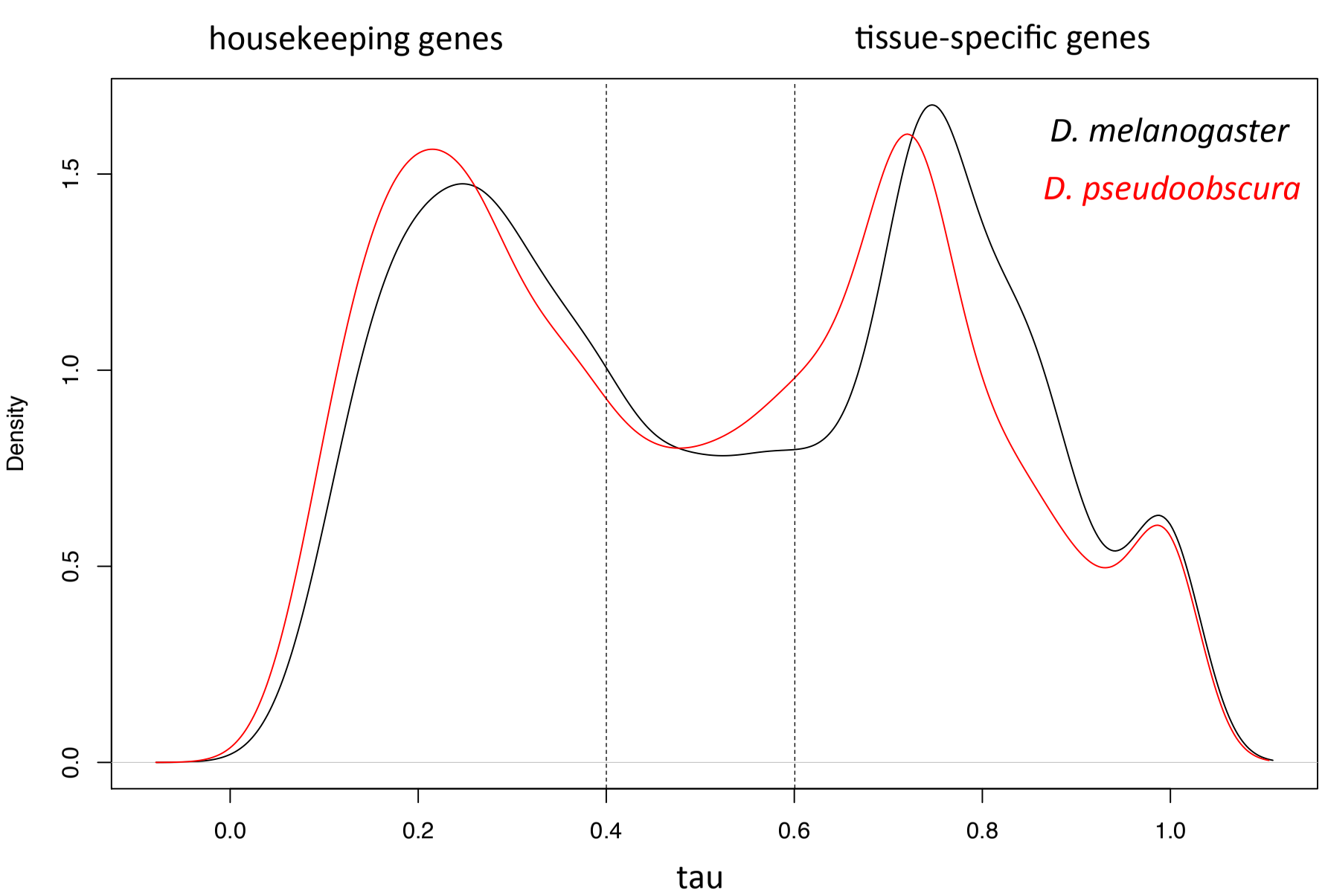

housekeeping genes
tissue-specific genes
D. melanogaster
D. pseudoobscura
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
tau

Figure S13

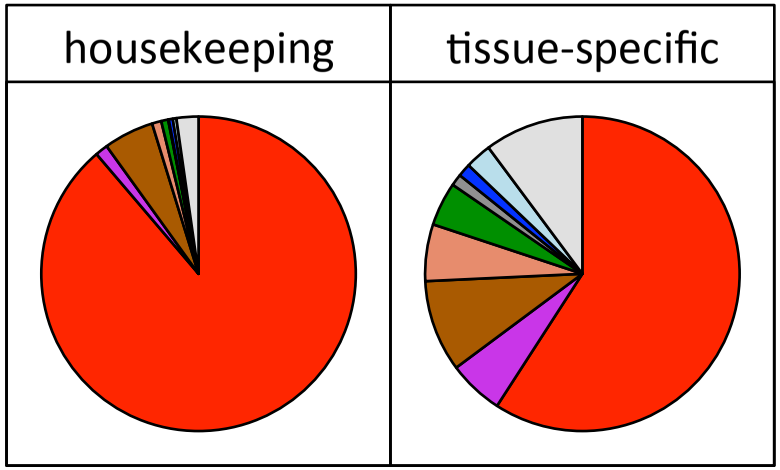


Figure S14

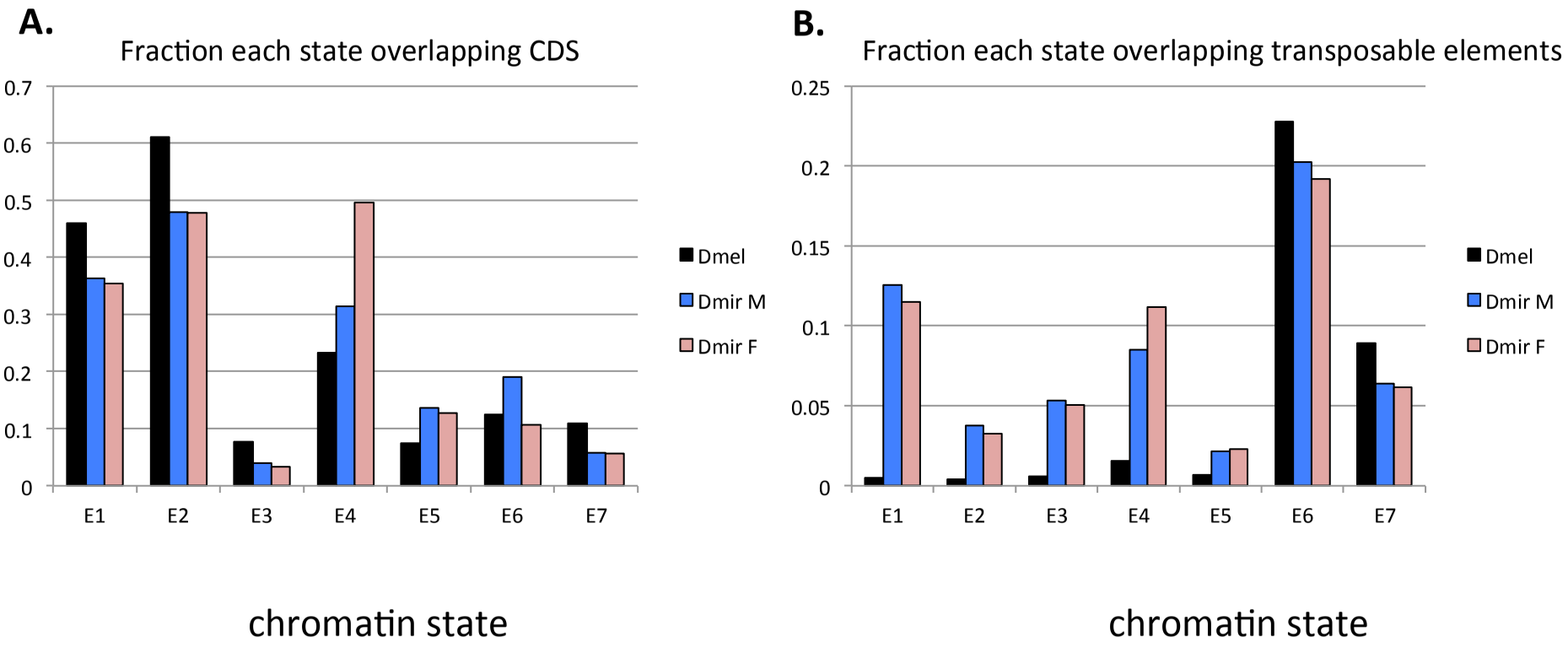

A.
Fraction each state overlapping CDS
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E1
E2
E3
E4
E5
E6
E7
Dmel
Dmir M
Dmir F
chromatin state
B.
Fraction each state overlapping transposable elements
0.25
0.2
0.15
0.1
0.05
0
E1
E2
E3
E4
E5
E6
E7
Dmel
Dmir M
Dmir F
chromatin state

Figure S15

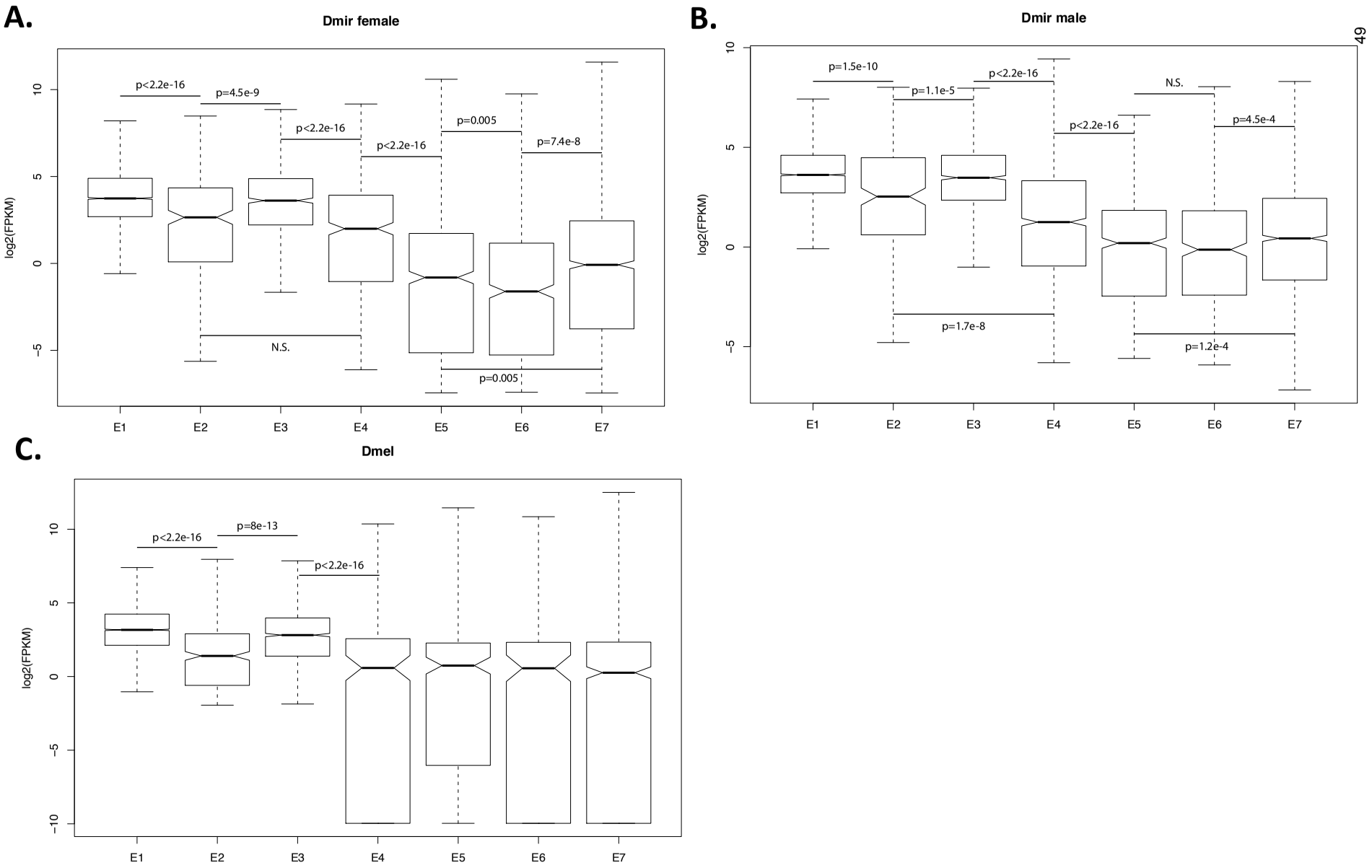

A.
Dmir female
log2(FPKM)
p<2.2e-16
p=4.5e-9
p<2.2e-16
p<2.2e-16
p=0.005
p=7.4e-8
N.S.
p=0.005
E1
E2
E3
E4
E5
E6
E7
B.
Dmir male
log2(FPKM)
p=1.5e-10
p=1.1e-5
p<2.2e-16
p<2.2e-16
N.S.
p=4.5e-4
p=1.7e-8
p=1.2e-4
E1
E2
E3
E4
E5
E6
E7
C.
Dmel
log2(FPKM)
p<2.2e-16
p=8e-13
p<2.2e-16
E1
E2
E3
E4
E5
E6
E7

Figure S16

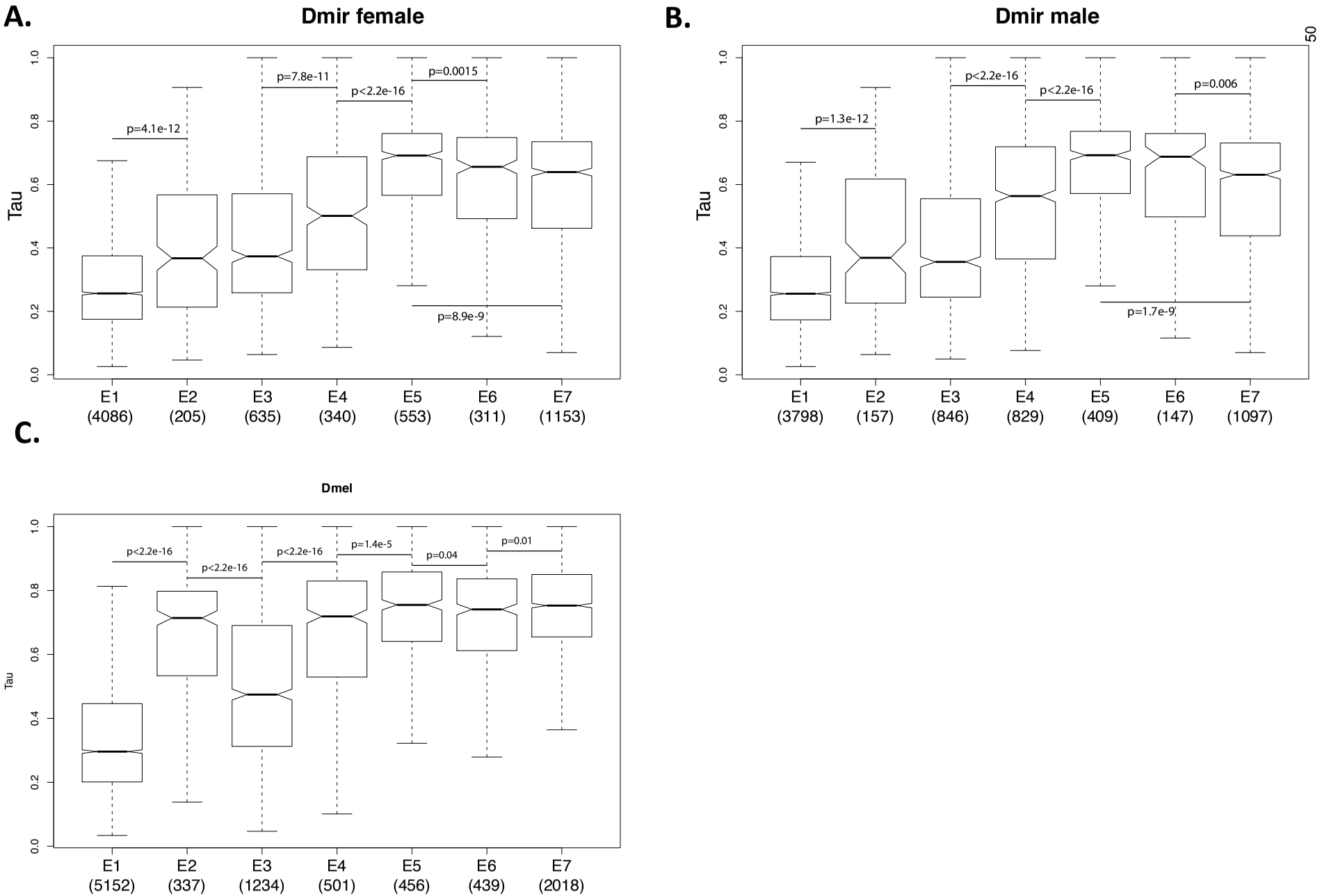

A.
Dmir female
Tau
p=4.1e-12
p=7.8e-11
p<2.2e-16
p=0.0015
p=8.9e-9
E1 (4086)
E2 (205)
E3 (635)
E4 (340)
E5 (553)
E6 (311)
E7 (1153)
B.
Dmir male
p=1.3e-12
p<2.2e-16
p<2.2e-16
p=0.006
p=1.7e-9
E1 (3798)
E2 (157)
E3 (846)
E4 (829)
E5 (409)
E6 (147)
E7 (1097)
C.
Dmel
p<2.2e-16
p<2.2e-16
p<2.2e-16
p=1.4e-5
p=0.04
p=0.01
E1 (5152)
E2 (337)
E3 (1234)
E4 (501)
E5 (456)
E6 (439)
E7 (2018)

Figure S17

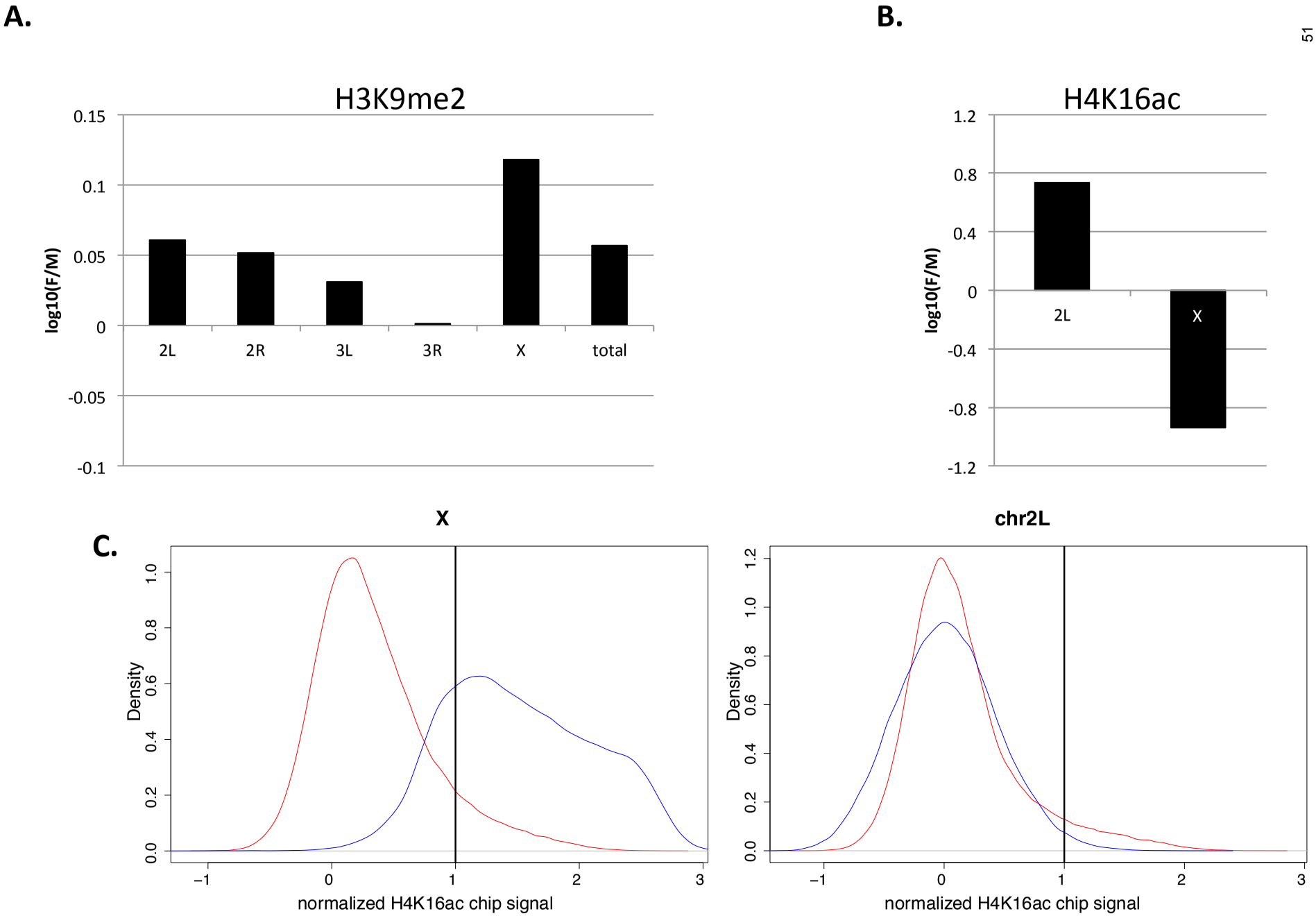

A.
H3K9me2
log10(F/M)
0.15
0.1
0.05
0
-0.05
-0.1
2L
2R
3L
3R
X
total
B.
H4K16ac
log10(F/M)
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
2L
X
C.
X
chr2L
Density
normalized H4K16ac chip signal

Figure S18

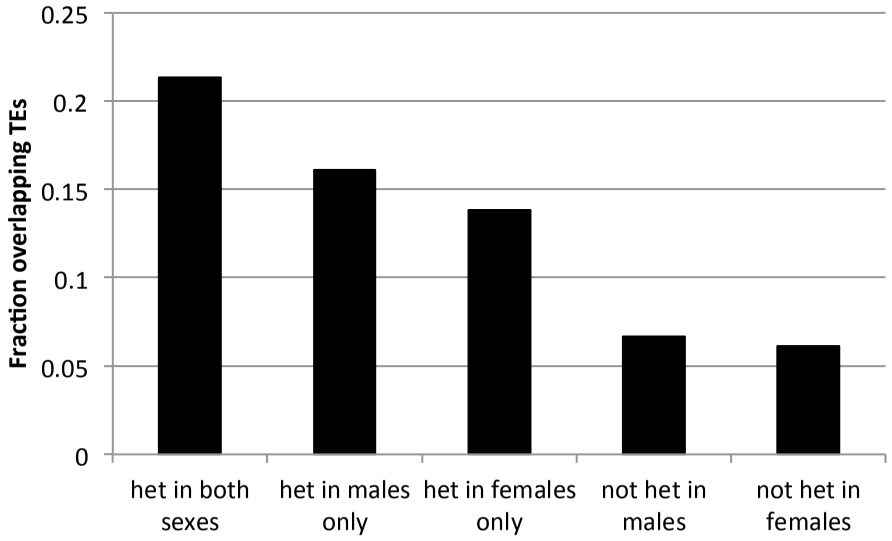

0.25
0.2
0.15
0.1
0.05
0
Fraction overlapping TEs
het in both sexes
het in males only
het in females only
not het in males
not het in females

Figure S19

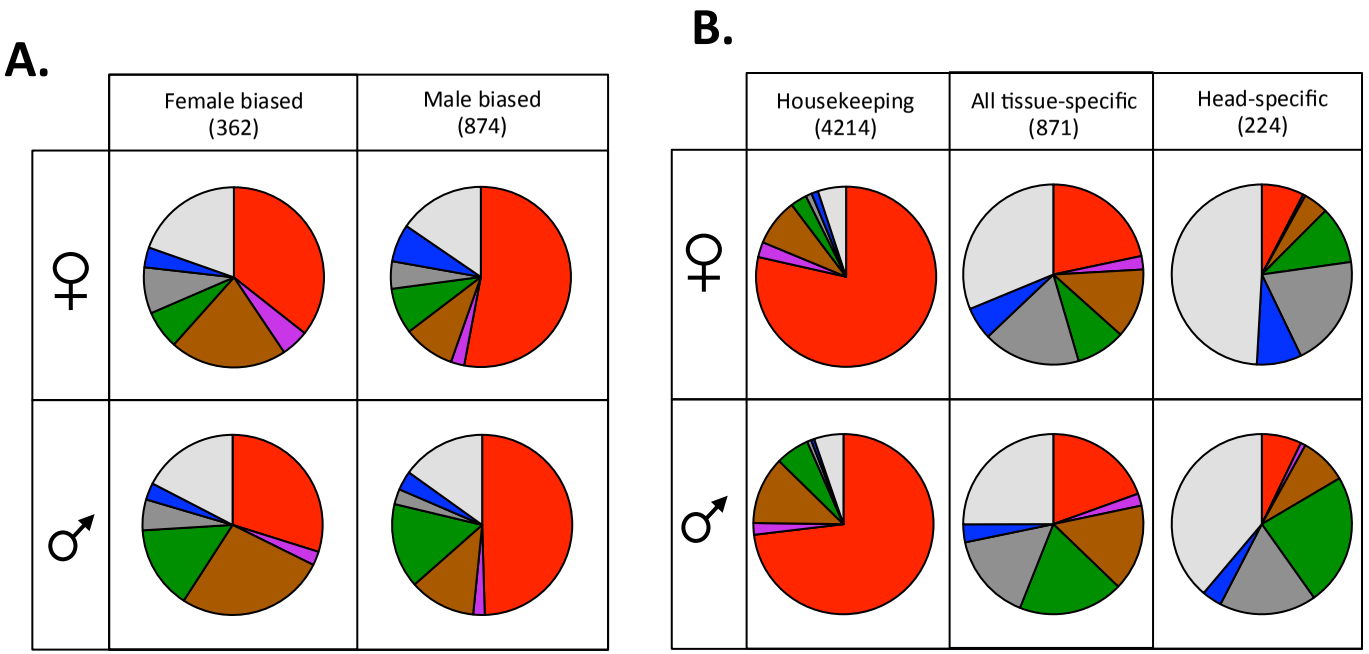

A.
Female biased
(362)
Male biased
(874)
♀
♂
B.
Housekeeping
(4214)
All tissue-specific
(871)
Head-specific
(224)
♀
♂

Figure S20

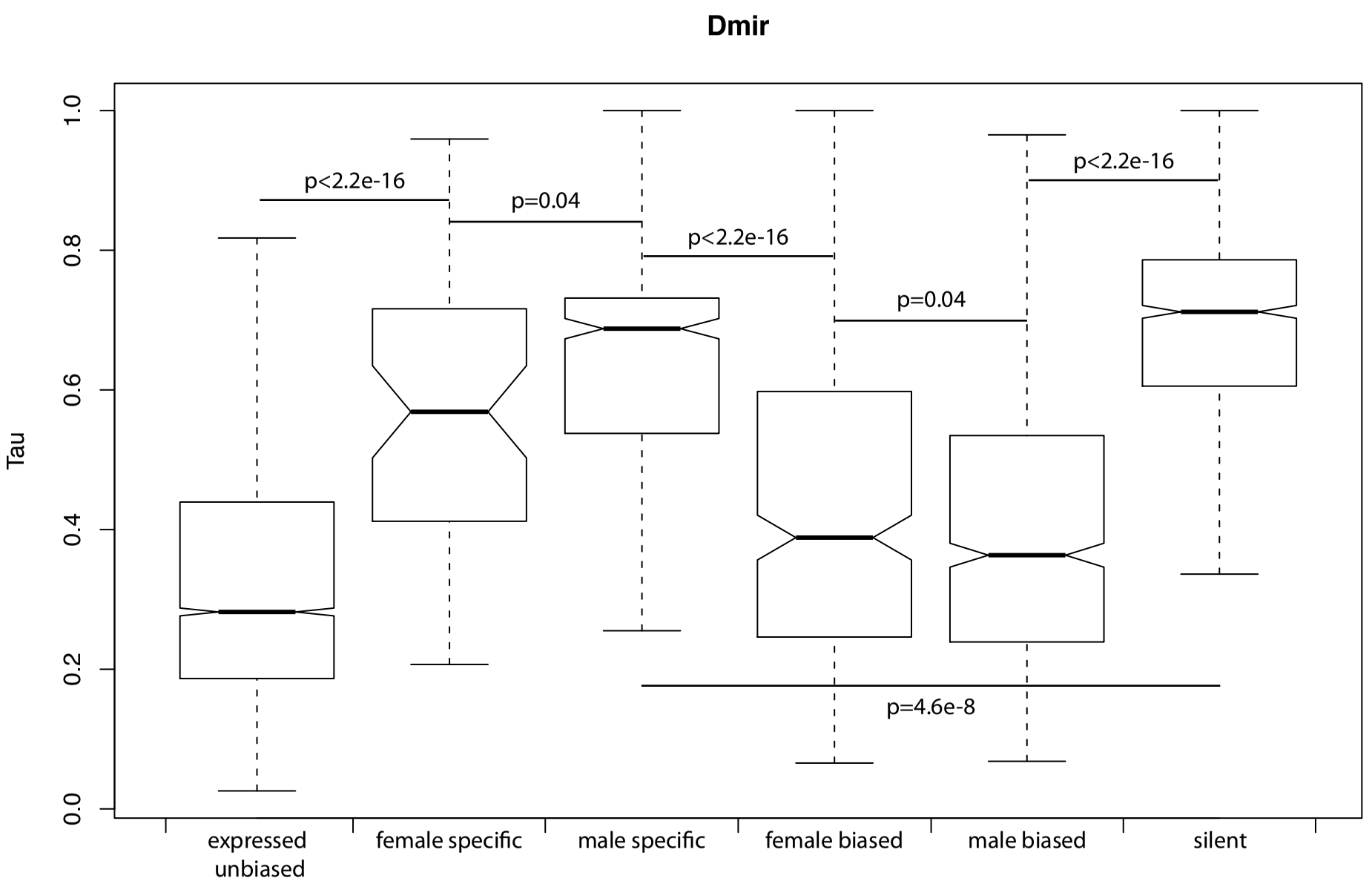
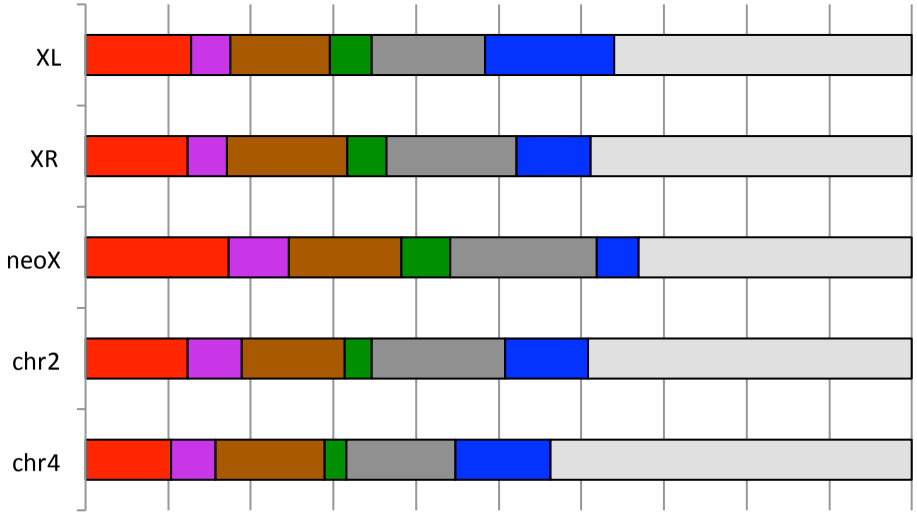
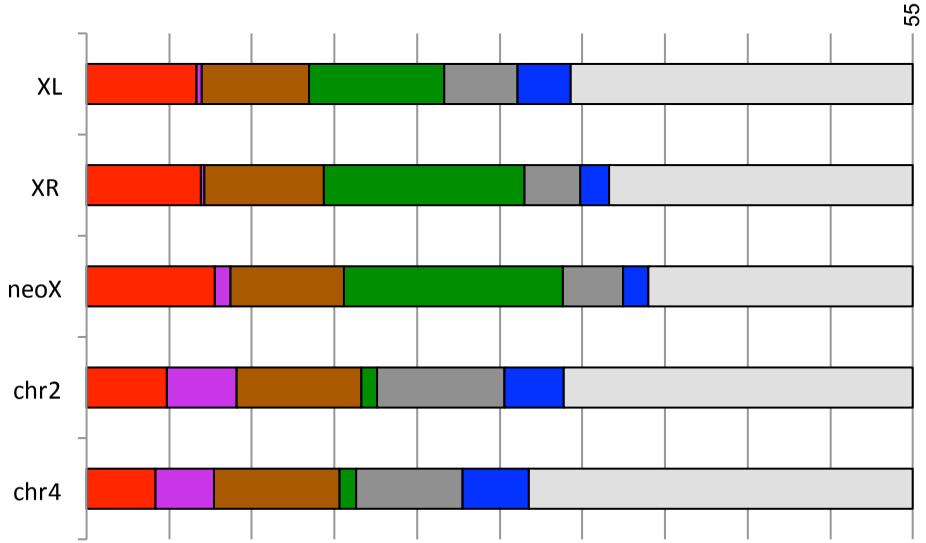

Dmir
Tau
0.0
0.2
0.4
0.6
0.8
1.0
p<2.2e-16
p=0.04
p<2.2e-16
p=0.04
p<2.2e-16
p=4.6e-8
expressed unbiased
female specific
male specific
female biased
male biased
silent

Figure S21

A. *D. miranda* female



B. *D. miranda* male



C. *D. melanogaster*

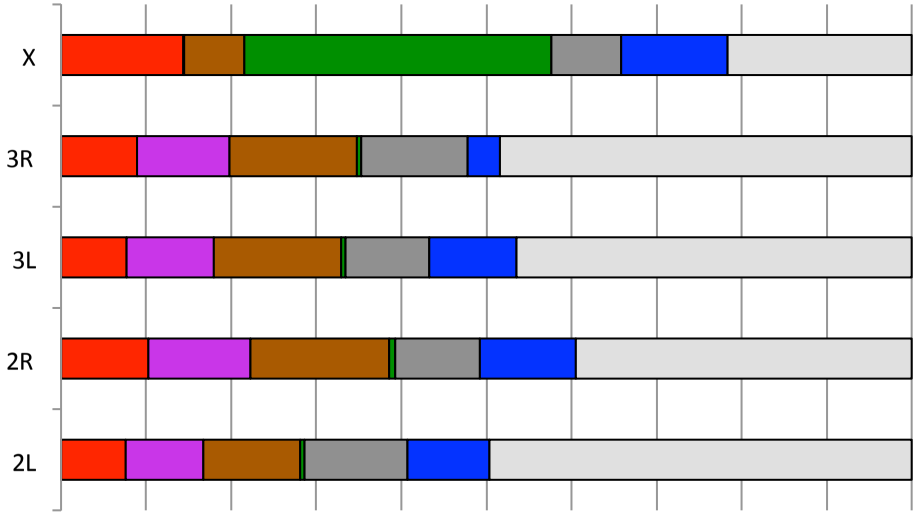
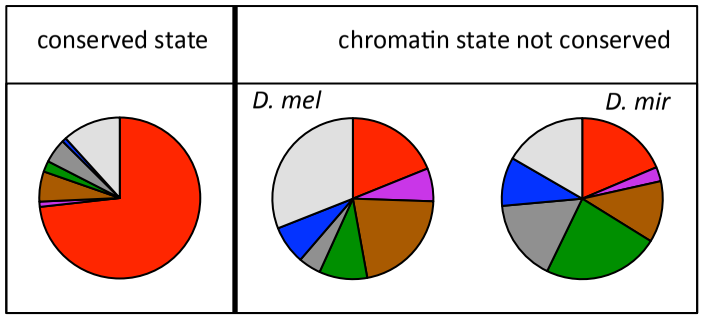
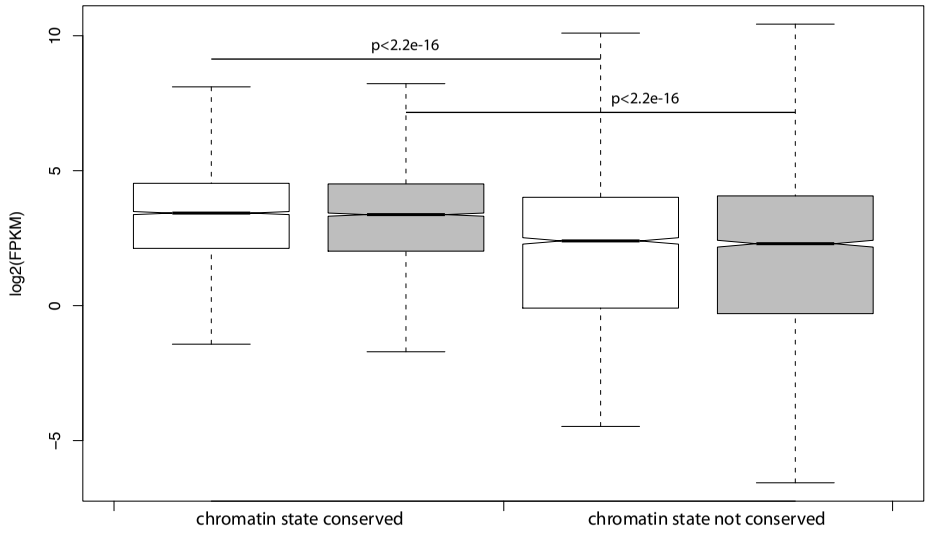
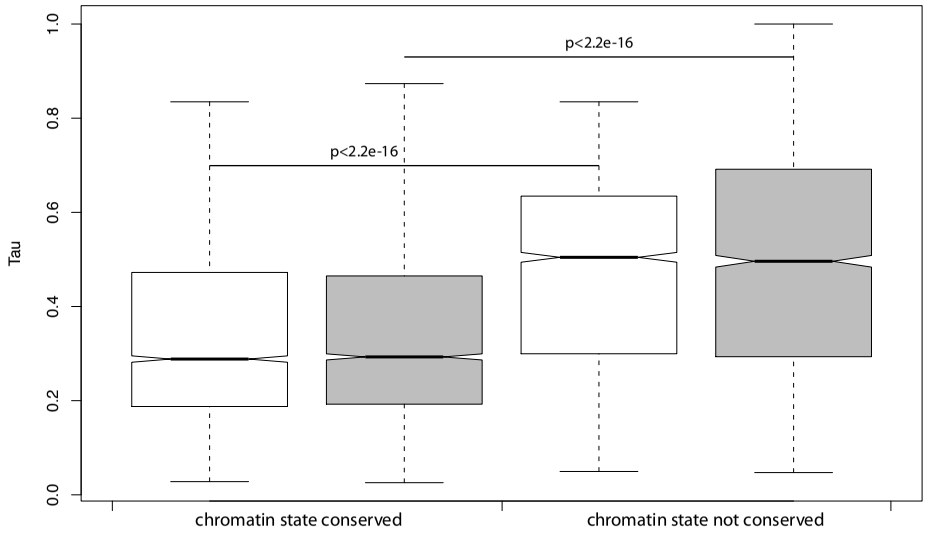


Figure S22


conserved state
chromatin state not conserved
D. mel
D. mir


p<2.2e-16
p<2.2e-16
log2(FPKM)
10
5
0
−5
chromatin state conserved
chromatin state not conserved


p<2.2e-16
p<2.2e-16
Tau
1.0
0.8
0.6
0.4
0.2
0.0
chromatin state conserved
chromatin state not conserved

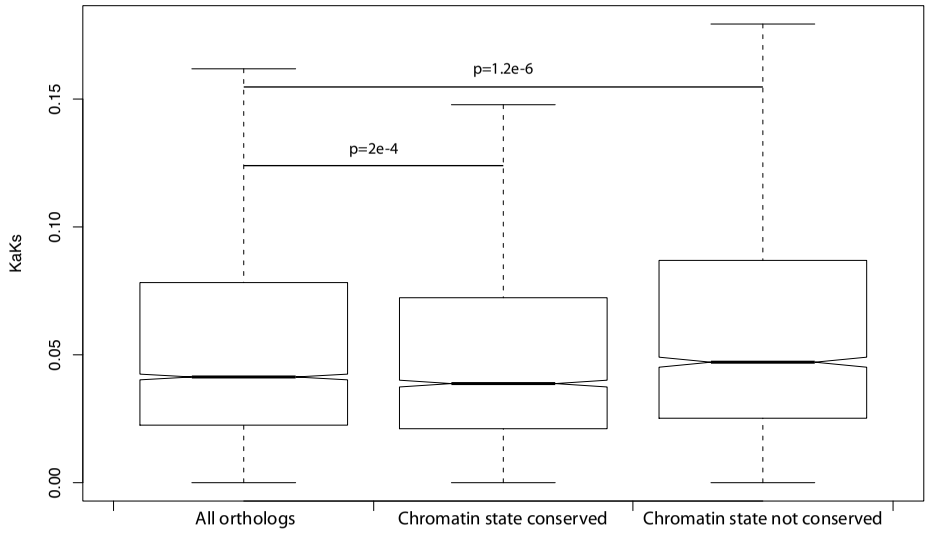

p=1.2e-6
p=2e-4
KaKs
0.15
0.10
0.05
0.00
All orthologs
Chromatin state conserved
Chromatin state not conserved

Figure S23

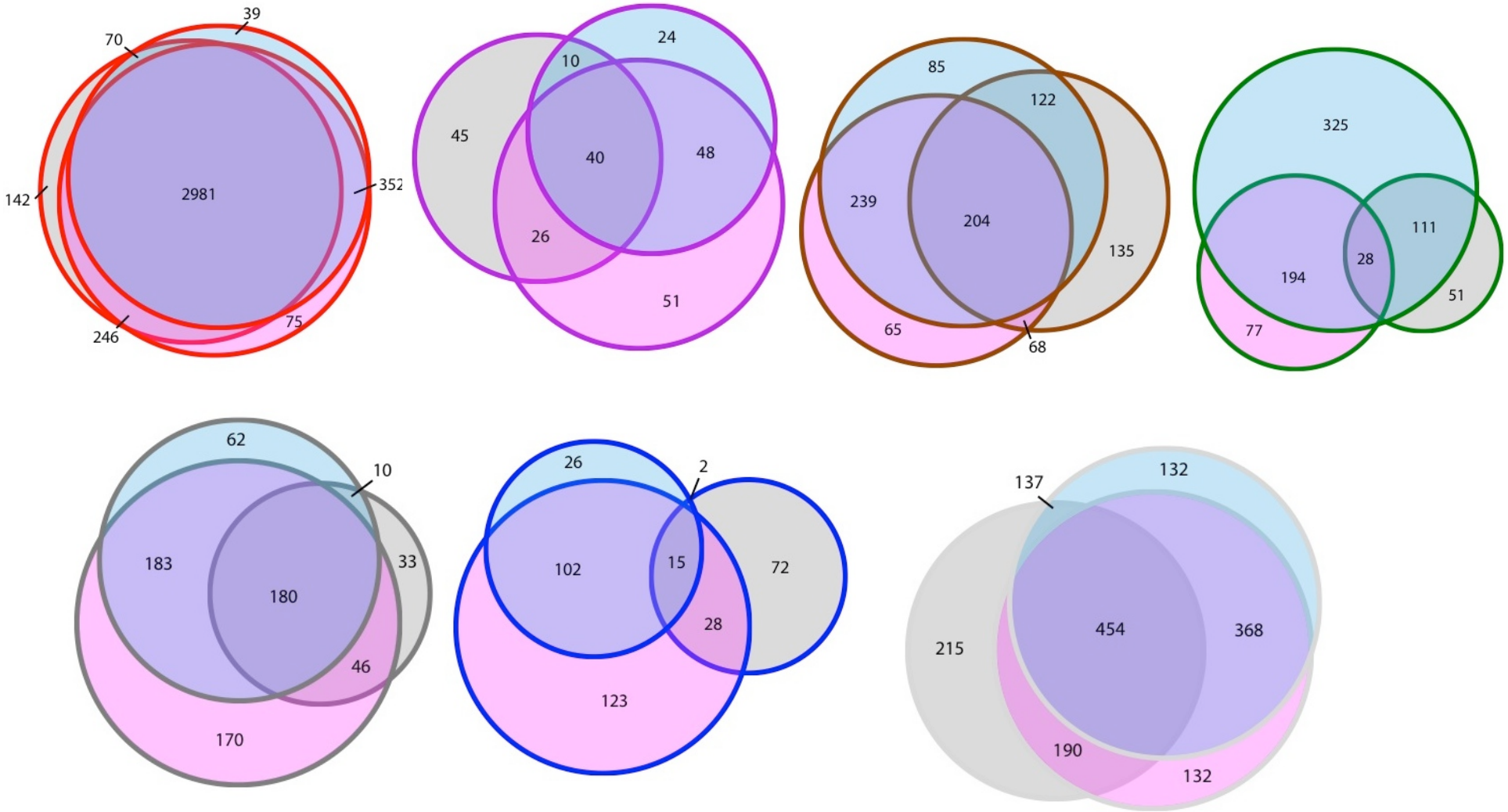

39
70
142
2981
352
246
75
24
10
45
40
48
26
51
85
122
239
204
135
65
68
325
111
28
194
51
77
62
10
183
33
180
46
170
26
2
102
15
72
28
123
137
132
215
454
368
190
132

