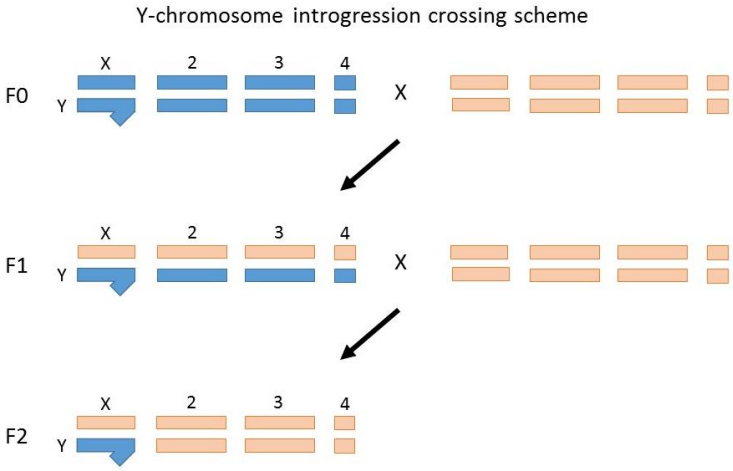
Katherine Silkaitis
Biophysics 205
Final proposal

Evaluating chromatin and protein alterations in Y-linked regulatory variation

Introduction: The Drosophila Y chromosome (ChrY) is a remarkable yet often-overlooked molecule. Notably, the *Drosophila melanogaster* Y chromosome is about the same size as the X chromosome but while the X chromosome contains more than 2,000 genes, ChrY contains only about 15, which are expressed exclusively during spermatogenesis[1,2]. Furthermore, ChrY is completely heterochromatic. While its genes are hypothesized to be monomorphic within species, there is significant structural polymorphism due to factors including copy number variation in repeated sequences[3]. Despite the paucity of protein-coding genes and effectively no nucleotide diversity within these genes, the Y chromosome has been demonstrated to have polymorphic effects on genome-wide gene regulation[4–6].

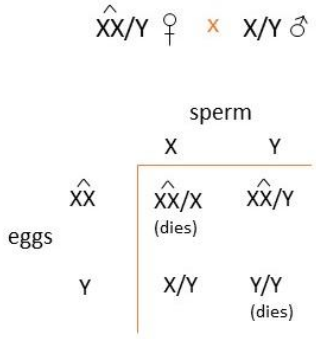
These Y-chromosome effects, known as Y-linked regulatory variation (YRV), have been observed in multiple Drosophilid species using whole-body and testes-specific microarray profiling[4–8] and RNA-seq of the testes and ovaries[9]. YRV is observed when divergent Y chromosomes are introgressed into an otherwise identical genetic background through repeated back-crossing of Y-carrying males to parental females of the desired genetic background (see figure). The varying Y chromosomes cause differential expression of thousands of autosomal and X-linked genes, including those involved in immune response, chromatin remodeling, and mitochondrial function[6]. Notably, the same effects are also apparent in Y-carrying females[5,6]: in Drosophila, sex determination is not initiated by the Y chromosome, but instead by the autosome:X chromosome ratio. The Y chromosome is only necessary for fertility, such that the X0 genotype results in viable but infertile males, while XXY genotypes are phenotypically normal females in which Y-chromosome protein-coding genes are transcriptionally inert.



