

ARTICLE



<https://doi.org/10.1038/s41467-020-19134-9>

OPEN

Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males

Kevin H.-C. Wei ^{1,2}, Lauren Gibilisco^{1,2} & Doris Bachtrog ¹ 

Large portions of eukaryotic genomes consist of transposable elements (TEs), and the establishment of transcription-repressing heterochromatin during early development safeguards genome integrity in *Drosophila*. Repeat-rich Y chromosomes can act as reservoirs for TEs ('toxic' Y effect), and incomplete epigenomic defenses during early development can lead to deleterious TE mobilization. Here, we contrast the dynamics of early TE activation in two *Drosophila* species with vastly different Y chromosomes of different ages. Zygotic TE expression is elevated in male embryos relative to females in both species, mostly due to expression of Y-linked TEs. Interestingly, male-biased TE expression diminishes across development in *D. pseudoobscura*, but remains elevated in *D. miranda*, the species with the younger and larger Y chromosome. The repeat-rich Y of *D. miranda* still contains many actively transcribed genes, which compromise the formation of silencing heterochromatin. Elevated TE expression results in more de novo insertions of repeats in males compared to females. This lends support to the idea that the 'toxic' Y chromosome can create a mutational burden in males when genome-wide defense mechanisms are compromised, and suggests a previously unappreciated epigenetic conflict on evolving Y chromosomes between transcription of essential genes and silencing of selfish DNA.

¹Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA. ²These authors contributed equally: Kevin H.-C. Wei, Lauren Gibilisco. ✉email: dbachtrog@berkeley.edu

In most animals, the zygotic genome is initially inactive and epigenetically “naive” (i.e., it contains few chromatin features)¹, and the first stages of embryonic development are solely controlled by maternal proteins and transcripts^{2,3}. Concordant with genome-wide activation of zygotic expression, the embryo also initiates silencing of genomic regions whose transcription would be harmful⁴. In particular, large fractions of eukaryotic genomes consist of repetitive DNA, including transposable elements (TEs)⁵, and transcriptional activation of TEs could result in their mobilization, causing insertional mutations and genomic instability⁶. Silencing of repeats is achieved in part through establishment of constitutive heterochromatin in all cells during early development at repetitive DNA at centromeres, telomeres, and along the repeat-rich Y chromosome^{7,8}. TEs are in a constant evolutionary battle with their host genome to avoid silencing, and may have evolved mechanisms to mobilize early in embryogenesis before widespread heterochromatin is in place⁹. In many species, including *Drosophila*, the Y chromosome consists almost entirely of repetitive DNA, and male embryos may thus be especially challenged to silence their TEs¹⁰.

Here, we contrast TE activation and heterochromatin formation during early embryogenesis in two closely related *Drosophila* species (Fig. 1). *Drosophila pseudoobscura* and *Drosophila miranda* diverged only 3 Ma (million years ago) and show a similar repeat complement, with most TEs being shared between species^{11,12}. Yet, the size of their Y chromosome differs dramatically: a fusion between an autosome and the ancestral Y created a neo-Y in *D. miranda* ~1.5 Ma (ref. 13; Fig. 1a), which has dramatically expanded in size (from ~25 Mb to almost 100 Mb; ref. 14). This drastic size expansion is driven almost entirely by the accumulation of TEs on the neo-Y¹⁴; for example, the most common TE on the neo-Y (the ISY element) is inserted roughly 22,000 times on the neo-Y (and occupies over 16 Mb of sequence), while it only shows ~1500 copies on the neo-X (<1 Mb of sequence)¹⁴. High-quality genome assemblies that contain

large amounts of highly repetitive DNA, including pericentromeric regions and Y-linked sequences, exist for both species^{12,14}, which allow us to study TE expression and heterochromatin formation at the molecular level, and the role of the Y chromosome.