Figure S24

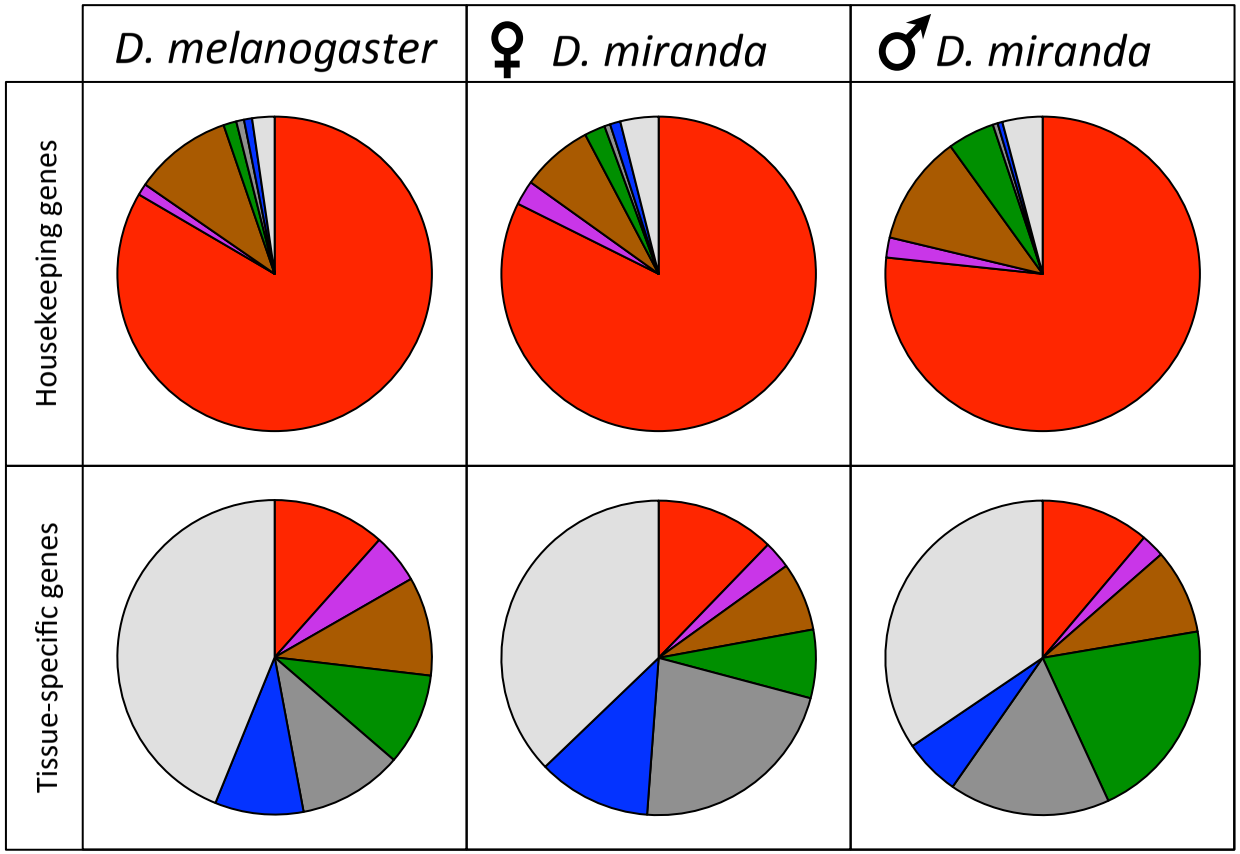

D. melanogaster
♀ D. miranda
♂ D. miranda
Housekeeping genes
Tissue-specific genes

Figure S25

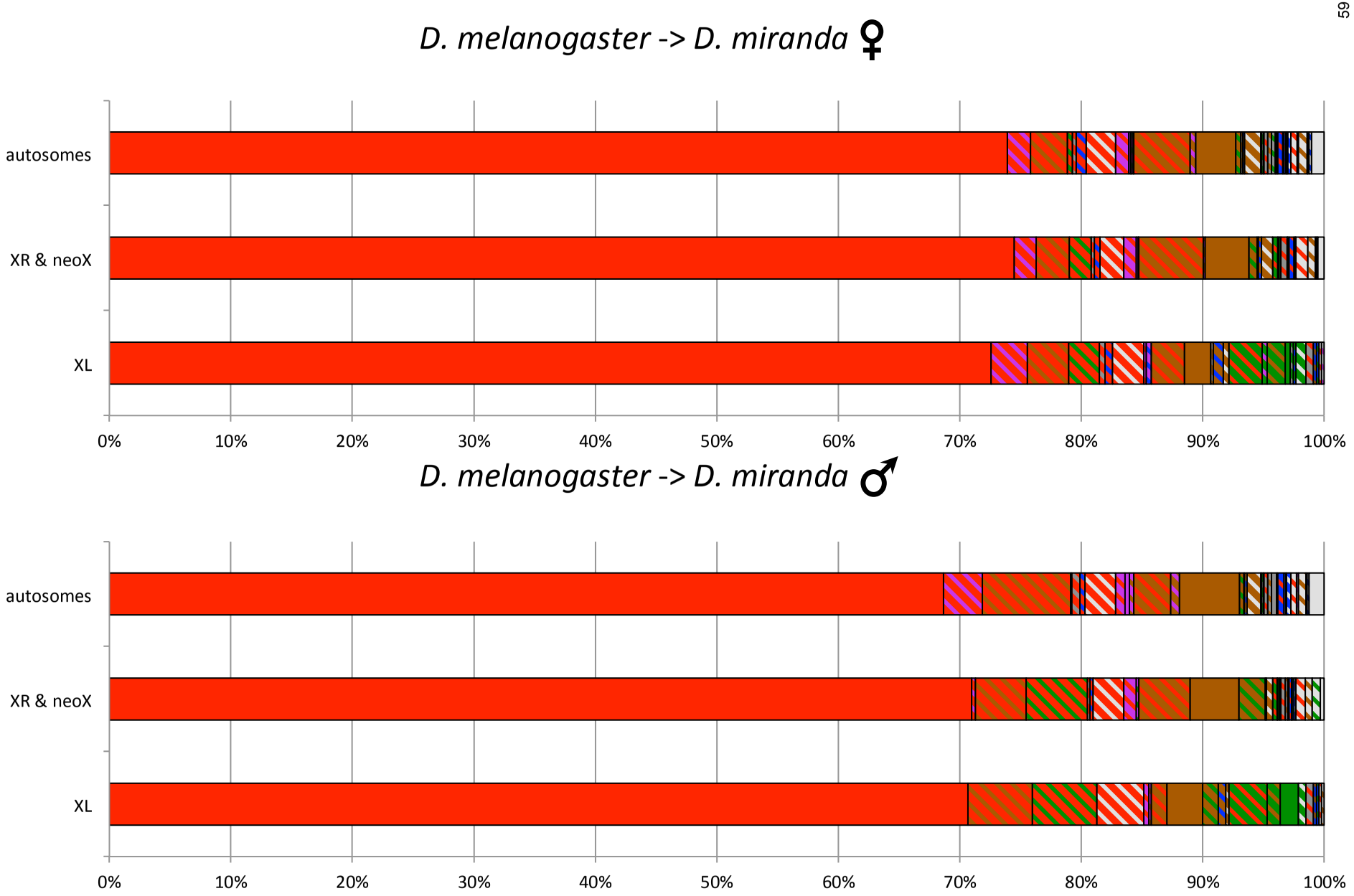

D. melanogaster -> D. miranda ♀
autosomes
XR & neoX
XL
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
D. melanogaster -> D. miranda ♂
autosomes
XR & neoX
XL
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Figure S26

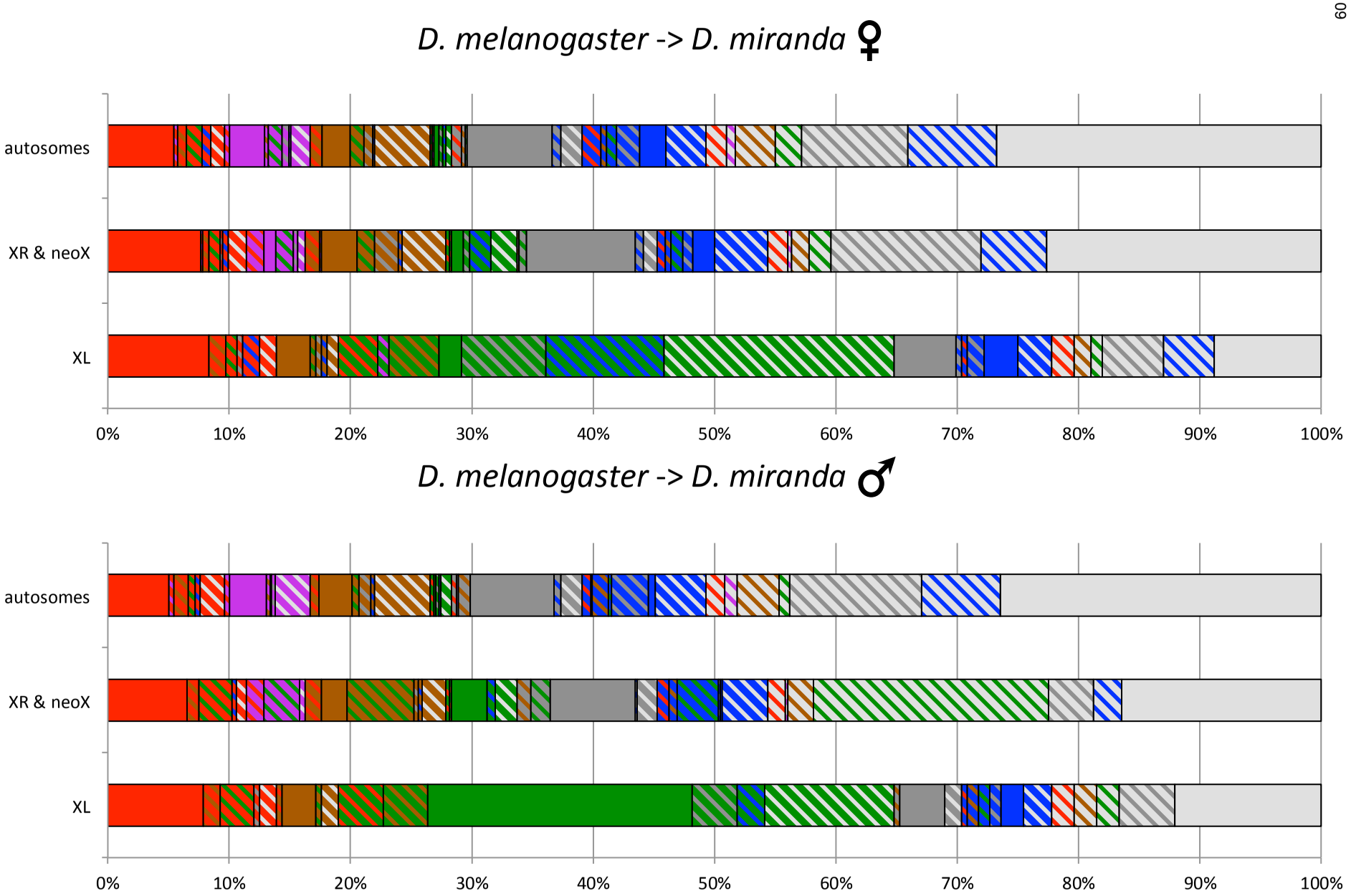

D. melanogaster -> D. miranda ♀
autosomes
XR & neoX
XL
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
D. melanogaster -> D. miranda ♂
autosomes
XR & neoX
XL
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Figure S27

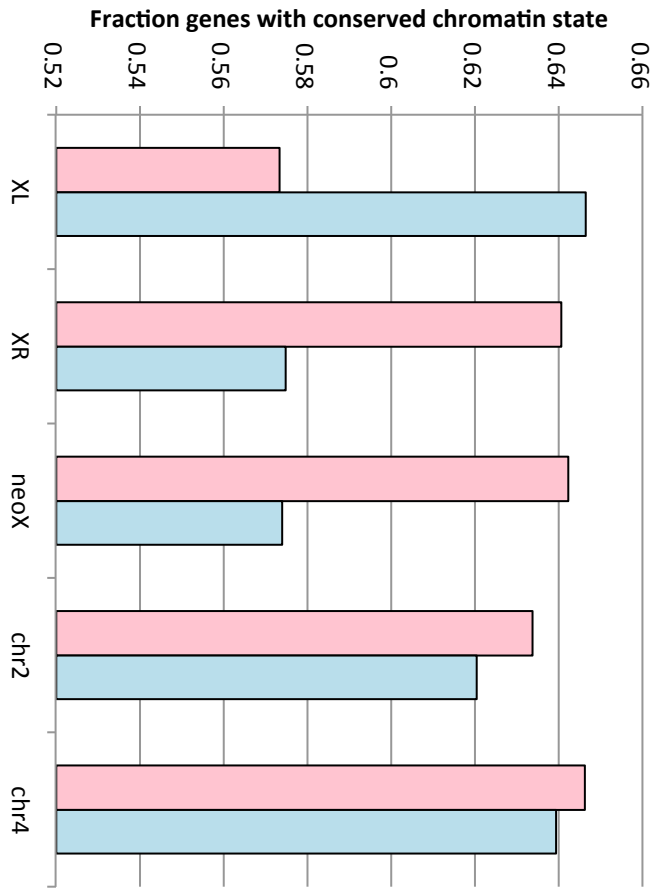

Fraction genes with conserved chromatin state
0.52
0.54
0.56
0.58
0.6
0.62
0.64
0.66
XL
XR
neoX
chr2
chr4

Figure S28

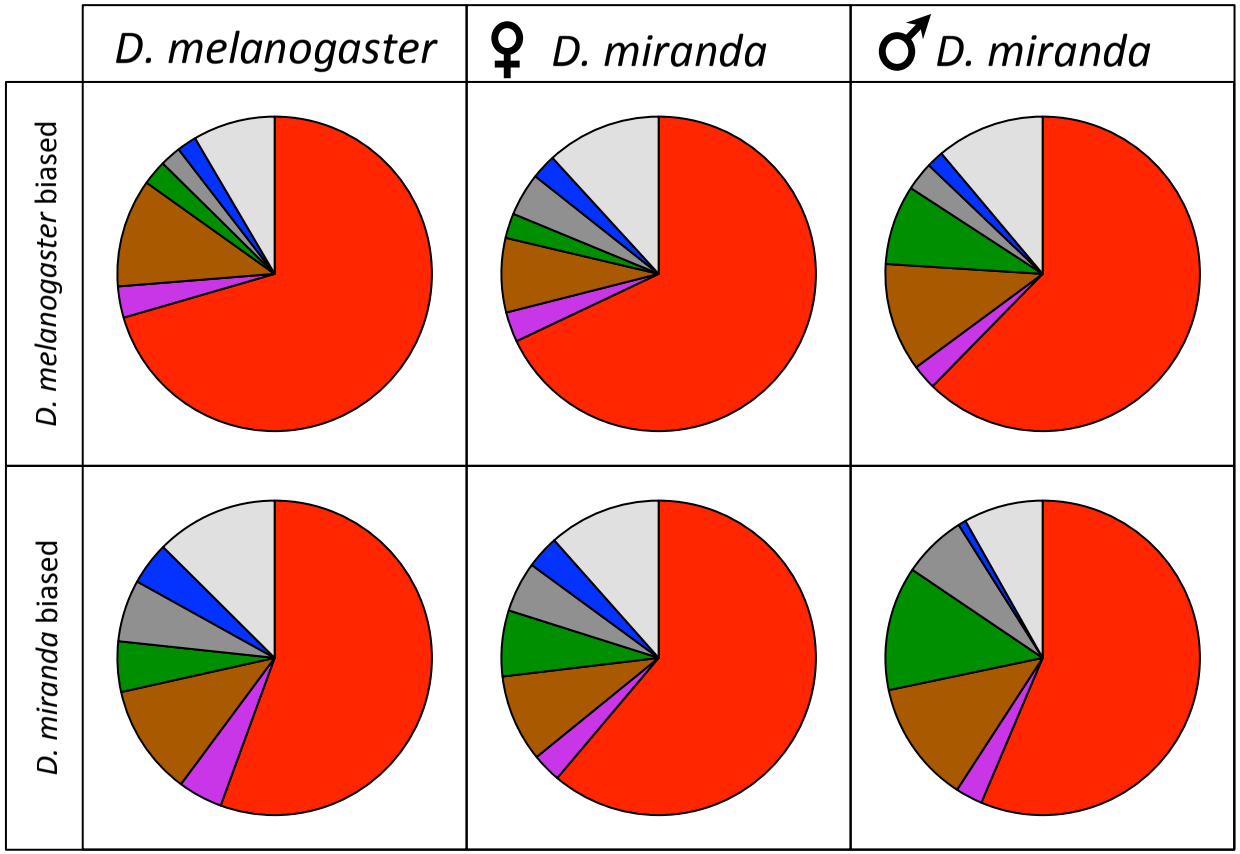

D. melanogaster
♀ D. miranda
♂ D. miranda
D. melanogaster biased
D. miranda biased

Figure S29

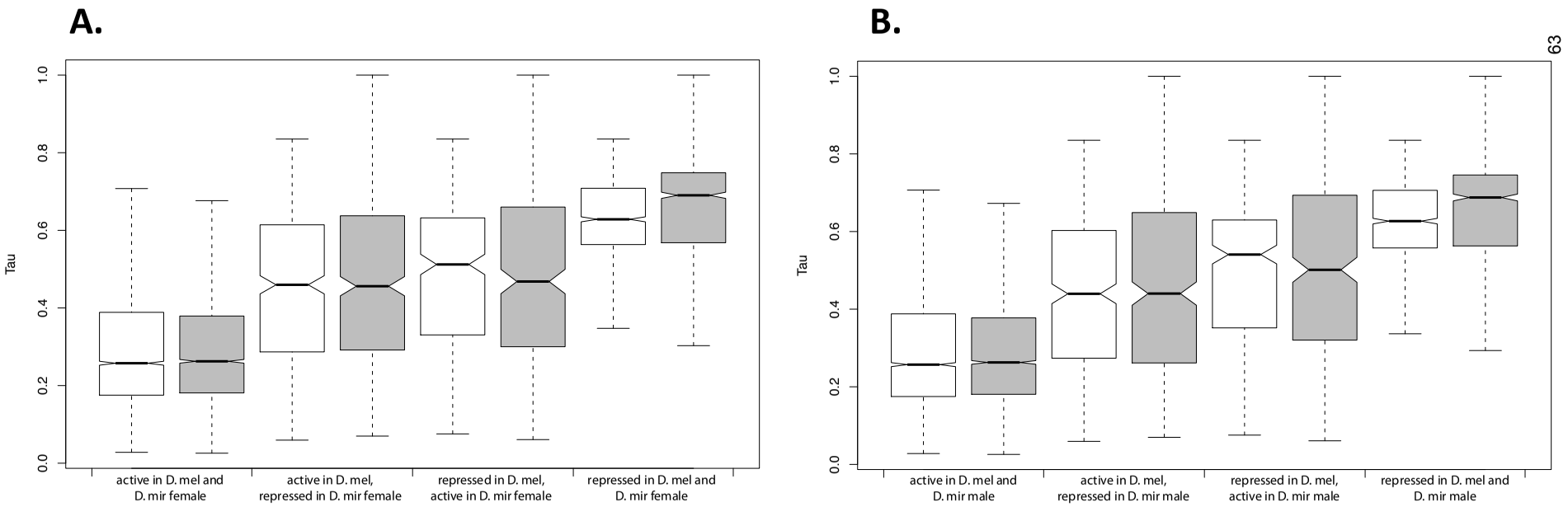

A.
B.
Tau
0.0
0.2
0.4
0.6
0.8
1.0
active in D. mel and D. mir female
active in D. mel, repressed in D. mir female
repressed in D. mel, active in D. mir female
repressed in D. mel and D. mir female
active in D. mel and D. mir male
active in D. mel, repressed in D. mir male
repressed in D. mel, active in D. mir male
repressed in D. mel and D. mir male
63

Figure S30

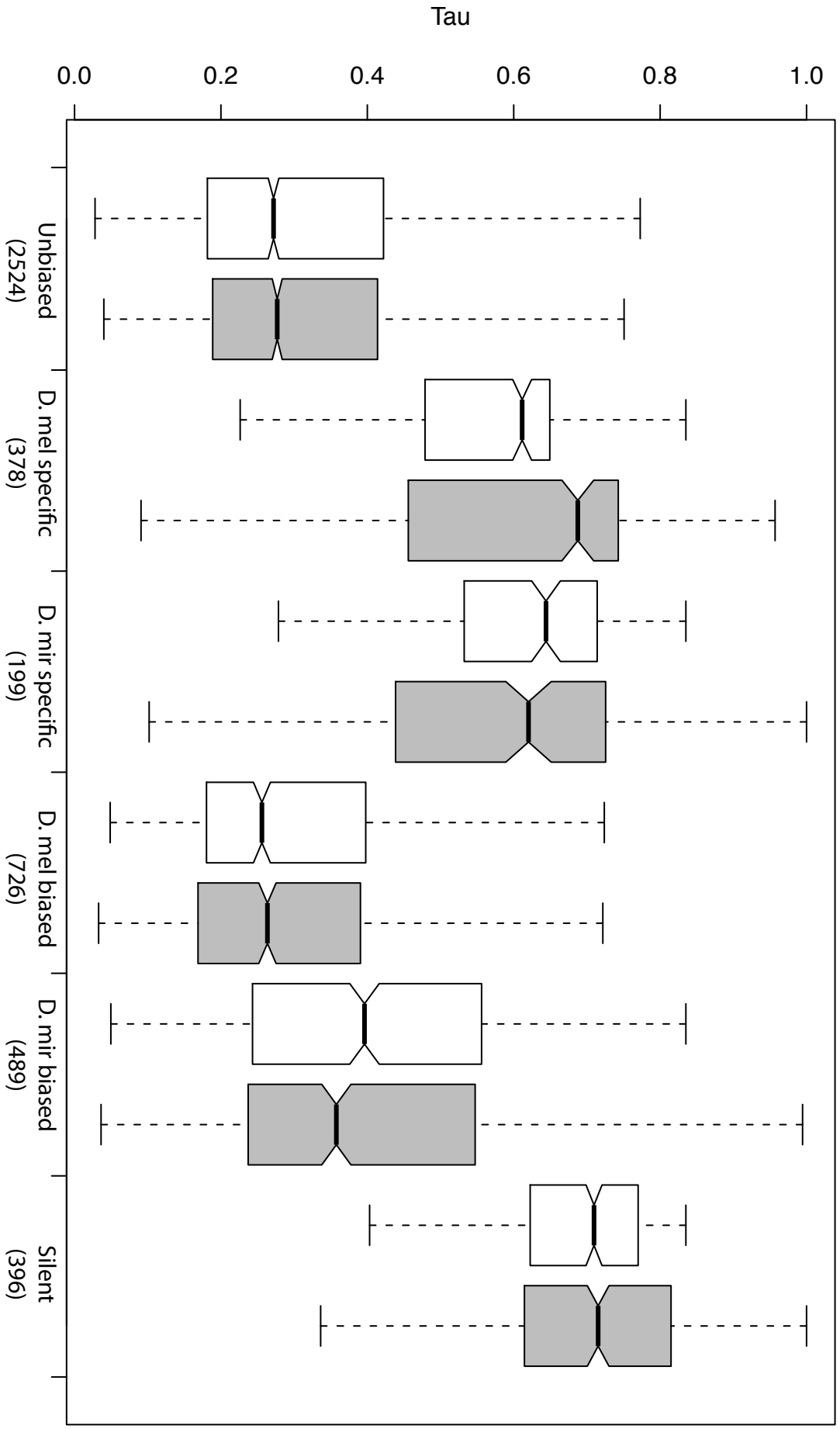
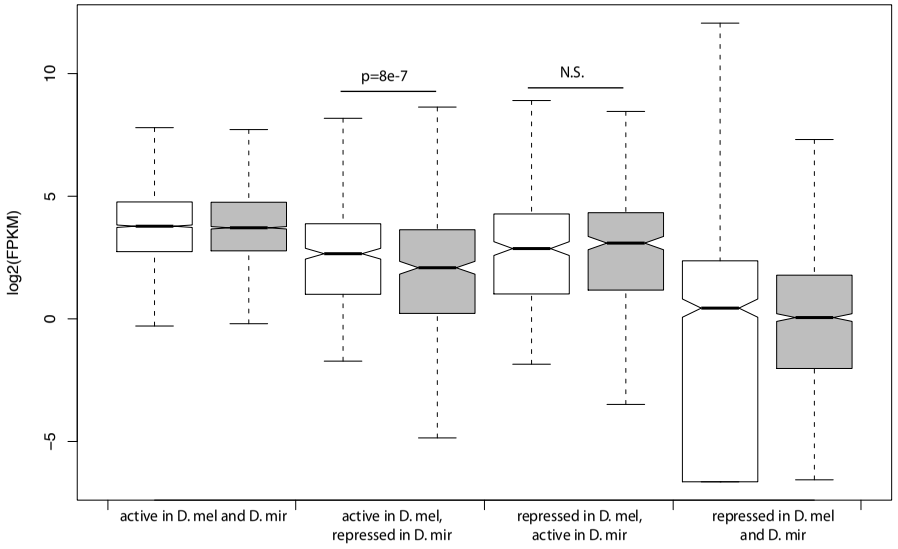

Tau
0.0
0.2
0.4
0.6
0.8
1.0
Unbiased (2524)
D. mel specific (378)
D. mir specific (199)
D. mel biased (726)
D. mir biased (489)
Silent (396)

Figure S31


log2(FPKM)
10
5
0
−5
p=8e-7
N.S.
active in D. mel and D. mir
active in D. mel, repressed in D. mir
repressed in D. mel, active in D. mir
repressed in D. mel and D. mir

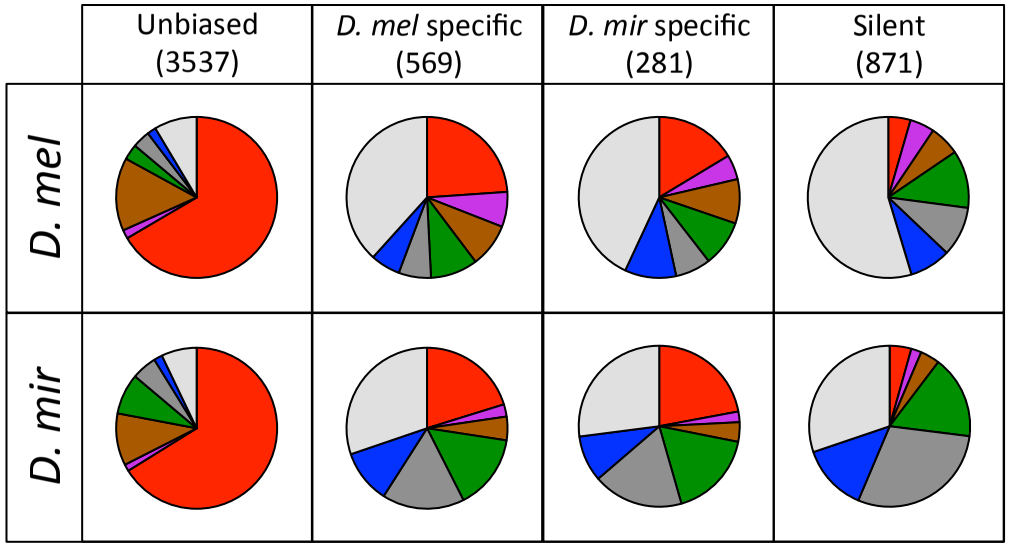

Unbiased (3537)
D. mel specific (569)
D. mir specific (281)
Silent (871)
D. mel
D. mir

Figure S32

Chapter 2

The Drosophila Y chromosome affects heterochromatin integrity genome-wide

Emily J. Brown & Doris Bachtrog*

Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

The Drosophila Y chromosome is gene poor and mainly consists of silenced, repetitive DNA. Nonetheless, the Y influences expression of hundreds of genes genome-wide, possibly by sequestering key components of the heterochromatin machinery away from other positions in the genome. To directly test the influence of the Y chromosome on the genome-wide chromatin landscape, we assayed the genomic distribution of histone modifications associated with gene activation (H3K4me3), or heterochromatin (H3K9me2 and H3K9me3) in fruit flies with varying sex chromosome complements (X0, XY and XYY males; XX and XXY females). Consistent with the general deficiency of active chromatin modifications on the Y, we find that Y gene dose has little influence on the genomic distribution of H3K4me3. In contrast, both the presence and the number of Y chromosomes strongly influence genome-wide enrichment patterns of repressive chromatin modifications. Highly repetitive regions such as the pericentromeres, the dot, and the Y chromosome (if present) are enriched for heterochromatic modifications in wildtype males and females, and even more strongly in X0 flies. In contrast, the additional Y chromosome in XYY males and XXY females diminishes the heterochromatic signal in these normally silenced, repeat-rich regions, which is accompanied by an increase in expression of Y-linked repeats. We find hundreds of genes that are expressed differentially between individuals with aberrant sex chromosome karyotypes, many of which also show sex-biased expression in wildtype Drosophila. Thus, Y chromosomes influence heterochromatin integrity genome-wide, and differences in the chromatin landscape of males and females may also contribute to sex-biased gene expression and sexual dimorphisms.

Introduction

The Drosophila Y is a degenerated, heterochromatic chromosome with only a few functional genes, primarily specialized in male reproductive function (Gatti and Pimpinelli 1983; Carvalho et al. 2000; Carvalho et al. 2001; Carvalho 2002). However, the *D. melanogaster* Y is about 40Mb in size and accounts for ~20% of the male haploid genome (Gatti and Pimpinelli 1983; Hoskins et al. 2002) (**Figure 1A**). Most of the Y chromosome is composed of repetitive satellite DNA, transposable elements (TEs), and Y-linked rDNA blocks (Bonaccorsi and Lohe 1991), and it is transcriptionally silenced through heterochromatin formation (Elgin and Reuter 2013). Despite its small number of genes, natural variation on the Y chromosome is associated with variation in several traits, including male fitness (Chippindale and Rice 2001) and position effect variegation (PEV;

the ability of spreading heterochromatin to induce partial silencing of reporter genes in some cells, resulting in mosaic expression patterns (Gowen and Gay 1934)). More recently, it was found that natural variation on the Y has substantial effects on regulation of hundreds of protein-coding genes genome-wide (Dimitri and Pisano 1989a; Lemos et al. 2008; Lemos et al. 2010; Sackton et al. 2011).

The molecular basis of this phenotypic variation is unclear. Single-nucleotide polymorphism in protein-coding genes is low on the Y chromosome (Zurovcova and Eanes 1999; Larracuente and Clark 2013), and it has been proposed that structural variation involving repetitive DNA is responsible for the observed phenotypic effects of different Y chromosomes (Francisco and Lemos 2014). Specifically, most of the highly repetitive Y chromosome is enriched for heterochromatic proteins and repressive histone modifications, and the Y may act as a 'heterochromatin sink' by sequestering core components of the heterochromatin machinery, thereby limiting the ability to silence other repetitive regions of the genome (Henikoff 1996; Francisco and Lemos 2014). Under the heterochromatin sink model, Y chromosomes vary in their ability to sequester heterochromatin components due to variations in the total amount or sequence content of their repetitive sequences. Protein-coding genes from the *D. melanogaster* Y chromosome are only expressed in germ cells of males, but the effects on global gene expression by different Y chromosomes also occur in XXY females and somatic cells of XY males (Lemos et al. 2008; Lemos et al. 2010; Sackton et al. 2011). This observation is consistent with the heterochromatin sink model, where the Y chromosome exerts its effect indirectly by depleting or redistributing chromatin regulators across the genome. However, these studies have been limited to reporter loci to assess the effect of the Y chromosome on heterochromatin formation, and the global chromatin landscapes of individuals with different amounts of heterochromatic sequence have not yet been directly examined. Here, we test the hypothesis that the Y chromosome acts to modulate heterochromatin integrity and gene expression genome-wide by contrasting the chromatin landscapes and expression profiles of X0 and XYY males and XXY females to that of wildtype *D. melanogaster*.

Results

Fly strains.

To compare the chromatin landscape between Drosophila that differ in their sex chromosome karyotype and their amount of repetitive DNA, we set up crosses between *D. melanogaster* stock number 2549 from the Bloomington Stock Center, which has a compound reversed metacentric X chromosome (C(1)RM) or a hetero-compound chromosome with the X chromosome inserted between the two arms of the Y chromosome (C(1;Y)), and the wildtype Canton-S stock (**Figure 1B**). We selected X0 males that contained a maternally transmitted X chromosome (as do wildtype males), and XXY females that contain a wildtype Y chromosome (rather than the C(1;Y) chromosome; see **Figure 1B**). Note that the resulting flies are not completely isogenic (and it is impossible to create completely isogenic flies using this crossing scheme), but different comparisons contrast flies with identical autosomal backgrounds. In particular, our wildtype male and female comparison share the same autosomal genotype (Canton-S), and our X0 males and XXY females both have one autosomal complement from Canton-S, and one from the 2549

stock. XYY males inherit 75% of autosomal genes from strain 2549. We used flow cytometry to estimate the genome sizes of the 5 karyotypes with different sex chromosome configurations, and using estimates of diploid euchromatic genome sizes of 231 Mb for individuals with two X chromosomes and 208 Mb for individuals with one X chromosome, we then estimated the amount of heterochromatic sequences in each karyotype. As expected, we found a gradient of heterochromatic sequence content for the 5 karyotypes, with XO males (~111Mb) < XX females (~124 Mb) < XY males (~151 Mb) < XXY females (~161 Mb) < XYY males (~181 Mb) (**Table S1**, **Figure 1A**).

Quantification of histone modifications.

We aged all flies for 8 days, and carried out chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) on head and thorax tissue using commercial antibodies against three post-translational histone modifications (H3K4me3, H3K9me2, H3K9me3). We employed a previously described normalization strategy to compare the genomic distribution and relative levels of chromatin marks across flies with different karyotypes. Specifically, we 'spiked in' a fixed amount of chromatin from female 3rd instar *Drosophila miranda* to each *D. melanogaster* chromatin sample prior to ChIP and sequencing. *D. miranda* chromatin served as an internal standard for the immunoprecipitation experiment (**Table S2**), and the relative recovery of *D. melanogaster* ChIP signal vs. *D. miranda* ChIP signal, normalized by their respective input counts, was used to quantity the relative abundance of the chromatin mark in *D. melanogaster* (see Methods for details). Note that this normalization strategy uses input coverage to account for differences in ploidy levels of sex chromosomes among the different karyotypes investigated and is agnostic to the total genome size of the sample (**Figure S1**). *D. miranda* is sufficiently diverged from *D. melanogaster* for sequencing reads to be unambiguously assigned to the correct species: even in repetitive regions, <4% of the reads cross-mapped between species; these regions were excluded from the analysis. Signal for H3K4me3 is highly correlated across samples, and H3K9me2 and H3K9me3 correlate well with each other for all samples, and replicate ChIP data without a *D. miranda* chromatin spike are highly correlated (**Table S3**, see Methods for details). We used the total normalized number of *D. melanogaster* reads to compare the genome-wide distribution of chromatin modifications in flies with different sex chromosome karyotypes. **Figure 2** shows the genomic distribution of the active H3K4me3 chromatin mark for the various karyotypes, and **Figures 3** and **4** show genomic distributions for the repressive H3K9me2/3 marks, respectively.

The genomic distribution of active chromatin is similar in flies with different karyotypes.

The histone modification H3K4me3 primarily associates with active genes (Guenther et al. 2007; Kharchenko et al. 2011) and is highly underrepresented in repeat-rich regions, including the Y chromosome; we thus expect that its relative abundance and genomic distribution is little influenced by the dose of Y chromosomes. Indeed, we find that H3K4me3 peaks are primarily located along the euchromatic chromosome arms, and highly deficient in pericentromeric regions, and along the Y chromosome (**Figure 2**). Genomic enrichment patterns of H3K4me3 are similar across sexes and flies with varying numbers of Y chromosomes (**Figure 2**), both when comparing the relative position of peaks, but also the absolute magnitude of signal across samples (**Figure 2**). This confirms

our expectation that Y dose should not dramatically influence the distribution of active chromatin marks, and also suggests that our normalization procedure is accurate in quantifying relative abundance of histone modifications across samples. Western blots confirm our inferences based on ChIP-seq, i.e. that H3K4me3 signal is similar across flies with different karyotypes (**Figure S2**).

Heterochromatic histone modifications in wildtype flies.

We investigated the genomic distribution of two histone marks that are associated with heterochromatin formation, H3K9me2 and H3K9me3 (Kharchenko et al. 2011). If the Y chromosome indeed acts as a sink for components of the heterochromatin machinery, we expect global differences in the enrichment patterns of heterochromatic histone modifications across strains with different numbers of Y chromosomes, or more generally, across flies with different amounts of repetitive DNA (see **Figure 1A**). In wildtype Drosophila, heterochromatin is highly enriched in pericentromeric regions, the small dot chromosome, and along the entire length of the Y chromosome (Hannah 1951; Gatti and Pimpinelli 1983; Bonaccorsi and Lohe 1991). Note that the *D. melanogaster* Y chromosome is estimated to be about 40Mb (i.e. 20% of the haploid male genome; (Gatti and Pimpinelli 1983; Hoskins et al. 2002)), but its assembled size is only 3.7Mb; thus our genome mapping analysis will underestimate the extent of heterochromatic histone modifications that are associated with the repetitive Y chromosome.

Overall, we find that levels of heterochromatin enrichment are similar for the H3K9me2 and H3K9me3 marks, but differ between flies with varying amount of repetitive DNA (**Figure 3, 4**). The male-specific Y chromosome is highly enriched for both of these repressive histone modifications in wildtype males, and we find that wildtype females have slightly higher levels of H3K9me2/3 enrichment than males in their pericentromeric regions, and on the dot chromosome, relative to euchromatic background levels (**Figure 3, 4**). Moreover, the heterochromatin / euchromatin boundary is slightly less clearly discernable from H3K9me2/3 enrichment patterns for males relative to females (**Figure 5, Figure S3-S5**). Western blots suggest that males harbor slightly more H3K9me2/3 compared to females (**Figure S2**). Thus, we find strong enrichment of the heterochromatic histone modifications on the Y and their relative deficiency at pericentromeric regions on autosomes and the X in wildtype males relative to females, despite similar amounts of overall H3K9me2/3. This observation is consistent with the hypothesis that the repeat-rich Y chromosome acts as a sink for components of the heterochromatic machinery, resulting in a relative paucity of heterochromatic histone modifications elsewhere in the genome. However, despite quantitative differences in levels of heterochromatic histone modifications, overall patterns of H3K9me2/3 enrichment are similar between sexes.

Heterochromatic histone modifications in X0, XXY & XYY flies.

To investigate the Y chromosome's role in the genome-wide distribution and enrichment for heterochromatic components, we studied histone modification profiles from female flies containing a Y chromosome (XXY females), and males with either zero or two Y chromosomes (X0 vs. XYY males). Female Drosophila that contain a wildtype Y chromosome show clear enrichment for both heterochromatic histone modifications on the Y chromosome, but an overall reduction in levels of H3K9me2/3 relative to wildtype

females, both at pericentromeric regions and along the dot (**Figure 3, 4**). The genomic distribution of H3K9me2/3 in XXY females is consistent with the model of the Y chromosome acting as a sink for components of the heterochromatin machinery, sequestering heterochromatic proteins to the Y chromosome and diluting them away from autosomal and X-linked targets. XXY females also show less heterochromatic histone modifications at pericentromeric regions and the dot relative to wildtype XY males (**Figure 3, 4**). This is consistent with the higher repeat content in XXY flies compared to XY flies - due to the large heterochromatic block on the X - contributing to the heterochromatin sink effect. This suggests that the effect of the Y chromosome on heterochromatin distribution is not a unique property of the Y but instead a result of a large amount of any additional repetitive sequence. XYY males harbor the highest amount of repetitive DNA and show severely decreased levels of H3K9me2/3 enrichment along repeat-rich, normally heterochromatic regions, including their Y chromosomes, pericentromeric regions, and along the dot, relative to levels found in other karyotypes investigated (**Figure 3, 4**).

X0 males, on the other hand, have the lowest repeat content of all flies, and show the strongest enrichment of heterochromatic histone modifications at pericentromeric regions and along the dot chromosome (**Figure 3, 4**). Enrichment levels of H3K9me2/3 at repetitive regions (pericentromere and the dot) relative to euchromatic background levels in X0 males is well above that of wildtype males and also wildtype females (or XXY females, which have the same autosomal background as X0 flies; **Figure 3, 4**). Together, our data provide clear evidence that Y chromosomes, and repetitive DNA in general, affect heterochromatin formation genome-wide, consistent with a model of the Y chromosome or other large blocks of repetitive sequences acting as heterochromatin sinks, possibly by redistributing heterochromatin components across the genome.

The depletion of heterochromatic histone modifications from pericentromeric regions causes the euchromatin/ heterochromatin boundaries to be significantly diluted in XXY and XYY individuals (**Figure 5**, **Figure S3, S4**). X0 males, in contrast, show spreading of their pericentromeric heterochromatin into chromosome arms that are normally euchromatic in wildtype flies, which is consistent with previous studies that found enhanced position effect variegation in XO males (**Figure 5, Figure S3, S4;** (Belyaeva et al. 1993; Wallrath and Elgin 1995b)).

Overall, we see that increasing the amount of repetitive DNA by changing the dose of both sex chromosomes corresponds with a decrease in the signal of heterochromatic histone modifications at pericentromeric regions and along the dot chromosome. This is consistent with a model of stoichiometric balance between protein components involved in the formation of heterochromatin and the amount of repetitive DNA sequences within a genome. Together, ChIP-seq profiles of histone modifications in wildtype flies, X0 and XXY males, and XXY females support the hypothesis that the Y chromosome acts as a heterochromatin sink in Drosophila.

Sex chromosome dose and gene expression.

Polymorphic Y chromosomes affect expression of hundreds of autosomal and X-linked genes in *D. melanogaster*, a phenomenon known as Y-linked regulatory variation (YRV)

(Dimitri and Pisano 1989a; Lemos et al. 2008; Lemos et al. 2010; Sackton et al. 2011). To test if genes that respond to YRV are also expressed differentially in flies with different sex chromosome configurations, we collected replicate RNA-seq data from heads for wildtype males and females, as well as from X0, XXY and XYY flies. As noted above, protein-coding Y-linked genes in Drosophila are only expressed in male germ line and thus cannot directly contribute to differences in expression profiles in head samples among flies with different numbers of Y chromosomes. Overall, we find that 100s of genes show differential expression among flies with different sex chromosome karyotypes (**Figure 6A**). GO analysis revealed that differentially expressed genes tend to be enriched for functions associated with reproductive processes (**Table S4**), and are not simply clustered around pericentromeric regions (**Figure S5, S6**). Genes that are expressed most differently between XO and XY males, and XX and XXY females, show significantly greater difference in H3K9me2 signal compared to all genes, while these genes have significantly less difference in H3K4me3 signal compared to all genes (**Figure S7**). This is consistent with the hypothesis that the Y chromosome re-distributes heterochromatin components genome-wide, and can thereby influence the expression of hundreds of genes.

We used a consensus set of 678 genes that were classified as susceptible to YRV (Sackton and Hartl 2013), and found that these genes were generally expressed more differently between different sex chromosome karyotypes compared to random genes (**Figure 6A**). This suggests that a similar mechanism is underlying both YRV and gene expression differences in flies with different sex chromosome configurations. Genes that are genetically defined to either suppress or enhance silencing in assays for PEV in *D. melanogaster*, i.e. Su(var) and E(var) genes (Elgin and Reuter 2013), are expressed at similar levels in flies with different karyotypes (**Figure 6**). This is consistent with our Western blots that reveal no consistent differences in H3K9me2/3 differences among flies with different sex chromosome configurations (**Figure S2**).

Interestingly, genes susceptible to YRV are more likely to be differentially expressed between wildtype sexes, and genes that are differentially expressed between males and females in head tissue tend to also be differentially expressed between X0 and XY males, or XX and XXY females ($p<1e-6$, permutation test, **Figure 6B, Figure S8**). In particular, 160 of the top 10% genes that are differentially expressed between wildtype XX females and XY males, vs. X0 and XY males vs. XX and XXY females overlap, while we only expect 9 by chance. This suggests that a substantial fraction of sex-biased expression in somatic tissues may simply be an indirect consequence of the absence or presence of the Y, i.e. the sink effect of the Y chromosome may contribute to sex-biased expression patterns in *D. melanogaster*.

Repeat reactivation in XXY and XYY flies.

Heterochromatin is established during early embryogenesis and leads to the transcriptional silencing of repetitive DNA and transposable elements (TEs) (Elgin and Reuter 2013). We used our RNA-seq data to assess whether changes in chromatin structure due to Y chromosome dose are associated with changes in gene expression patterns of repetitive elements. We first used consensus sequences of known TEs annotated by FlyBase (flybase.org), and found that overall repeat content correlated negatively with

H3K9me2/3 enrichment at TEs: X0 flies had the highest level of H3K9me2/3 enrichment across TE families, followed by XX and XY wildtype flies, and XXY and XYY flies having the lowest amount of heterochromatin marks at their TEs ($p<0.01$ for each comparison; **Figure 7A, Figure S9A**; note that these estimates are corrected for differences in copy numbers between repeats, by looking at the enrichment of H3K9me2/3 enrichment over input for each karyotype). Despite dramatic differences in overall levels of repressive histone marks across repeat families, levels of expression for the various TEs between karyotypes are very similar ($p>0.05$, **Figure 7A, Figure S10**). A subset of TEs show an increase in expression in XYY males compared to other samples, including at least 5 retroviral elements (1731, 297, Max element, mdg1, and mdg3, **Figure S10**). Increased expression of these repeats appears in part be driven by an increased copy number in the XYY male genome; if we correct for genomic copy number, we find that only three of these repeats (1731, 297, and Max element) are expressed more highly in XYY males compared to the other karyotypes (**Figure S11**). Thus, despite global differences in heterochromatin formation associated with repeats across karyotypes, this does not manifest itself in a global de-repression of TEs, but seems to instead involve de-repression of just a subset of TE families.