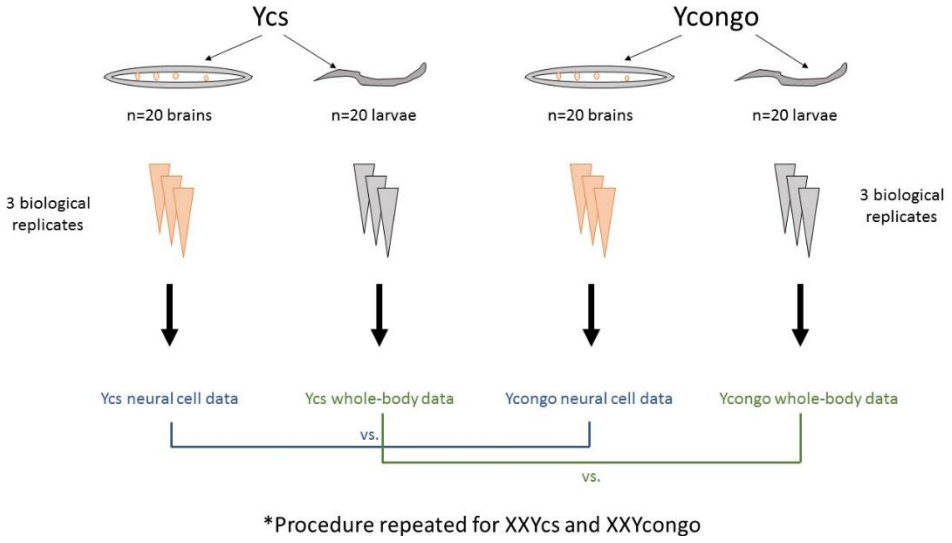
The mechanisms underlying YRV remain to be established, but it is hypothesized that differential sequestration of heterochromatin proteins may play a role. Known as the heterochromatin sink model, it hypothesizes that the Y chromosome heterochromatin serves as a sink for chromatin regulators and/or transcription factors, causing their cellular abundance to differ depending on the Y chromosome. HP1, for example, is a component of heterochromatin that also directly binds to RNA molecules and interacts with RNA polymerase II and proteins involved in RNA processing[10]. Lemos et al identified many chromatin components as differentially expressed in Y-chromosome introgressions[6], and some chromatin-associated proteins (GAGA factor and ORC2) bind to sequence repeats in the Y chromosome[11,12]. In order to gain a better understanding of the molecular forces impacting YRV, I propose to (i) probe genome-wide nucleosome occupancy in Y chromosome introgressions; (ii) search for motifs among the differentially occupied regions; and (iii) identify differential localization of heterochromatin proteins within ChrY and genome-wide.

Methods overview: All flies will be raised in a light-, humidity-, and temperature-controlled environment to minimize extraneous variation. I will utilize two existing Y-chromosome introgressions and one existing attached-X strain. The Y introgressions contain geographically distinct Y

chromosomes in an identical autosomal and X-chromosome background. The Y chromosomes originate from populations in either the Democratic Republic of Congo (Ycongo) or Ohio (Ycs). These two strains were selected because past research has shown that they display extreme responses to YRV phenotypes, including gene expression and eye color changes caused by alterations to chromatin structure at the pigment-producing *white* gene[6]. The attached-X chromosome is a genetic tool that allows the creation of Y-chromosome bearing, phenotypically normal female Drosophila. The attached-X is a compound chromosome formed by the fusion of two X chromosomes and is inherited as a unit. When the attached-X strain is crossed with a Y-chromosome introgression strain, all female offspring will bear the ChrY of interest, and each cross will have an identical genetic background (excepting ChrY; see figure). Since females are fertile and do not transcribe the ~15 Y-linked genes, XXY females allow the examination of ChrY-induced chromatin conformation and gene expression changes independent of Y-linked gene expression. For clarity, I will refer to ChrY introgressions as consomic Y strains and to female XXY with variant Y chromosomes as consomic XXY strains.



Tissue-specific YRV has not been probed other than in reproductive tissues. However, since large-scale effects of YRV have been observed and cluster into overrepresented Gene Ontology categories, it is reasonable to expect that multiple tissues will display YRV. To investigate tissue-specific and global gene expression changes between consomic strains, I will perform experiments using cultured larval neuroblasts and whole-body larval samples of XY males and XXY females. Larval neural cells will be cultured using established protocols[13] with slight modifications; briefly, I will sex third instar larvae and isolate brains from ChrY-carrying larvae. After each culture reaches ~80% confluency (~2-3 days), RNA or DNA will be extracted. For whole body samples, Y-carrying larvae will be pooled and RNA or DNA immediately extracted. Biological triplicates will be performed for all sets of extractions, such that there will be a total of 24 samples (see figure). RNA extraction, library preparation, sequencing, and mapping will be done according to established protocols using the Illumina HiSeq 2500 platform with 150 bp paired-end reads[9]. Differential gene expression will be assessed with DESeq2 in Bioconductor/R[14].