Most of the Y chromosome has not yet been assembled (Hoskins et al. 2015), including its repetitive elements, and we were interested in whether expression of Y-linked repeats would be particularly sensitive to Y chromosome dosage. We thus used a *de novo* approach to identify male-specific Y-linked repeats that does not rely on a genome assembly, but instead uses kmer abundances from next generation sequencing reads to produce a repeat library (Koch et al. 2014). We then mapped male and female genomic reads from the Canton-S strain back to our *de novo* assembled repeat library, in order to infer Y-linkage for repeats that were only covered by male genomic reads (**Figure S12, S13**). Male-specific repeats are highly enriched for H3K9me2/3 in wildtype males, and transcriptionally silenced (**Figure 7B**). However, while Y-linked repeats show similar enrichment for the H3K9me3 mark in all karyotypes (**Figure S9B**), XXY females and XYY males are highly deficient for H3K9me2 at Y-linked repeats and expression of Y-linked repeats is de-repressed relative to wildtype males (**Figure 7B, Figure S14**). If we account for differences in copy number of the Y-linked repeats, we still find that Y-linked repeats are expressed more highly in XXY females and XYY males compared to wildtype males (**Figure S15**). Thus, consistent with the ChIP-seq data that showed low levels of heterochromatic histone modifications (especially H3K9me2) along the Y of XXY females or the two Y chromosomes of XYY males, relative to wildtype males, our gene expression data demonstrate that Y-linked repeats become transcriptionally activated in female flies that normally do not have a Y chromosome, or male flies with double the dose of Y-linked repeats, and this is not simply a consequence of an increased copy number of Y-linked repeats.

Discussion

Dosage effects of chromatin components and repetitive DNA.

Many eukaryotic genomes contain large amounts of selfish, repetitive DNA, and transcriptional silencing of repeats through heterochromatin formation is one way to alleviate the deleterious effects of repetitive DNA (Elgin and Reuter 2013). Studies of PEV

in *D. melanogaster* have yielded important insights into the biology of heterochromatin (Muller 1930; Schultz 1936; Zhimulev et al. 1986), and frequently found dose-dependent effects of chromatin proteins and trans-activating factors (Weiler and Wakimoto 1995). For example, depletion of HP1, an important protein involved in both the recruitment and maintenance of heterochromatic histone modifications, suppresses variegation (Eissenberg et al. 1990) (i.e. it results in less heterochromatin formation and thus less suppression at a reporter gene), whereas increased dosage of HP1 enhances variegation (Eissenberg et al. 1992) (i.e. it increases silencing through increased heterochromatin formation). In *D. melanogaster*, the Y chromosome is a potent suppressor of variegation, i.e. it induces less heterochromatin at a reporter gene (Gowen and Gay 1934), and *D. melanogaster* males with different Y chromosomes in otherwise identical genetic backgrounds vary in their propensity to silence a heterochromatin-sensitive reporter gene in PEV assays (Lemos et al. 2010). Highly repetitive Y chromosomes are thought to sequester heterochromatic factors that are present in only limited amounts (Dimitri and Pisano 1989b), and different Y chromosomes vary in their repeat content and thus the extent to which they sequester those heterochromatin components, thereby influencing PEV.

In our study, we directly demonstrate that the Y chromosome, and repeat-rich DNA in general, can act to globally affect heterochromatin formation in *D. melanogaster*. We find that increasing the amount of repetitive DNA generally decreases the amount of H3K9me2/3 enrichment at repeat-rich regions, such as pericentromeres, the dot, or the Y chromosome. Individuals with the lowest repeat content (X0 males in our experiment) show the highest enrichment of H3K9me2/3 in repeat-rich regions, and the pericentromeric heterochromatin on the autosomes of X0 flies clearly extends into genomic regions that are normally euchromatic in wildtype *D. melanogaster*. Wildtype females show slightly higher H3K9me2/3 levels at their pericentromeric regions and the dot chromosome and a slightly sharper euchromatin/ heterochromatin boundary at autosomes compared to wildtype males. Indeed, females generally show a higher degree of silencing in assays for PEV, suggesting that normally euchromatic regions are more prone to acquire a heterochromatic conformation in females (Wallrath and Elgin 1995a; Girton and Johansen 2008).

XYY males and XXY females, on the other hand, show a dramatic reduction of H3K9me2/3 enrichment at repeat-rich regions, and the boundaries between the heterochromatic pericentromere and the euchromatic chromosome arms become blurry. Thus, this dosage sensitivity of H3K9me2/3 enrichment in repetitive regions suggests that there is a stoichiometric balance among protein components and total repeat content of the genome to maintain proper heterochromatic silencing.

Functional consequences of the Y chromosome's global effects on heterochromatin. Analyses of gene expression profiles suggest that global changes in heterochromatic histone modifications can have broad functional consequences for the organism. Specifically, we show that hundreds of genes are differentially expressed in individuals that differ in their sex chromosome karyotype, and genes that are susceptible to YRV are more prone to be differentially expressed in individuals with different sex chromosome complements. We find that increasing the amount of repetitive DNA leads to a decrease in

heterochromatic histone modification signal at TEs. XYY males and XXY females have low levels of H3K9me2 signal in TEs, and especially so in male-specific repeats, and we show that this deficiency of heterochromatin is associated with a de-repression of Y-linked repeats that we detect as an increase in expression levels of these repeats. Thus, while fruit flies have efficient mechanisms in place to silence wildtype levels of repetitive DNA, a large increase in the amount of repetitive sequences, caused by introducing additional Y chromosomes, limits the organism's ability to form heterochromatin and those additional repeats apparently cannot be efficiently silenced.

Whole-genome sequencing studies can provide information on the genome size by estimating the amount of euchromatic DNA, but cannot reliably estimate the amount of repetitive, heterochromatic sequences. Cytogenetic studies suggest that individuals within a population can differ greatly in how much repetitive heterochromatic DNA they contain. The size of the pericentromeric heterochromatic block on the *D. melanogaster* X chromosome, for example, varies by about 2-fold among strains (Halfer 1981), and dramatic variation in size and morphology of the Y chromosome has been reported in natural populations of *D. pseudoobscura* (Dobzhansky 1937). Moreover, haploid genome size estimates of different *D. melanogaster* strains using flow cytometry differ by almost 100Mb, and the vast majority of this variation is thought to result from differences in repetitive heterochromatin (Bosco et al. 2007). Similarly, a recent bioinformatics analysis that identified and quantified simple sequence repeats from whole genome sequences also found a 2.5-fold difference in their abundance between *D. melanogaster* strains (Wei et al. 2014). Thus, natural variation in repetitive DNA among individuals may in fact span a wider range than that across sex chromosome karyotypes investigated here. This implies that repetitive DNA might serve as an important determinant of global chromatin dynamics in natural populations, and may be an important modifier of the differential expression of genes and TEs between individuals.

Heterochromatin/ euchromatin balance between sexes.

Males contain a Y chromosome that is highly repetitive and heterochromatic, and which may shift the genome-wide heterochromatin/ euchromatin balance between the sexes (Brown and Bachtrog 2014). In particular, if the Y chromosome sequesters proteins required for heterochromatin formation, males may be more sensitive to perturbations of the balance between repetitive sequence content and heterochromatic protein components, and might have lower levels of heterochromatin-like features in the rest of their genome, as compared to females. Indeed, RNAi knockdown of the heterochromatin protein HP1 preferentially reduces male viability (Liu et al. 2005), and the presence of Y-linked heterochromatin is thought to underlie this differential sensitivity. Female Drosophila are also more tolerant of heat shock, survive heat-induced knock-down better, and become sterile at higher temperatures than males (Yamamoto and Ohba 1982), and it is possible that differences in the chromatin landscape may contribute to sex-specific differences in heat stress response. As mentioned, female flies show stronger silencing in assays for PEV (Wallrath and Elgin 1995a; Girton and Johansen 2008), consistent with having more heterochromatin protein components relative to repetitive sequences, which can then spread into reporter genes more readily.

Many recent studies in animals have shown that a large portion of the transcriptome in animals is sex-biased (Ranz et al. 2003; Mank et al. 2008). Sex-biased expression patterns are typically seen as an adaptation to form the basis of sexually dimorphic phenotypes (Parsch and Ellegren 2013). In Drosophila, most sex-biased expression patterns are due to differences in expression in sex-specific tissues (i.e. gonads; (Meisel 2011; Assis et al. 2012)); however, hundreds of genes also show differential expression in shared, somatic tissues (Meisel 2011; Assis et al. 2012). Interestingly, we find that a similar set of genes that show differences in expression patterns between males and females (in head) are also differentially expressed between XY and X0 males, or XX and XXY females. This suggests that not sex *per se*, but the absence or presence of the Y chromosome is responsible for much of the differences in expression patterns between sexes. Thus, while sex-biased expression is normally interpreted as a sex-specific adaptation to optimize expression levels of genes in males and females, it is also possible that sex-biased expression patterns are simply an indirect consequence of males having to silence a large repetitive Y chromosome, thereby changing the chromatin structure genome-wide as compared to females.

Materials & Methods

Drosophila strains

Fly strains were obtained from the Bloomington Stock Center. The following strains were used: Canton-S and 2549 (C(1;Y),$y^1cv^1v^1B/0$ & C(1)RM,$y^1v^1/0$). The crossing scheme used to obtain X0 and XYY males and XXY females is depicted in **Figure 1B**. For chromatin and gene expression analyses, flies were grown in incubators at 25°C, 32% of relative humidity, and 12h light. Newly emerged adults were collected and aged for 8 days under the same rearing condition before they were flash-frozen in liquid nitrogen and stored at -80°C.

Genome size estimation

We estimated genome size of the 5 karyotypes of interest using flow cytometry methods similar to those described in (Ellis et al. 2014). Briefly, samples were prepared by using a 2mL Dounce to homogenize one head each from an internal control (*D. virilis* female, 1C=328 Mb) and one of the 5 karyotypes in Galbraith buffer (44mM magnesium chloride, 34mM sodium citrate, 0.1% (v/v) Triton X-100, 20mM MOPS, 1mg/mL RNAse I, pH 7.2). After homogenizing samples with 15-20 strokes, samples were filtered using a nylon mesh filter, and incubated on ice for 45 minutes in 25 ug/mL propidium iodide. Using a BD Biosciences LSR II flow cytometer, we measured 10,000 cells for each unknown and internal control sample. We ran samples at 10-60 events per second at 473 voltage using a PE laser at 488 nm. Fluorescence for each *D. melanogaster* karyotype was measured using the FACSDiva 6.2 software and recorded as the mode of the sample's fluorescent peak interval. We calculated the genome size of the 5 karyotypes by multiplying the known genome size of *D. virilis* (328 Mb) by the ratio of the propidium iodide fluorescence in the unknown karyotype to the *D. virilis* control.

Western blotting

We performed Western blots from acid-extracted histones, probing for H3K9me2, H3K9me3, H3K4me3, and total H3. Briefly, approximately 30 flies of each karyotype were dissected on dry ice to remove the abdomen. The resulting heads and thoraces were ground in PBS plus 10mM sodium butyrate, and were acid-extracted overnight at 4°C. Samples were then run on a 4-12% gradient bis-tris gel and transferred to a nitrocellulose membrane using Invitrogen's iBlot Dry Transfer Device. After blocking with 5% milk in PBS, we incubated membranes overnight with either 1:1000 H3K9me2 antibody (Abcam ab1220), 1:2000 H3K9me3 antibody (Abcam ab8898), 1:2000 H3K4me3 antibody (Abcam ab8580), or 1:2000 H3 antibody (Abcam ab1791) in Hikari Signal Enhancer (Nacalai 02272). We then incubated membranes with 1:2500 secondary antibody (Licor 68070 and 32213), imaged bands on a Licor Odyssey CLx Imager, and quantified intensity using ImageJ.

Chromatin-IP and sequencing

We performed ChIP-seq experiments using a standard protocol adapted from (Alekseyenko et al. 2006). Briefly, approximately 2 mL of adult flash-frozen flies were dissected on dry ice, and heads and thoraces were used to fix and isolate chromatin. Following chromatin isolation, we spiked in 60uL of chromatin prepared from female *Drosophila miranda* larvae (approximately 1ug of chromatin). We then performed immunoprecipitation using 4uL of the following antibodies: H3K9me2 (Abcam ab1220), H3K9me3 (Abcam ab8898), and H3K4me3 (Abcam ab8580).

After reversing the cross-links and isolating DNA, we constructed sequencing libraries using the BIOO NextFlex sequencing kit. Sequencing was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. We performed 50bp single-read sequencing for our input and H3K4me3 libraries, and 100bp paired-end sequencing for the H3K9me2 and H3K9me3 libraries, due to their higher repeat content.

For H3K4me3, Pearson correlation values between the 5 karyotypes is very high, and the magnitude of difference between the samples is low (**Table S2**). For the two heterochromatin marks, Pearson correlation values between the two marks were generally high for all samples, and overlap of the top 40% of 5kb windows was similarly high for all samples (**Table S2**). Additionally, we obtained replicates for H3K9me3 for all samples except XX female, which has extremely high correlation values between H3K9me2 and H3K9me3. The unspiked replicate data for H3K9me3 correlate well with the *D. miranda* chromatin spike data that was used for the bulk of our analyses (**Table S2**).

RNA extraction and RNA-seq

We collected mated males and females of the various karyotypes, aged them for 8 days, and dissected and pooled 5 heads from each karyotype. We then extracted RNA and prepared stranded total RNA-seq libraries using Illumina's TruSeq Stranded Total RNA Library Prep kit with Ribo-Zero ribosomal RNA reduction chemistry, which depletes the highly abundant ribosomal RNA transcripts (Illumina RS-122-2201). We performed 50bp single-read sequencing for all total RNA libraries at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

Mapping of sequencing reads, and data normalization
For all *D. melanogaster* alignments, we used Release 6 of the genome assembly and annotation (Hoskins et al. 2015). For all ChIP-seq datasets, we used Bowtie2 (Langmead and Salzberg 2012) to map reads to the genome, using the parameters "-D 15 –R 2 –N 0 –L 22 –i S,1,0.50 --no-1mm-upfront", which allowed us to reduce cross-mapping to the *D. miranda* genome to approximately 2.5% of 50bp reads, and 1% of 100bp paired-end reads. We also mapped all ChIP-seq datasets to the *D. miranda* genome assembly (Ellison and Bachtrog 2013) to calculate the proportion of each library that originated from the spiked-in *D. miranda* chromatin versus the *D. melanogaster* sample.

To calculate ChIP signal, we first calculated the coverage across 5kb windows for both the ChIP and the input, and then normalized by the total library size, including reads that map to both *D. melanogaster* and the *D. miranda* spike. We then calculated the ratio of ChIP coverage to input coverage, and normalized by the ratio of *D. melanogaster* reads to *D. miranda* reads in the ChIP library, and then by the ratio of *D. melanogaster* reads to *D. miranda* reads in the input, to account for differences in the ratio of sample to spike present before immunoprecipitation. Note that this normalization procedure is accounts for differences in ploidy as well as genome size by using a ratio of ChIP coverage to input coverage (see **Figure S1**).

Gene expression analysis
We first mapped RNA-seq reads to the ribosomal DNA scaffold in the Release 6 version of the genome, and removed all reads that mapped to this scaffold, as differences in rRNA expression are likely to be technical artifacts from the total RNA library preparation. We then mapped the remaining reads to the Release 6 version of the *D. melanogaster* genome using Tophat2 (Kim et al. 2013), using default parameters. We then used Cufflinks and Cuffnorm to calculate normalized FPKMs for all samples. GO analysis was performed using GOrilla using ranked lists of differentially expressed genes (Eden et al. 2009).

Repeat libraries
We used two approaches to quantify expression of repeats. Our first approach was based on consensus sequences of known repetitive elements that were included in the Release 6 version of the *D. melanogaster* genome and are available on FlyBase. These included consensus sequences for 125 TEs. We also added the consensus sequences of three known satellite sequences, (Dodeca, Responder, and 359), to include larger non-TE repetitive sequences in our repeat analyses.

We were particularly interested in mis-regulation of the Y chromosome, which is poorly assembled. We therefore assembled repetitive elements *de novo* from male and female genomic DNA reads using RepARK (Koch et al. 2014), setting a manual threshold for abundant kmers of 5 times the average genome coverage, which corresponds to a repetitive sequence occurring at least 5 times in the genome. To identify male-specific repeats, we mapped male and female genomic reads back to our *de novo* assembled repeats, and identified repeats that had high coverage in males and either no coverage or significantly lower coverage in females (**Figure S9**). After filtering in this way, we obtained

101 male-specific repeats comprising 13.7kb of sequence, with a median repeat size of 101bp.

References:

Alekseyenko A, Larschan E, Lai W, Park P, Kuroda M. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**(7): 848 - 857.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. *Genome biology and evolution* **4**(11): 1189-1200.

Belyaeva ES, Demakova Ov Fau - Umbetova GH, Umbetova Gh Fau - Zhimulev IF, Zhimulev IF. 1993. Cytogenetic and molecular aspects of position-effect variegation in Drosophila melanogaster. V. Heterochromatin-associated protein HP1 appears in euchromatic chromosomal regions that are inactivated as a result of position-effect variegation. *Chromosoma* **102**(8): 583-590.

Bonaccorsi S, Lohe A. 1991. Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: relationships between satellite sequences and fertility factors. *Genetics* **129**(1): 177-189.

Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* **177**(3): 1277-1290.

Brown EJ, Bachtrog D. 2014. The chromatin landscape of Drosophila: comparisons between species, sexes, and chromosomes. *Genome Res* **24**(7): 1125-1137.

Carvalho AB. 2002. Origin and evolution of the Drosophila Y chromosome. *Curr Opin Genet Dev* **12**(6): 664-668.

Carvalho AB, Dobo BA, Vibranovski MD, Clark AG. 2001. Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **98**(23): 13225-13230.

Carvalho AB, Lazzaro BP, Clark AG. 2000. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. *Proc Natl Acad Sci U S A* **97**(24): 13239-13244.

Chippindale AK, Rice WR. 2001. Y chromosome polymorphism is a strong determinant of male fitness in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **98**(10): 5677-5682.

Dimitri P, Pisano C. 1989a. Position effect variegation in Drosophila melanogaster: relationship between suppression effect and the amount of Y chromosome. *Genetics* **122**(4): 793-800.

-. 1989b. Position effect variegation in Drosophila melanogaster: relationship between suppression effect and the amount of Y chromosome. *Genetics* **122**(4): 793-800.

Dobzhansky T. 1937. Further Data on the Variation of the Y Chromosome in Drosophila Pseudoobscura. *Genetics* **22**(3): 340-346.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48.

Eissenberg JC, James TC, Foster-Hartnett DM, Hartnett T, Ngan V, Elgin SC. 1990. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression

of position-effect variegation in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **87**(24): 9923-9927.
Eissenberg JC, Morris GD, Reuter G, Hartnett T. 1992. The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. *Genetics* **131**(2): 345-352.
Elgin SC, Reuter G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harbor perspectives in biology* **5**(8): a017780.
Ellis L, Huang W, Quinn A, Ahuja A, Alfrejd B, Gomez F, Hjelmen C, Moore K, Mackay T, Johnston J et al. 2014. Intrapopulation genome size variation in D. melanogaster reflects life history. *PLoS Genet* **10**(7): e1004522.
Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**(6160): 846-850.
Francisco FO, Lemos B. 2014. How do y-chromosomes modulate genome-wide epigenetic States: genome folding, chromatin sinks, and gene expression. *J Genomics* **2**: 94-103.
Gatti M, Pimpinelli S. 1983. Cytological and genetic analysis of the Y-chromosome of *Drosophila melanogaster*. 1. Organization of the fertility factors. *Chromosoma* **88**: 349-373.
Girton JR, Johansen KM. 2008. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. *Adv Genet* **61**: 1-43.
Gowen JW, Gay EH. 1934. Chromosome Constitution and Behavior in Eversporting and Mottling in Drosophila Melanogaster. *Genetics* **19**(3): 189-208.
Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**(1): 77-88.
Halfer C. 1981. Interstrain heterochromatin polymorphisms in Drosophila melanogaster. *Chromosoma* **84**(2): 195-206.
Hannah A. 1951. Localization and function of heterochromatin in Drosophila melanogaster. *Adv Genet* **4**(0065-2660 (Print)): 87-125.
Henikoff S. 1996. Dosage-dependent modification of position-effect variegation in Drosophila. *Bioessays* **18**(5): 401-409.
Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R et al. 2015. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res* **25**(3): 445-458.
Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG et al. 2002. Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome biology* **3**(12): RESEARCH0085.
Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**(7339): 480-485.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**(4): R36.
Koch P, Platzer M, Downie BR. 2014. RepARK--de novo creation of repeat libraries from whole-genome NGS reads. *Nucleic Acids Res* **42**(9): e80.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**(4): 357-359.

Larracuente AM, Clark AG. 2013. Surprising differences in the variability of Y chromosomes in African and cosmopolitan populations of Drosophila melanogaster. *Genetics* **193**(1): 201-214.
Lemos B, Araripe LO, Hartl DL. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**(5859): 91-93.
Lemos B, Branco AT, Hartl DL. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **107**(36): 15826-15831.
Liu LP, Ni JQ, Shi YD, Oakeley EJ, Sun FL. 2005. Sex-specific role of Drosophila melanogaster HP1 in regulating chromatin structure and gene transcription. *Nat Genet* **37**(12): 1361 - 1366.
Mank JE, Hultin-Rosenberg L, Webster MT, Ellegren H. 2008. The unique genomic properties of sex-biased genes: insights from avian microarray data. *BMC genomics* **9**: 148.
Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Molecular biology and evolution* **28**(6): 1893-1900.
Muller HJ. 1930. Types of visible variations induced by X-rays in Drosophila. *Journal of genetics* **22**: 299-334.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nature reviews Genetics* **14**(2): 83-87.
Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science* **300**(5626): 1742 - 1745.
Sackton TB, Hartl DL. 2013. Meta-analysis reveals that genes regulated by the Y chromosome in Drosophila melanogaster are preferentially localized to repressive chromatin. *Genome biology and evolution* **5**(1): 255-266.
Sackton TB, Montenegro H, Hartl DL, Lemos B. 2011. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **108**(41): 17046-17051.
Schultz J. 1936. Variegation in Drosophila and the Inert Chromosome Regions. *Proc Natl Acad Sci U S A* **22**(1): 27-33.
Wallrath LL, Elgin SC. 1995a. Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes & development* **9**(10): 1263-1277.
-. 1995b. Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes & development* **9**(10): 1263-1277.
Wei KH, Grenier JK, Barbash DA, Clark AG. 2014. Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **111**(52): 18793-18798.
Weiler KS, Wakimoto BT. 1995. Heterochromatin and gene expression in Drosophila. *Annual review of genetics* **29**: 577-605.
Yamamoto A, Ohba S. 1982. Strategic differences in thermal adaptation between two Drosophila species, D. virilis and D. immigrans. *Oecologia* **52**: 333-339.
Zhimulev IF, Belyaeva ES, Fomina OV, Protopopov MO, Bolshakov VN. 1986. Cytogenetic and molecular aspects of position effect variegation in Drosophila melanogaster.

Chromosoma **94**(6): 492-504.
Zurovcova M, Eanes WF. 1999. Lack of Nucleotide Polymorphism in the Y-Linked Sperm Flagellar Dynein Gene Dhc-Yh3 of *Drosophila melanogaster* and *D. simulans*. *Genetics* **153**(4): 1709-1715.

Figure legends

Figure 1. Chromosome structure of *Drosophila melanogaster*, and crossing scheme utilized. **A.** The left and right arms of chromosome 2 (2L, 2R) and 3 (3L, 3R), the small chromosome 4 (the dot chromosome), and the sex chromosomes X and Y are shown (adapted from (Hoskins et al. 2002)). The numbers correspond to approximate lengths in megabases, and the approximate heterochromatin content for flies with different sex chromosome karyotypes is indicated, based on flow cytometry estimates of the genome size of the five karyotypes. **B.** Crossing scheme utilized to obtain X0 and XYY males, and XXY females (only sex chromosomes are shown). Wildtype Canton-S males and females where crossed to the 2549 strain whose females have C(1)RM and males have C(1;Y). Circled karyotypes were used for the analyses.

Figure 2. Genome-wide enrichment of H3K4me3 for *D. melanogaster* strains with different karyotypes along each chromosome arm. For each karyotype, the enrichment in 5kb windows is shown in red lines (normalize ratio of ChIP to input, see Materials & Methods), and the enrichment in 20kb windows is shown in gray scale according to the scale in the upper left. Note that the enrichment profiles for all 5 karyotypes are plotted on the same scale to allow for direct comparisons. Below the enrichment profiles for each chromosome arm, subtraction plots show the absolute difference in signal of 50kb windows between pairs of karyotypes along the chromosome arms; these plots are smoothed by subtracting the median signal in the pericentromere from all windows. The box plots show the smoothed ChIP signal for all 5kb windows in different chromosomal regions, with p-values calculated relative to XX females (Wilcoxon test).

Figure 3. Genome-wide enrichment of H3K9me2 for *D. melanogaster* strains with different karyotypes along each chromosome arm. These plots were made in the same manner as those for H3K4me3 (see **Figure 2**), except we smoothed the subtraction plots by subtracting the median signal in the euchromatic portion of the genome, rather than the pericentromere. Box plots show the smoothed ChIP signal for all 5kb windows in heterochromatic chromosomal regions, with p-values calculated relative to XX females (Wilcoxon test).

Figure 4. Genome-wide enrichment of H3K9me3 for *D. melanogaster* strains with different karyotypes along the different chromosome arms. These plots were made in the same manner as those for H3K9me2 (see **Figure 3**).

Figure 5. Enrichment of H3K9me2 within 1Mb of the heterochromatin/ euchromatin boundaries (as defined in the Release 6 of the *D. melanogaster* genome (Hoskins et al. 2015)). The upper panel shows H3K9me2 signal in 5kb windows for each chromosome

arm, and the bottom panel shows scaled heatmaps for the same 5kb windows, to allow direct comparisons of H3K9me2 signal across samples. For H3K9me3 plots, see **Figure S3**. Box plots shows H3K9me2 signal of 5kb windows in euchromatic regions within 500-kb of the heterochromatin/ euchromatin boundary. Significance values are all calculated using the Wilcoxon test.

Figure 6. Gene expression variation between flies with different sex chromosome karyotypes. **A.** Pairwise expression comparisons for flies with different karyotypes. Genes marked in red are susceptible to YRV (from (Sackton and Hartl 2013)), and genes in yellow are genetically defined Su(var) and E(var) genes in *D. melanogaster* (from (Elgin and Reuter 2013)). **B.** Overlap of top 10% differentially expressed genes between wildtype male and female, and males and females with and without Y chromosomes, i.e. XX vs. XXY females and XY vs. X0 males.

Figure 7. Chromatin and expression patterns at TE families. Shown is enrichment of H3K9me2 at different TE families and their expression levels for the various karyotypes. For H3K9me3 plots, see **Figure S8**. **A.** All repeats from the library of consensus transposable elements and satellites from FlyBase. **B.** Putatively Y-linked (male-specific) *de novo* assembled repeats only.

Supplementary Figure legends

Figure S1. Normalization accounts for differences in ploidy of sex chromosomes. In the upper boxplot, we show the raw H3K4me3 signal in genes on the X chromosome, where we know that there is a two-fold difference in ploidy between male and female karyotypes. In the lower boxplot, we plot the normalized H3K4me3 signal in genes on the X chromosome, demonstrating that our normalization method corrects for differences in signal driven by differences in ploidy.

Figure S2. Western blots for H3K9me2, H3K9me3 and H3K4me3. We normalized the Western signal for all histone modifications by signal for total H3, and averaged across dilutions that were within the linear signal range.

Figure S3. Enrichment of H3K9me3 within 1Mb of the heterochromatin/ euchromatin boundaries. The plots were made in the same manner as for H3K9me2 (see **Figure 5**).

Figure S4. H3K9me2 and H3K9me3 signal of 5kb windows in euchromatic regions within 1Mb of the heterochromatin/ euchromatin boundary. Significance values are all calculated relative to XX females using the Wilcoxon test.

Figure S5. Distance to the pericentromere boundary of the top 10% of differentially expressed genes compared to all genes. For each pair of boxplots, the top 10% genes are on the left, and all genes are on the right. None of the comparisons are significant (Wilcoxon test).

Figure S6. Pairwise expression comparisons for flies with different karyotypes. Genes marked in red are located in the pericentromere.
Figure S7. Signal of H3K9me2 and H3K4me3 in the top 10% of differentially expressed genes compared to all genes. P-values were calculated for all comparisons using the Wilcoxon test.

Figure S8. Overlap of genes categorized as significantly differently expressed by Cuffdiff between wildtype male and female, XX vs. XXY females, and XY vs. XO males.

Figure S9. Enrichment of H3K9me3 at different TE families. **A.** All repeats from the library of consensus transposable elements and satellites from FlyBase. **B.** Putatively Y-linked (male-specific) *de novo* assembled repeats only.

Figure S10. Comparisons of expression of consensus transposable elements, color-coded by type of transposable element. Outlier transposable elements are identified.

Figure S11. Comparisons of expression of consensus transposable elements normalized by genomic copy number, as estimated by genomic coverage, color-coded by type of transposable element. Outlier transposable elements are identified.

Figure S12. Categorization of *de novo* assembled repeats from RepARK as male-specific, male-biased, or unbiased based on coverage of female and male genomic reads. After *de novo* assembling repeats, we mapped male and female genomic reads to the repeats and removed all repeats that did not have at least 5 times the average genome coverage in one sex. We categorized male-specific repeats as those with at least 10 times the average male genome coverage and less than 2 times the average female genome coverage, and male-biased repeats as those with at least 10 times the average male genome coverage, at least 2 times the average female genome coverage, and male coverage at least 2.5 times higher than female coverage.

Figure S13. Comparisons of genomic coverage of *de novo* assembled male-biased and male-specific repeats for karyotypes containing at least one Y chromosome (XY, XXY, and XYY).

Figure S14. Comparisons of total transcript levels of *de novo* assembled male-biased and male-specific repeats for karyotypes containing at least one Y chromosome (XY, XXY, and XYY).

Figure S15. Normalized expression variation of male-specific/male-biased repeats between flies with different sex chromosome karyotypes.

Supplementary Tables

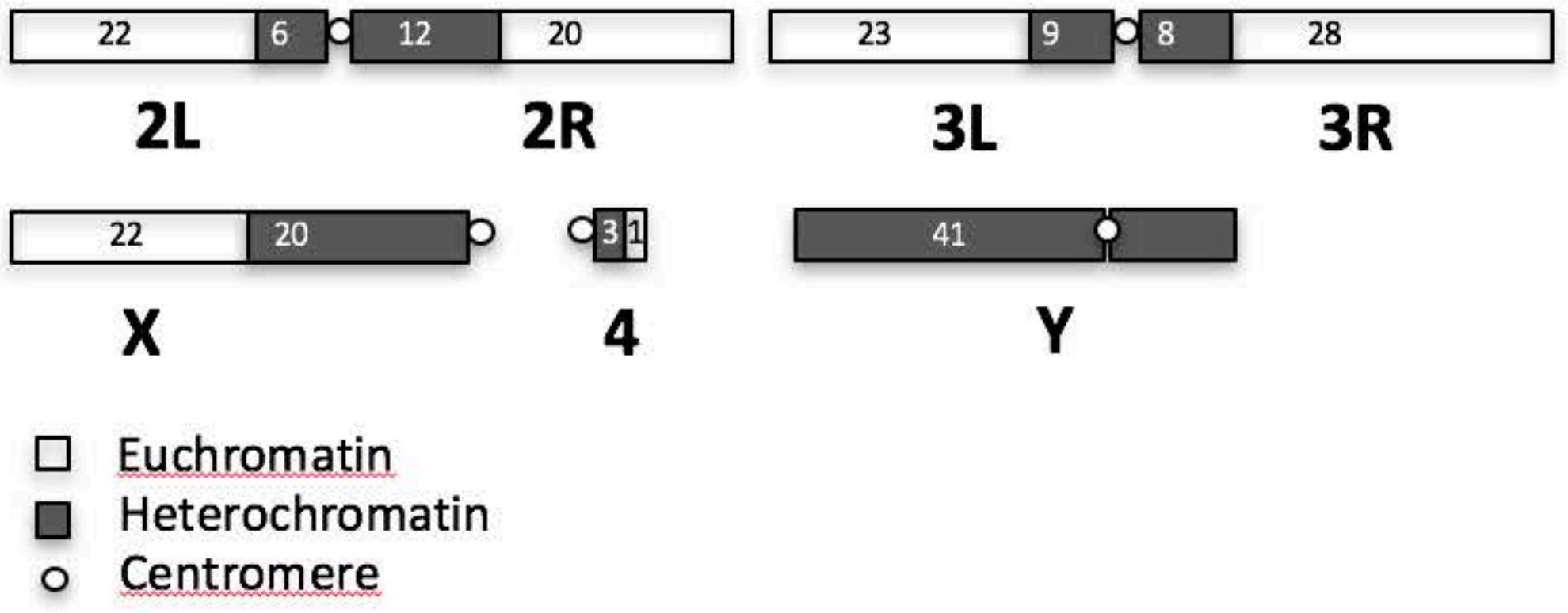
Table S1. Flow cytometry estimates of genome sizes for flies with different karyotypes.

Table S2. Pearson correlation coefficients of signal of *D. miranda* spike.

Table S3. Pearson correlation coefficients for different ChIP experiments.

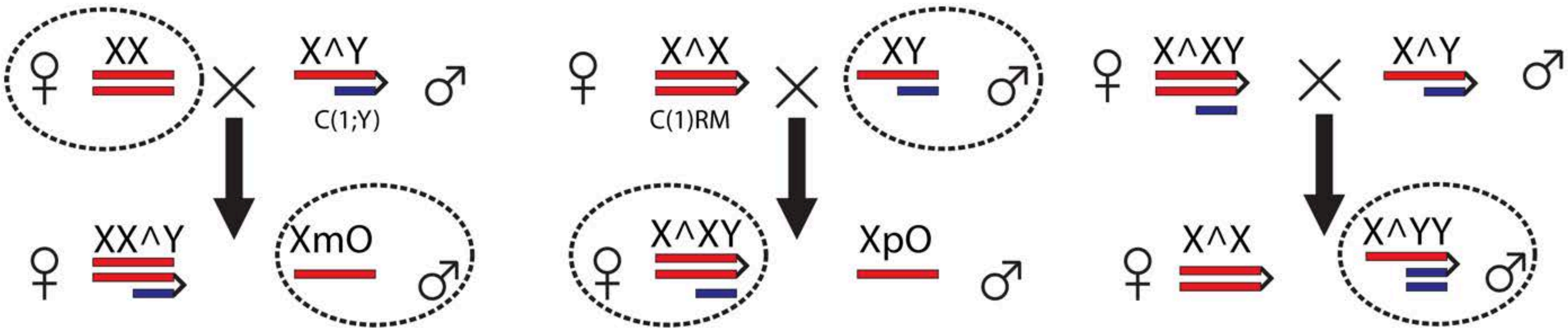
Table S4. GO categories of differentially expressed genes.

A.



Approximate heterochromatin content (diploid genome):
X0 males: 111 Mb
XX females: 124 Mb
XY males: 151 Mb
XXY females: 161 Mb
XYY males: 181 Mb

B.



H3K4me3

2L

2R

XY male
XX female
XO male
XXY female
XYY male

5.0
0.0

XX minus XY
XX minus XXY
XX minus XO
XY minus XO
XY minus XYY

3L

3R

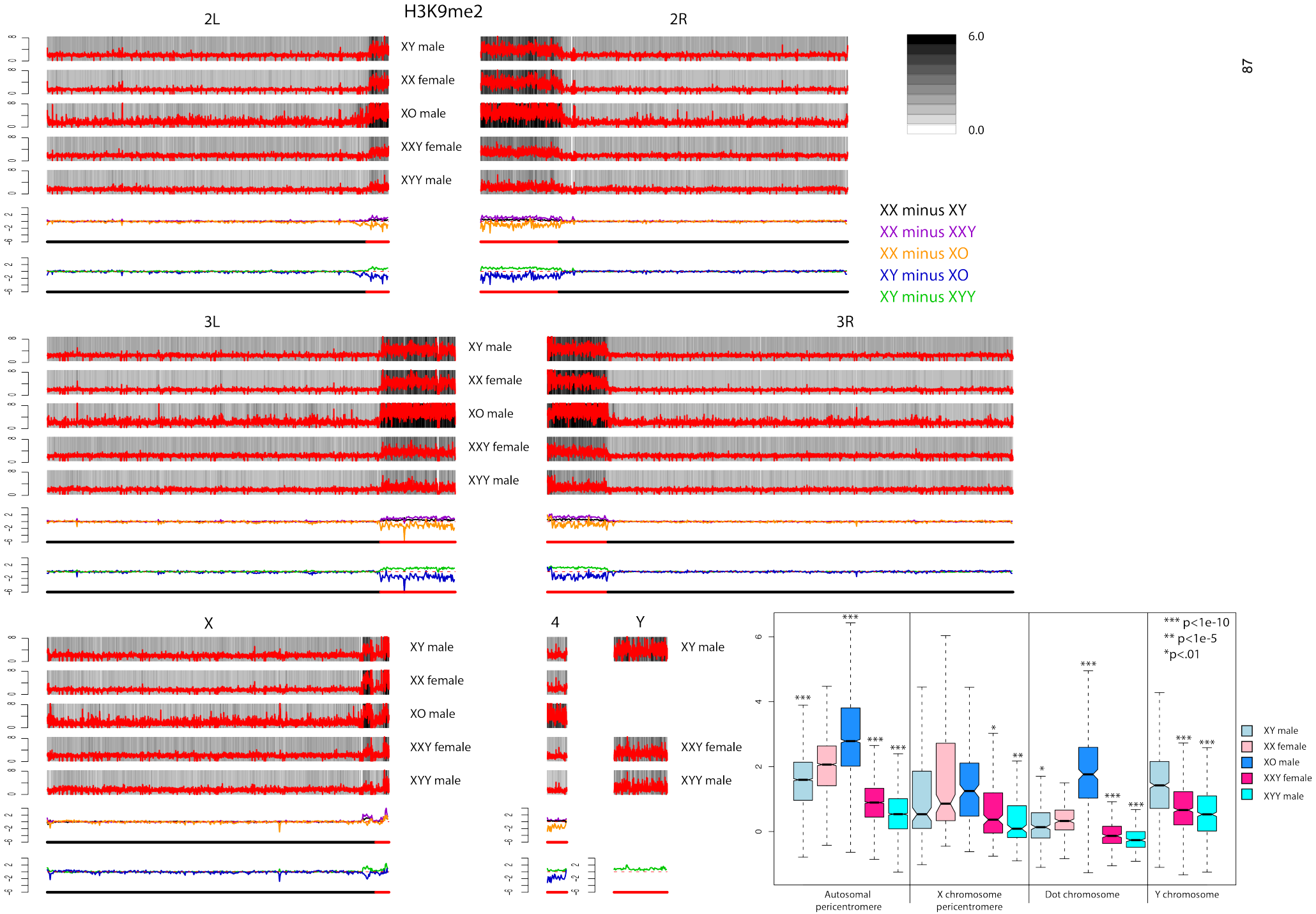
X

4

Y

*** p<1e-10
** p<1e-5
*p<.01

Autosomal euchromatin
X chromosome euchromatin
Dot chromosome
Y chromosome


H3K9me2
2L
2R
3L
3R
X
4
Y
XY male
XX female
XO male
XXY female
XYY male
6.0
0.0
XX minus XY
XX minus XXY
XX minus XO
XY minus XO
XY minus XYY
*** p<1e-10
** p<1e-5
*p<.01
Autosomal pericentromere
X chromosome pericentromere
Dot chromosome
Y chromosome

H3K9me3

2L

2R

XY male
XX female
XO male
XXY female
XYY male

8.0
0.0

XX minus XY
XX minus XXY
XX minus XO
XY minus XO
XY minus XYY

3L

3R

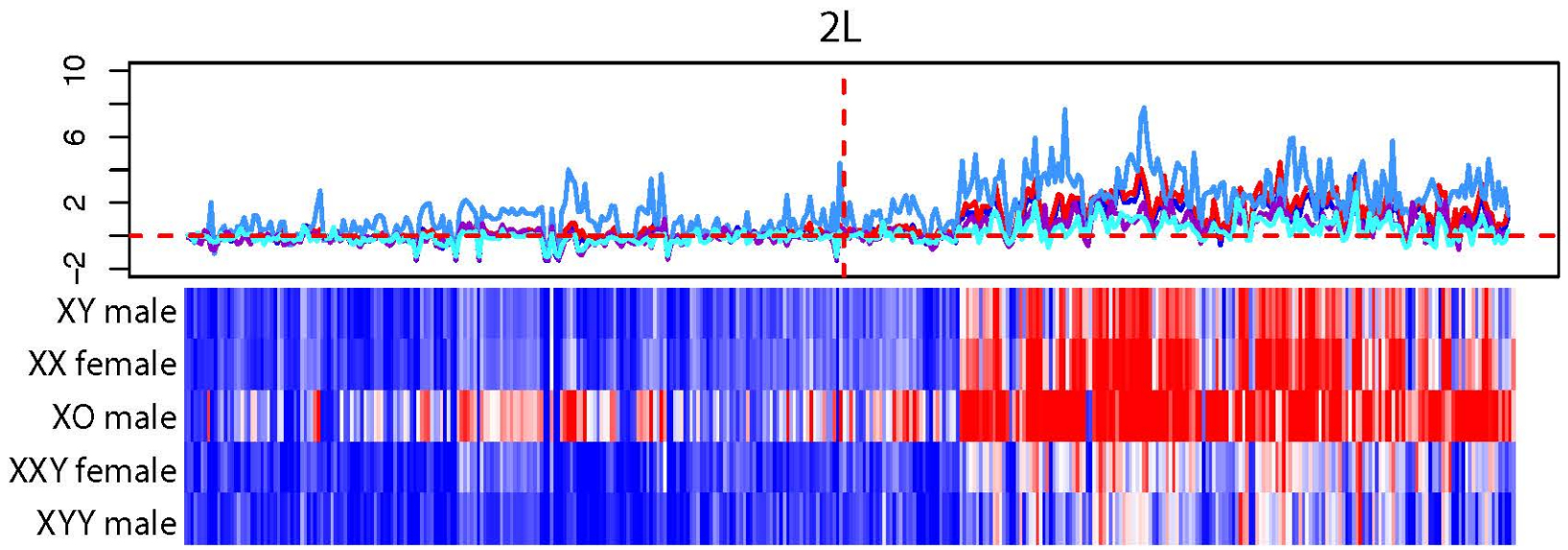
X

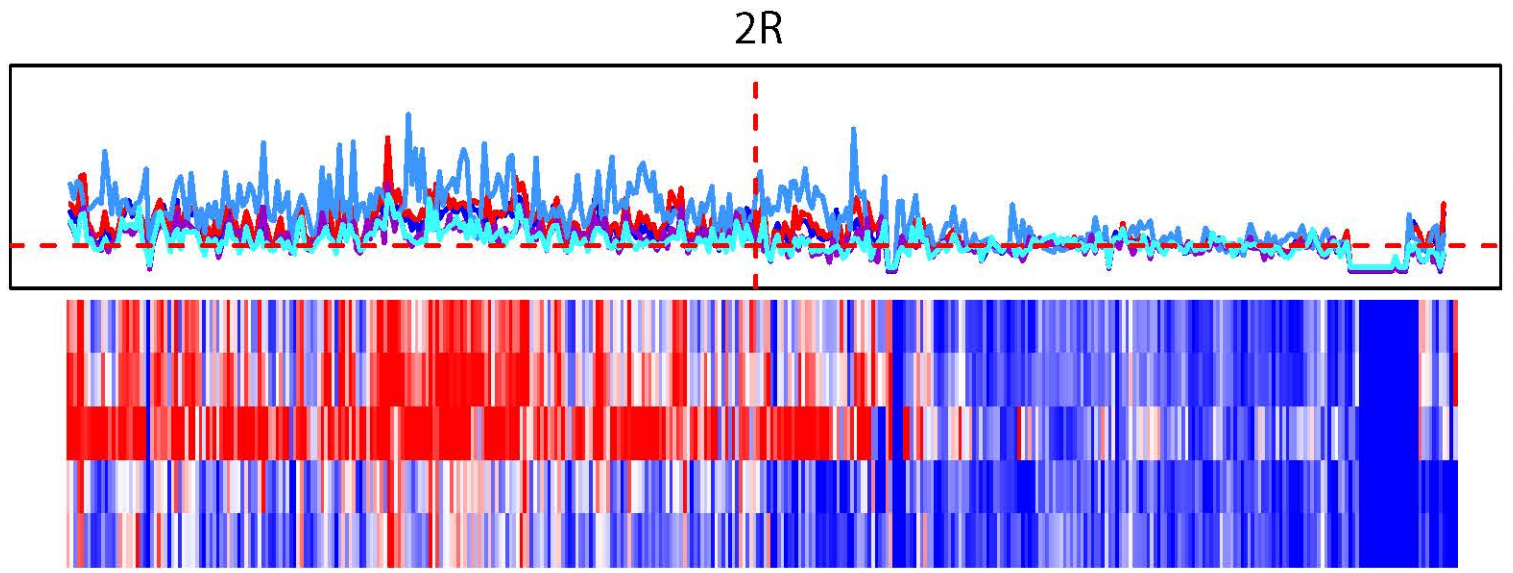
4

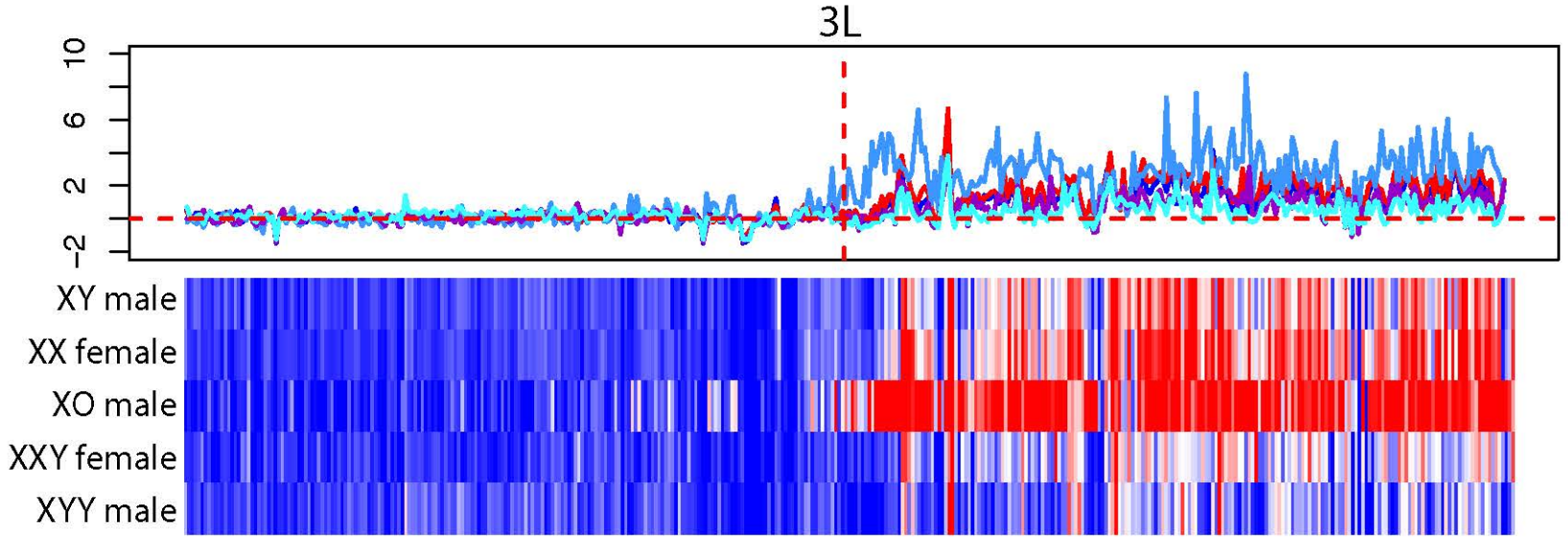
Y

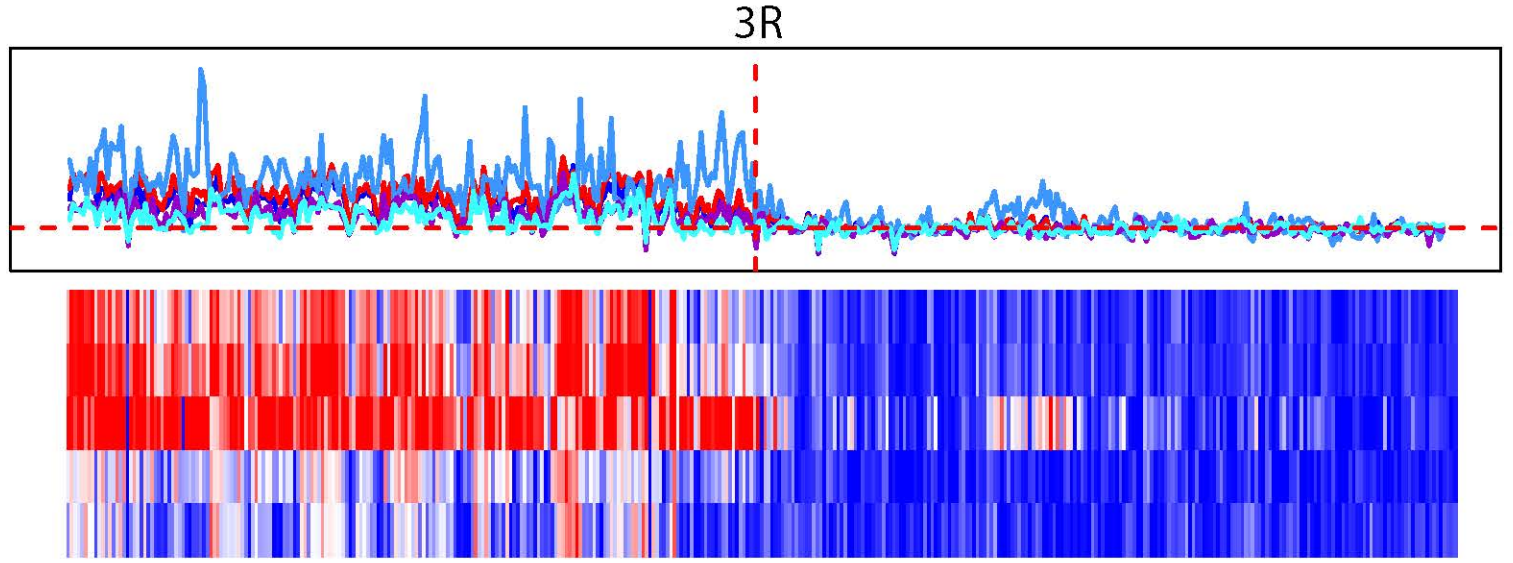
*** p<1e-10
** p<1e-5
*p<.01

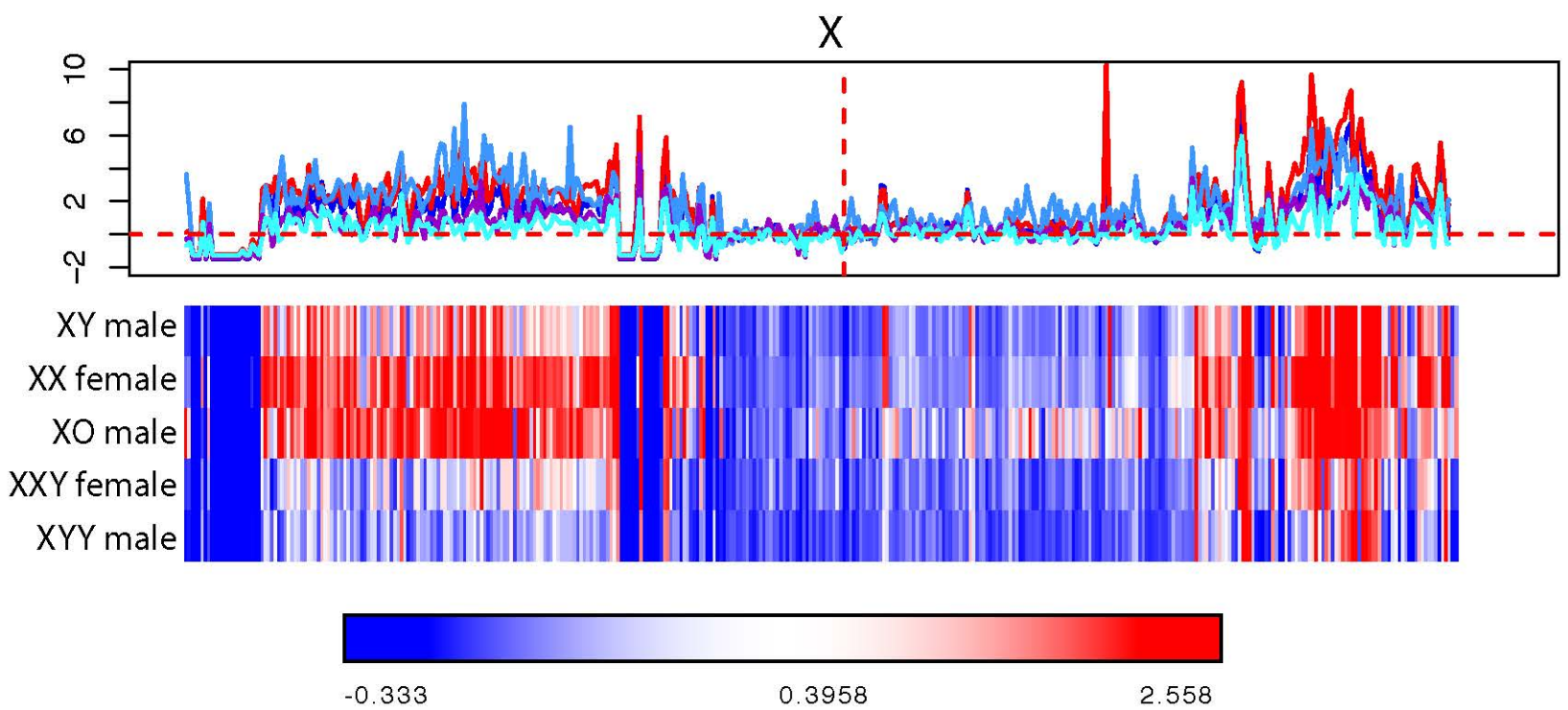
Autosomal pericentromere
X chromosome pericentromere
Dot chromosome
Y chromosome


2L
10
6
2
−2
XY male
XX female
XO male
XXY female
XYY male

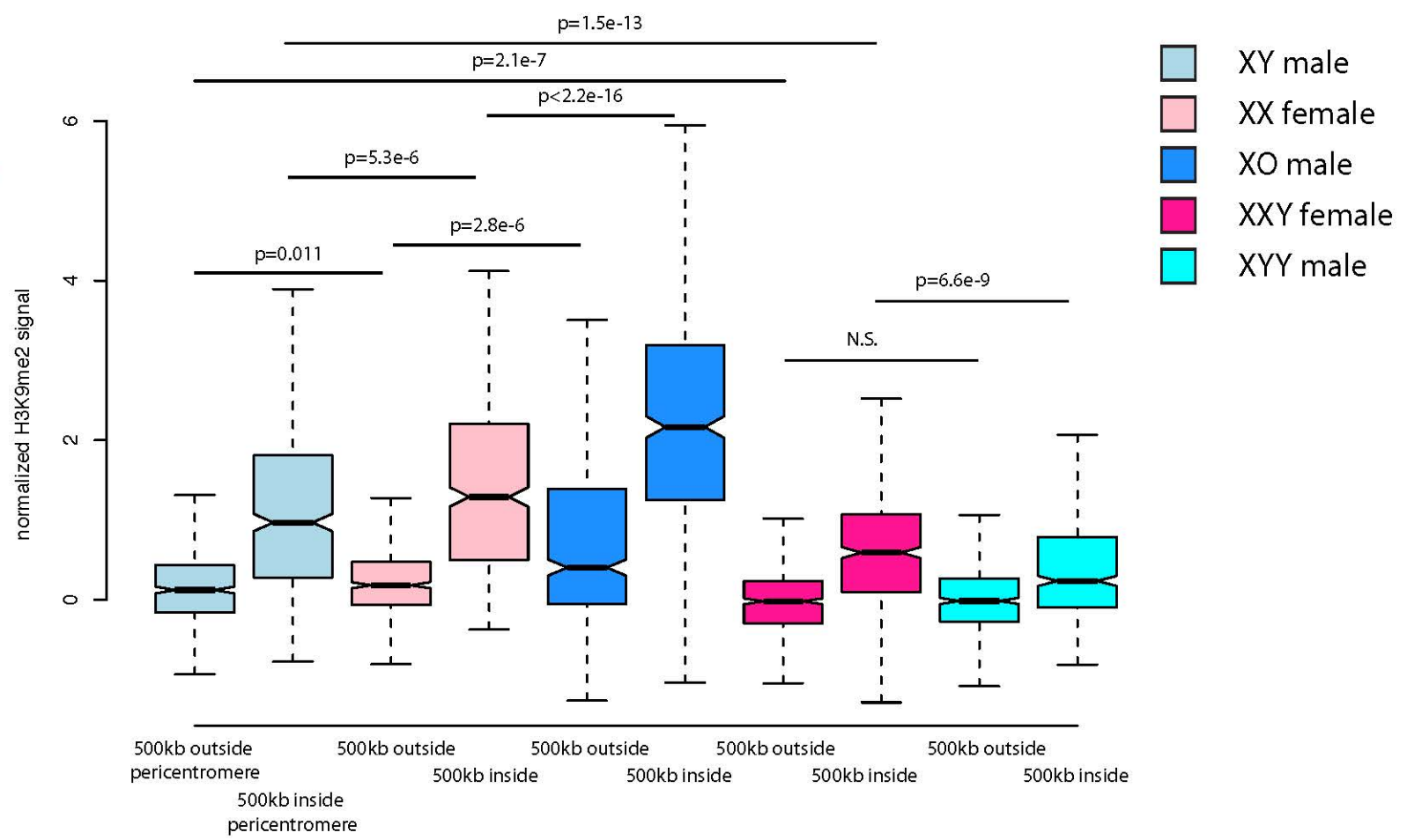

2R


3L
10
6
2
−2
XY male
XX female
XO male
XXY female
XYY male

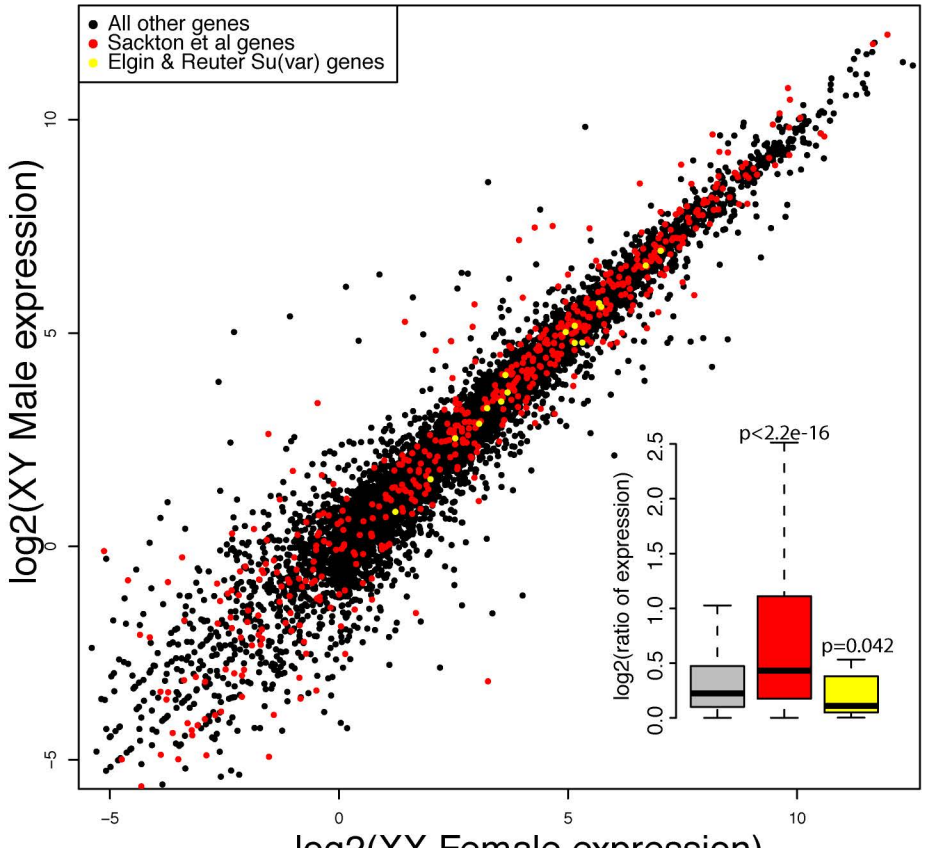

3R

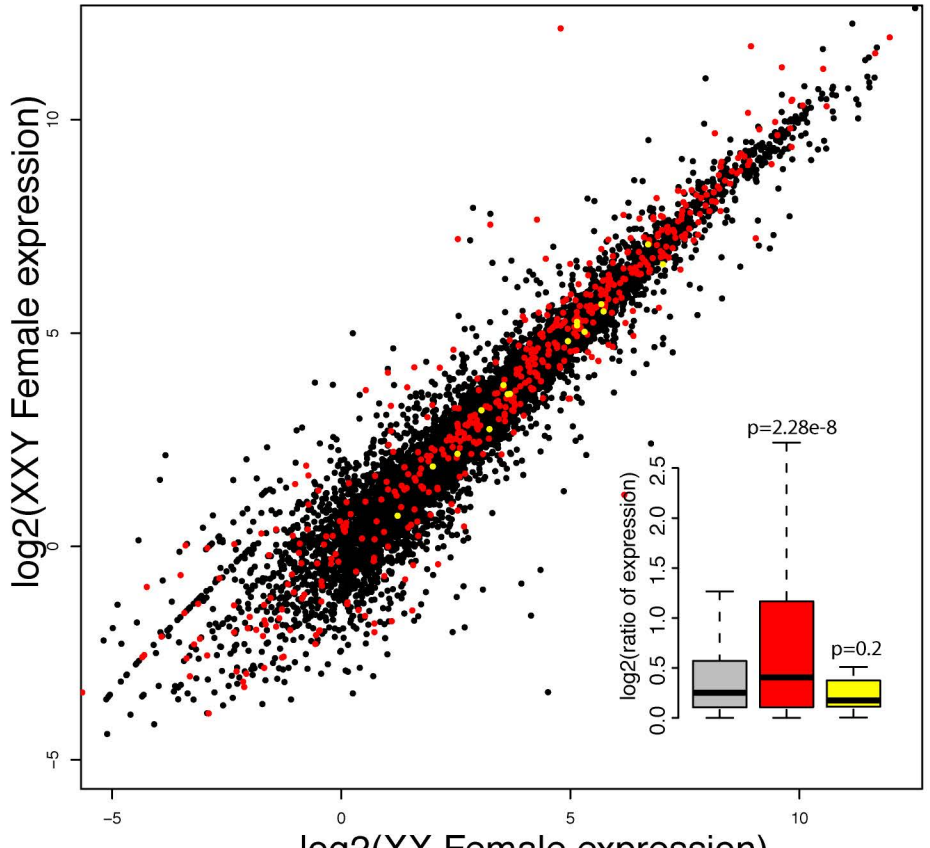

X
10
6
2
−2
XY male
XX female
XO male
XXY female
XYY male
-0.333
0.3958
2.558

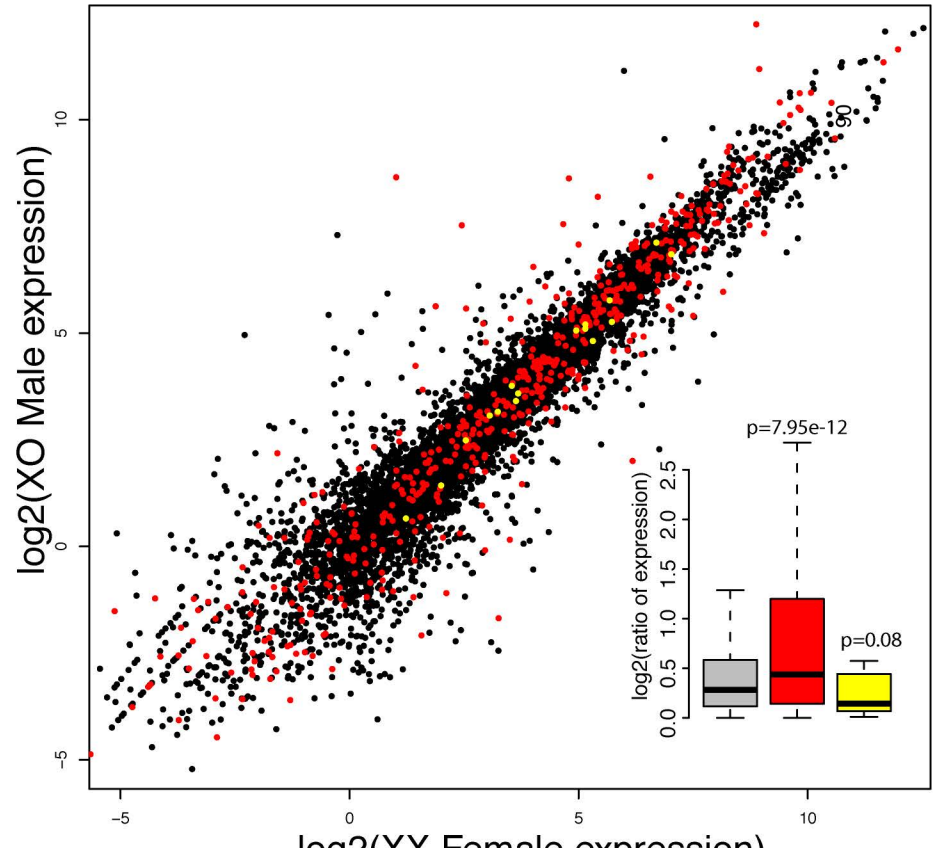

XX female
XY male
XO male
XXY female
XYY male

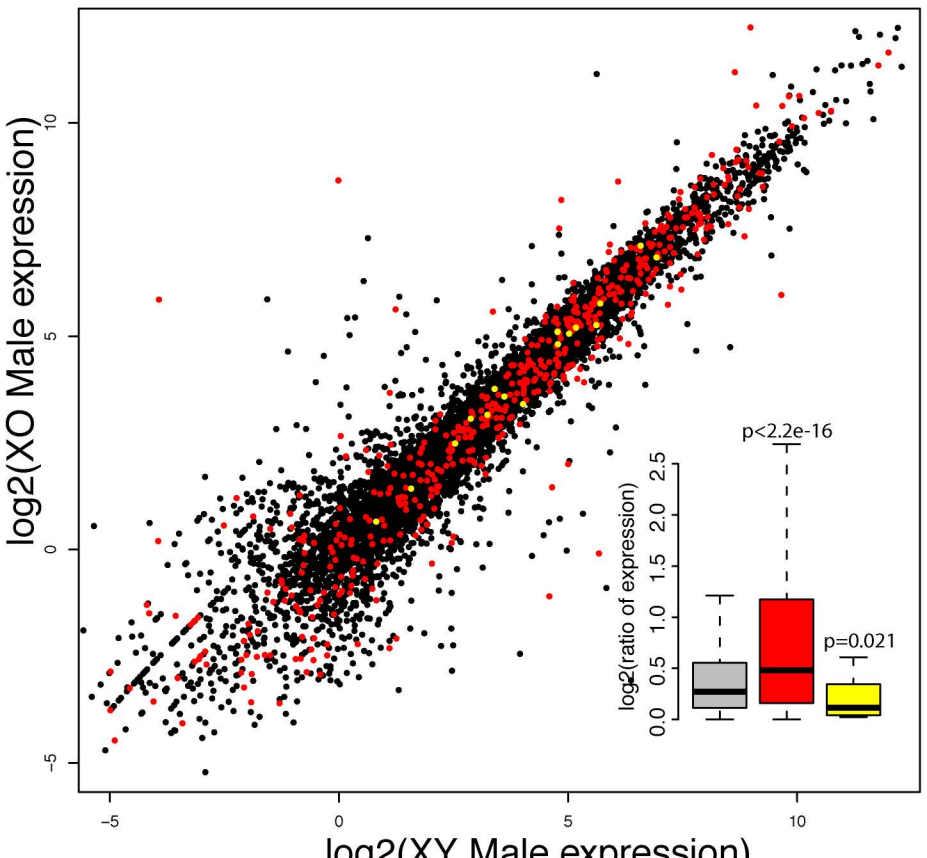

p=1.5e-13
p=2.1e-7
p<2.2e-16
p=5.3e-6
p=2.8e-6
p=0.011
p=6.6e-9
N.S.
normalized H3K9me2 signal
0
2
4
6
500kb outside pericentromere
500kb inside pericentromere
500kb outside
500kb inside
500kb outside
500kb inside
500kb outside
500kb inside
500kb outside
500kb inside
XY male
XX female
XO male
XXY female
XYY male

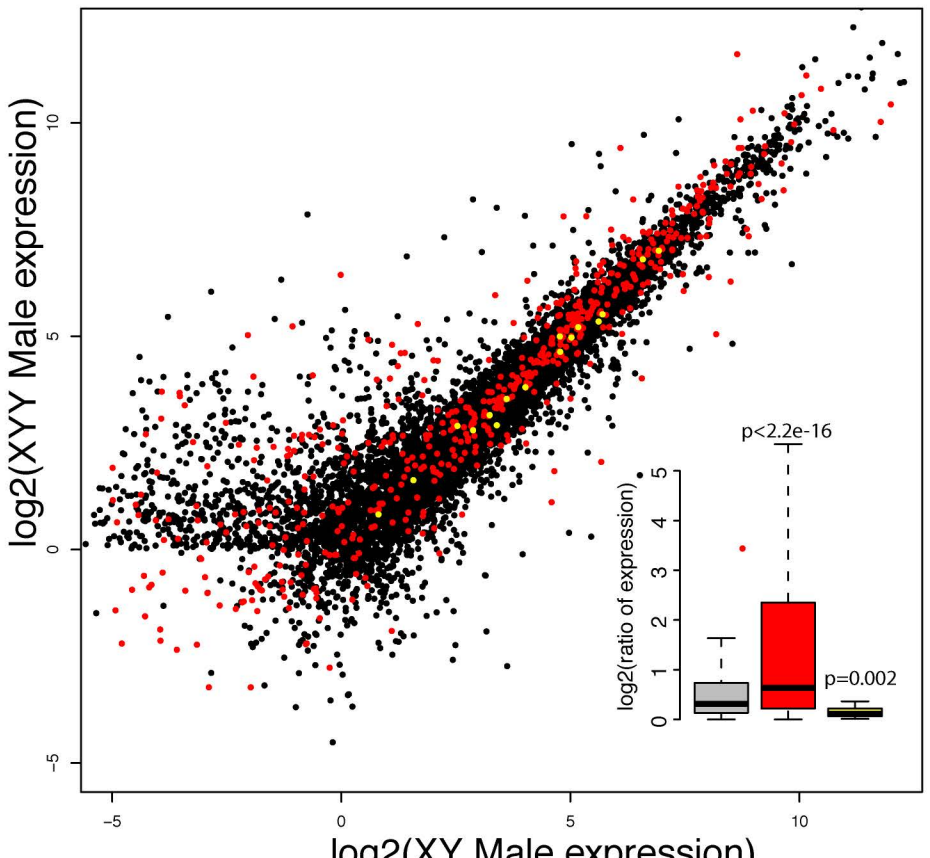
A.


All other genes
Sackton et al genes
Elgin & Reuter Su(var) genes
log2(XY Male expression)
log2(XX Female expression)
log2(ratio of expression)
p<2.2e-16
p=0.042

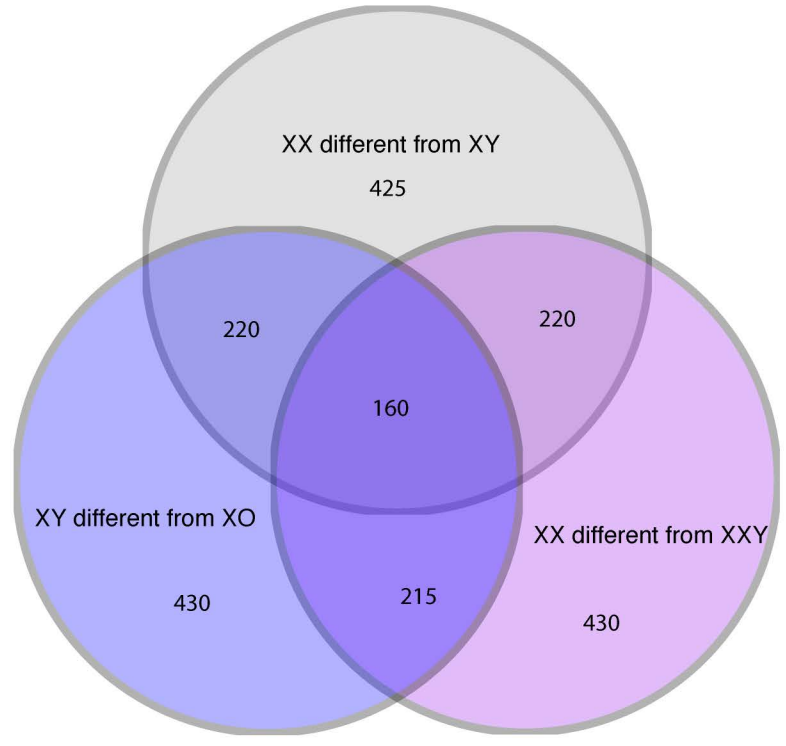

log2(XXY Female expression)
log2(XX Female expression)
log2(ratio of expression)
p=2.28e-8
p=0.2

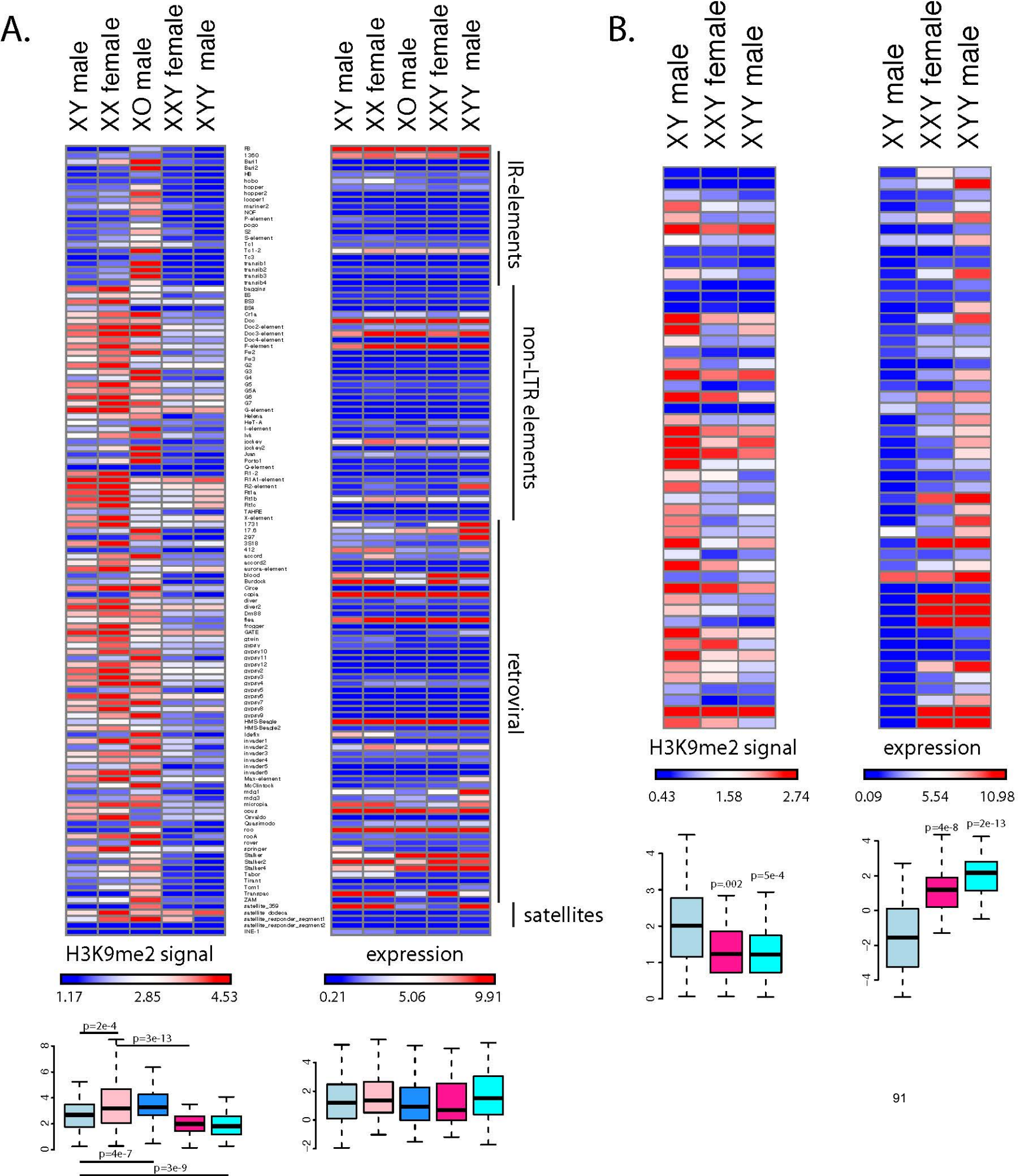

log2(XO Male expression)
log2(XX Female expression)
log2(ratio of expression)
p=7.95e-12
p=0.08


log2(XO Male expression)
log2(XY Male expression)
log2(ratio of expression)
p<2.2e-16
p=0.021


log2(XXY Male expression)
log2(XY Male expression)
log2(ratio of expression)
p<2.2e-16
p=0.002

B.


XX different from XY
425
220
220
160
XY different from XO
430
215
XX different from XXY
430


A.
XY male
XX female
XO male
XXY female
XYY male
IR-elements
non-LTR elements
retroviral
satellites
H3K9me2 signal
1.17
2.85
4.53
expression
0.21
5.06
9.91
p=2e-4
p=3e-13
p=4e-7
p=3e-9
B.
H3K9me2 signal
0.43
1.58
2.74
p=.002
p=5e-4
expression
0.09
5.54
10.98
p=4e-8
p=2e-13

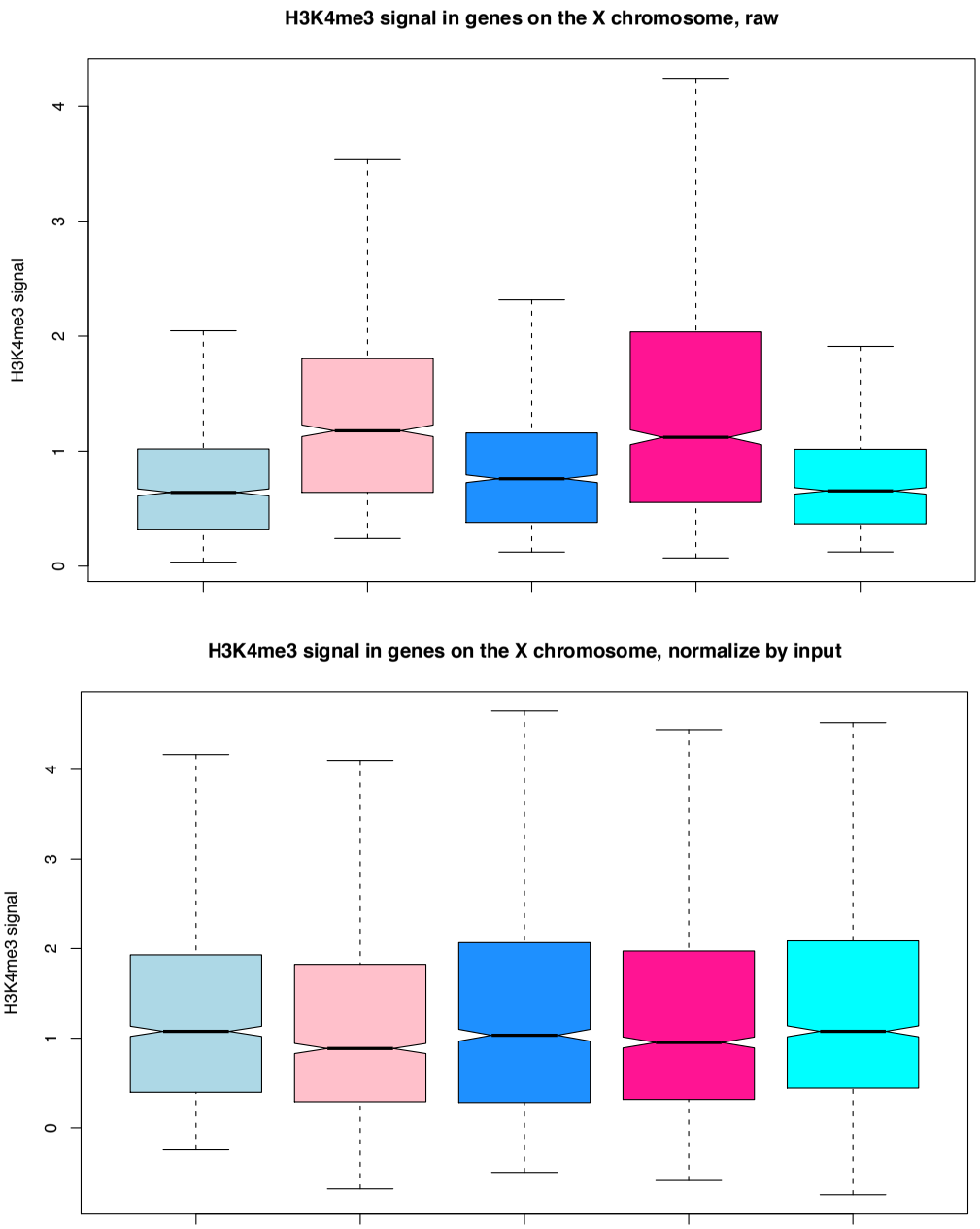
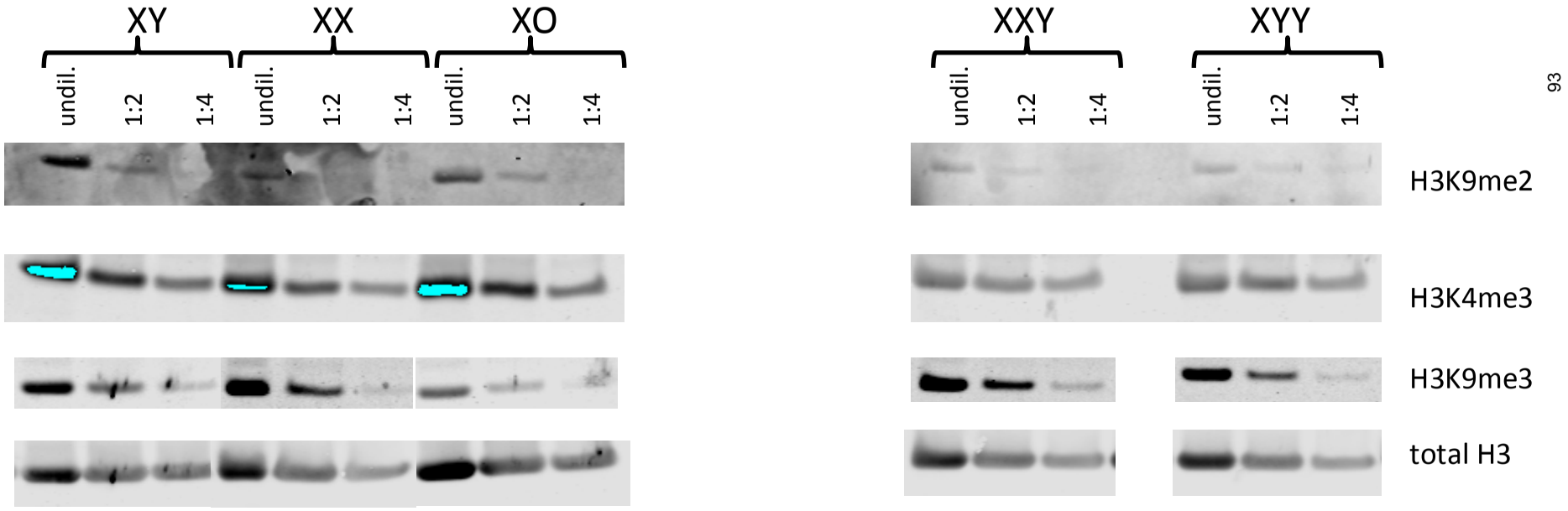

H3K4me3 signal in genes on the X chromosome, raw
H3K4me3 signal
0
1
2
3
4
H3K4me3 signal in genes on the X chromosome, normalize by input
H3K4me3 signal
0
1
2
3
4
XY male
XX female
X0 male
XXY female
XYY male

Figure S1


XY
XX
XO
XXY
XYY
undil.
1:2
1:4
H3K9me2
H3K4me3
H3K9me3
total H3

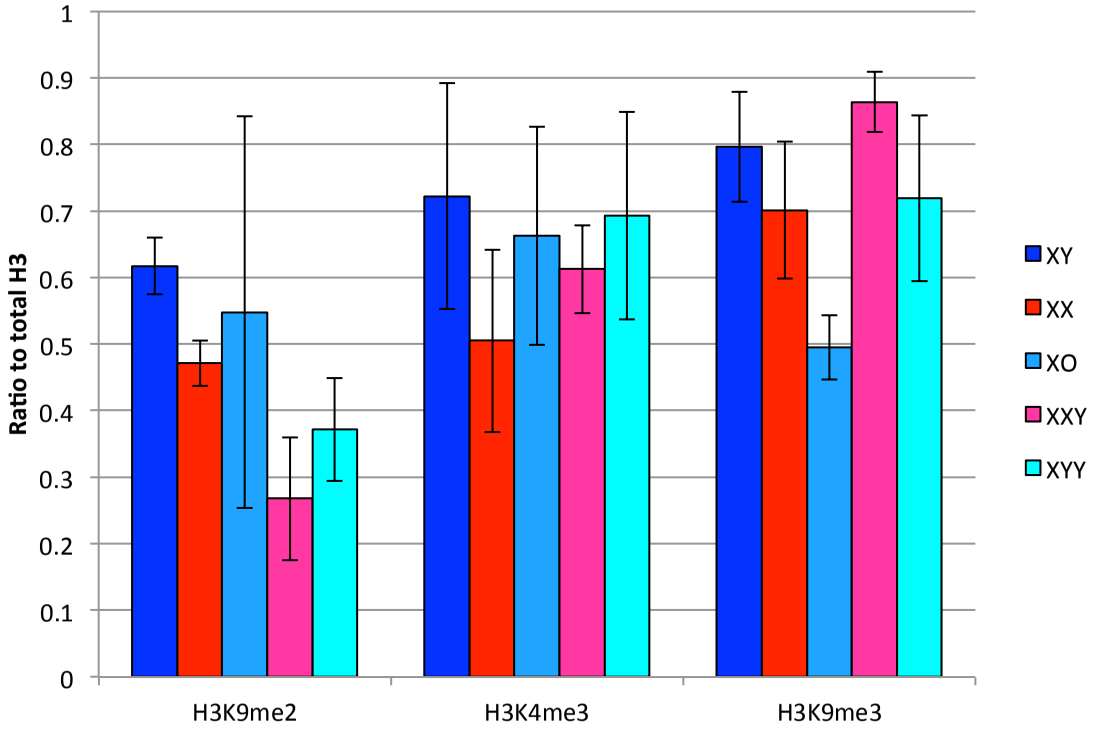

Ratio to total H3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
H3K9me2
H3K4me3
H3K9me3
XY
XX
XO
XXY
XYY