Specific Aim 1: Assess differences in nucleosome occupancy among Y and XXY consomic strains. Since I hypothesize that the Y chromosome causes variable chromatin states and thus, differential expression of genes, I will assess chromatin state using nucleosome occupancy as a proxy. I will identify regions of differential occupancy between consomic Y strains using MNase-seq. I

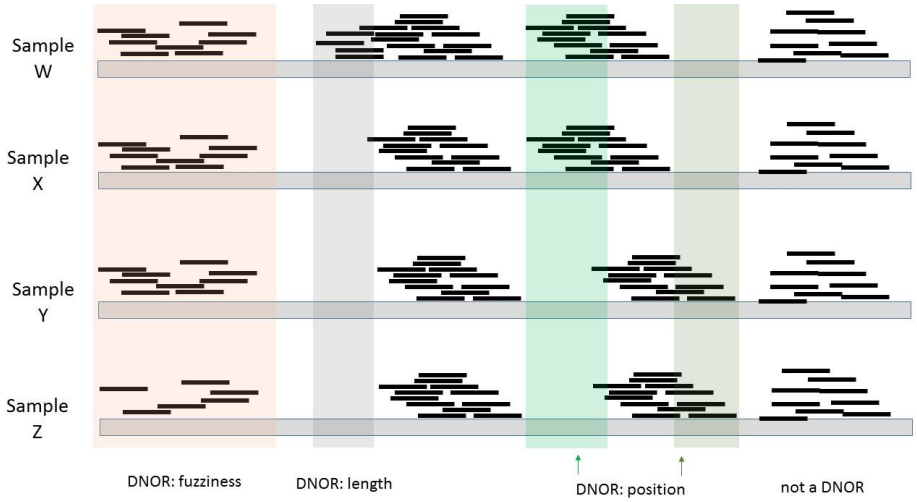
will repeat this procedure with XXY consomic strains to validate that differential nucleosome occupancy is the result of the Y chromosome structure, and not possible differential transcription of ChrY protein-coding sequences.

The preparation of nuclei, MNase digestion, and ChIP assays will be performed according to established protocol[15] and then sequenced. I will assess differential nucleosome occupancy using DiNuP, which captures changes in location, occupancy, and fuzziness between two sets of samples and accounts for biological replicates[16]. The results of this assay will yield information on genomic regions that are susceptible to changes in chromatin state in response to YRV. Since two tissue types will be assessed, neural cells and whole body, this assay will also provide insight into the potential for tissue-specific chromatin structures, while also providing information on the gross trends of YRV. I anticipate that the results from Y consomic strains will be more pronounced than those from XXY consomic strains, as YRV disproportionately affects genes that have higher expression in males[6], but that XXY females will show congruent results but at less statistical significance and/or at fewer loci.

Since MNase digestion can also identify transcription factors and regions actively transcribed by RNA polymerase III[17], I will select for fragments between 130-160 bp using the Agilent TapeStation System. An additional concern in this assay is the bias induced by the differences in chromatin structure that I am trying to asses. MNase is biased toward AT-rich regions, so if the heterochromatin sink model is valid and genomes of Y consomic strains have differential amounts and location of heterochromatin, the differential accessibility of MNase to sequences that are AT-rich will impact the ability to detect these differences. However, the AT bias has been found to have minimal influence in practice, and there are computational methods to adjust for it if needed[17].

Specific Aim 2: Identify putative regulatory motifs modulated by the Y chromosome. Differential gene expression has been observed in Drosophila, but the molecular underpinnings of the differential expression have not been investigated. One possibility is that the differential peaks identified in Specific Aim 1 contain regulatory motifs susceptible to YRV, which can alter the downstream expression of associated genes. If motifs are present, it may indicate that YRV-responsive genes are regulated cohesively based on biological function or common transcription factors; if motifs are not present, it may indicate that YRV-induced differential expression is induced at random but consistently reproducible locations in the genome, or is induced by other factors, such as proximity to heterochromatic regions or spatial arrangement in the nucleolus. To identify putative motifs, I will use the four sets of differential nucleosome occupancy peaks identified in Specific Aim 1 (Ycs vs Ycongo and XXYcs vs XXYcongo for neural cells and larval whole body) and search for enriched motifs using the de novo motif finder HOMER (Hypergeometric Optimization of Motif EnRichment)[18].