Figure S2

H3K9me3 1Mb around het/eu boundary

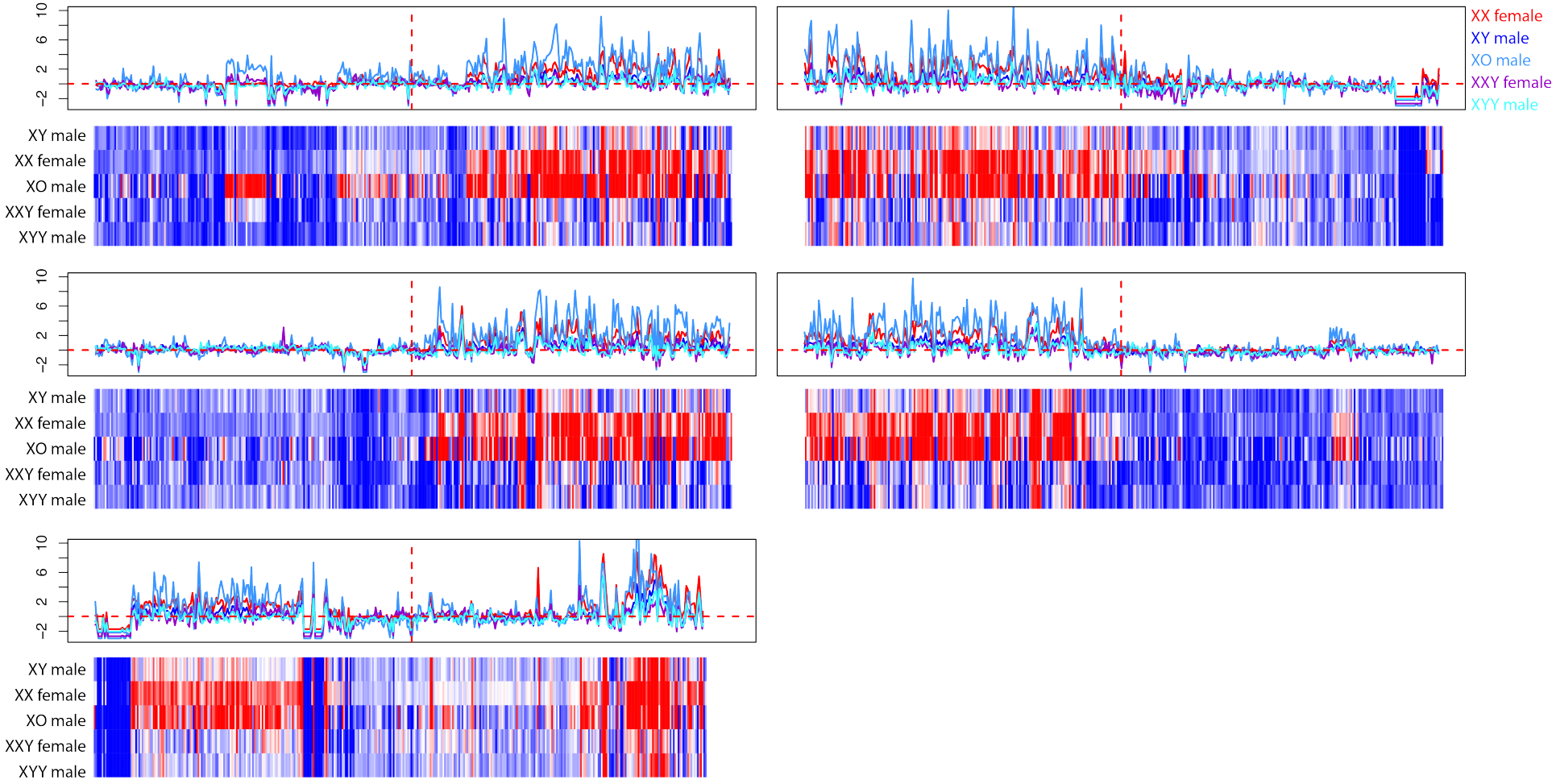


Figure S3

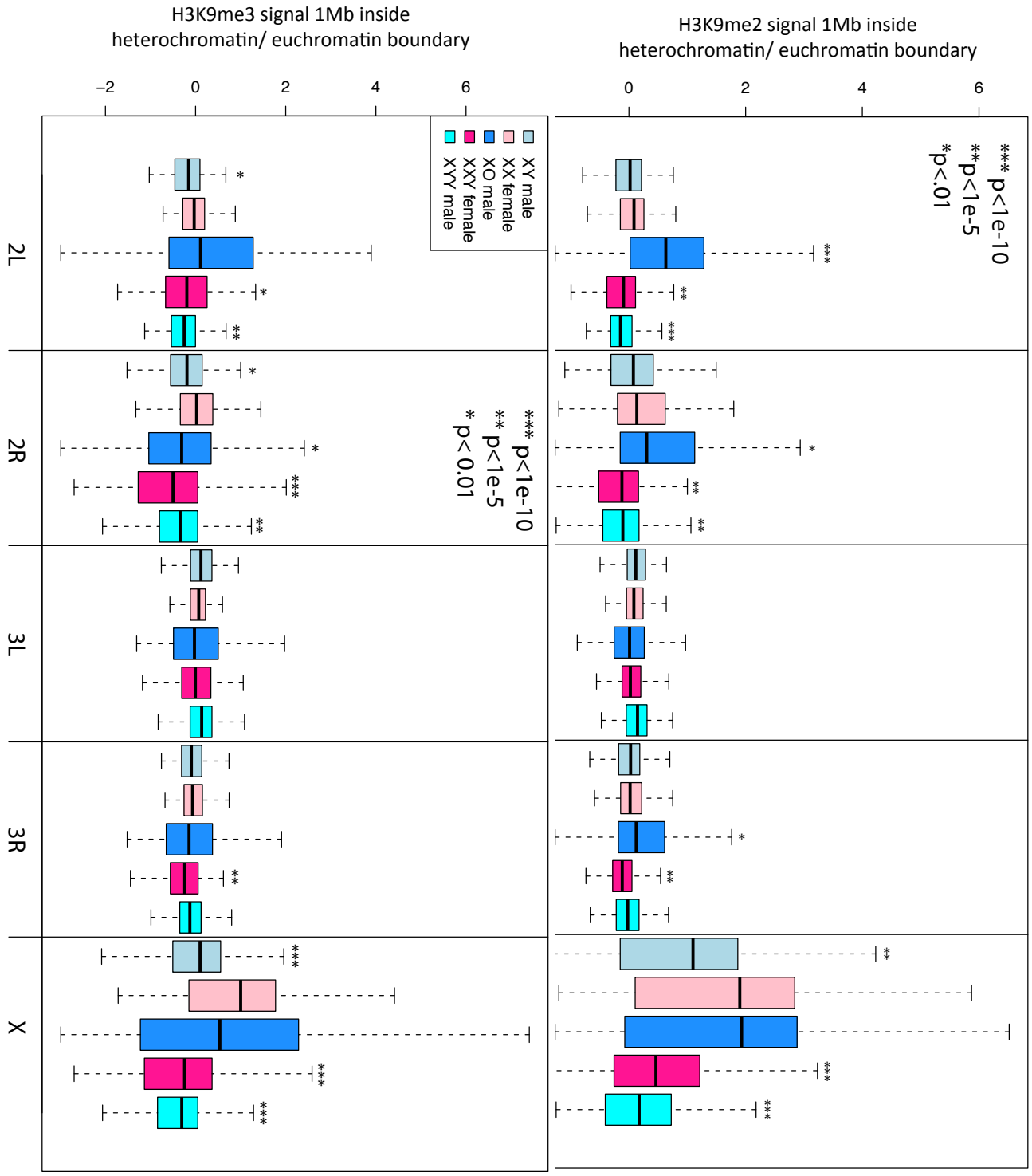

H3K9me3 signal 1Mb inside
heterochromatin/ euchromatin boundary
H3K9me2 signal 1Mb inside
heterochromatin/ euchromatin boundary
XY male
XX female
XO male
XXY female
XYY male
*** p<1e-10
** p<1e-5
* p< 0.01
2L
2R
3L
3R
X

Figure S4

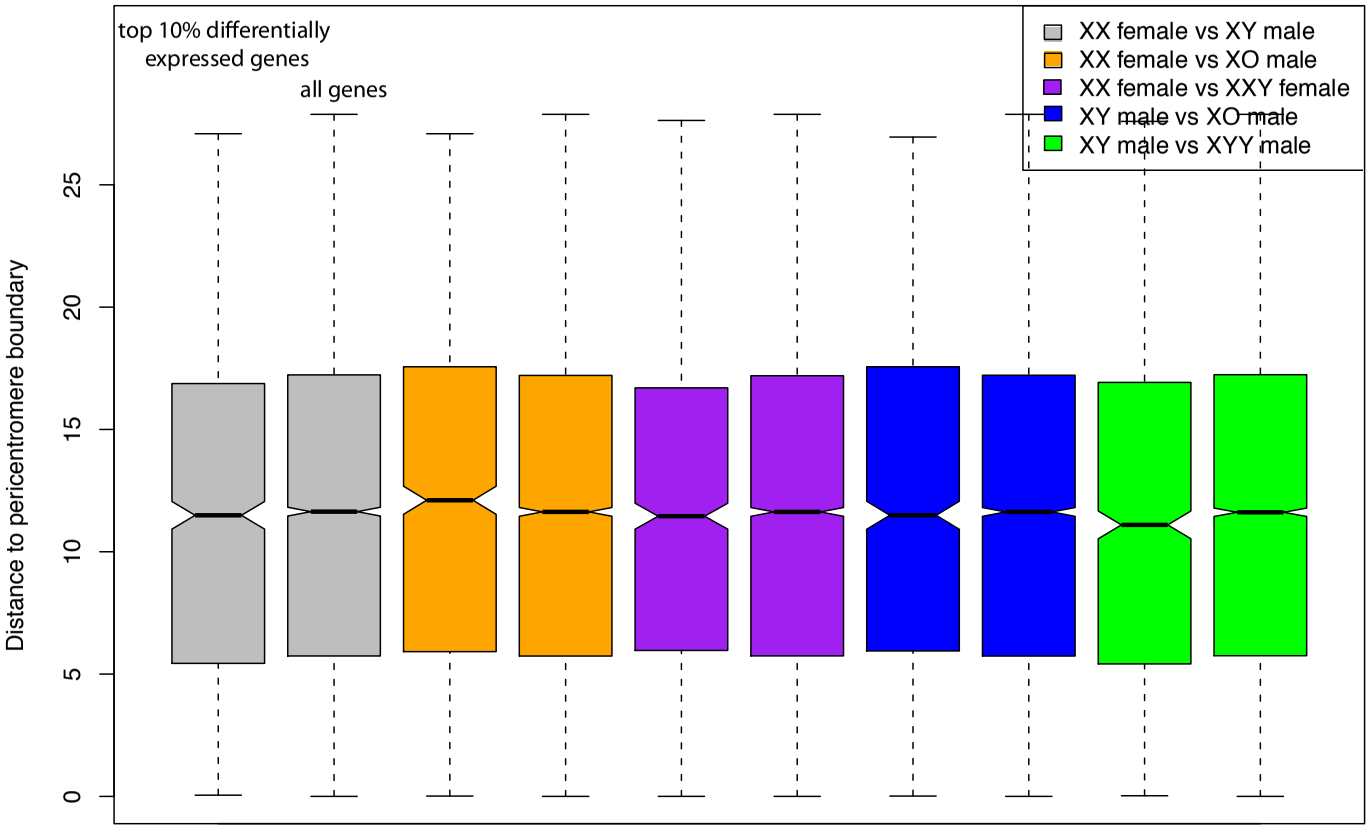

top 10% differentially
expressed genes
all genes
XX female vs XY male
XX female vs XO male
XX female vs XXY female
XY male vs XO male
XY male vs XYY male
Distance to pericentromere boundary
0
5
10
15
20
25

Figure S5

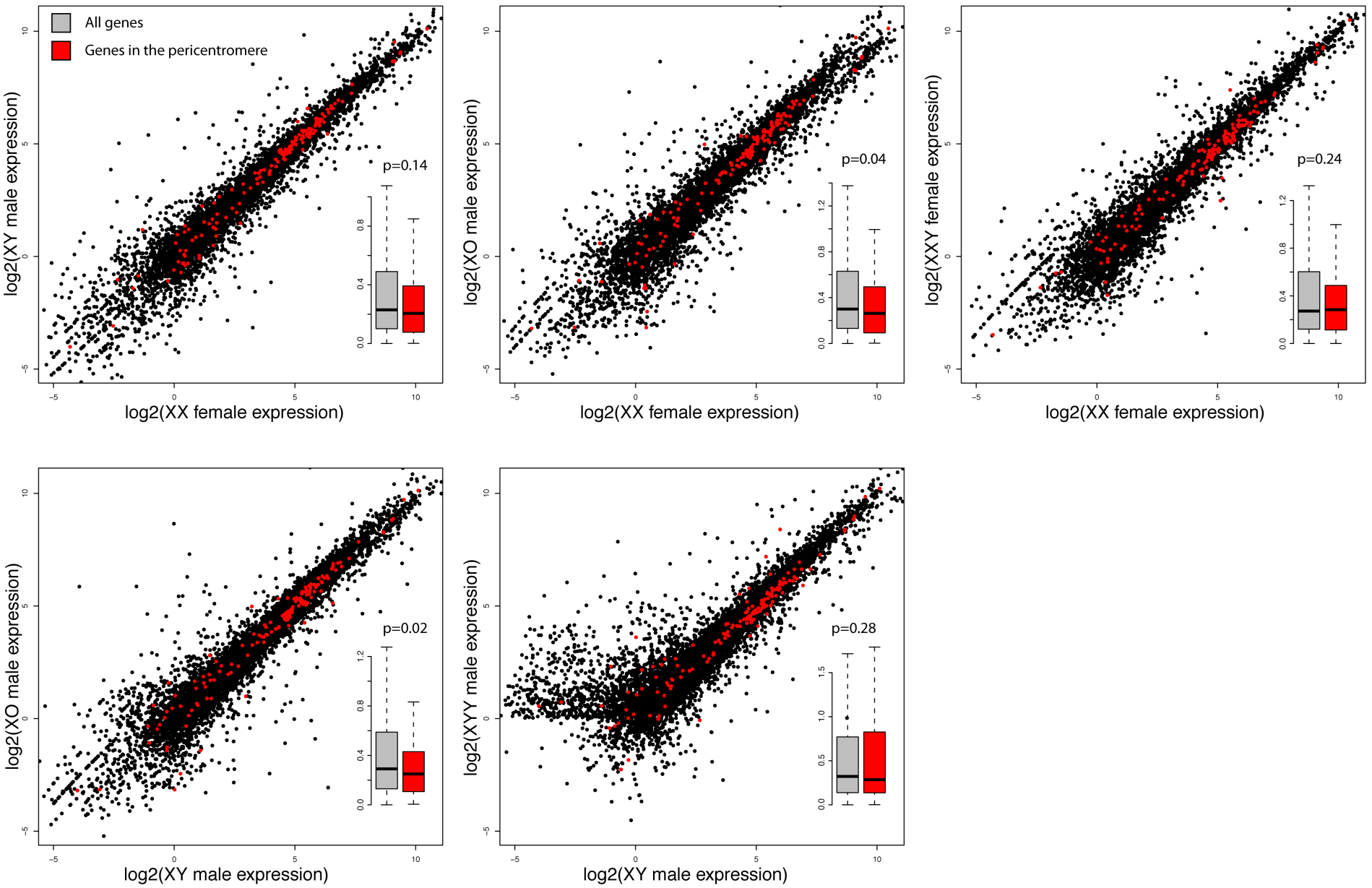
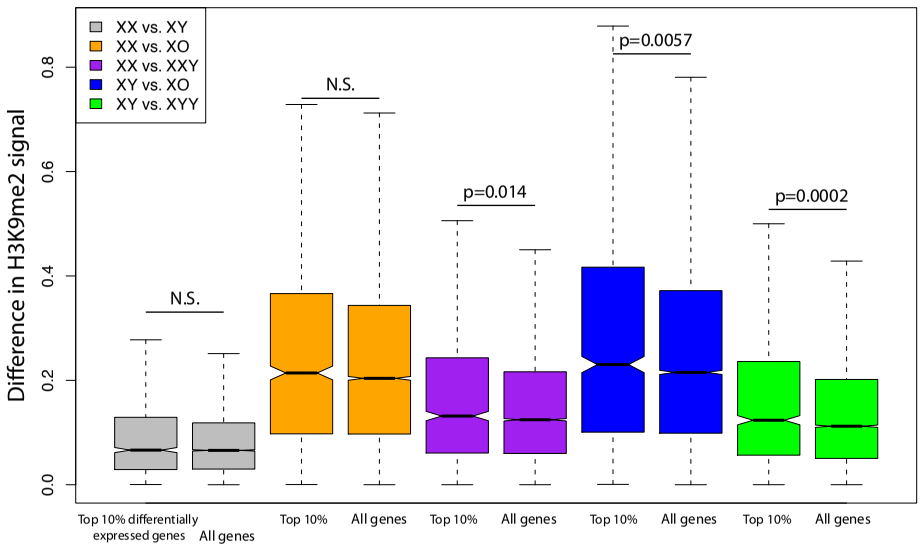
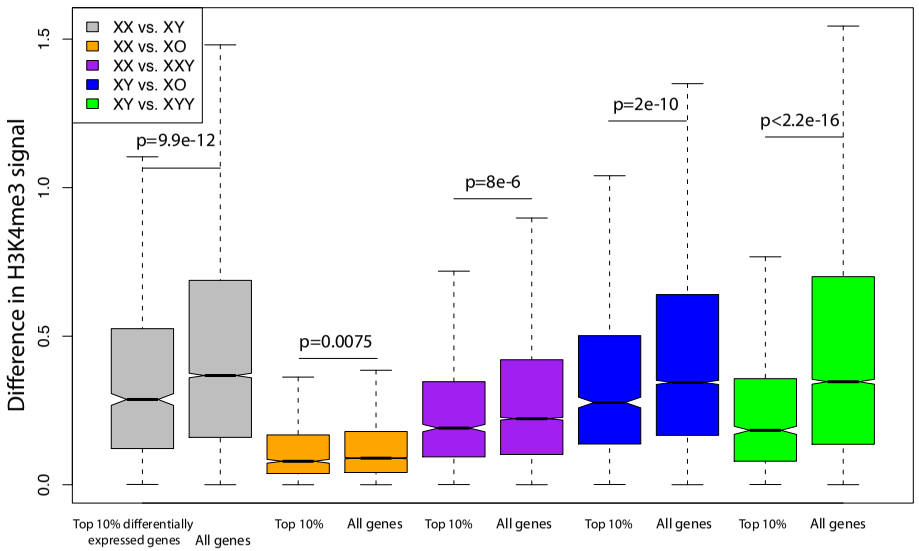

All genes
Genes in the pericentromere
p=0.14
log2(XY male expression)
log2(XX female expression)
p=0.04
log2(XO male expression)
log2(XX female expression)
p=0.24
log2(XXY female expression)
log2(XX female expression)
p=0.02
log2(XO male expression)
log2(XY male expression)
p=0.28
log2(XYY male expression)
log2(XY male expression)

Figure S6



difference in H3K9me2 signal of genes differentially expressed vs. all genes



difference in H3K4me3 signal of genes differentially expressed vs. all genes

Figure S7

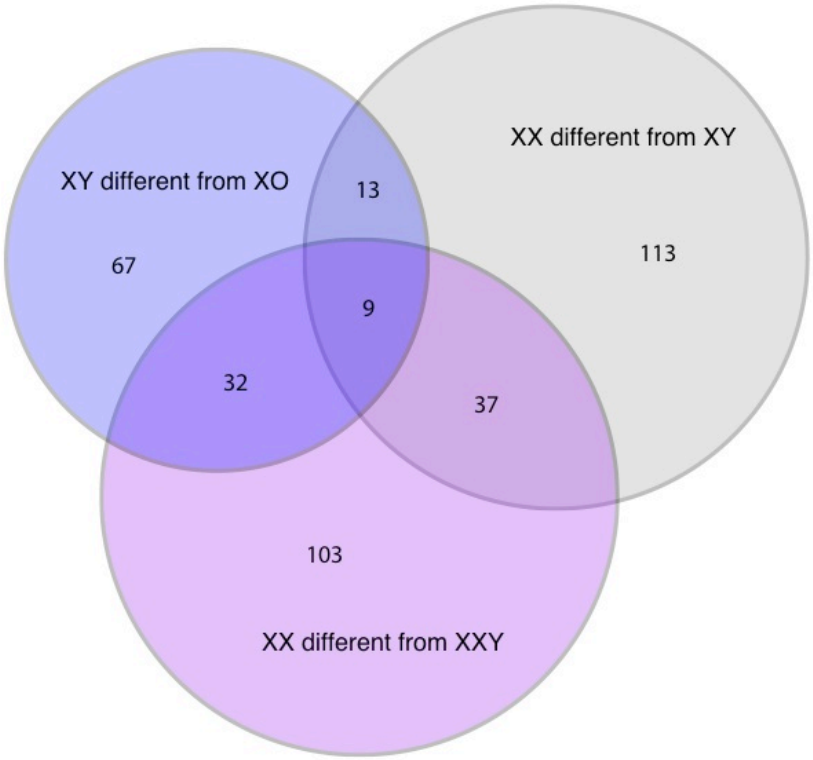

XY different from XO
67
13
XX different from XY
113
9
32
37
103
XX different from XXY

Figure S8

A.
XY male
XX female
XO male
XXY female
XYY male
retroviral
non-LTR elements
IR elements
B.
XY male
XXY female
XYY male

Figure S9

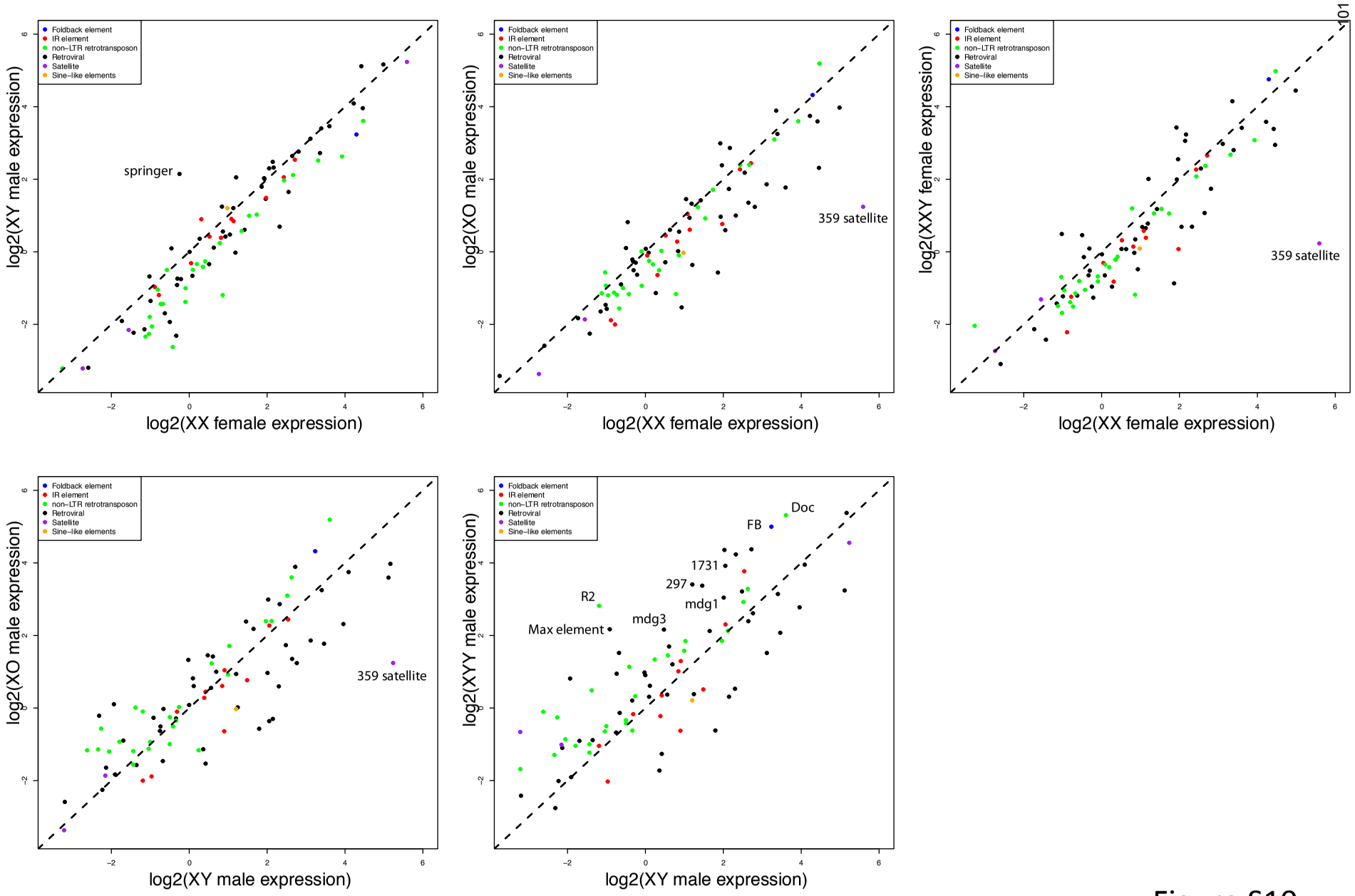

Foldback element
IR element
non−LTR retrotransposon
Retroviral
Satellite
Sine−like elements
springer
log2(XY male expression)
log2(XX female expression)
359 satellite
log2(XO male expression)
log2(XXY female expression)
101
log2(XXY male expression)
Doc
FB
1731
297
mdg1
mdg3
R2
Max element

Figure S10

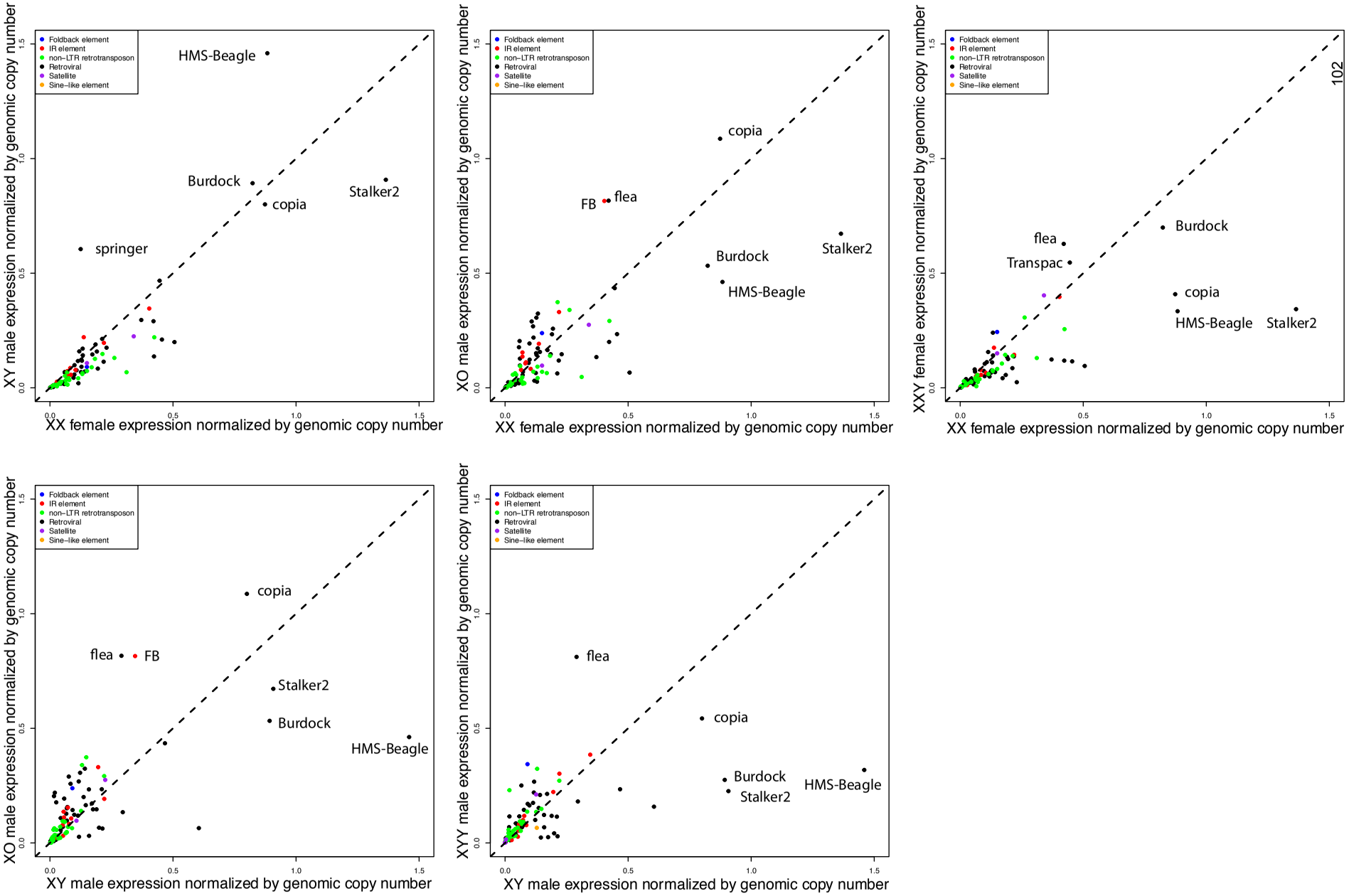

Foldback element
IR element
non−LTR retrotransposon
Retroviral
Satellite
Sine−like element
HMS-Beagle
Burdock
copia
Stalker2
springer
XY male expression normalized by genomic copy number
XX female expression normalized by genomic copy number
FB
flea
XO male expression normalized by genomic copy number
Transpac
XXY female expression normalized by genomic copy number
102
XYY male expression normalized by genomic copy number
0.0
0.5
1.0
1.5

Figure S11

Male specific:
Mcovg>10
Fcovg<2

Male biased:
Mcovg>10
Fcovg>2
Mcovg>2.5xFcovg

Unbiased

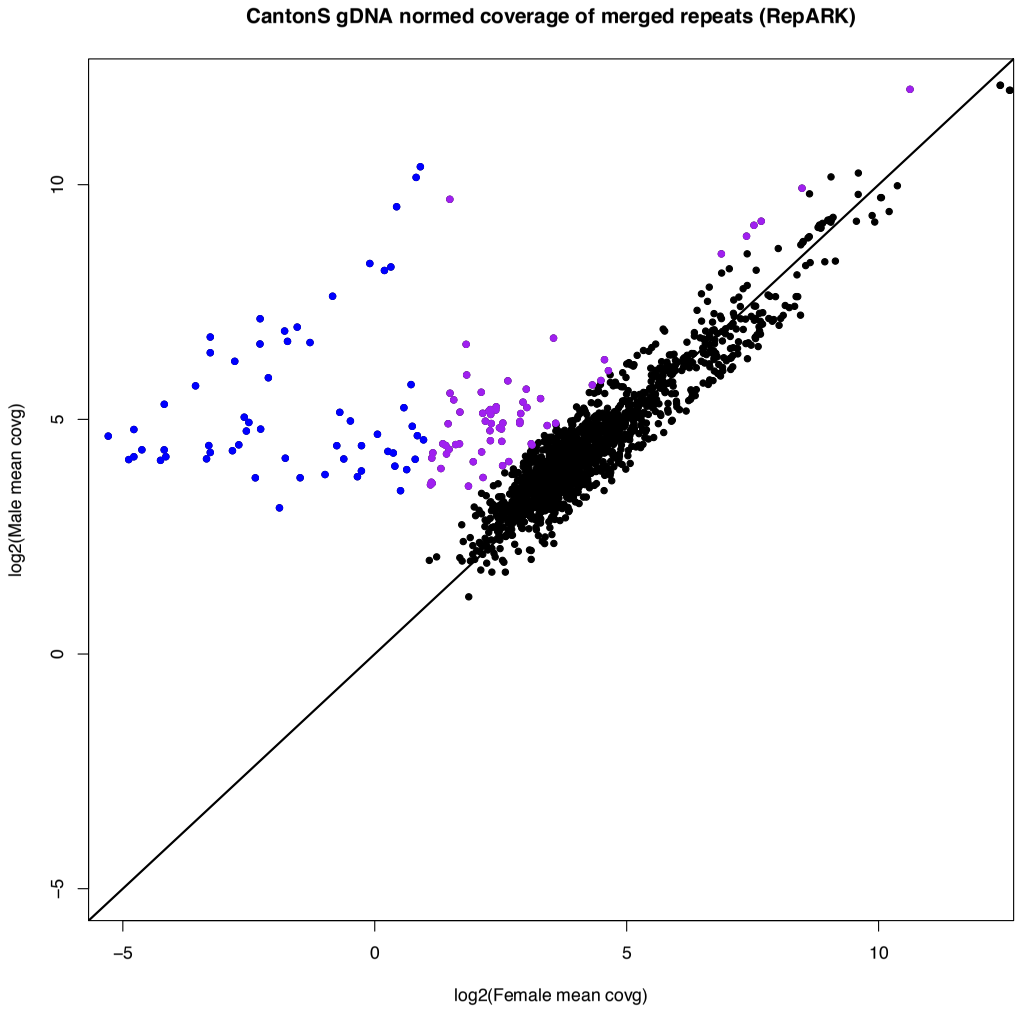


Figure S12

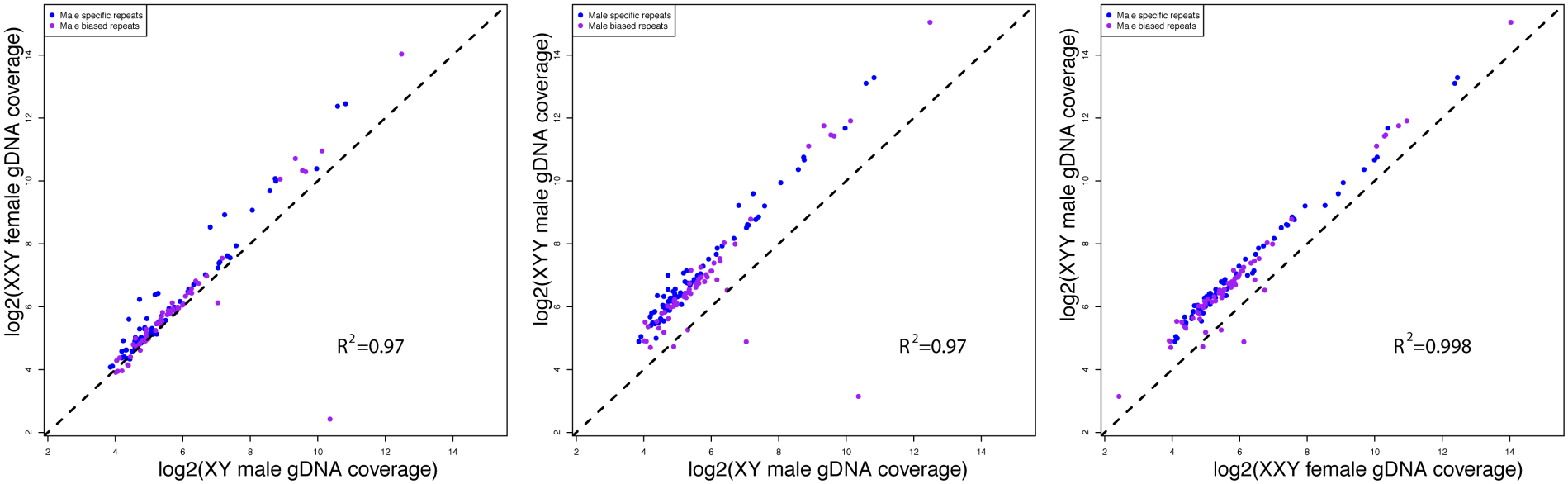

Male specific repeats
Male biased repeats
log2(XXY female gDNA coverage)
log2(XY male gDNA coverage)
R2=0.97
Male specific repeats
Male biased repeats
log2(XYY male gDNA coverage)
log2(XY male gDNA coverage)
R2=0.97
Male specific repeats
Male biased repeats
log2(XYY male gDNA coverage)
log2(XXY female gDNA coverage)
R2=0.998

Figure S13

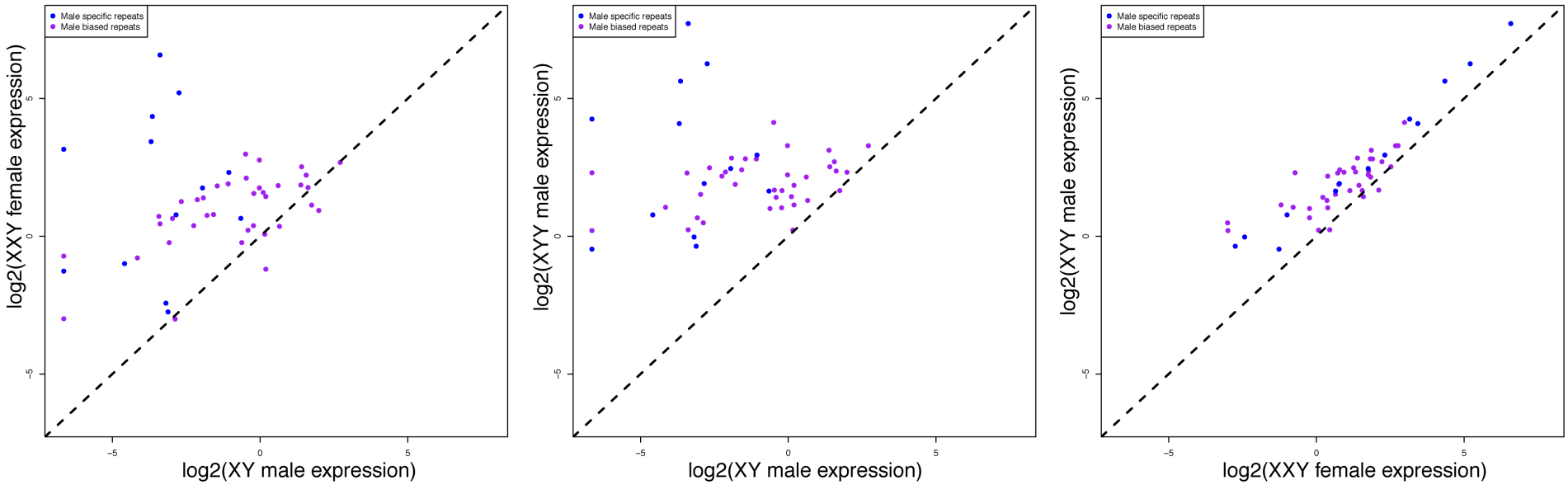

Male specific repeats
Male biased repeats
log2(XXY female expression)
log2(XY male expression)
Male specific repeats
Male biased repeats
log2(XYY male expression)
log2(XY male expression)
Male specific repeats
Male biased repeats
log2(XYY male expression)
log2(XXY female expression)

Figure S14

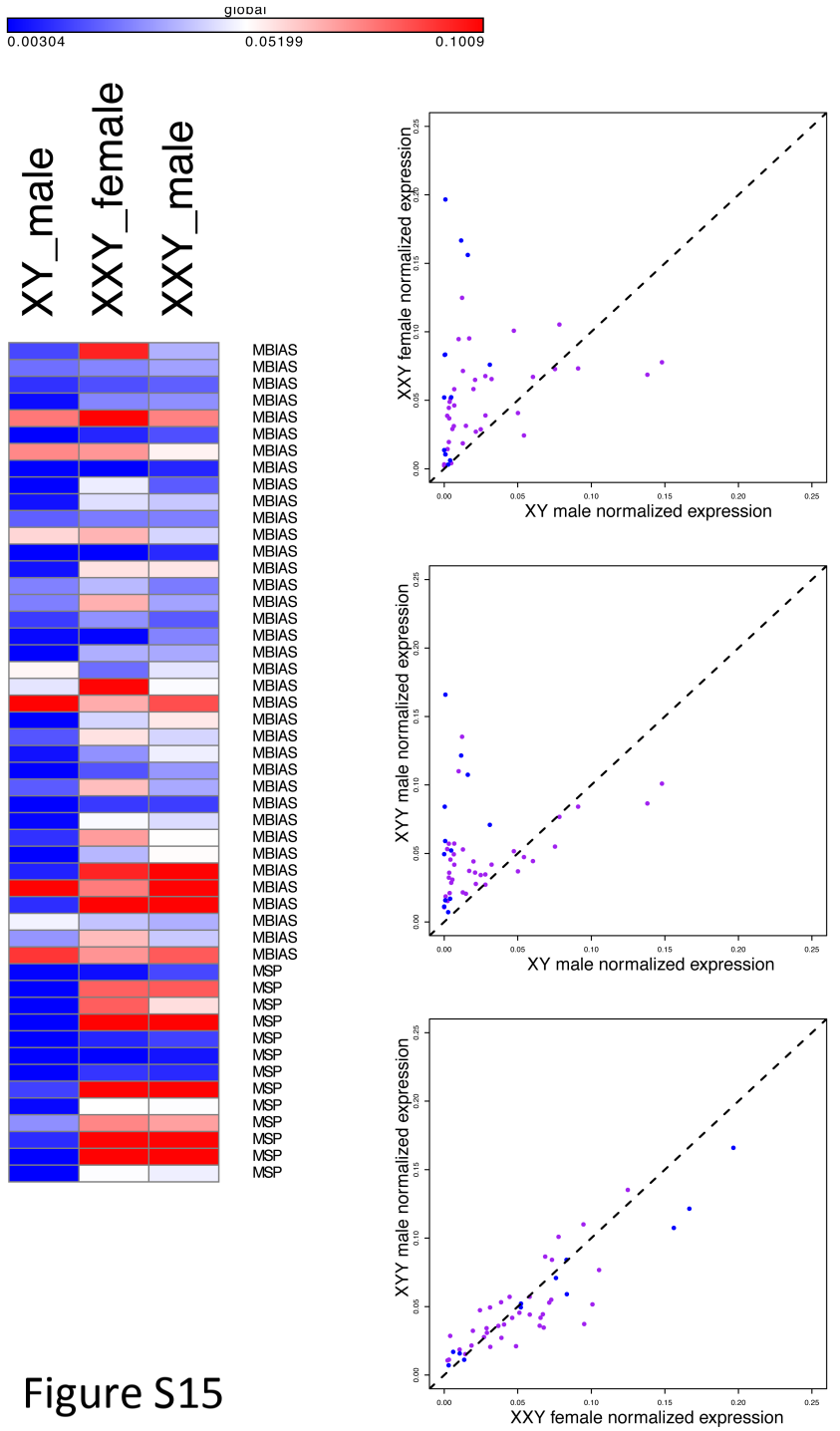

global
0.00304
0.05199
0.1009
XY_male
XXY_female
XXY_male
MBIAS
MSP
XXY female normalized expression
XY male normalized expression
XXY male normalized expression
XY male normalized expression
XXY male normalized expression
XXY female normalized expression

Figure S15

Table S1. Flow cytometry estimates of genome sizes for flies with different karyotypes

	Replicate 1	Replicate 2	SE
CantonS F	177.596	#VALUE!	N/A
CantonS M	179.869	#VALUE!	N/A
X^X	174.08	174.5395439	0.324946591
X^Y	173.766	174.3436929	0.408490546
XO	159.684	157.9259259	1.2431461
XXY	195.803	192.7714374	2.143638491
XYY	194.498	200.4503704	4.208961453

Table S2. Pearson correlation coefficients of signal of *D. miranda* spike

H3K9me2

	XY	XX	XO	XXY	XYY
XY		0.9865341	0.7172258	0.8047425	0.9471368
XX	0.9865341		0.7202507	0.8043829	0.9417158
XO	0.7172258	0.7202507		0.9539047	0.8605764
XXY	0.8047425	0.8043829	0.9539047		0.923939
XYY	0.9471368	0.9417158	0.8605764	0.923939	

H3K9me3

	XY	XX	XO	XXY	XYY
XY		0.9388051	0.793487	0.8037694	0.9419146
XX	0.9388051		0.8131611	0.7626068	0.9210255
XO	0.793487	0.8131611		0.950172	0.9049075
XXY	0.8037694	0.7626068	0.950172		0.9019085
XYY	0.9419146	0.9210255	0.9049075	0.9019085	

H3K4me3

	XY	XX	XO	XXY	XYY
XY		0.872109	0.8219456	0.8217578	0.9357865
XX	0.872109		0.9621275	0.7084921	0.8511443
XO	0.8219456	0.9621275		0.6723546	0.835176
XXY	0.8217578	0.7084921	0.6723546		0.7832339
XYY	0.9357865	0.8511443	0.835176	0.7832339	

Table S3

correlation of H3K4me3 signal across samples

	XY	XX	XO	XXY	XYY
XY		0.898	0.926	0.866	0.893
XX	0.898		0.963	0.816	0.879
XO	0.926	0.963		0.832	0.89
XXY	0.866	0.816	0.832		0.748
XYY	0.893	0.879	0.89	0.748	

H3K9me2 signal vs. H3K9me3 signal

Pearson correlation

XY	XX	XO	XXY	XYY
0.7356	0.9015	0.674	0.5335	0.7557

overlap of top 40% of 5kb windows

XY	XX	XO	XXY	XYY
0.6873	0.7166	0.5998	0.657	0.6866

Unspiked vs. spiked H3K9me3 replicates

Pearson correlation

XY	XO	XXY	XYY
0.7559	0.675	0.5665	0.8967

GO term	XX vs. XY	XX vs. XO	XX vs. XXY	XY vs. XO	XY vs. XYY
amide biosynthetic process	*	***	**	*	
amino sugar metabolic process		***	*	***	
aminoglycan metabolic process	*	***		***	
antibacterial humoral response			*		
axoneme assembly					*
behavioral response to starvation				*	
biosynthetic process			*		
body morphogenesis				*	
carbohydrate derivative metabolic process		*		*	
cellular amide metabolic process	*	***	**	*	
cellular biosynthetic process			*		
cellular macromolecule biosynthetic process	*	*	**	*	
cellular nitrogen compound biosynthetic process	*		*	*	
cellular protein metabolic process			*		
cellular response to heat	*	**	**	*	
cellular response to UV	*				*
centrosome duplication		**			
chitin metabolic process	*	***	*	***	*
chitin-based cuticle development		**	*	***	
coagulation		*			
cold acclimation		*	*	*	
cuticle chitin metabolic process				*	
cuticle development		**	*	***	
cytopasm organization	**	**			
defense response	*				
defense response to bacterium		*			
defense response to Gram-positive bacterium	**	*	*	*	
dicarboxylic acid catabolic process		*			
glucosamine-containing compound metabolic process		***	*	***	

***p<10-9
**p<10-6
*p<10-3

heat shock-mediated polytene chromosome puffing		*	*	*	
hemolymph coagulation		*			
hemostasis		*			
humoral immune response			*		
lipid catabolic process	*				
macromolecule biosynthetic process	*		*	*	
mating plug formation					*
melanin biosynthetic process		*			
microtubule bundle formation					*
microtubule organizing center organization		*			
mitotic spindle organization		*			
multi-organism process	*				*
multicellular organism reproduction	***			***	***
negative regulation of female receptivity					*
nitrogen compound metabolic process			*		
organic substance biosynthetic process			*		
organonitrogen compound biosynthetic process	*	**	**	*	
organonitrogen compound metabolic process	*	***	*		
oxidation-reduction process	*		*		
peptide biosynthetic process	**	***	**	*	
peptide metabolic process	*	***	**	*	
polytene chromosome puffing		*	*		
post-mating behavior					***
regulation of rhdopsin mediated signaling pathway					*
reproduction	***			**	***
reproductive behavior					*
response to bacterium	*	**	*	*	
response to biotic stimulus	*	*			
response to chemical			*		
response to cold		*		*	

response to external biotic stimulus	*	*			
response to heat		*	**	*	
response to organic substance			*		
response to other organism	*	*			
response to pheromone	*	*	*		*
response to temperature stimulus		*	**	*	
rRNA 2'-O-methylation		*		*	
rRNA metabolic process		*			
rRNA methylation				*	
rRNA processing		*			
sensory perception of chemical stimulus	*	*	**	**	
sex differentiation	**	**			
sperm axoneme assembly					*
sperm chromatin condensation					*
sperm competition					*
sperm displacement					*
spermatogenesis					*
translation	**	***	**	*	
vitellogenesis	***	***			

Chapter 3

The Y chromosome contributes to sex-specific aging in Drosophila

Emily J. Brown & Doris Bachtrog*
Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

Heterochromatin suppresses repetitive DNA, and a loss of heterochromatin has been observed in aged cells of several species, including humans and Drosophila. Males often contain substantially more heterochromatic DNA than females, due to the presence of a large, repeat-rich Y chromosome, and male flies generally have shorter average life spans than females. Here we show that repetitive DNA becomes de-repressed more rapidly in old male flies relative to females, and repeats on the Y chromosome are disproportionally mis-expressed during aging. This is associated with a loss of heterochromatin at repetitive elements during aging in male flies, and a general loss of repressive chromatin in aged males away from pericentromeric regions and the Y. By generating flies with different sex chromosome karyotypes (XXY females; X0 and XYY males), we show that repeat de-repression and average lifespan is directly correlated with the number of Y chromosomes. Thus, sex-specific chromatin differences contribute to sex-specific aging in flies.