I will also correlate differential gene expression with differential nucleosome occupancy by performing RNA-seq on the 24 samples and assessing differential gene expression in the four comparisons. For each comparison, I will identify differentially expressed genes (DEGs) that overlap with differential nucleosome occupancy regions (DNORs). I will also look for DNORs within 500 bp of the transcription start site of DEGs. These two metrics will correlate gene expression differences with nucleosome occupancy; I expect to see agreement between the two, but more comprehensive analysis of the perturbations can be further probed by looking at the detailed DiNuP output. Each DNOR has metrics relating to location, length, and fuzziness; identifying patterns among these characteristics can provide insight into whether DEGs are the result of one primary nucleosome-associated factor, or result from many factors (see figure on the next page for cartoon of DNOR parameters).

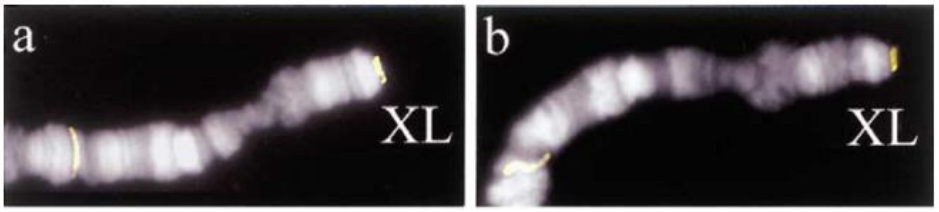


I will also use HOMER to identify motifs in DNORs associated with DEG expression and evaluate their relative enrichment or depletion in the same regions of all protein-coding genes. Careful parameters will need to be defined to define an appropriate level of acceptable degeneracy and to define a null distribution on which to determine statistically significant enrichment or depletion of DNOR-DEG associated motifs. An enrichment score $E = \frac{\#\ DEGs\ with\ motif\ /\ total\ \#\ DEGs}{\#\ genes\ in\ genome\ with\ motif\ /\ total\ \#\ genes\ in\ genome}$ and computed threshold score for each motif will identify de novo and known motifs that may be directly or indirectly affected by YRV.

Specific Aim 3: Identify differential localization, transcription and synthesis of heterochromatin proteins. As described earlier in this proposal, expression of genes in chromatin-associated Gene Ontology categories are differentially enriched in consomic strains[6]. To gain further insight into whether or not the localization and intracellular concentration of these proteins also differs, I will perform a series of immunostains, Western blots, ChIP-seq, and qPCR on cells from three biological replicates of larval neural cells. Since I anticipate the results of specific aims 1 and 2 will demonstrate that YRV effects are not due to cryptic translation of Y-linked protein coding genes, I will use only Y consomic strains in this aim.

I will focus on three heterochromatin-associated proteins: HP1, Su(var)3-9, and the histone variant H2Av. HP1 is a constitutive component of heterochromatin and generally represses gene transcription. There is strong evidence that heterochromatin expansion into euchromatic regions requires HP1, but binding of HP1 to DNA is also sensitive to the dosage of the histone methyltransferase Su(var)3-9[19]. Additionally, Lemos et al observed variable expression of HP1 and Su(var)3-9 in Y consomic strains[6]. H2Av is the Drosophila homolog of the conserved histone variant H2A.Z and it plays a role in Polycomb-mediated silencing and in establishing centromeric heterochromatin[20]. Primary and secondary antibodies for these proteins are detailed in references[21,22].

To qualitatively assess HP1, Su(var)3-9, and H2Av localization, I will treat cultures from the six cell populations (three biological replicates of Ycs and Ycongo larval neural cells) with colchicine to disrupt spindle microtubules and induce metaphase arrest[23]. I will transfer cells to a siliconized slide and stain with DAPI and the antibody against the protein of interest. The final results will appear similar the images in the figure, which illustrates HP1 localization at the long arm of the X



chromosome in two different *D. melanogaster* strains[24]. I anticipate identifying regions of differential protein occupancy between Ycs and Ycongo, but the qualitative nature of the assay will not allow for conclusions regarding the amount of bound protein. I hypothesize that HP1 will show the largest variation, and Su(var)3-9 localization will mirror HP1, but to a lesser degree since it is a histone methyltransferase and has not been observed to be a constitutive component of heterochromatin. While H2Av is associated with heterochromatin, I am less certain about its potential for differential localization; its association with DNA is more defined than that of HP1 and it is mostly found at centromeres and Polycomb-silenced chromatin, regions at which YRV effects are uncharacterized.