The chronic deterioration of chromatin structure has been implicated as one of the molecular signatures of aging (Wood et al. 2010; O'Sullivan and Karlseder 2012), and an overall loss of heterochromatin and repressive histone marks is observed in many old animals (Haithcock et al. 2005; Larson et al. 2012; Zhang et al. 2015). Heterochromatin is enriched at repetitive DNA, and its loss can result in de-repression and mobilization of silenced transposable elements (TEs) (De Cecco et al. 2013a; De Cecco et al. 2013b; Li et al. 2013; Wood and Helfand 2013; Van Meter et al. 2014; Wood et al. 2016). The amount of repetitive DNA can differ substantially between sexes, due to the presence of a highly repetitive (and normally poorly assembled) Y or W chromosome in the heterogametic sex. In the fruit fly *Drosophila melanogaster*, males contain a ~40-Mb large completely repetitive Y chromosome, implying that almost 50% of the male genome is heterochromatic compared to only 33% of the female genome (Hoskins et al. 2002). Males have a shorter average lifespan in many taxa, including humans and most Drosophila species (Yoon et al. 1990). Indeed, the genetic sex determination system predicts adult sex ratios in tetrapods, with the heterogametic sex being less frequent (Pipoly et al. 2015). Lower survivorship of the sex with the repetitive Y or W chromosome may suggest a link between sex-specific mortality, chromatin and sex chromosomes.

To test for sex-specific heterochromatin loss and a de-repression of repetitive DNA during aging, we assayed chromatin and gene expression profiles in young and aged individuals of a standard lab strain of *D. melanogaster*. Lifespan assays confirm that Canton-S males live significantly shorter than females (Kaplan and Meier 1958) (**Figure 1A**), consistent with multiple studies on sex-specific lifespan in Drosophila (Yoon et al. 1990; Tower and Arbeitman 2009) (**Figure S1**). We gathered replicate stranded RNA-seq data and ChIP-seq data for a repressive histone modification typical of heterochromatin (H3K9me2) from

young 8-day and old 64-68-day *D. melanogaster* males and females, using a 'spike in' normalization method to compare the genomic distribution of chromatin marks across samples (Brown and Bachtrog 2016).

Figure 1B shows the genomic distribution of the repressive histone modification H3K9me2 for young and old male and female flies. As expected, heterochromatin is enriched at repetitive regions, including pericentromeres, the small dot chromosome and the repeat-rich Y. While the genomic distribution of H3K9me2 looks similar between young males and females (Brown and Bachtrog 2016), heterochromatin enrichment changes dramatically in old male but less so in old female flies. In particular, we see a general loss of heterochromatin at repetitive regions in aged males, (**Figure 1B-D; Figure S2**), and males show significantly more regions that lose H3K9me2 signal (1.5-fold or more) during aging compared to females (232 vs. 73, p<2.2e-16, Fisher's exact test; **Figure S3**). Almost all regions that lose heterochromatin are located within the pericentromere or the Y chromosome (**Figure S3,4**). On the other hand, fewer regions in males gain H3K9me2 signal (1.5-fold or more) during aging relative to females (6 vs. 120, p<2.2e-16, Fisher's exact test; **Figure S3**). Genomic regions that gain H3K9me2 signal are enriched on the X of females (p<2.2e-16, Fisher's exact test; **Figure S3,4**), and tend to be located close to the pericentromeric boundary (p=0.04 Fisher's exact test; **Figure 1D**, **Figure S3,4**). This suggests that heterochromatin / euchromatin boundaries are less efficiently maintained in old flies, resulting in spreading of heterchromatin from the repeat-rich pericentromere into neighboring regions.

Sex-specific chromatin changes during aging are associated with sex-specific expression changes. We find that genes located in pericentromeric regions change their expression more during aging compared with genes in chromosomal arms, in both sexes (**Figure S5**). While heterochromatin typically has a repressive effect on gene expression, genes located in normally heterochromatic regions (such as the pericentromere) are known to depend on this repressive chromatin environment for proper transcription (Lu et al. 2000). Indeed, the global loss of heterochromatin in pericentromeric regions is associated with reduced expression levels of pericentromeric genes in aged males and females (**Figure S5**). Genes that gain the H3K9me2 mark during aging tend to decrease in expression (**Figure S6,7**).

Overall, we find that gene expression during aging differs between males and females. Of the top 10% of genes that are differentially expressed during aging in males and females, 35.5% show expression changes in both sexes (**Figure S8**). Chromosomal location does not appear to be the main determinant influencing gene expression during aging; of the 464 genes that are most differentially expressed in both males and females, only 17 are located inside or within 1Mb of the pericentromere (we expect 22 by chance). The most differentially expressed genes affect similar functional categories in males and females (including GO categories "antibacterial humoral response", "macromolecule biosynthetic process", "reproduction", and "translation", **Figure S9**), but many GO terms are unique to one sex (13 highly significantly enriched (p<10-9) GO terms shared by both sexes; 13 male-specific and 1 female-specific GO term).

In addition, we find sex-specific differences in repeat reactivation during aging. We mapped our transcriptome data to the consensus repeat library of *D. melanogaster* and detect low levels of expression of repetitive elements in young male and female flies (**Figure 2A**). Aged females maintain efficient repression of TEs, while expression for the major classes of annotated TEs increases during aging for males (**Figure 2A,C**). De-repression of TEs is more pronounced in males both in terms of the number of individual elements that show a significant increase in expression during aging, as well as the fraction of the transcriptome that consists of repetitive transcripts across all repetitive elements. Overall, we find that in females, 6 repetitive elements show a significant increase in expression during aging and 14 a significant decrease (**Figure 2C**), but the total fraction of transcripts derived from repeats increases during aging (the fraction of repetitive reads in all RNA-seq reads is 2.0% at 8 days, vs. 4.6% at 68 days, **Table S1, Figure S10**). The increase in repeat expression is much more pronounced in males, with 32 repetitive elements showing a significant increase in expression during aging and 4 showing a significant decrease (**Figure 2C, Figure S10**), and the total fraction of repetitive reads increases from 1.6% to 5.8% (**Table S1**). The TE showing the highest level of de-repression in both sexes is *copia*, which is expressed 28-fold more in old versus young males, and expressed 15-fold more in old versus young females (**Figure 2C**). H3K9me2 profiles at TE families show that there is a general enrichment of this repressive mark in young male and female flies (**Figure 2B**). Consistent with genome-wide expression profiles showing efficient silencing of repeats in old females, there is no global loss of the repressive chromatin mark at repetitive elements in 68-day old females (in fact, there is a slight increase, **Figure 2B**). However, aged *D. melanogaster* males undergo a general loss of the H3K9me2 histone modification in repetitive elements (**Figure 2B**).

Thus, chromatin and gene expression profiles show that TEs lose their epigenetic silencing and become mis-expressed in old male flies. Males have approximately 20% more repetitive sequence than females, due to the repeat-rich Y chromosome. The sex-specific increase in repeat expression may be triggered by the presence of the heterochromatic Y chromosome in males, and the Y indeed shows the strongest loss of heterochromatin during aging (**Figure 1**). To see if Y-linked repeats are especially prone to mis-regulation during aging, we used *de novo* assembled male-specific and male-biased (putatively Y-linked) repetitive sequences (**Figure S11**) (Brown and Bachtrog 2016). Indeed, we find that in males, putatively Y-linked repeats are up-regulated more strongly during aging, relative to repeats present in both sexes (p=5.9e-12, Wilcoxon test, **Figure 2D,E**). Overall, we find that 42 Y-linked repeats show a significant increase in expression during aging (and only one a significant decrease; **Figure 2D,E, Figure S12**), and the total fraction of transcripts derived from Y-linked repeats increases more than 9-fold in old males (**Table S2**). Additionally, putatively Y-linked repeats disproportionately lose the repressive histone modification H3K9me2 during aging compared to other repeats (p=3.4e-11, Wilcoxon test, **Figure 2D**). Thus, male-biased and male-specific repeats, i.e. repeats that are located on the Y chromosome, are especially prone to de-repression during aging in males.

To directly test whether the Y chromosome contributes to sex-specific TE de-repression and aging, we generated *D. melanogaster* females containing a Y chromosome (XXY flies), and males with either zero or two Y chromosomes (X0 and XYY flies), by crossing Canton-S

flies to different strains with attached-X and attached-X-Y chromosomes (**Figure 3A**, see **Methods** for strain information). We compared sex-specific lifespans of wildtype Canton-S *D. melanogaster* flies, and XXY female and XO and XYY male (**Figure 3B, C**). Note that XO and XXY females of a given cross are isogenic on the autosomes, but differ in their genomic background from other crosses and from Canton-S, which can contribute to lifespan variation among strains (**Figure 3B,C**). Cumulative survival probabilities show that life span of females that contain a Y chromosome (XXY females) is reduced relative to wildtype females or males that lack a Y chromosome (X0 males) for all crosses assayed (**Figure 3B, C**). Indeed, X0 males show a dramatic increase in life span relative to wildtype males, and even outlive wildtype females (**Figure 3B, C**). X0 males are sterile and have the least amount of repetitive DNA of all karyotypes investigated (~10Mb less than Canton-S females and ~40Mb less than Canton-S males (Brown and Bachtrog 2016)); both of these factors may contribute to increased lifespan. Males with two Y chromosomes (XYY), in contrast, live the shortest (**Figure 3B, C**), and their lifespan is reduced considerably relative to wildtype males, despite both karyotypes being fertile. Thus, survivorship data suggest that the number of Y chromosomes influences organismal survival.

Gene expression changes during aging in the aberrant karyotypes show many of the same patterns as wildtype flies, with similar networks of GO terms enriched in both XO and XYY males, and XXY females, including "reproduction" or "sensory perception of chemical stimulus" (**Figure S9**). Overall, we find that 101 of the top 10% of genes mis-expressed during aging are shared among all 5 karyotypes ($p<<1e-5$, permutation test). These genes do not show any enrichment for a particular GO term, and only 6 of them are located inside or within 1Mb of the pericentromere (expect 4.8 genes).

Genomic location also influences gene expression changes during aging in flies with aberrant karyotypes. As in wildtype males and females, genes located in the pericentromere show a decrease in expression during aging in XXY females and X0 males (**Figure S13;** XYY flies show no significant expression change at pericentromeric genes). Y-linked genes in wild-type males are expressed almost exclusively in male reproductive tissues (Carvalho et al. 2009), and we do not detect any expression of Y-linked genes even in very old XY male heads (**Figure S13**). In contrast, we find that Y-linked genes are inefficiently silenced in heads of XXY females and XYY males and are becoming de-repressed even more during aging (6 Y genes are expressed in old XXY females, and 14 in old XYY males, **Figure S13**).

Thus, wildtype males maintain efficient silencing at their Y-linked genes during aging, despite global heterochromatin loss on the Y and mis-expression of Y-linked repeats. Silencing mechanisms on the Y chromosome of XXY and XYY flies, on the other hand, appear to be generally compromised, even in young individuals. Indeed, we previously showed that the Y chromosome affects global heterochromatin integrity (Brown and Bachtrog 2016). Young flies with additional Y chromosomes (XXY females or XYY males) show lower levels of H3K9me2 enrichment at their TEs, and a de-repression of Y-linked repeats relative to wildtype flies, while X0 flies showed increased levels of H3K9me2 at repeats (Brown and Bachtrog 2016). Expression profiles in XXY females and XYY/X0 males demonstrate that the absence or presence of the Y chromosome modulates expression of

TEs during aging (**Figure 4**). Expression profiles from aged flies with aberrant sex chromosome karyotypes confirm our expectation that X0 males show less de-repression of TEs during aging relative to wildtype males (**Figure 4A,B**). XXY females, on the other hand, show more mis-expression of repeats during aging compared to wildtype females. In XXY females, 7 elements show a significant increase in expression during aging and 3 elements show a significant decrease in expression (**Figure 4A,B**, compared to 6 /14 elements that increase/decrease expression in wild-type females). The fraction of repetitive transcripts increases more during aging for XXY females (3.1-fold increase in XXY females vs. 2.2-fold in wildtype females, **Table S1**), and a greater proportion of the transcriptome is comprised of repeats in old XXY females relative to wildtype (4.8% in XXY females vs. 4.6% in wild-type females). XYY males show the greatest number of repeats with significantly increased expression during aging (33 elements, **Figure 4A,B**), but not the highest fold change in total fraction of repetitive reads (**Table S1**), partly because young XYY males show the highest expression of repeats of any of the 5 karyotypes (**Table S1**), and partly because old XYY males are approximately 30 days younger than the other karyotypes (**Figure 3B**).

Mis-expression of repetitive elements in XXY females and XYY males is especially pronounced for repeats found on the Y chromosome. Y-linked repeats show reduced silencing even in young XXY females and XYY males relative to wildtype males (**Figure 4C,D**) (Brown and Bachtrog 2016), and become de-repressed even more during aging in XXY and XYY individuals (**Figure 4C,D**). Overall, wildtype males express 64 putatively Y-linked repeats during their lifespan, and XXY females express 86 and XYY males express 102. In XXY females, 13 repeats significantly increase in expression during aging (and 5 decrease), and 71 elements significantly increase in expression in XYY males (and 8 decrease; **Figure S12**). Indeed, even at just 37 days old, XYY males already show higher, presumably aberrant, expression of Y-linked repeats compared to 68-day-old wild-type males (**Figure 4C,D**).

To conclude, our data demonstrate that the repeat-rich Y chromosome decreases life span in Drosophila. Loss of heterochromatin in repetitive regions during aging is more pronounced in male flies, and is accompanied by a de-repression of TEs. Y-linked repeats disproportionally lose their repressive marks and become reactivated, and analysis of flies with aberrant sex chromosome configurations demonstrates that the Y has a direct influence on organismal survival. Age-related heterochromatin loss on the repetitive, sex-limited Y or W-chromosome and repeat re-activation can contribute to lower survivorship of the heterogametic sex across taxa (Pipoly et al. 2015), including humans. Y chromosomes of Drosophila and humans are known to harbor structural polymorphism in heterochromatic sequences and copy-number variation in repeats (Lyckegaard and Clark 1989; Repping et al. 2003), and polymorphism on the *D. melanogaster* Y has been shown to affect lifespan (Griffin et al. 2015), the formation of heterochromatin (Lemos et al. 2010), and the regulation of TEs and 100s of genes genome-wide (Lemos et al. 2008; Sackton et al. 2011). More generally, individual humans and flies show extensive variation in their repeat content (Bosco et al. 2007; Ewing and Kazazian 2010), and our results raise the question whether natural variation in repetitive sequences can contribute to genetic variation in longevity among individuals.

Materials & Methods

Drosophila strains. Fly strains were obtained from the Bloomington Stock Center and the Kyoto Stock Center. The following strains were used: Canton-S; Oregon-R; 2549 (C(1;Y),$y^1cv^1v^1B$/0 & C(1)RM,y^1v^1/0); 4248 (C(1)RM, y^1 pn^1 v^1/C(1;Y)1, y^1 B^1/0; $sv^{spa\text{-}pol}$) from the Bloomington Stock Center, and 100950 (0/ C(1)RM, y^1w^{str}/C(t;Y)1, y^1 y^+ ac^1 sc^1 w^1) from the Kyoto Stock Center. The crossing scheme used to obtain XO and XYY males and XXY females is depicted in **Fig. 3A**. For chromatin and gene expression analyses, flies were grown in incubators at 25°C, 60% relative humidity, and 12h light for the indicated number of days following eclosion, and were then flash-frozen in liquid nitrogen and stored at -80°C.

Lifespan assays. Lifespan data was collected for all karyotypes in the same rearing conditions as described above. The lifespan assays were conducted as described (Linford et al. 2013). Briefly, synchronized embryos were collected on agar plates, mobilized with a cotton swab, washed 3 times with PBS pH 7.4, and 10µl of embryos were pipetted to a fresh vial of standard fly medium. Adult flies were collected over 2 days, and were allowed to mate for 2 more days. Flies were then sexed, and 30 flies were counted into each vial. Vials were then flipped, without using CO_2, every 2-3 days, and fly deaths were recorded. Flies that were observed escaping the vial were censored. To collect samples for the RNA-seq and ChIP-seq experiment, we censored the entire lifespan experiment once it reached 50% survivorship and flash-froze the remaining flies in liquid nitrogen. In total, 8,829 flies in 297 vials were counted for the lifespan assays reported here.

Chromatin Immunoprecipitation and Sequencing. We performed ChIP-seq experiments using a standard protocol adapted from (Alekseyenko et al. 2006). Briefly, approximately 2ml of adult flash-frozen flies were dissected on dry ice, and heads and thoraces were used to fix and isolate chromatin. Following chromatin isolation, we spiked in 60µl of chromatin prepared from female *Drosophila miranda* larvae (approximately 1µg of chromatin). We then performed immunoprecipitation using 4µl of the H3K9me2 (Abcam ab1220) antibody.

After reversing the cross-links and isolating DNA, we constructed sequencing libraries using the BIOO NextFlex sequencing kit. Sequencing was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. We performed 50bp single-read sequencing for our input libraries, and 100bp paired-end sequencing for H3K9me2 libraries, due to their higher repeat content.

We collected replicate datasets for H3K9me2 enrichment in aged males and females to confirm differences seen between the sexes and between young and old samples (**Fig S2**). Replicate H3K9me2 data for young flies are from (Brown and Bachtrog 2016).

RNA extraction and RNA-seq. We collected replicate RNA samples for aged individuals of all five karyotypes of interest; replicate RNA data for young flies are from (Brown and Bachtrog 2016). Additionally, we collected 3 replicate samples for aged male Canton-S, aged XXY females, and aged XYY males, and 4 replicate samples for aged female Canton-S

and aged XO males. After flash-freezing in liquid nitrogen, we dissected and pooled 5 heads from each sample, extracted RNA, and prepared stranded total RNA-seq libraries using Illumina's TruSeq Stranded Total RNA Library Prep kit with Ribo-Zero ribosomal RNA reduction chemistry, which depletes the highly abundant ribosomal RNA transcripts (Illumina RS-122-2201). We performed single-read sequencing for all total RNA libraries at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

Mapping of sequencing reads and data normalization. For all *D. melanogaster* alignments, we used Release 6 of the genome assembly and annotation (Hoskins et al. 2015). For all ChIP-seq datasets, we used Bowtie2 (Langmead and Salzberg 2012) to map reads to the genome, using the parameters "-D 15 –R 2 –N 0 –L 22 –i S,1,0.50 --no-1-mm-upfront", which allowed us to reduce cross-mapping to the *D. miranda* genome to approximately 2.5% of 50bp reads, and 1% of 100bp-paired end reads. We also mapped all ChIP-seq datasets to the *D. miranda* genome assembly (Ellison and Bachtrog 2013) to calculate the proportion of each library that originated from the spiked-in *D. miranda* chromatin versus the *D. melanogaster* sample.

To calculate the ChIP signal we first calculated the coverage across 5kb windows for both the ChIP and the input, and then normalized by the total library size, including reads that map to both *D. melanogaster* and the *D. miranda* spike. We then calculated the ratio of ChIP coverage to input coverage for each of the 5kb windows, and normalized by the ratio of *D. melanogaster* reads to *D. miranda* reads in the ChIP library, and then by the ratio of *D. melanogaster* reads to *D. miranda* reads in the input, to account for differences in the ratio of sample to spike present before immunoprecipitation. We describe the validation of this normalization method in (Brown and Bachtrog 2016).

Gene expression analysis. For each replicate of RNA-seq data, we first mapped RNA-seq reads to the ribosomal DNA scaffold in the Release 6 version of the *D. melanogaster* genome, and removed all reads that mapped to this scaffold, as differences in rRNA transcript abundance are likely to be technical artifacts from the total RNA library preparation, which aims to remove the bulk of rRNA transcripts. We then mapped the remaining reads to the Release 6 version of the *D. melanogaster* genome using Tophat2 (Kim et al. 2013), using default parameters. We then used Cufflinks (Trapnell et al. 2012) and Cuffdiff (Trapnell et al. 2013) to merge replicates and calculate normalized FPKMs for all samples. GO analysis was performed using GOrilla, using ranked lists of differentially expressed genes (Eden et al. 2009).

Repeat libraries. We used two approaches to quantify expression of repeats. Our first approach was based on consensus sequences of known repetitive elements that were included in the Release 6 version of the *D. melanogaster* genome and are available on FlyBase. These included consensus sequences for 125 TEs and the 3 largest satellites (359, dodeca, and responder).

Our second approach aimed to specifically assess the repeat content of the Y chromosome. Since the Y chromosome is poorly assembled and repetitive elements on the Y are not annotated, we previously assembled repetitive elements *de novo* from male and female

genomic DNA reads using RepARK and identified 101 male-specific repeats comprising 13.7kb of sequence, based on male-specific coverage analysis (Koch et al. 2014; Brown and Bachtrog 2016).

To assess expression of repetitive elements, we mapped RNA-seq reads to each of the repeat libraries (consensus TEs and putatively Y-linked repetitive elements) using Bowtie2 and the parameters "-D 15 –R 2 –N 0 –L 22 –i S,1,0.50 --no-1-mm-upfront". We then calculated the mean coverage across each repetitive element using Bedtools, and normalized the coverage by the number of uniquely-mapping reads in the sequencing library. We made this calculation independently for each replicate for each time sample and karyotype, and then calculated both the average expression value as well as the standard deviation, to assess statistical significance and reproducibility.

To assess H3K9me2 signal in repetitive elements, we took a similar approach as we did for calculating ChIP enrichment profiles across the genome. First, we mapped both ChIP and input sequencing reads to each of the repeat libraries using Bowtie2 and the parameters "-D 15 –R 2 –N 0 –L 22 –i S,1,0.50 --no-1-mm-upfront". We then calculated the mean coverage across each repetitive element using Bedtools, and normalized the coverage by the total library size, including reads that mapped to both the *D. melanogaster* and *D. miranda* genomes. We then calculated the ratio of ChIP coverage to input coverage for each repetitive element, and then normalized by the ratio of *D. melanogaster* reads to *D. miranda* reads in the ChIP library, and then by the ratio of *D. melanogaster* reads to *D. miranda* reads in the input, as described above and in (Brown and Bachtrog 2016). This method accounts for differences in copy number of the repetitive elements by dividing the ChIP coverage by each repeat's coverage in the input.

Author contributions: DB and EB conceived the study and wrote the paper. EB collected and analyzed the data.

References:

Alekseyenko A, Larschan E, Lai W, Park P, Kuroda M. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**(7): 848 - 857.

Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* **177**(3): 1277-1290.

Brown EJ, Bachtrog D. 2016. The Drosophila Y chromosome affects heterochromatin integrity genome-wide submitted.

Carvalho AB, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: Drosophila tales. *Trends in Genetics* **25**(6): 270-277.

De Cecco M, Criscione SW, Peckham EJ, Hillenmeyer S, Hamm EA, Manivannan J, Peterson AL, Kreiling JA, Neretti N, Sedivy JM. 2013a. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of

transposable elements. *Aging Cell* **12**(2): 247-256.
De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA. 2013b. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5**(12): 867-883.
Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48.
Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**(6160): 846-850.
Ewing AD, Kazazian HH, Jr. 2010. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* **20**(9): 1262-1270.
Griffin RM, Le Gall D, Schielzeth H, Friberg U. 2015. Within-population Y-linked genetic variation for lifespan in Drosophila melanogaster. *J Evol Biol* **28**(11): 1940-1947.
Haithcock E, Dayani Y Fau - Neufeld E, Neufeld E Fau - Zahand AJ, Zahand Aj Fau - Feinstein N, Feinstein N Fau - Mattout A, Mattout A Fau - Gruenbaum Y, Gruenbaum Y Fau - Liu J, Liu J. 2005. Age-related changes of nuclear architecture in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **102**(46): 16690-16695.
Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R et al. 2015. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res* **25**(3): 445-458.
Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG et al. 2002. Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome biology* **3**(12): RESEARCH0085.
Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *J Amer Statist Assn* **53**(282): 457–481.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**(4): R36.
Koch P, Platzer M, Downie BR. 2014. RepARK--de novo creation of repeat libraries from whole-genome NGS reads. *Nucleic Acids Res* **42**(9): e80.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**(4): 357-359.
Larson K, Yan SJ, Tsurumi A, Liu J, Zhou J, Gaur K, Guo D, Eickbush TH, Li WX. 2012. Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet* **8**(1): e1002473.
Lemos B, Araripe LO, Hartl DL. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**(5859): 91-93.
Lemos B, Branco AT, Hartl DL. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **107**(36): 15826-15831.
Li W, Prazak L, Chatterjee N, Gruninger S, Krug L, Theodorou D, Dubnau J. 2013. Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat Neurosci* **16**(5): 529-531.
Linford NJ, Bilgir C, Ro J, Pletcher SD. 2013. Measurement of lifespan in Drosophila

melanogaster. *J Vis Exp*(71).
Lu BY, Emtage PC, Duyf BJ, Hilliker AJ, Eissenberg JC. 2000. Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. *Genetics* **155**(2): 699-708.
Lyckegaard EM, Clark AG. 1989. Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **86**(6): 1944-1948.
O'Sullivan RJ, Karlseder J. 2012. The great unravelling: chromatin as a modulator of the aging process. *Trends in biochemical sciences* **37**(11): 466-476.
Pipoly I, Bokony V, Kirkpatrick M, Donald PF, Szekely T, Liker A. 2015. The genetic sex-determination system predicts adult sex ratios in tetrapods. *Nature* **527**(7576): 91-94.
Repping S, Skaletsky H, Brown L, van Daalen S, Korver C, Pyntikova T, Kuroda-Kawaguchi T, de Vries J, Oates R, Silber S et al. 2003. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* **35**(3): 247 - 251.
Sackton TB, Montenegro H, Hartl DL, Lemos B. 2011. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **108**(41): 17046-17051.
Tower J, Arbeitman M. 2009. The genetics of gender and life span. (1475-4924 (Electronic)).
Trapnell C, Hendrickson D, Sauvageau M, Goff L, Rinn J, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**(1): 46-53.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **7**(3): 562-578.
Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V. 2014. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun* **5**: 5011.
Wood JG, Helfand SL. 2013. Chromatin structure and transposable elements in organismal aging. *Frontiers in genetics* **4**: 274.
Wood JG, Hillenmeyer S, Lawrence C, Chang C, Hosier S, Lightfoot W, Mukherjee E, Jiang N, Schorl C, Brodsky AS et al. 2010. Chromatin remodeling in the aging genome of Drosophila. *Aging Cell* **9**(6): 971-978.
Wood JG, Jones BC, Jiang N, Chang C, Hosier S, Wickremesinghe P, Garcia M, Hartnett DA, Burhenn L, Neretti N et al. 2016. Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. *Proc Natl Acad Sci U S A*: in press.
Yoon JS, Gagen KP, Zhu DL. 1990. Longevity of 68 species of Drosophila. . *Ohio J Sci* **90**: 16-32.
Zhang W, Li J, Suzuki K, Qu J, Wang P, Zhou J, Liu X, Ren R, Xu X, Ocampo A et al. 2015. Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* **348**(6239): 1160-1163.

Figure Legends

Figure 1. Aging and the sex-specific chromatin landscape in Drosophila. **A.** Kaplan-Meier survivorship curves (Kaplan and Meier 1958) for Canton-S males (blue) and females (red), with the shaded region indicating the upper and lower 95% confidence interval calculated from the Kaplan-Meier curves. Karyotypes of male and female *D. melanogaster* are shown, with heterochromatic regions indicated in blue, and euchromatic regions in gray. **B.** Genome-wide enrichment of H3K9me2 for young (8 days) and old (64 or 68 days) *D. melanogaster* males and females along the different chromosome arms. Enrichment in 5kb windows is shown in red lines (normalized ratio of ChIP to input, see Materials & Methods), and the enrichment in 20kb windows is shown in gray scale according to the scale in the upper left, with the darkest gray corresponding to the highest 5% of values across all windows from all samples, and the lightest gray corresponding to the lowest 10% of values across all windows from all samples. Subtraction plots show the absolute difference in signal of 50kb windows between young and aged flies along the chromosome arms, with each sample further smoothed by subtracting out the median autosomal euchromatin signal, with females in red and males in blue, and the pericentromeric region of each chromosome indicated by the red segment of the line beneath each chromosome. **C.** Box plot showing the smoothed ChIP signal for all 5kb windows in different chromosomal regions (* $p<0.05$, ** $p<1e-6$, *** $p<1e-12$, Wilcoxon test) for males (blue) and females (red), with pericentromere boundaries defined by the Release 6 version of the *D. melanogaster* genome. **D.** Enrichment of H3K9me2 (in 5kb windows) for 1Mb upstream and downstream of the euchromatin/ pericentromere boundary, indicated by the dotted red line, on the 5 major chromosome arms. Subtraction plots show higher H3K9me2 signal in young (blue) or old (red) flies.

Figure 2. Enrichment of H3K9me2 at different repeat families and their expression levels for young and old females and males. **A.** Expression of all repeats from FlyBase consensus library from Release 6 of the *D. melanogaster* genome in young (8day) and old (~68day) male and female Canton-S, averaged across replicates, with significance values calculated using the Wilcoxon test (* $p< 0.05$, ** $p<0.01$, *** $p<1e-5$). Heatmaps are visualized globally according to the scale, with dark red corresponding to the top 5% of all values across all samples and dark blue corresponding to the bottom 5% of all values across all samples. **B.** H3K9me2 enrichment in repeats from FlyBase consensus library in young (8day) and old (64 or 68day) male and female Canton-S, averaged across replicates, with significance values calculated using the Wilcoxon test (* $p<0.05$, ** $p<0.01$, *** $p<1e-5$). The heatmap is scaled in the same manner as in (A.). **C.** Expression for repeat families for old and young males and females, with lines indicating the standard deviation for each estimate of expression across replicates and colors indicating the class of repetitive element. **D.** Expression and H3K9me2 signal in putatively Y-linked repeats in young (8day) and old (64 or 68day) Canton-S males, with significance values calculated using the Wilcoxon test (* $p<0.01$, ** $p<1e-4$, *** $p<1e-10$). Heatmaps are scaled in the same manner as in (A). **E.** Expression of putatively Y-linked repeats for old and young Canton-S males, with lines indicating the standard deviation for each estimate of expression across replicates.

Figure 3. Survivorship of XXY females and X0 and XXY males. **A.** Crossing scheme used to generate flies with aberrant sex chromosomes, with Canton-S used as the wild-type males and females for all crosses, and various lines with C(1)RM and C(1;Y) indicated by the attached X/ XY karyotypes. **B.** Kaplan-Meier survivorship curves for flies with aberrant sex chromosome karyotype, generated with various C(1)RM and C(1;Y) lines as indicated at the top of each survivorship curve. Shaded areas indicate the upper and lower 95% confidence interval calculated from the Kaplan-Meier curves. **C.** Median lifespan for each of the different karyotypes measured, with error bars indicating the upper and lower 95% confidence intervals (estimated by the Kaplan-Meier curves). Significance is compared to the wild-type Canton-S of the same sex for each aberrant karyotype, and was calculated using the survdiff package in R (* $p<0.01$, ** $p<1e-6$, *** $p<1e-12$).

Figure 4. Expression of repetitive elements in XXY females and X0 and XYY males during aging. **A.** Expression of all repeats from FlyBase consensus library from the Release 6 of the *D. melanogaster* genome, averaged across replicates, with significance values calculated using a Wilcoxon test (* $p<0.1$, ** $p<0.05$, *** $p<0.01$). **B.** Expression of all repeats from the FlyBase consensus library in XO males, XXY females, and XYY males, with lines indicating the standard deviation of each expression value calculated from replicates, and color indicating the class of repetitive element. **C.** Expression for putatively Y-linked (male-specific) repeats in karyotypes with a Y chromosome, averaged across replicates, with significance calculated using the Wilcoxon test (* $p<0.05$, ** $p<0.001$, *** $p<1e-5$) **D.** Like (B.) for putatively Y-linked repeats.

Figure S1. Kaplan-Meier survivorship curves of line 2549 males and females ((C(1;Y),$y^1cv^1v^1B/0$ & C(1)RM,$y^1v^1/0$) and Oregon-R wild-type males and females.

Figure S2. Pearson correlation coefficients for replicate H3K9me2 datasets for old males and females, and boxplots of normalized enrichment values for the replicates. Genome-wide plots were generated using replicate data as in Figure 1B. and 1D.

Figure S3. Chromosomal locations of 50kb windows that gain (red) or lose (blue) at least 1.5-fold H3K9me2 signal during aging for males and females. Pericentromeric regions are indicated by the red portion of the line beneath each chromosome.

Figure S4. Chromosomal locations of the top 10% of 50kb windows that gain (red) or lose (blue) H3K9me2 enrichment during aging for males and females. Pericentromeric regions are indicated by the red portion of the line beneath each chromosome.

Figure S5. Expression values of all genes, normalized across replicates, of young and old males and females by chromosome location, as annotated in the Release 6 of the *D. melanogaster* genome. We consider expressed genes as those with FPKM>1, as determined by median intronic FPKM. Significance values are calculated using the Wilcoxon test.

Figure S6. Expression values of genes located in 50kb windows that show either a 1.5-fold loss or 1.5-fold gain of H3K9me2 during aging in males and females. Significance values are calculated using the Wilcoxon test.

Figure S7. Expression values of genes located in the top 10% of 50kb windows that either gain or lose H3K9me2 during aging in males and females. Significance values are calculated using the Wilcoxon test.

Figure S8. Overlap of the top 10% of differentially expressed genes during aging, normalized across replicates, for various combinations of the 5 sex chromosome karyotypes examined.

Figure S9. GO category enrichment of genes differentially expressed during aging in wild-type Canton-S males (**A**.) and Canton-S females (**B**.), and XO males (**C**.), XXY females (**D**.), and XYY males (**E**.). Genes were ranked by their fold change in expression, averaged across replicates, regardless of direction, and submitted to GOrilla (Eden et al. 2009) for GO category enrichment analysis.

Figure S10. Number of repeats that show a significant increase (red) or decrease (blue) in expression during aging as a fraction of all repeats from the FlyBase consensus repeat library, with significance estimated using standard errors from replicate datasets. Significance is calculated using Fisher's exact test, with red stars indicating significance for repeats that increase in expression, and blue stars indicating significance for repeats that decrease in expression during aging. We also show the estimates of the total fraction of RNA-seq reads that map to the FlyBase consensus repeat library, with error bars calculated from replicate datasets, for young and old samples from each of the 5 karyotypes.

Figure S11. Male vs. female genomic coverage of *de novo* assembled repeats, with putatively Y-linked repeats indicated in blue and purple as those with male-specific or highly male-biased genomic coverage patterns (Brown and Bachtrog 2016).

Figure S12. Number of putative Y-linked repeats that show a significant increase (red) or decrease (blue) in expression during aging as a fraction of all repeats from a male-specific repeat library (see **Figure S11**), with significance estimated using standard errors from replicate datasets. Significance is calculated using Fisher's exact test, with red stars indicating significance for repeats that increase in expression, and blue stars indicating significance for repeats that decrease in expression during aging. We also show the estimates of the total fraction of RNA-seq reads that map to the Y-specific consensus repeat library, with error bars calculated from replicate datasets, for young and old samples from each of the 5 karyotypes.

Figure S13. Expression values of all genes, normalized across replicates, of young and old XO males, XXY females, and XYY males by chromosome location, as annotated in the Release 6 of the *D. melanogaster* genome. Significance values are calculated using the Wilcoxon test.

Figure S14. Estimates of expression values for the FlyBase consensus library for all replicates for all karyotypes.

Figure S15. Estimates of H3K9me2 signal for the FlyBase consensus library for all replicates for all karyotypes.

Table S1. Average estimation, across all replicates, of the fraction of all RNA-seq reads that are derived from the FlyBase consensus repeat library, as well as the fold change in repetitive content during aging.

Table S2. Average estimation, across all replicates, of the fraction of all RNA-seq reads that are derived from the putative Y-linked consensus repeat library, as well as the fold change in repetitive content during aging.

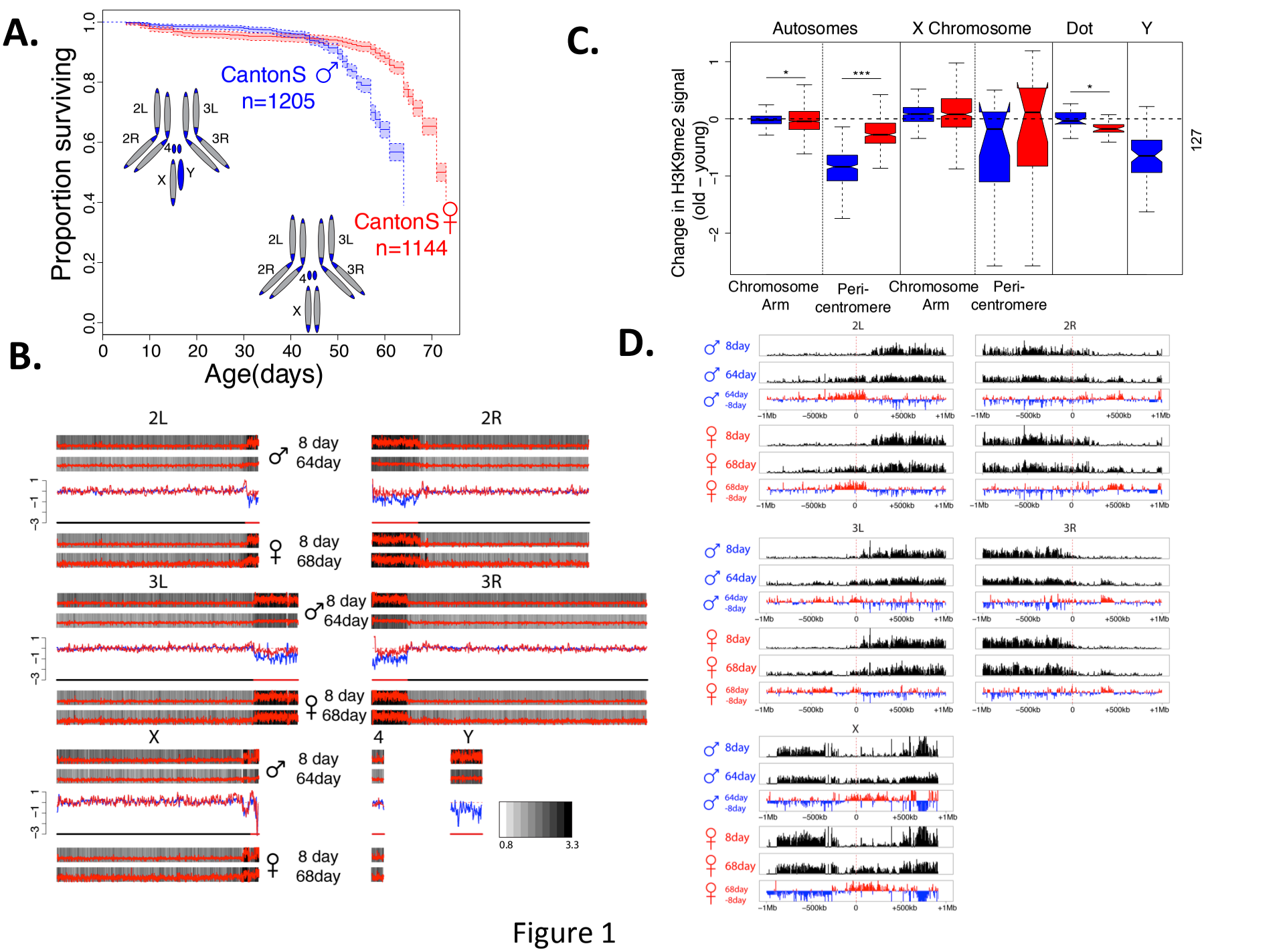

A.
Proportion surviving
Age(days)
CantonS ♂
n=1205
CantonS ♀
n=1144
2L
2R
3L
3R
4
X
Y
B.
2L
2R
3L
3R
X
4
Y
8 day
64day
8 day
68day
0.8
3.3
C.
Autosomes
X Chromosome
Dot
Y
Change in H3K9me2 signal
(old – young)
Chromosome Arm
Peri-centromere
Chromosome Arm
Peri-centromere
127
D.
8day
64day
64day -8day
68day
68day -8day
−1Mb
−500kb
0
+500kb
+1Mb

Figure 1

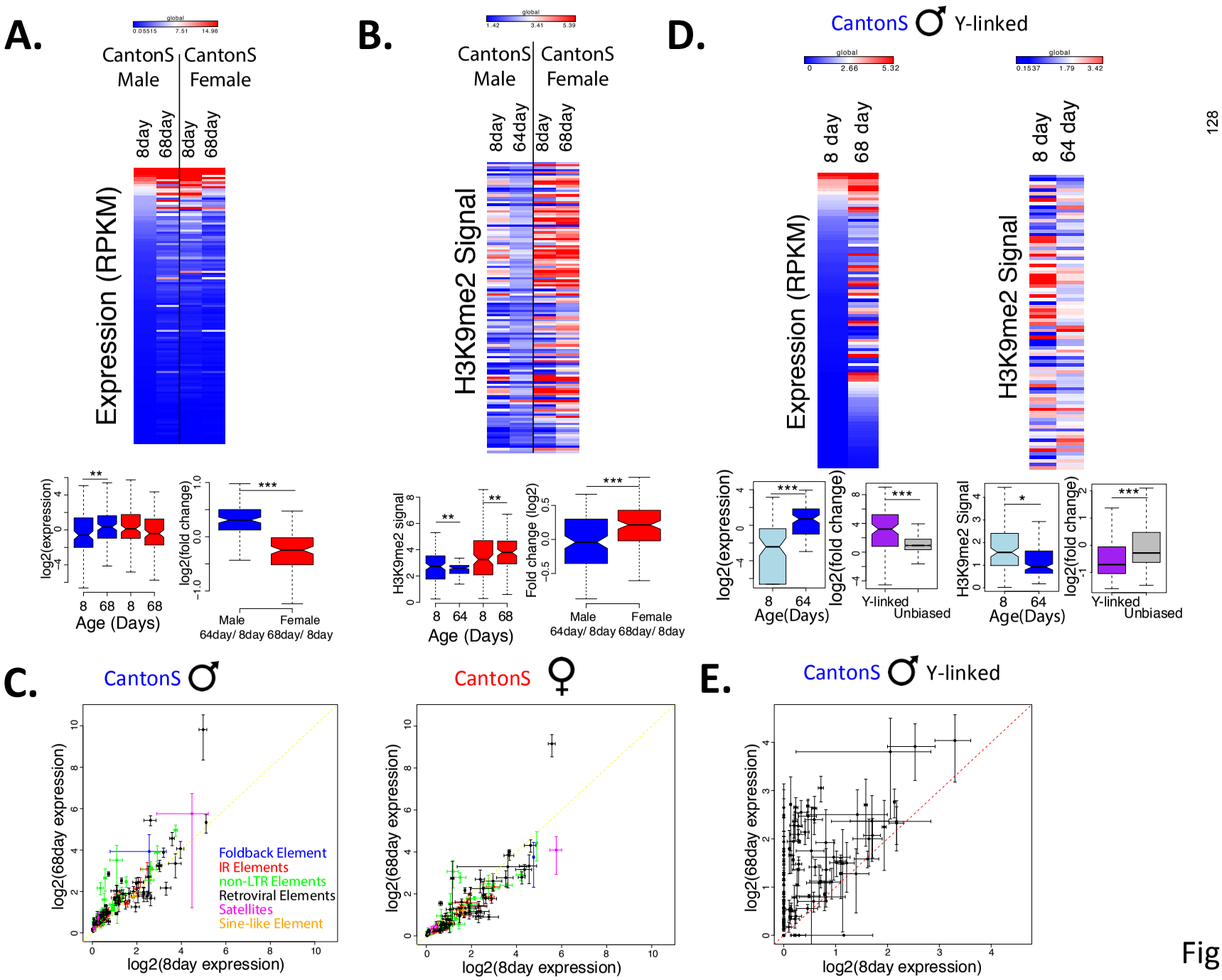

A.
CantonS Male
CantonS Female
8day
68day
8day
68day
Expression (RPKM)
B.
CantonS Male
CantonS Female
8day
64day
8day
68day
H3K9me2 Signal
D.
CantonS ♂ Y-linked
8 day
68 day
Expression (RPKM)
8 day
64 day
H3K9me2 Signal
Age (Days)
Male
Female
64day/ 8day 68day/ 8day
Age(Days)
Y-linked
Unbiased
C.
CantonS ♂
CantonS ♀
log2(68day expression)
log2(8day expression)
Foldback Element
IR Elements
non-LTR Elements
Retroviral Elements
Satellites
Sine-like Element
E.
CantonS ♂ Y-linked

Figure 2

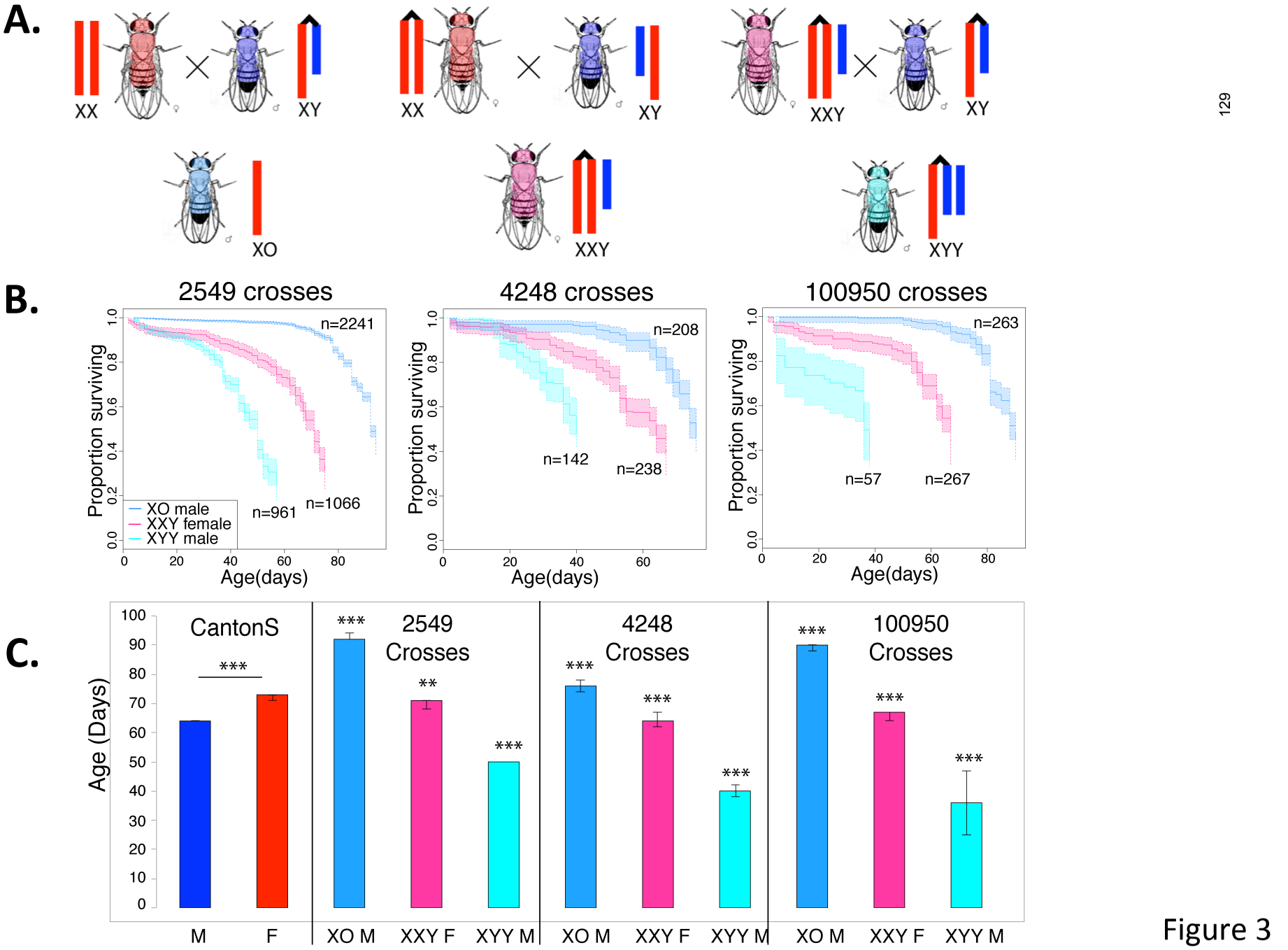

A.
XX
XY
XX
XY
XXY
XY
XO
XXY
XYY
B.
2549 crosses
4248 crosses
100950 crosses
Proportion surviving
Age(days)
n=2241
n=961
n=1066
n=208
n=142
n=238
n=263
n=57
n=267
XO male
XXY female
XYY male
C.
Age (Days)
CantonS
2549 Crosses
4248 Crosses
100950 Crosses
M
F
XO M
XXY F
XYY M

Figure 3

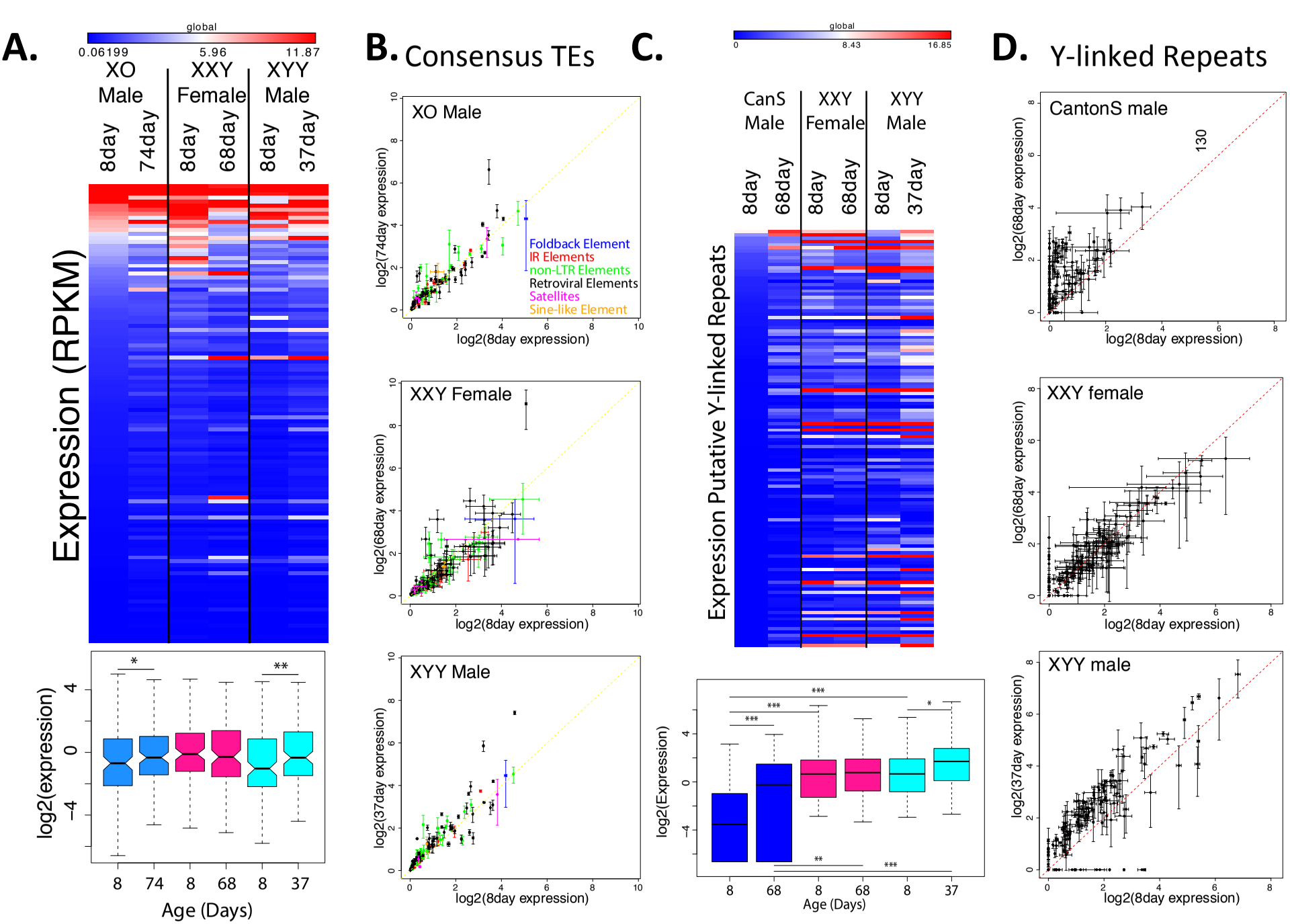
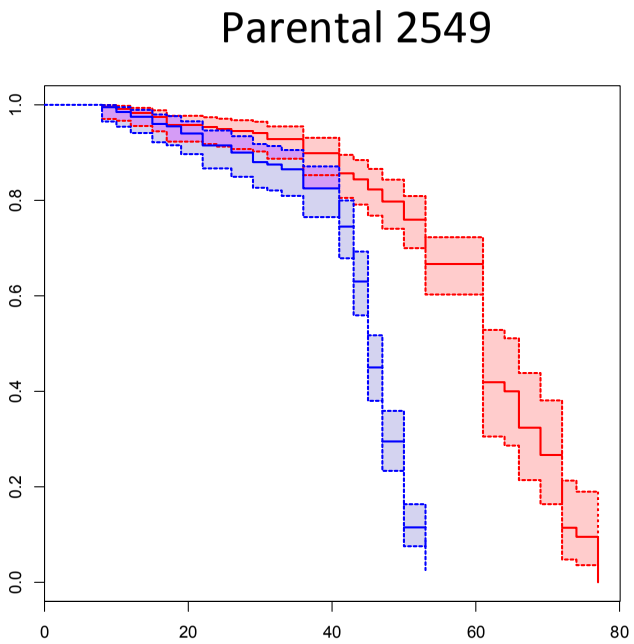

A.
global
0.06199
5.96
11.87
XO Male
XXY Female
XYY Male
8day
74day
8day
68day
8day
37day
Expression (RPKM)
log2(expression)
Age (Days)
B. Consensus TEs
XO Male
XXY Female
XYY Male
log2(74day expression)
log2(68day expression)
log2(37day expression)
log2(8day expression)
Foldback Element
IR Elements
non-LTR Elements
Retroviral Elements
Satellites
Sine-like Element
C.
global
0
8.43
16.85
CanS Male
XXY Female
XYY Male
Expression Putative Y-linked Repeats
log2(Expression)
Age (Days)
D. Y-linked Repeats
CantonS male
XXY female
XYY male
130
log2(68day expression)
log2(37day expression)
log2(8day expression)

Figure 4


Parental 2549
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80

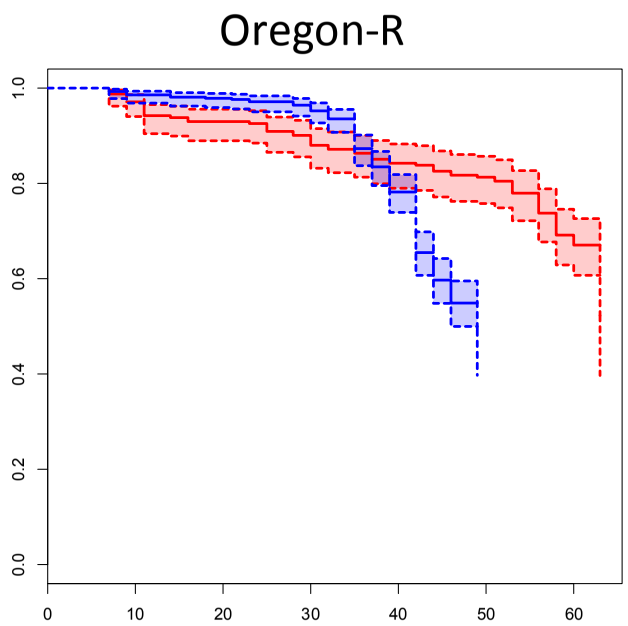
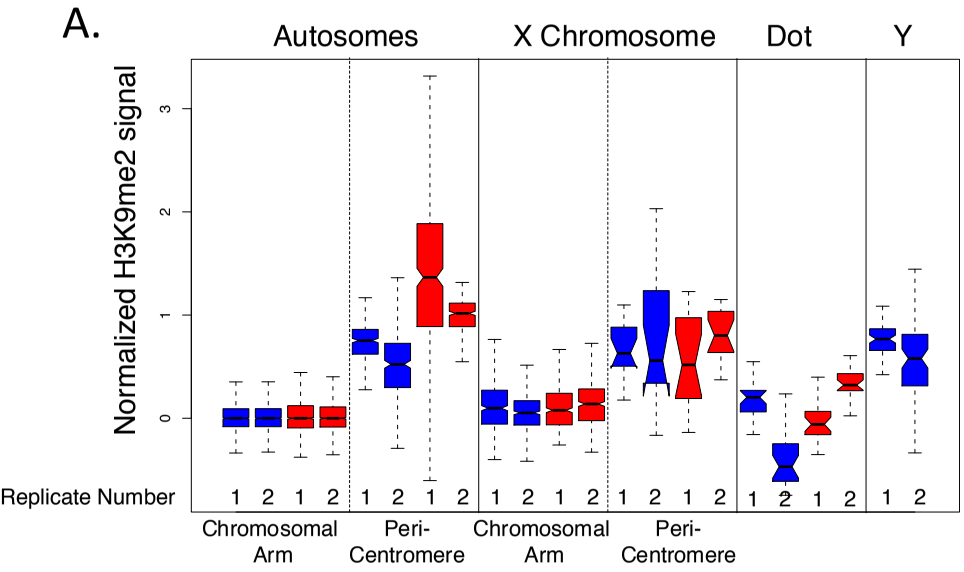
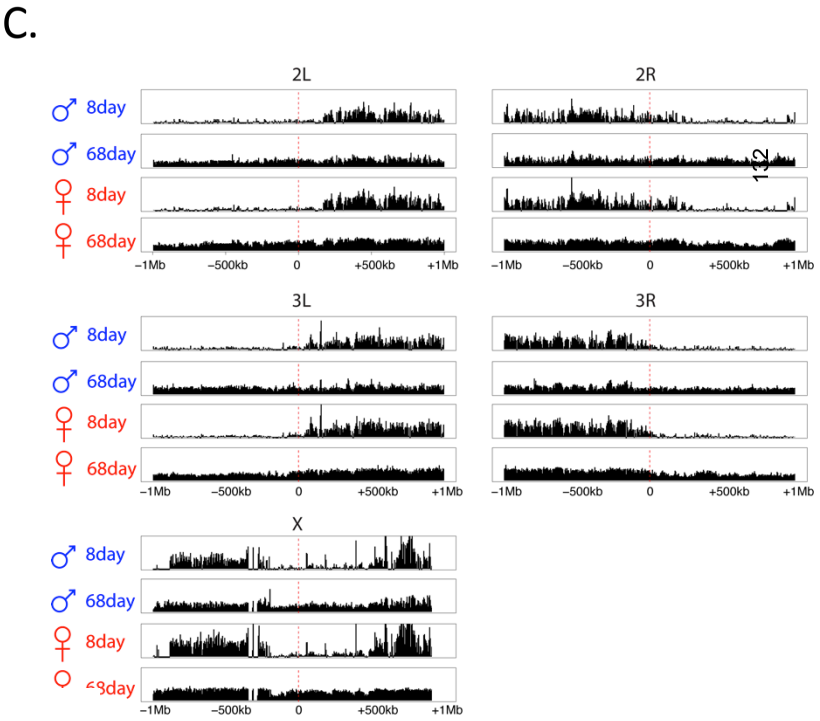

Oregon-R
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60

Figure S1


A.
Autosomes
X Chromosome
Dot
Y
Normalized H3K9me2 signal
0
1
2
3
Replicate Number
1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
Chromosomal Arm
Peri-Centromere
Chromosomal Arm
Peri-Centromere


C.
2L
2R
3L
3R
X
8day
68day
8day
68day
−1Mb
−500kb
0
+500kb
+1Mb

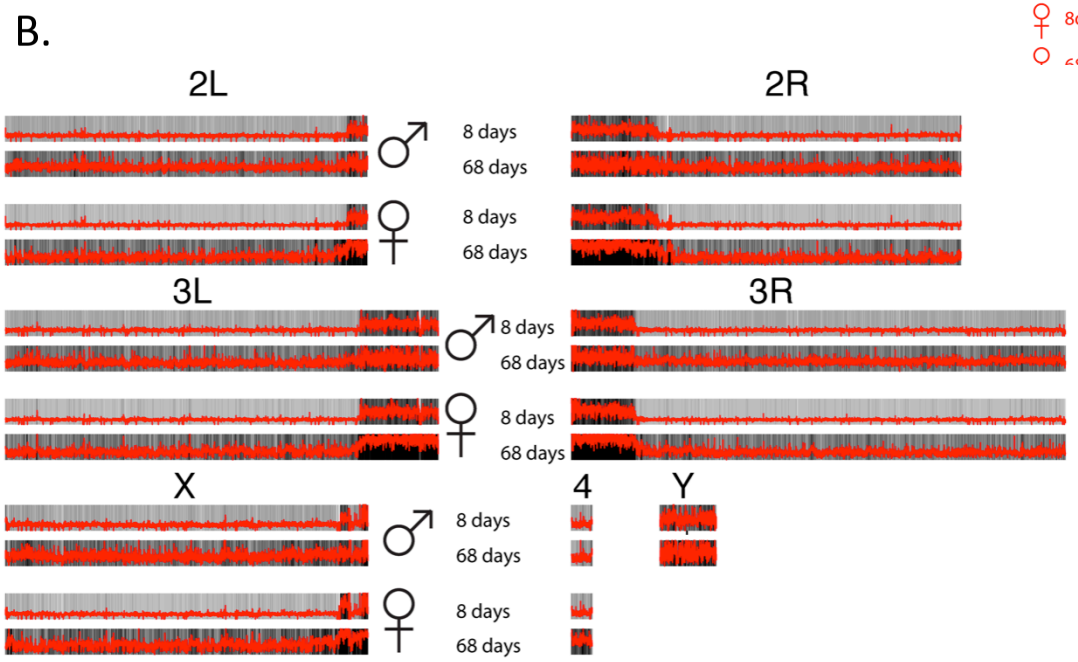

B.
2L
2R
3L
3R
X
4
Y
8 days
68 days

Figure S2

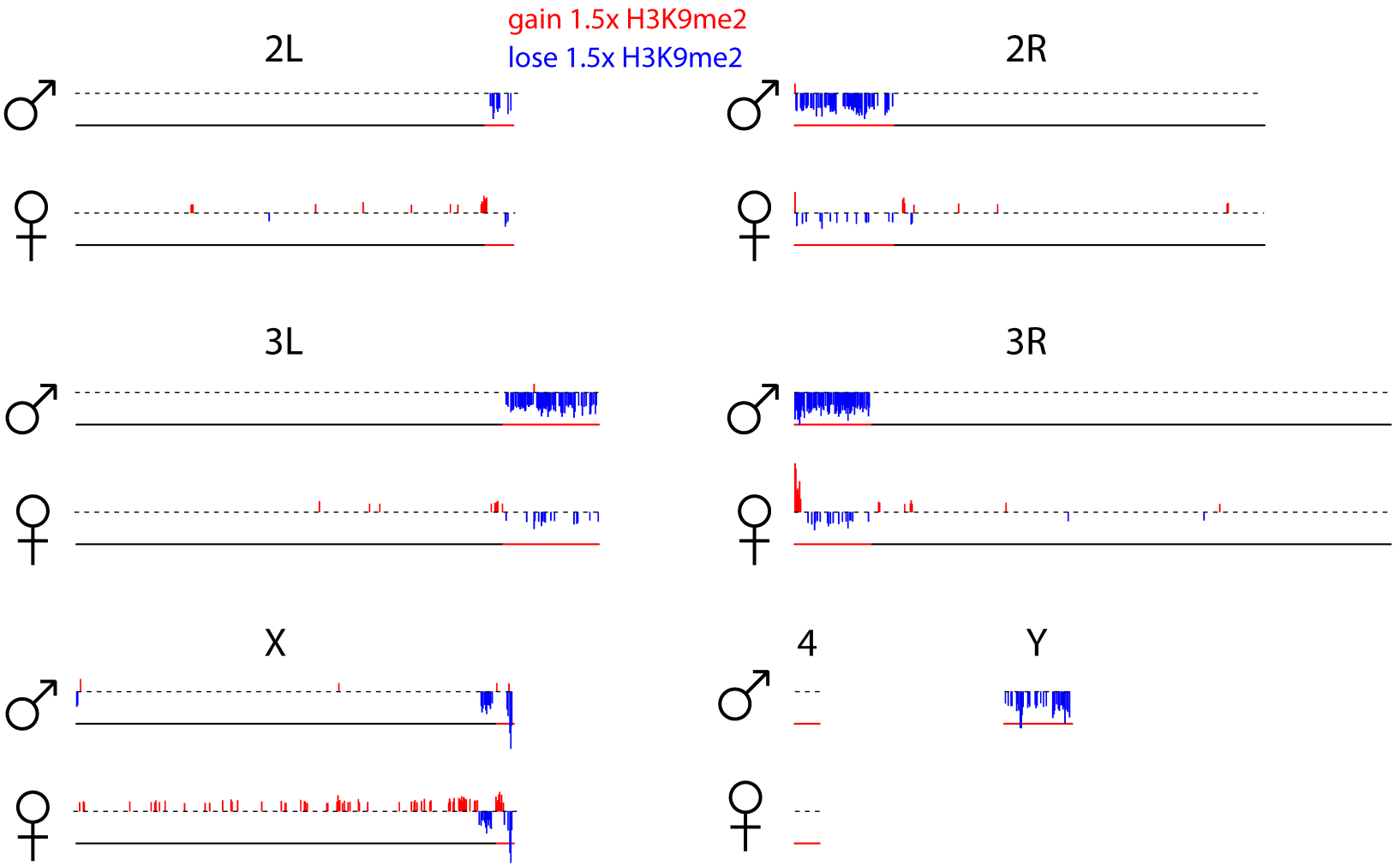


Figure S3

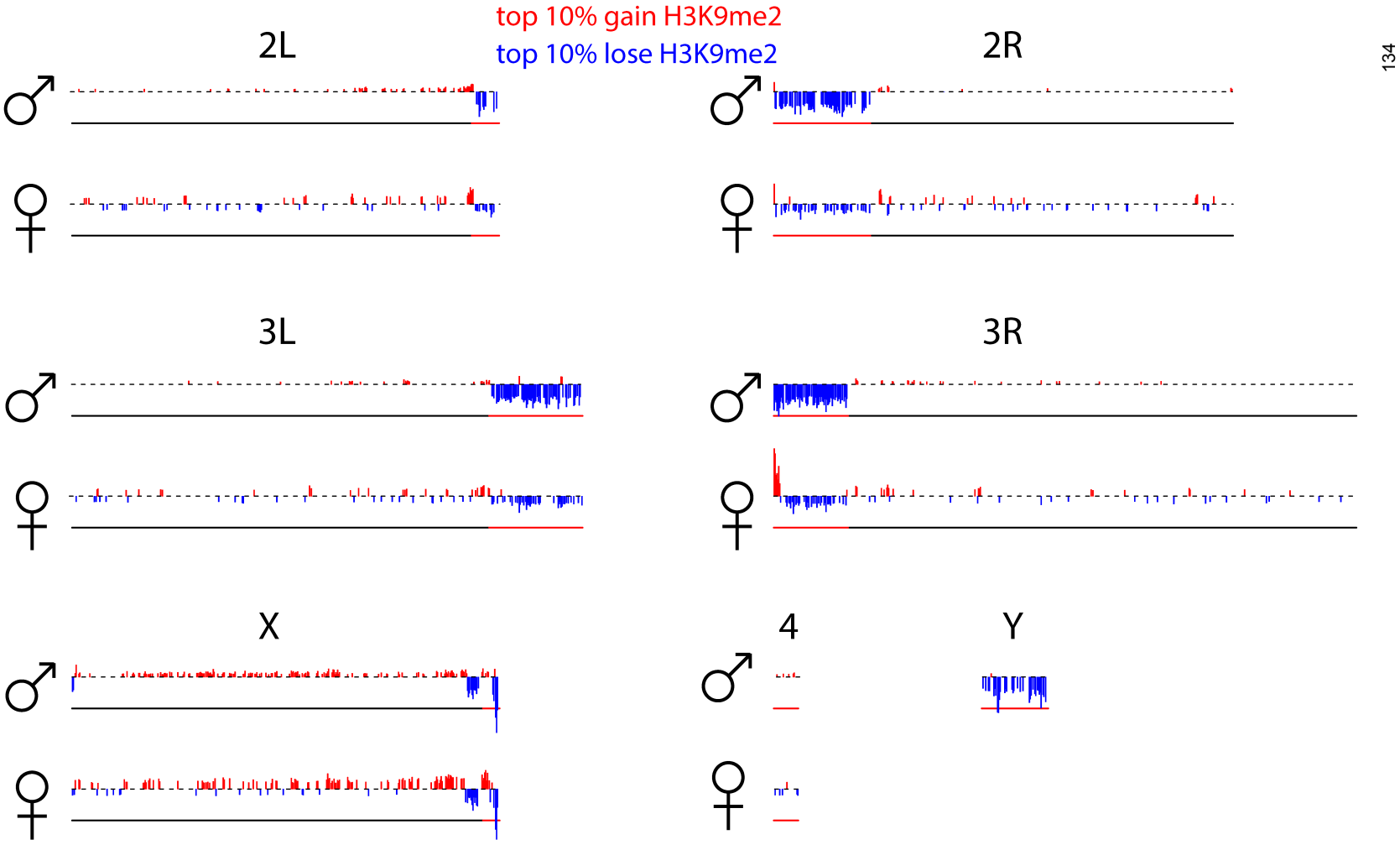
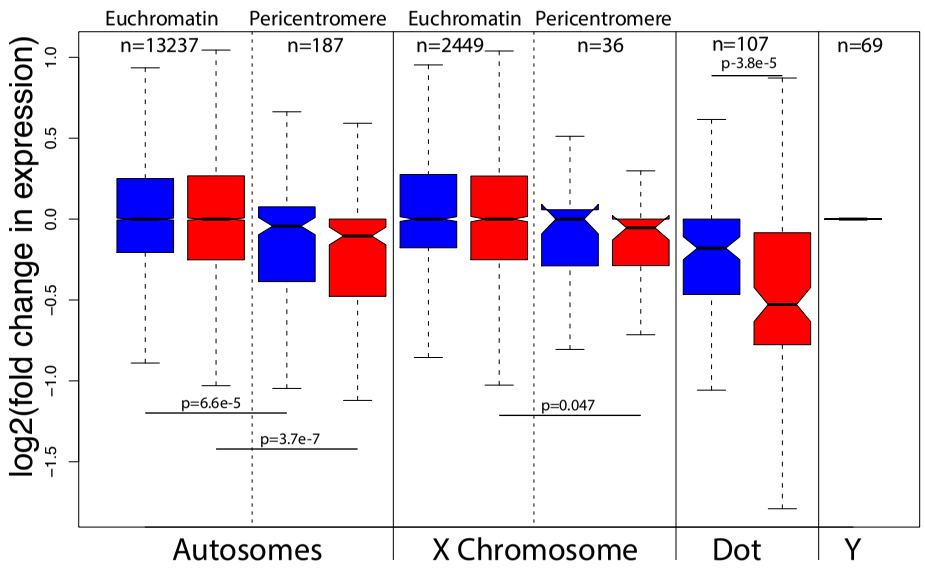
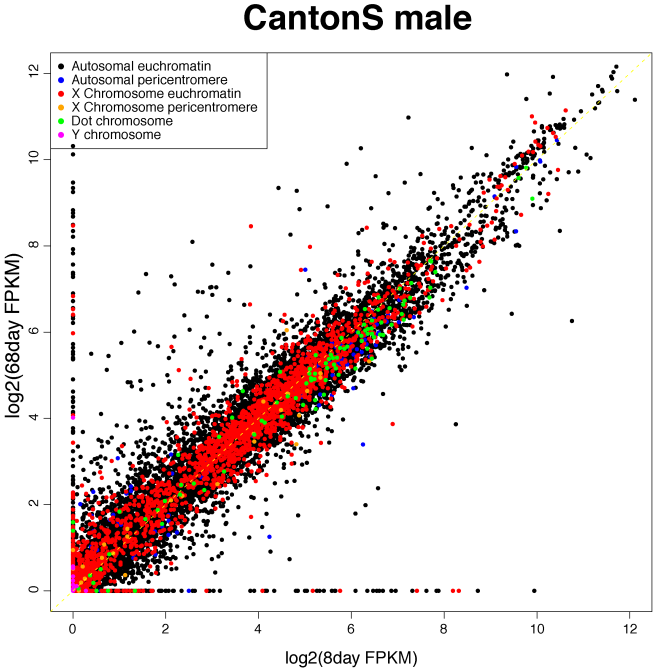
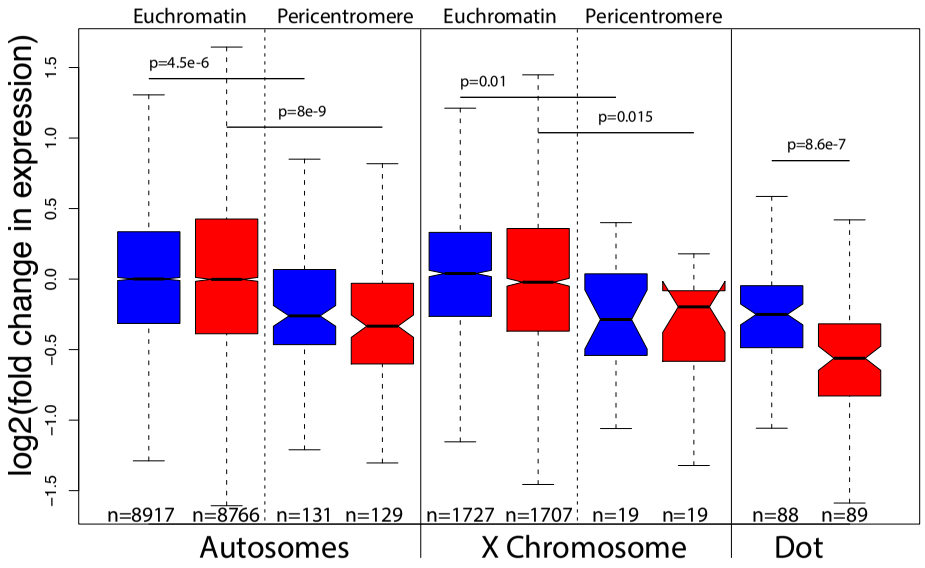


Figure S4





Expressed genes only



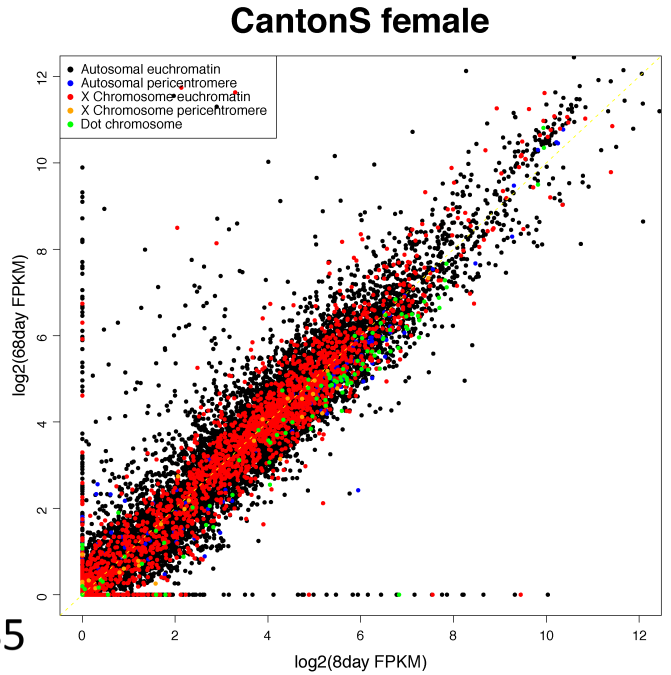


Figure S5

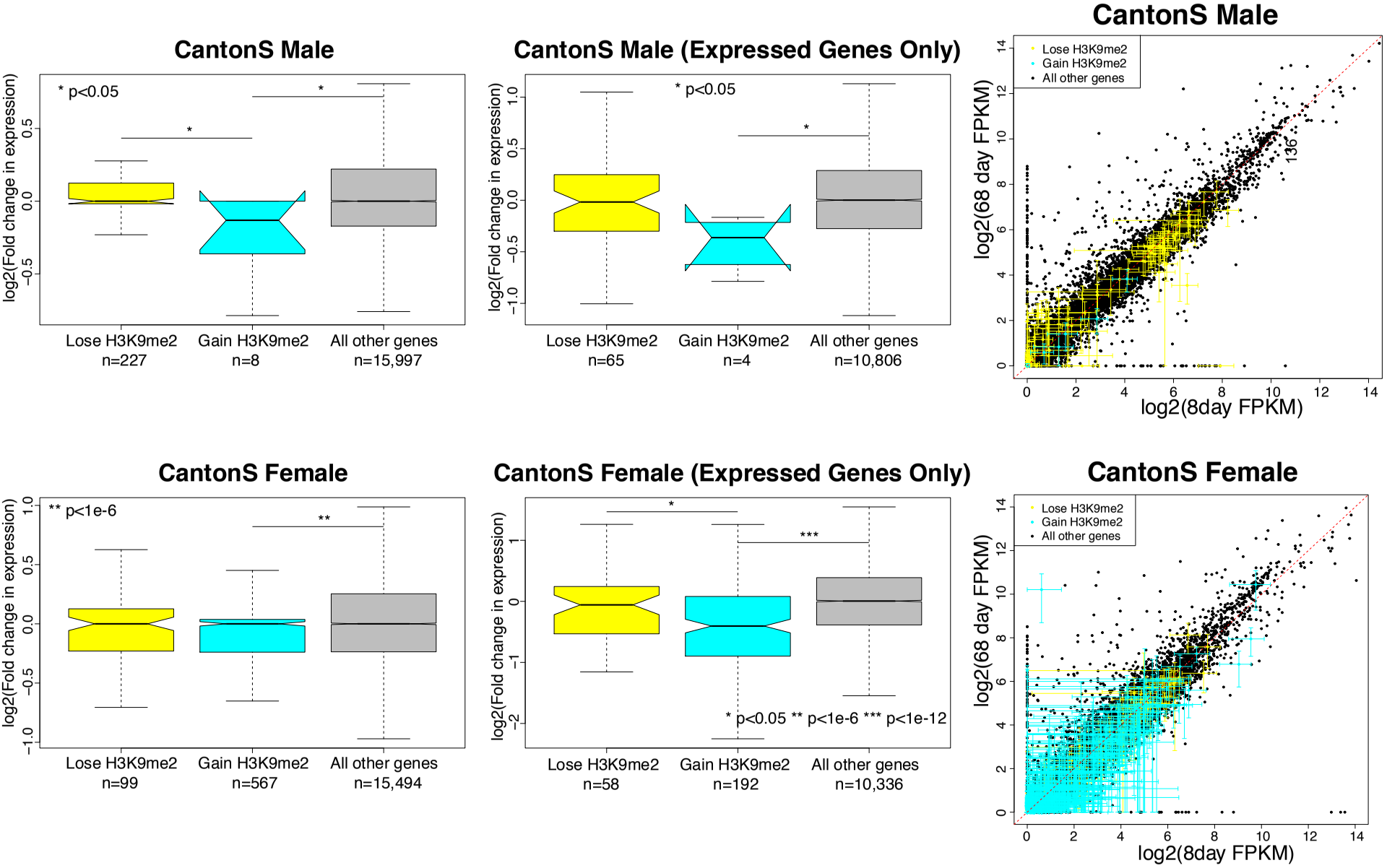
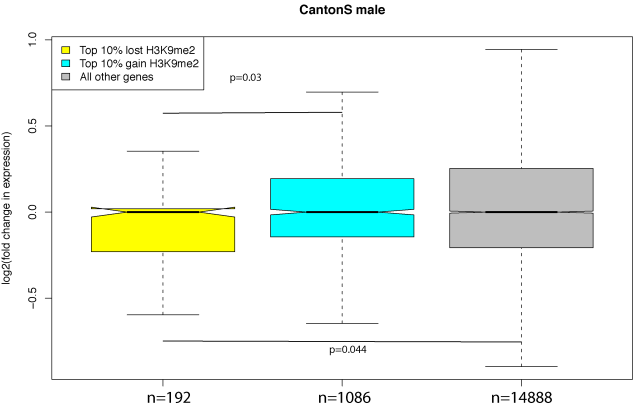
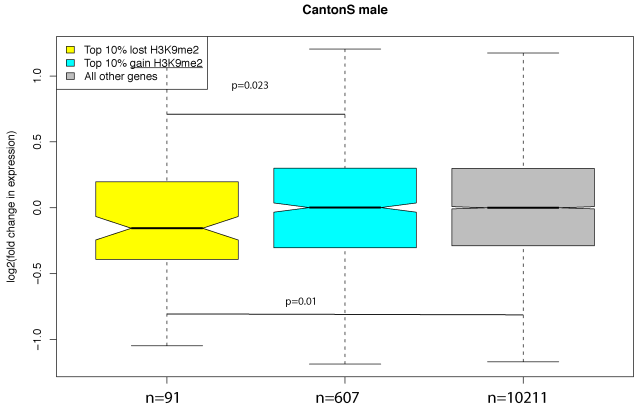
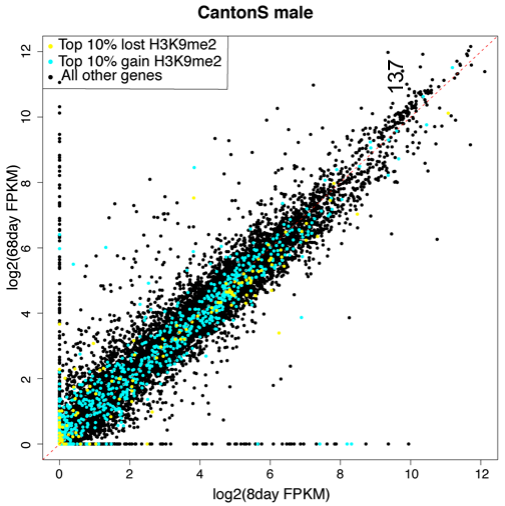
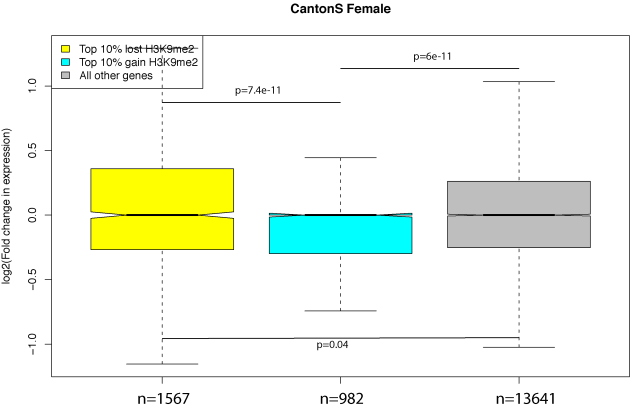

CantonS Male
* p<0.05
*
*
log2(Fold change in expression)
Lose H3K9me2
n=227
Gain H3K9me2
n=8
All other genes
n=15,997
CantonS Male (Expressed Genes Only)
* p<0.05
*
log2(Fold change in expression)
Lose H3K9me2
n=65
Gain H3K9me2
n=4
All other genes
n=10,806
CantonS Male
Lose H3K9me2
Gain H3K9me2
All other genes
136
log2(68 day FPKM)
log2(8day FPKM)
CantonS Female
** p<1e-6
**
log2(Fold change in expression)
Lose H3K9me2
n=99
Gain H3K9me2
n=567
All other genes
n=15,494
CantonS Female (Expressed Genes Only)
*
***
* p<0.05 ** p<1e-6 *** p<1e-12
log2(Fold change in expression)
Lose H3K9me2
n=58
Gain H3K9me2
n=192
All other genes
n=10,336
CantonS Female
Lose H3K9me2
Gain H3K9me2
All other genes
log2(68 day FPKM)
log2(8day FPKM)

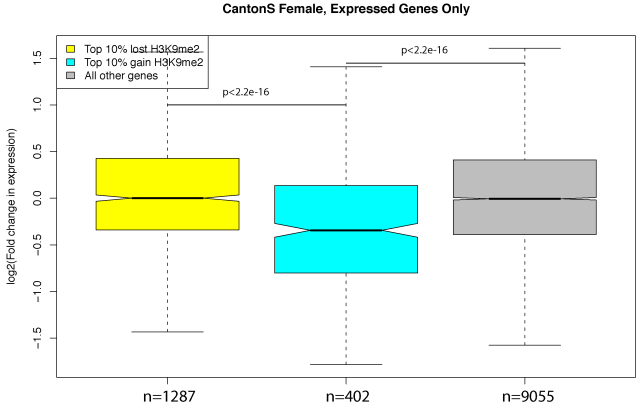
Figure S6


CantonS male
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
p=0.03
p=0.044
log2(fold change in expression)
n=192
n=1086
n=14888


CantonS male
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
p=0.023
p=0.01
log2(fold change in expression)
n=91
n=607
n=10211


CantonS male
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
137
log2(66day FPKM)
log2(8day FPKM)


CantonS Female
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
p=7.4e-11
p=6e-11
p=0.04
log2(Fold change in expression)
n=1567
n=982
n=13641


CantonS Female, Expressed Genes Only
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
p<2.2e-16
p<2.2e-16
log2(fold change in expression)
n=1287
n=402
n=9055

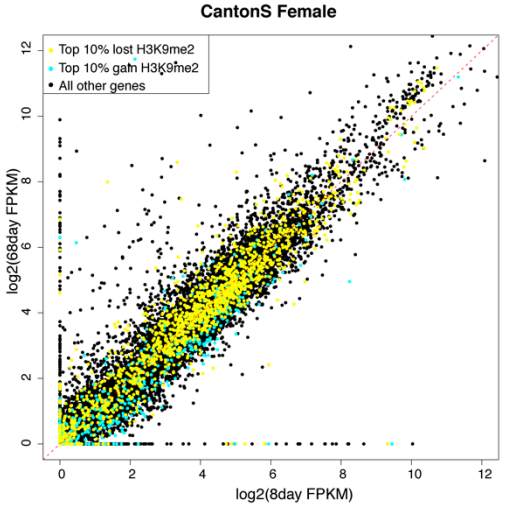
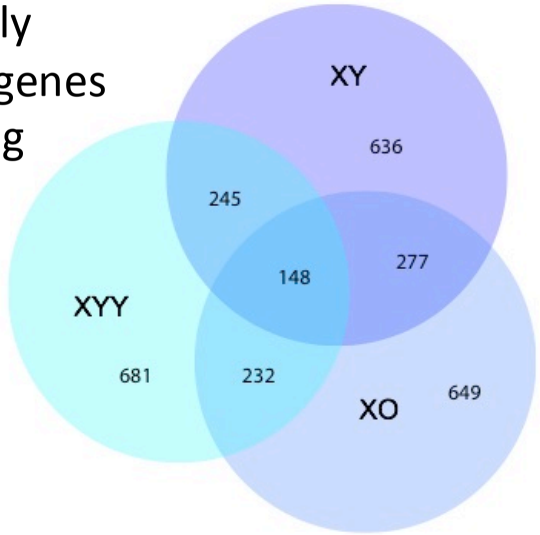

CantonS Female
Top 10% lost H3K9me2
Top 10% gain H3K9me2
All other genes
log2(66day FPKM)
log2(8day FPKM)

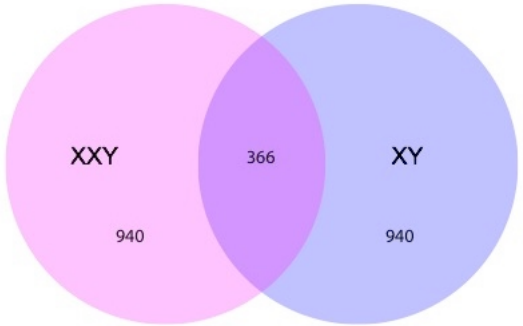
Figure S7

overlap top 10% differentially expressed genes during aging

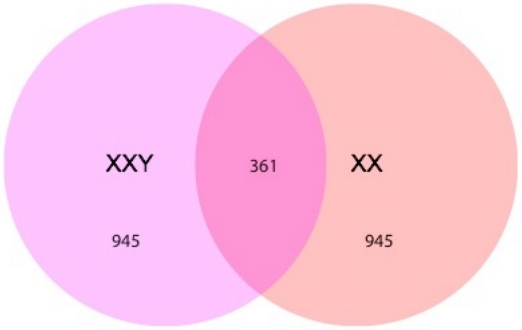
males



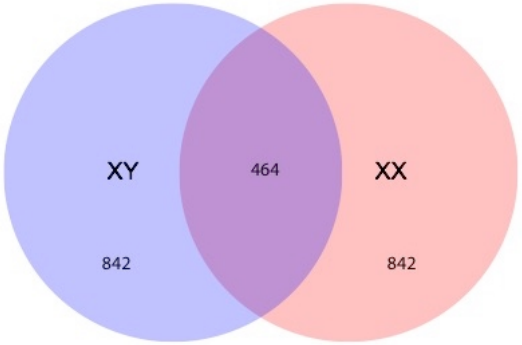
1 Y



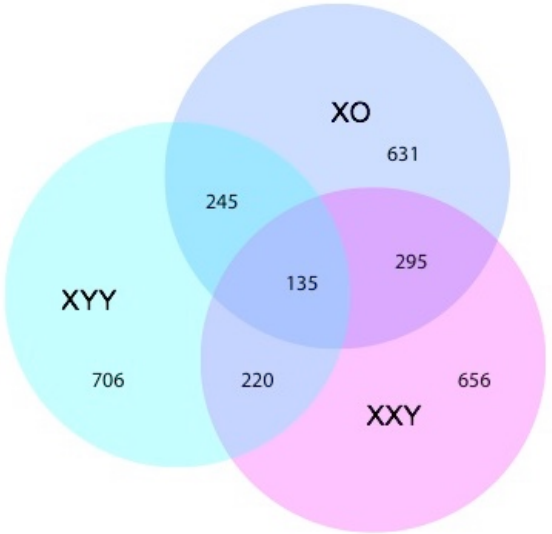
no Y



females



CantonS



2549 crosses

Figure S8

A.