To quantitatively assess protein localization at DNA sequences, I will perform ChIP-seq on the same cultures using the same antibodies as for immunostaining, with the caveat that the antibodies have not been validated for ChIP and may not work. I will use Model-based Analysis of ChIP-Seq data (MACS) software to analyze the resulting short-read sequences to detect peaks[25] and will use the R package DiffBind to identify sites of differential binding between Ycs and Ycongo. These results will yield important information on whether HP1, Su(var)3-9 and H2Av associate with different sequences, and if so, what sequences they differentially bind. The latter may yield insight into functional gene groups that are regulated by the proteins, or regulatory regions or uncharacterized DNA that is affected by YRV. However, assessment of differential binding to repetitive sequences, which are often heterochromatic and may exert regulatory forces within the genome[26,27], will not be assessed due difficulties in quantifying reads from repetitive sequences.

I will perform Western blots from the same cultures to quantitatively evaluate the amount of each protein in each sample. I will use the same antibodies as for immunostaining, with the caveat that some antibodies may not work for both procedures. I will also perform RT-qPCR on RNA isolated from the cultures to quantify the amount of mRNA produced from each gene, and to identify possible differences in the mRNA:protein ratios. If the heterochromatin sink model is correct, I anticipate that protein concentrations derived from Western blot analysis will not show a significant differences between Ycs and Ycongo. Similarly, mRNA levels will also be equivalent between the two strains.

Conclusions: The broader impacts of ChrY research are beginning to be observed and appreciated in higher organisms. Thus far, the Y chromosome has been implicated in susceptibility to a mouse model of multiple sclerosis, the transcriptomes of murine immune cells, and cardiac regulation and chromatin remodeling in male mice[28–30]. Y-chromosome haplogroups have been associated with differential risk of coronary artery disease in a human population[31]. Thus, gaining a better understanding of how these effects arise may help explain disease susceptibility and progression.

YRV is a nascent field and while this research proposal will significantly add to the breadth of knowledge on the topic, it does not address or validate the causative mechanisms underlying the phenomenon. It will, however, help distinguish between potential modes of YRV and will be foundational to future studies, such as point mutations in the catalytic domains of YRV-associated proteins and resulting effects on nucleosome occupancy and differential gene expression.

There are other limitations to consider, as well. While experiments with XXY females demonstrate that differences in Y-linked protein coding gene transcripts are not the underlying cause of YRV, it is very possible that there is cryptic transcription of other Y-linked sequences. These transcripts could regulate gene expression in trans via RNA-DNA or RNA-RNA interactions. Additionally, while I use nucleosome occupancy as a proxy for chromatin state, this is a rough metric and there are estimated to be 9 distinct states defined by combinations of histone marks and protein occupancy[32]. A finer analysis of how YRV affects different types of chromatin will need to be done to make conclusive statements about its effects. Yet, given the exploratory nature of this proposal, even null results such as no correlation between DNOR peaks and differential gene expression, or no differences in protein localization and concentration, will be informative for the field.

Contributions:

- Alan T. Branco provided feedback on the one-page topic proposal
- Bernardo Lemos provided insight into the merits of DNase vs. MNase

References

1. Krsticevic, F. J., Santos, H. L., Januário, S., Schrago, C. G. & Carvalho, A. B. Functional copies of the Mst77F gene on the Y chromosome of Drosophila melanogaster. *Genetics* **184,** 295–307 (2010).

2. Carvalho, A. B., Koerich, L. B. & Clark, A. G. Origin and evolution of Y chromosomes: Drosophila tales. *Trends Genet.* **25,** 270–277 (2009).

3. Jiang, P.-P., Hartl, D. L. & Lemos, B. Y Not a Dead End: Epistatic Interactions Between Y-Linked Regulatory Polymorphisms and Genetic Background Affect Global Gene Expression in Drosophila melanogaster. *Genetics* **186,** 109–118 (2010).

4. Lemos, B., Araripe, L. O. & Hartl, D. L. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319,** 91–3 (2008).

5. Paredes, S., Branco, A. T., Hartl, D. L., Maggert, K. A. & Lemos, B. Ribosomal DNA deletions modulate genome-wide gene expression: 'rDNA–sensitive' genes and natural variation. *PLoS Genet.* **7,** 10 (2011).