CanS Male GO enrichment during aging

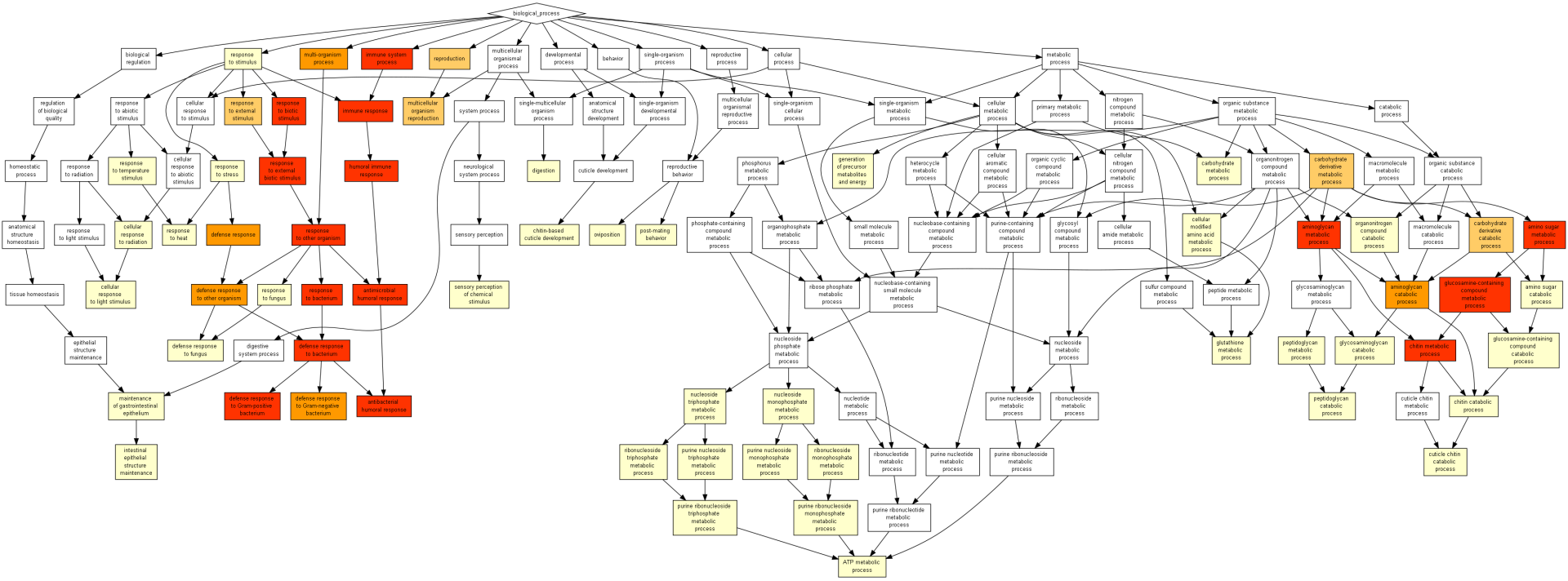


Figure S9

B.

CanS Female GO enrichment during aging

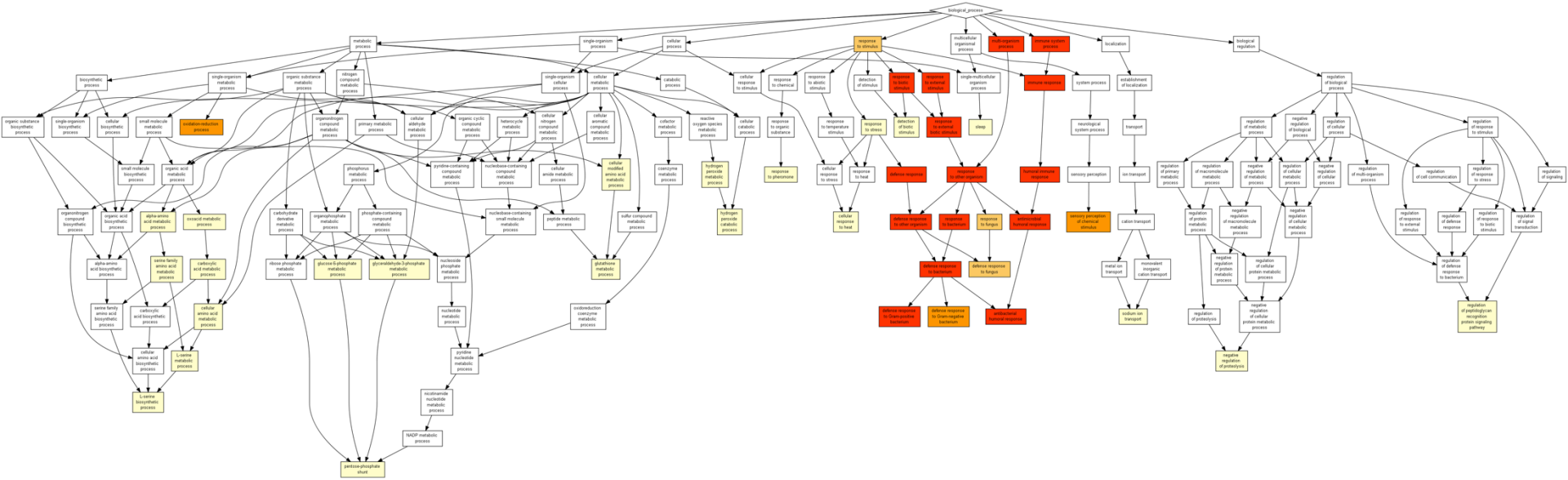


Figure S9

C. XO male GO enrichment during aging

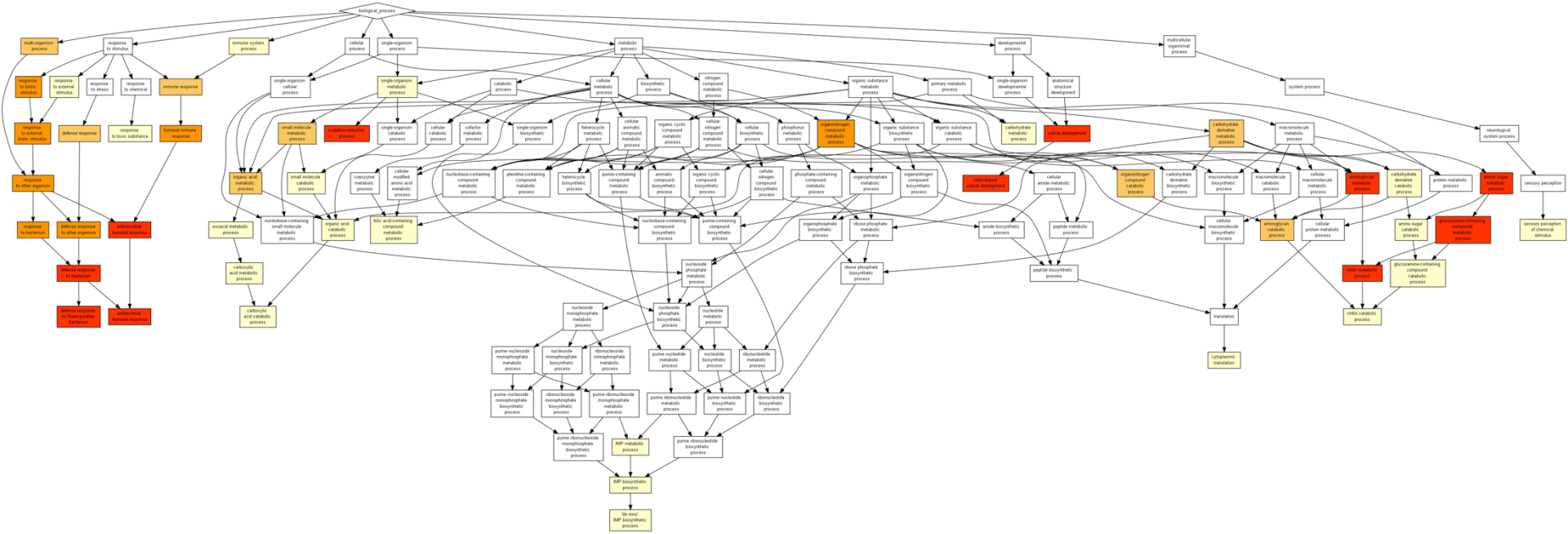


Figure S9

D. XXY female GO enrichment during aging

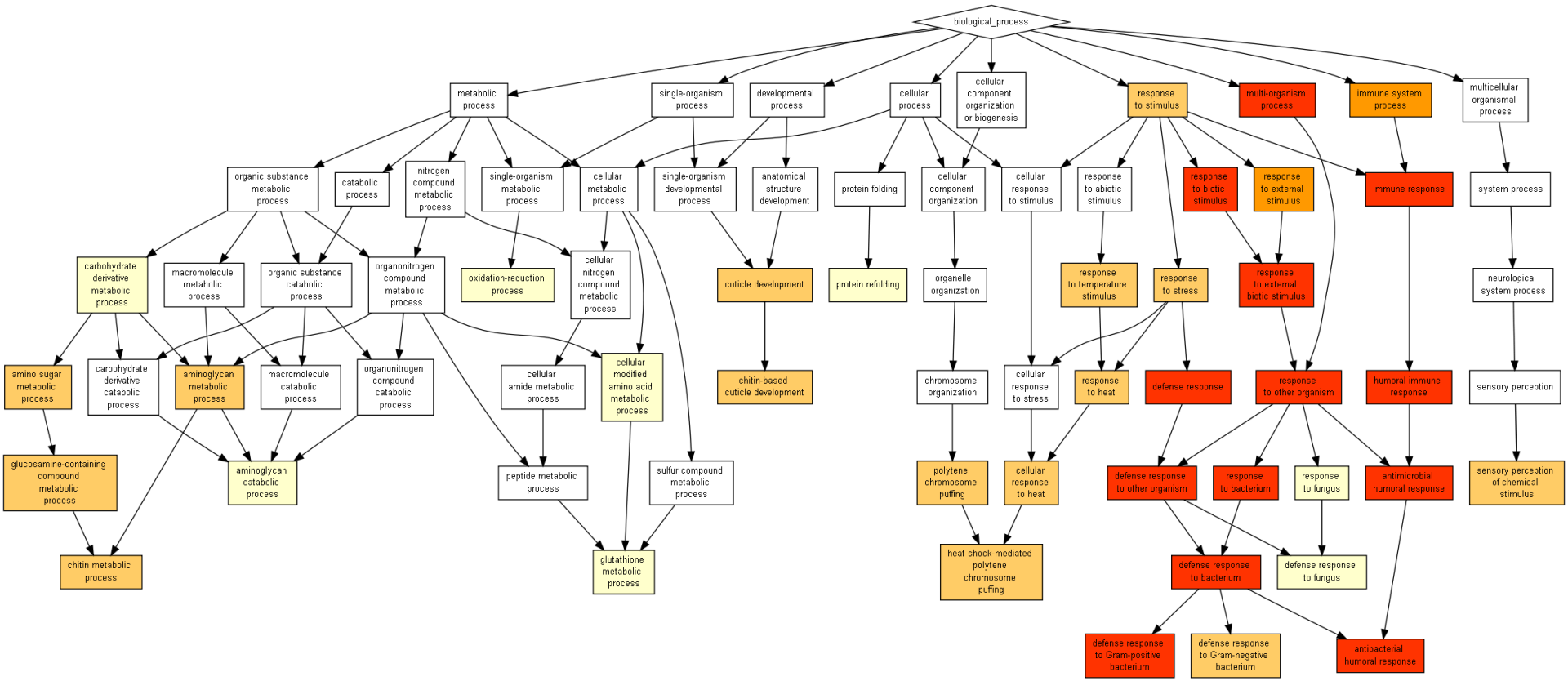


Figure S9

E.

XYY male GO enrichment during aging

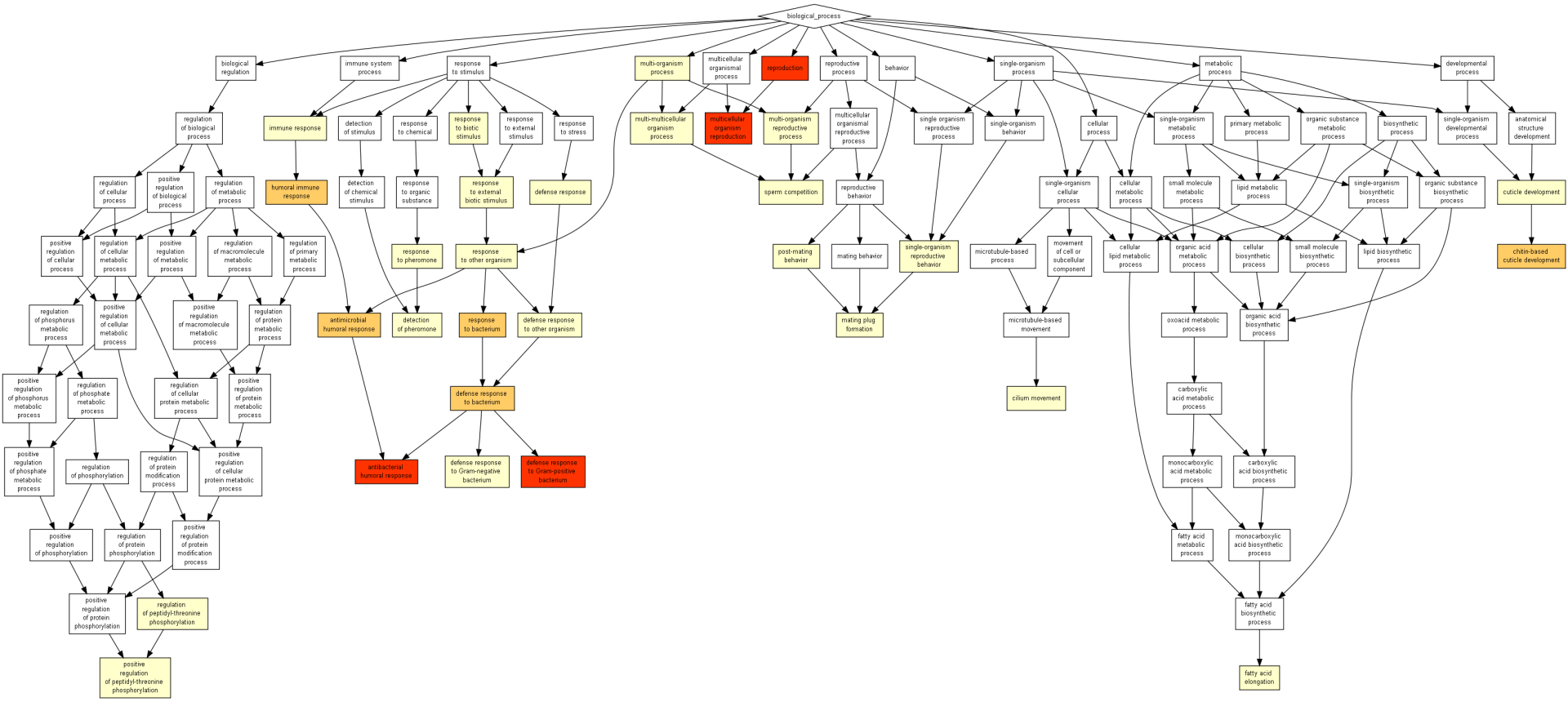
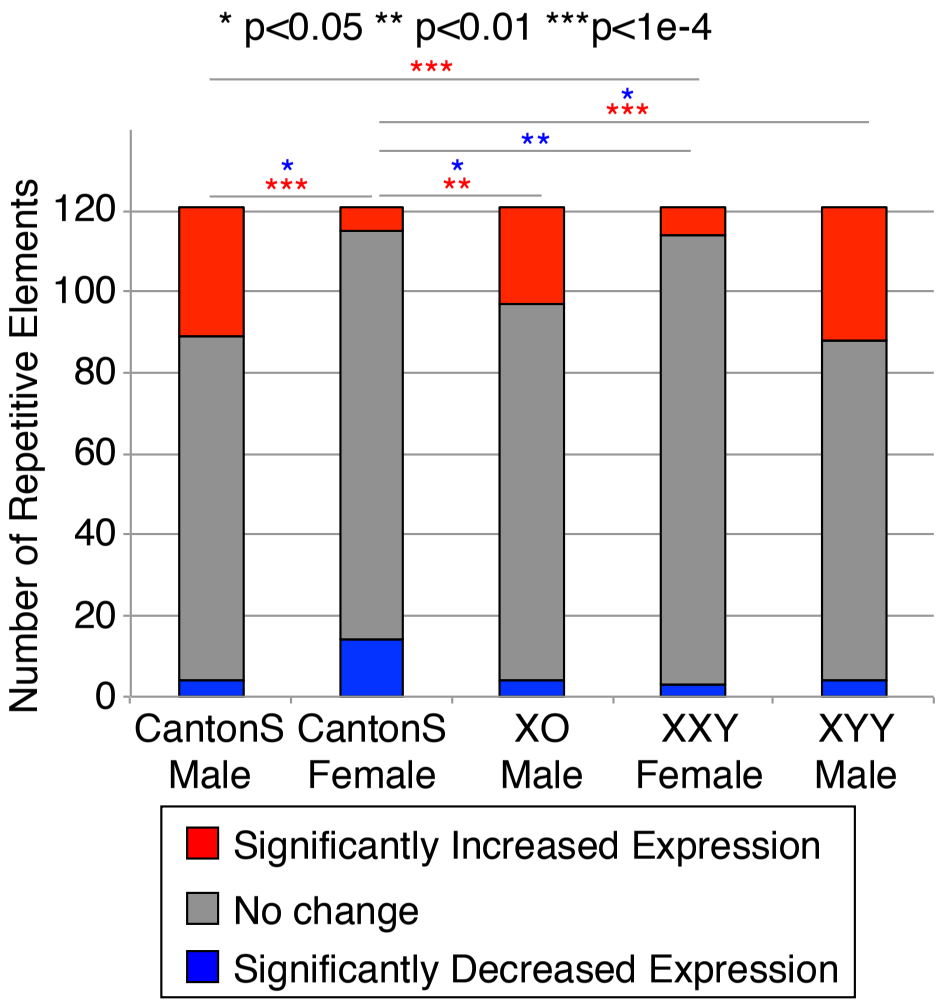


Figure S9


* p<0.05 ** p<0.01 ***p<1e-4
***
*
***
**
*
***
*
**
Number of Repetitive Elements
0
20
40
60
80
100
120
CantonS Male
CantonS Female
XO Male
XXY Female
XYY Male
Significantly Increased Expression
No change
Significantly Decreased Expression

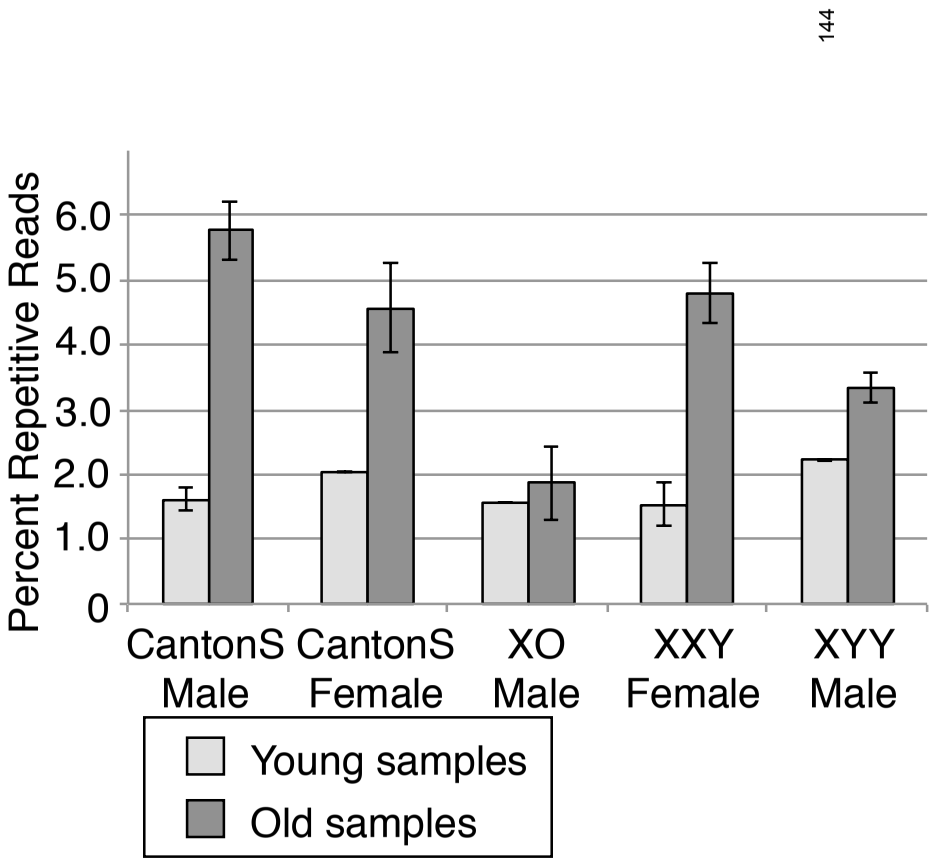

144
Percent Repetitive Reads
0
1.0
2.0
3.0
4.0
5.0
6.0
CantonS Male
CantonS Female
XO Male
XXY Female
XYY Male
Young samples
Old samples

Figure S10

Genomic coverage of *de novo* assembled repeats

Male specific:
Mcovg>10
Fcovg<2

Male biased:
Mcovg>10
Fcovg>2
Mcovg>2.5xFcovg

Both Male specific and Male biased repeats are putatively Y-linked

Unbiased

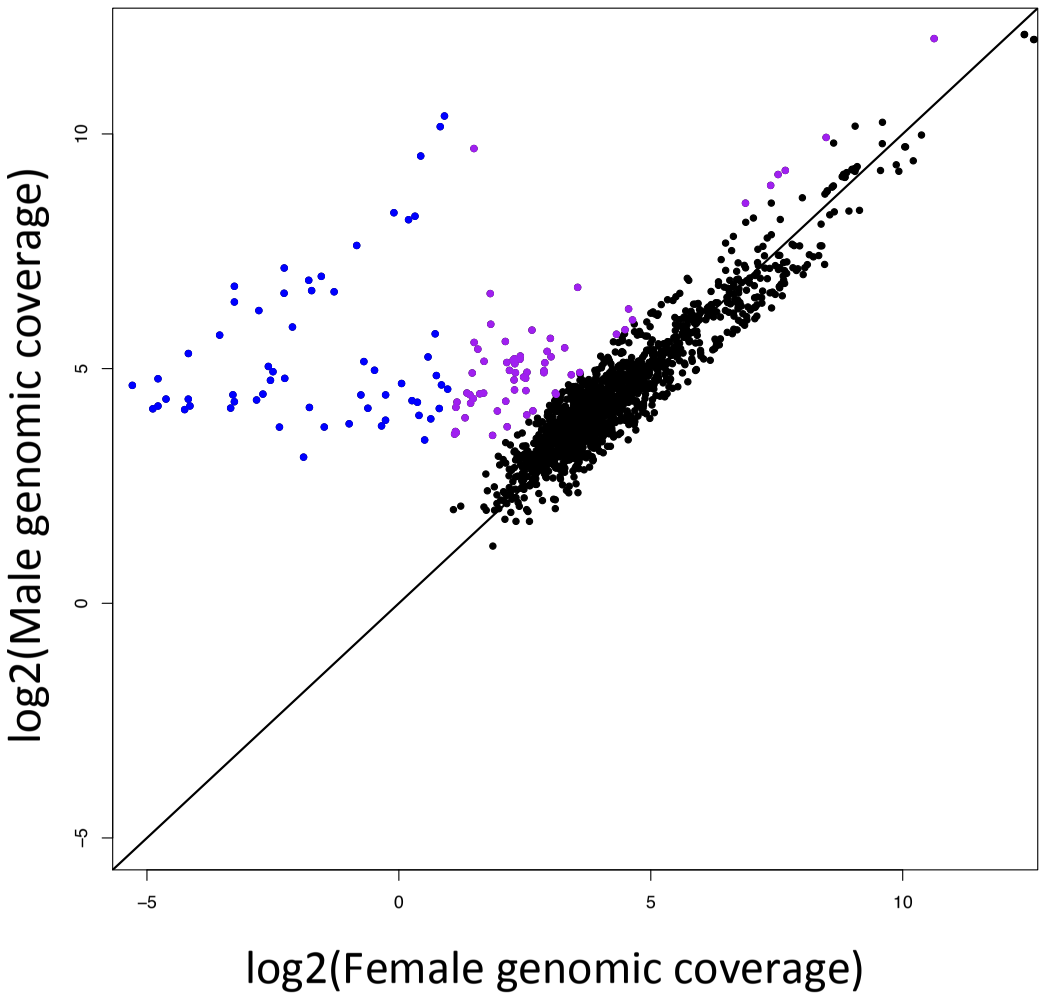


Figure S11

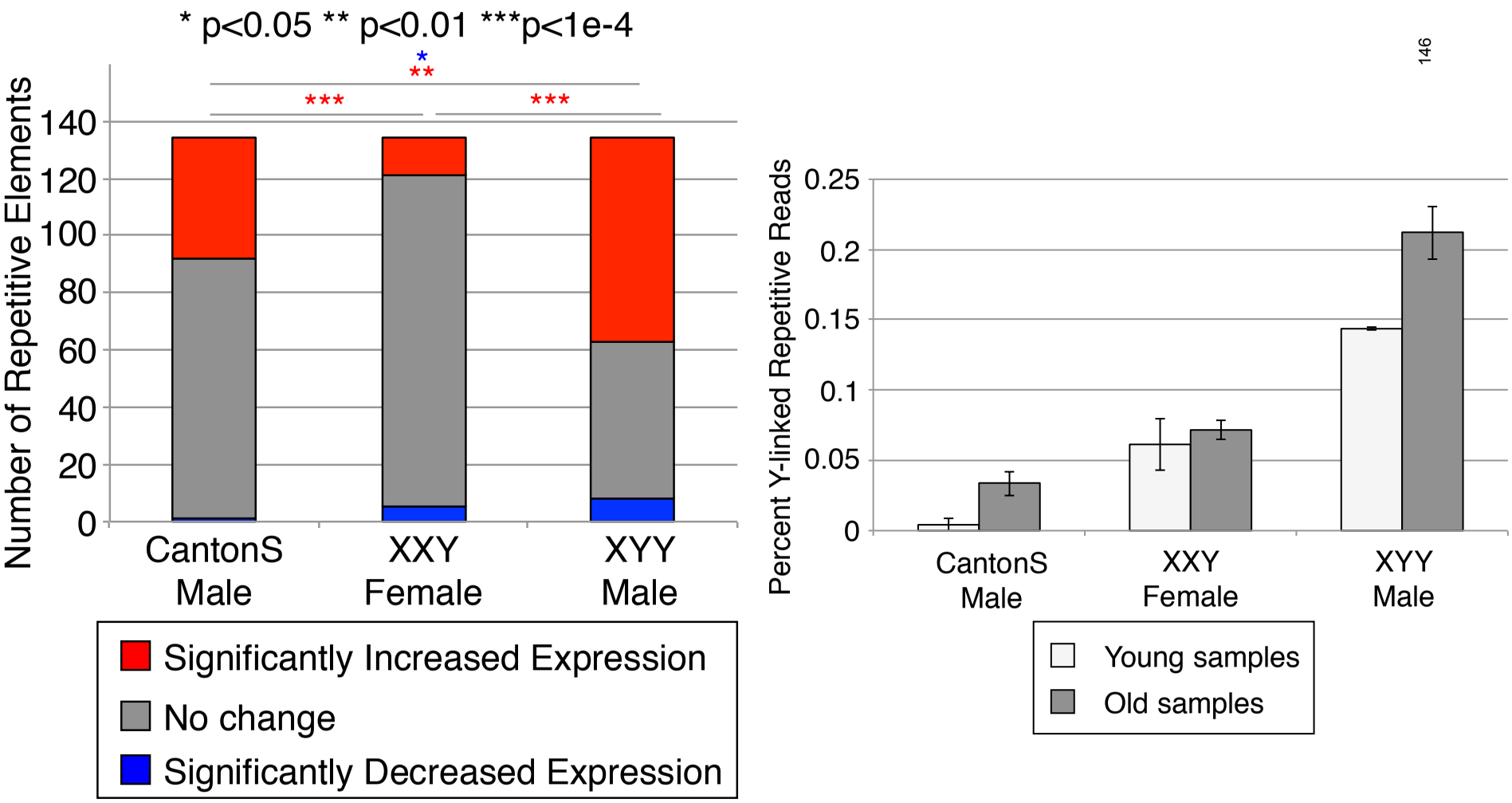
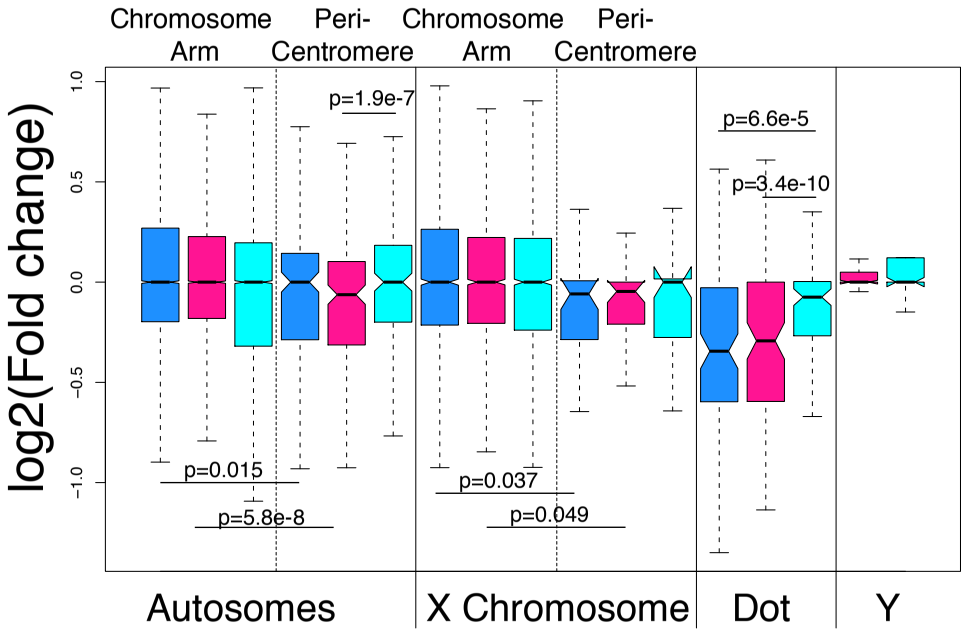
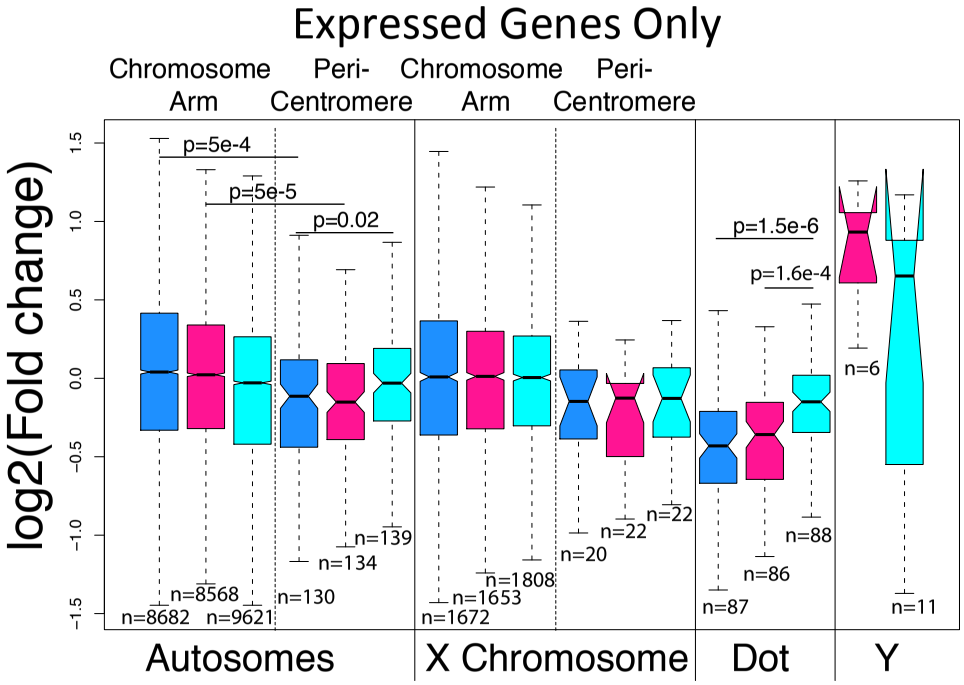
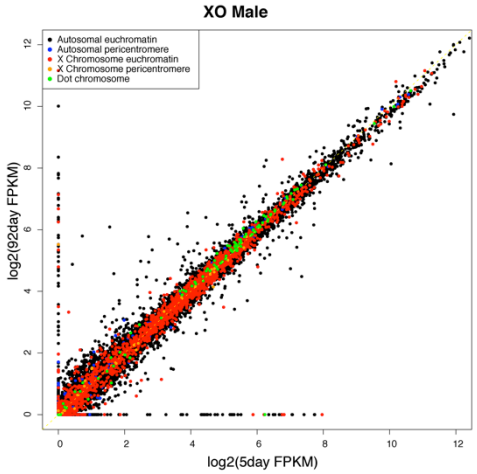
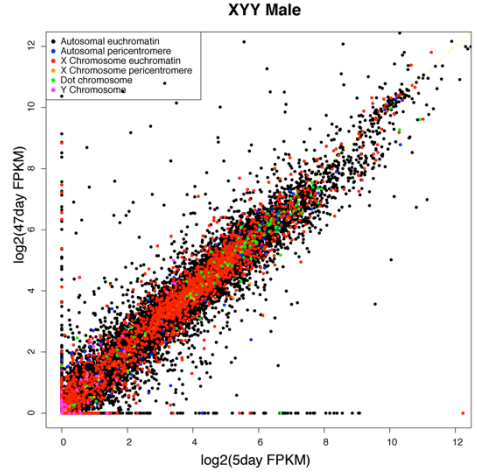

* p<0.05 ** p<0.01 ***p<1e-4
*
**
***
***
Number of Repetitive Elements
0
20
40
60
80
100
120
140
CantonS Male
XXY Female
XYY Male
Significantly Increased Expression
No change
Significantly Decreased Expression
146
Percent Y-linked Repetitive Reads
0
0.05
0.1
0.15
0.2
0.25
CantonS Male
XXY Female
XYY Male
Young samples
Old samples

Figure S12


Chromosome Arm
Peri-Centromere
Chromosome Arm
Peri-Centromere
log2(Fold change)
p=1.9e-7
p=6.6e-5
p=3.4e-10
p=0.015
p=5.8e-8
p=0.037
p=0.049
Autosomes
X Chromosome
Dot
Y


Expressed Genes Only
Chromosome Arm
Peri-Centromere
Chromosome Arm
Peri-Centromere
log2(Fold change)
p=5e-4
p=5e-5
p=0.02
p=1.5e-6
p=1.6e-4
n=8682
n=9621
n=8568
n=130
n=134
n=139
n=1672
n=1653
n=1808
n=20
n=22
n=22
n=87
n=86
n=88
n=6
n=11
Autosomes
X Chromosome
Dot
Y


XO Male
Autosomal euchromatin
Autosomal pericentromere
X Chromosome euchromatin
X Chromosome pericentromere
Dot chromosome
log2(92day FPKM)
log2(5day FPKM)


XYY Male
Autosomal euchromatin
Autosomal pericentromere
X Chromosome euchromatin
X Chromosome pericentromere
Dot chromosome
Y Chromosome
log2(47day FPKM)
log2(5day FPKM)

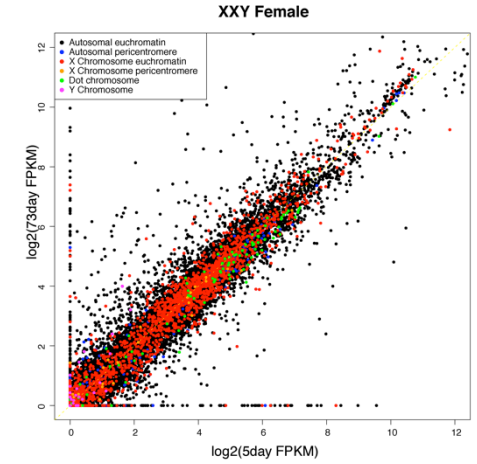

XXY Female
Autosomal euchromatin
Autosomal pericentromere
X Chromosome euchromatin
X Chromosome pericentromere
Dot chromosome
Y Chromosome
log2(73day FPKM)
log2(5day FPKM)

Figure S13

Replicate values for consensus TEs

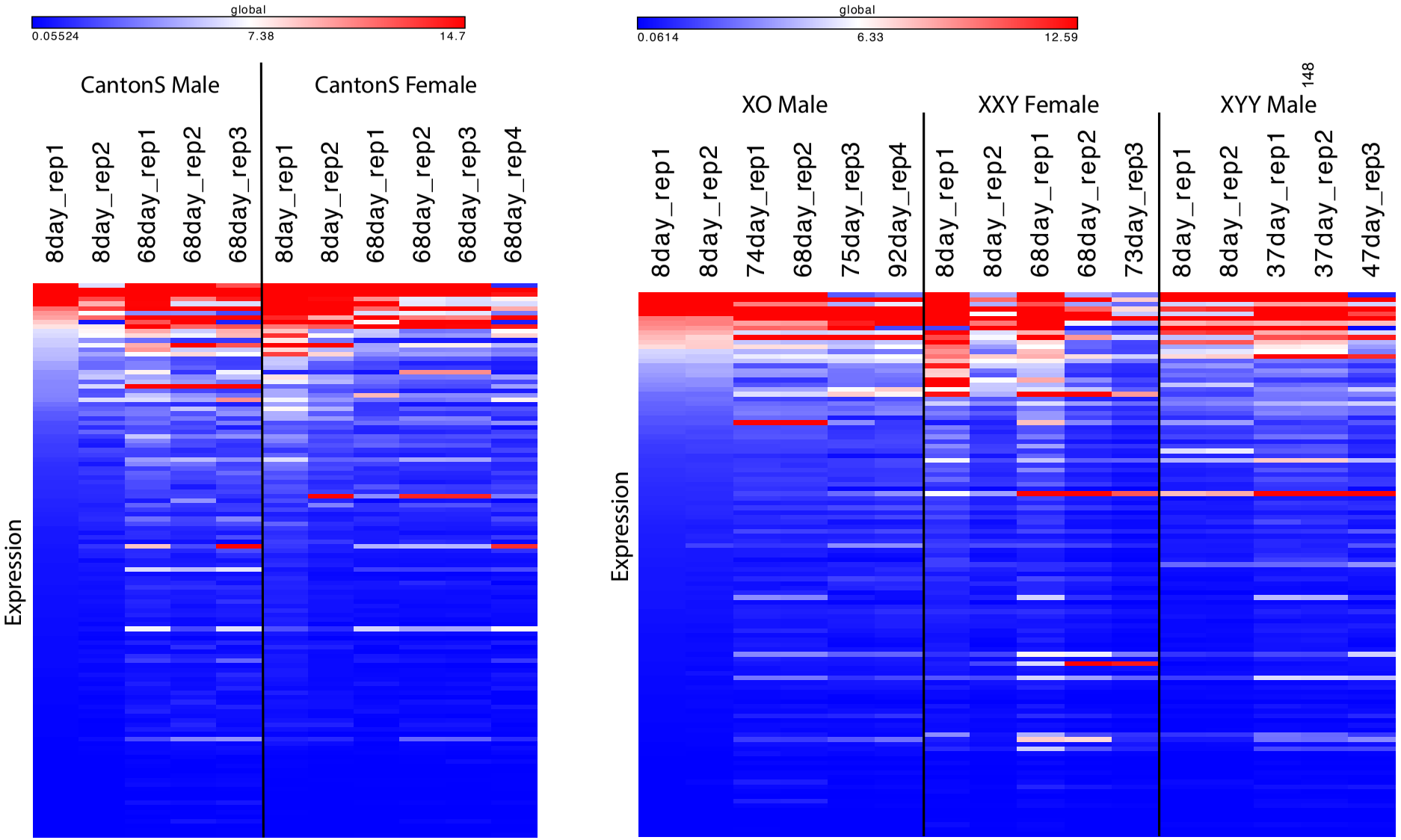


Figure S14

Replicate H3K9me2 datasets, signal in consensus TEs

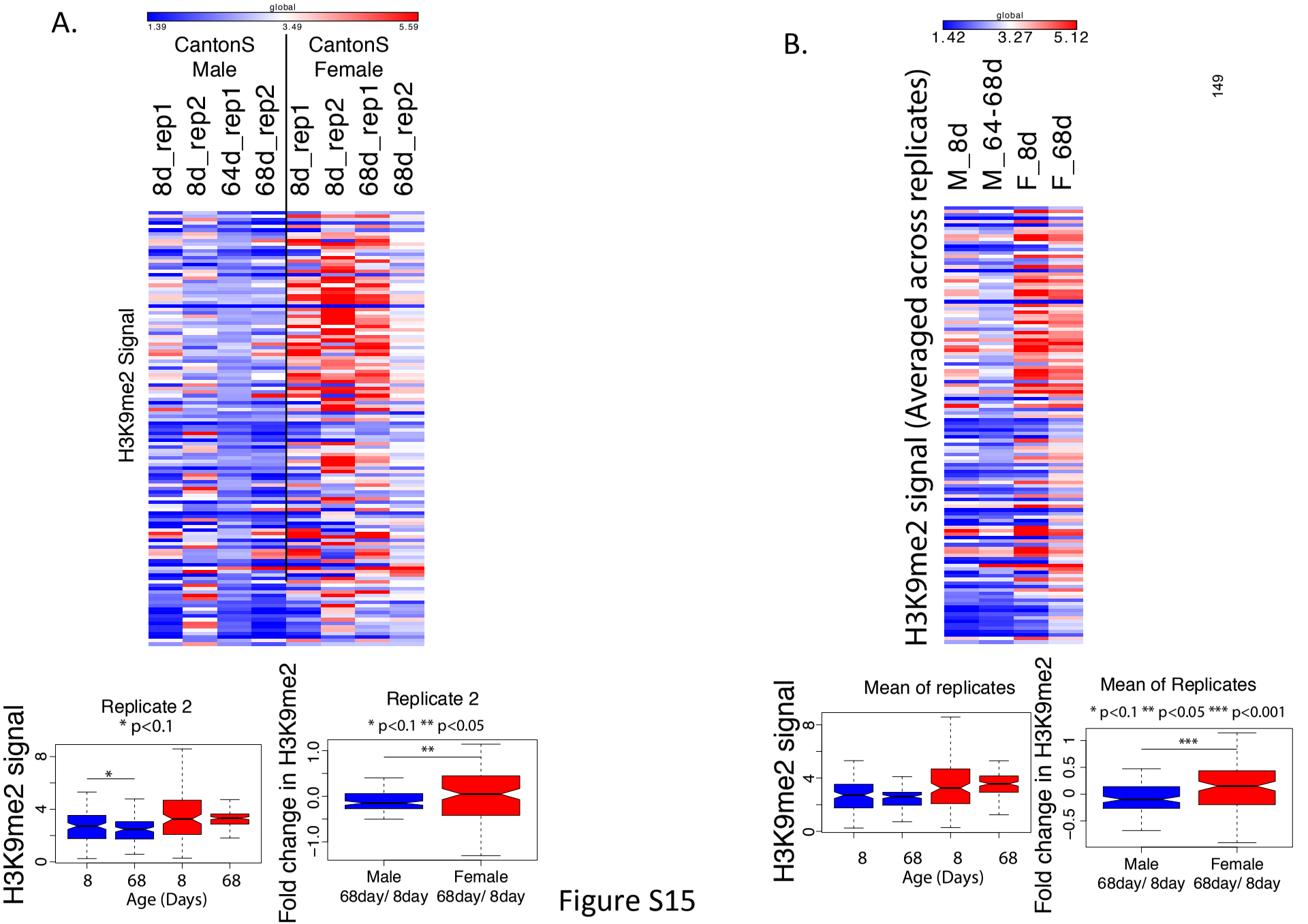


Figure S15

Table S1

	Wild-type male	Wild-type female	XO male	XXY female	XYY male
young	1.61	2.04	1.57	1.54	2.21
old	5.76	4.56	1.87	4.80	3.33
fold change	3.57	2.23	1.19	3.12	1.51

Table S2

	Wild-type male	XXY Female	XYY Male
young	0.00	0.06	0.14
old	0.03	0.07	0.21
fold change	9.12	1.17	1.48