6. Lemos, B., Branco, A. T. & Hartl, D. L. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15826–15831 (2010).

7. Branco, A. T., Tao, Y., Hartl, D. L. & Lemos, B. Natural variation of the Y chromosome suppresses sex ratio distortion and modulates testis-specific gene expression in Drosophila simulans. *Heredity (Edinb).* **111,** 8–15 (2013).

8. Branco, A. T., Hartl, D. L. & Lemos, B. Chromatin-associated proteins HP1 and Mod(mdg4) modify Y-linked regulatory variation in the drosophila testis. *Genetics* **194,** 609–18 (2013).

9. Unpublished data.

10. Piacentini, L. *et al.* Heterochromatin Protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. *PLoS Genet.* **5,** (2009).

11. Platero, S. J., Csink, A. K., Quintanilla, A. & Henikoff, S. Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in Drosophila. *J. Cell Biol.* **140,** 1297–1306 (1998).

12. Pak, D. T. S. *et al.* Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* **91,** 311–323 (1997).

13. Egger, B., van Giesen, L., Moraru, M. & Sprecher, S. G. In vitro imaging of primary neural cell culture from Drosophila. *Nat. Protoc.* **8,** 958–65 (2013).

14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 1–21 (2014).

15. Teves, S. S. & Henikoff, S. in *Chromatin Remodel. Methods Mol. Biol.* (Morse, R. H.) **833,** 421–432 (Humana Press, 2012).

16. Fu, K., Tang, Q., Feng, J., Liu, X. S. & Zhang, Y. DiNuP: A systematic approach to identify regions of differential nucleosome positioning. *Bioinformatics* **28,** 1965–1971 (2012).

17. Zentner, G. E. & Henikoff, S. High-resolution digital profiling of the epigenome. *Nat. Rev. Genet.* **15,** 814–827 (2014).

18. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38,** 576–589 (2010).

19. Vogel, M. J. *et al.* High-resolution mapping of heterochromatin redistribution in a Drosophila position-effect variegation model. *Epigenetics Chromatin* **2,** 1 (2009).

20. Swaminathan, J., Baxter, E. M. & Corces, V. G. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. *Genes Dev.* **19,** 65–76 (2005).

21. Eskeland, R., Eberharter, A. & Imhof, A. HP1 binding to chromatin methylated at H3K9 is enhanced by auxiliary factors. *Mol. Cell. Biol.* **27,** 453–465 (2007).

22. Ma, X. *et al.* Piwi is required in multiple cell types to control germline stem cell lineage development in the Drosophila ovary. *PLoS One* **9,** e90267 (2014).

23. Pimpinelli, S., Bonaccorsi, S., Fanti, L. & Gatti, M. Preparation and orcein staining of Mitotic Chromosomes from Drosophila larval brain. *Cold Spring Harb. Protoc.* **5,** 1–4 (2010).

24. Cenci, G., Siriaco, G. & Gatti, M. The role of HeT-A and TART retrotransposons in Drosophila telomere capping. *Genetica* **117,** 311–318 (2003).

25. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

26. Shapiro, J. a & von Sternberg, R. Why repetitive DNA is essential to genome function. *Biol. Rev. Camb. Philos. Soc.* **80,** 227–250 (2005).

27. Muotri, A. R., Marchetto, M. C. N., Coufal, N. G. & Gage, F. H. The necessary junk: New functions for transposable Elements. *Hum. Mol. Genet.* **16,** 159–167 (2007).

28. Case, L. K. *et al.* Copy number variation in Y chromosome multicopy genes is linked to a paternal parent-of-origin effect on CNS autoimmune disease in female offspring. *Genome Biol.* **16,** (2015).

29. Case, L. K. *et al.* The Y chromosome as a regulatory element shaping immune cell transcriptomes and susceptibility to autoimmune disease. *Genome Res.* **23,** 1474–85 (2013).

30. Praktiknjo, S. D. *et al.* Novel effects of chromosome Y on cardiac regulation, chromatin remodeling, and neonatal programming in male mice. *Endocrinology* **154,** 4746–56 (2013).

31. Charchar, F. J. *et al.* Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet* **379,** 915–22 (2012).

32. Kharchenko, P. V *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471,** 480–485 (2011).