Results and discussion

Elevated zygotic TE transcripts in male embryos. To study the dynamics of early repeat activation in *D. pseudoobscura* and *D. miranda*, we combine gene expression and chromatin profiles in single embryos during early development. In *Drosophila*, zygotic transcription begins at the pre-blastoderm (stage 2) and gradually increases; by the end of stage 4 (syncytial blastoderm), widespread zygotic transcription is observed (the maternal-to-zygotic (MZ) transition, see Fig. 1b)^{15–19}. We analyzed transcriptomes of replicate single male and female embryos of *D. pseudoobscura* and *D. miranda* that have been developmentally staged from embryonic stage 2 through 12 (Fig. 2a, Supplementary Figs. 1 and 2, and Supplementary Data 1 and 2 (ref. 19)). Hierarchical clustering of the samples by their TE transcript abundance divides the embryos into two distinct groups that coincide with the MZ transition (Fig. 2b). Prior to the onset of widespread zygotic transcription (stages 2 and 4), TE transcript profiles are highly correlated between stages of the same species (Fig. 2b). As expected, female and male embryos are highly similar as the transcript pool is predominated by maternally deposited RNAs. After the MZ transition, samples form clades separated by species and sex (Fig. 2b). Interestingly, while *D. pseudoobscura* female and male samples are highly correlated and clustered, TE transcription profiles in *D. miranda* are less similar between sexes, and females of *D. miranda* group more closely to *D. pseudoobscura* (Fig. 2b).

The sex differences after the MZ transition appear to be driven in part by higher TE expression in males in both species (Fig. 2a,

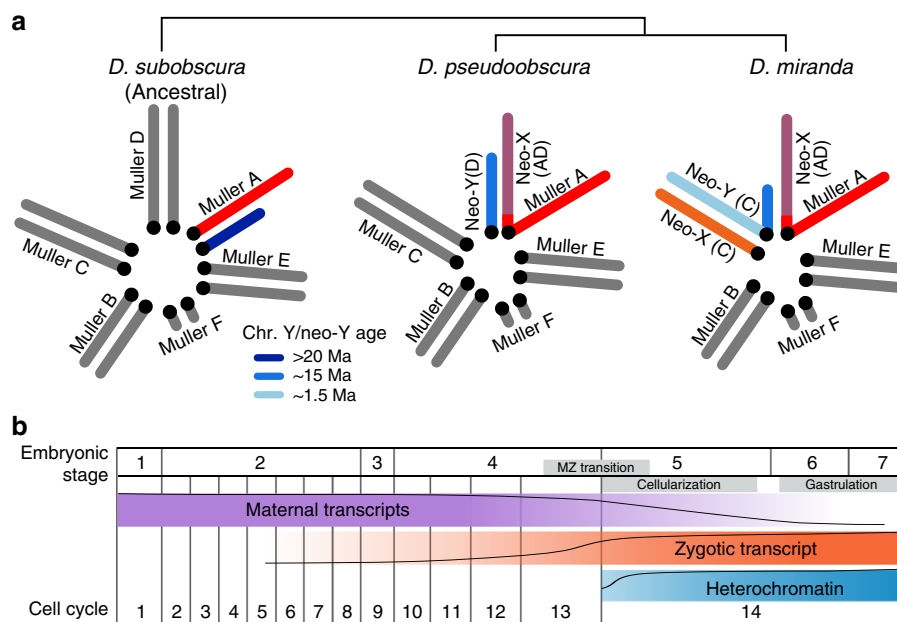


Fig. 1 Neo-Y chromosome emergence and transcriptional activity during early embryogenesis in *Drosophila*. **a** The male karyotypes of three species in the obscura species group showing three different Y and neo-Y (blue) chromosome configurations. Chromosome arms are labeled as Muller Elements. The X/neo-X chromosomes are in warm colors and autosomes in gray. *Drosophila subobscura* represents the ancestral karyotype. *D. pseudoobscura* has a highly degenerate and old neo-Y that formed 15 Ma after the fusion of the X (Muller A) with Muller D, creating a neo-X chromosome. *D. miranda* has a young neo-Y that formed after Muller C fused with the Y ~1.5 Ma. The unfused Muller C then became the neo-X. **b** Schematic representation of transcriptional changes and heterochromatin regulation in early embryonic development^{15–19}. The approximate abundance of maternal transcripts (purple), zygotic expression (orange), and repressive heterochromatin (blue) are depicted.

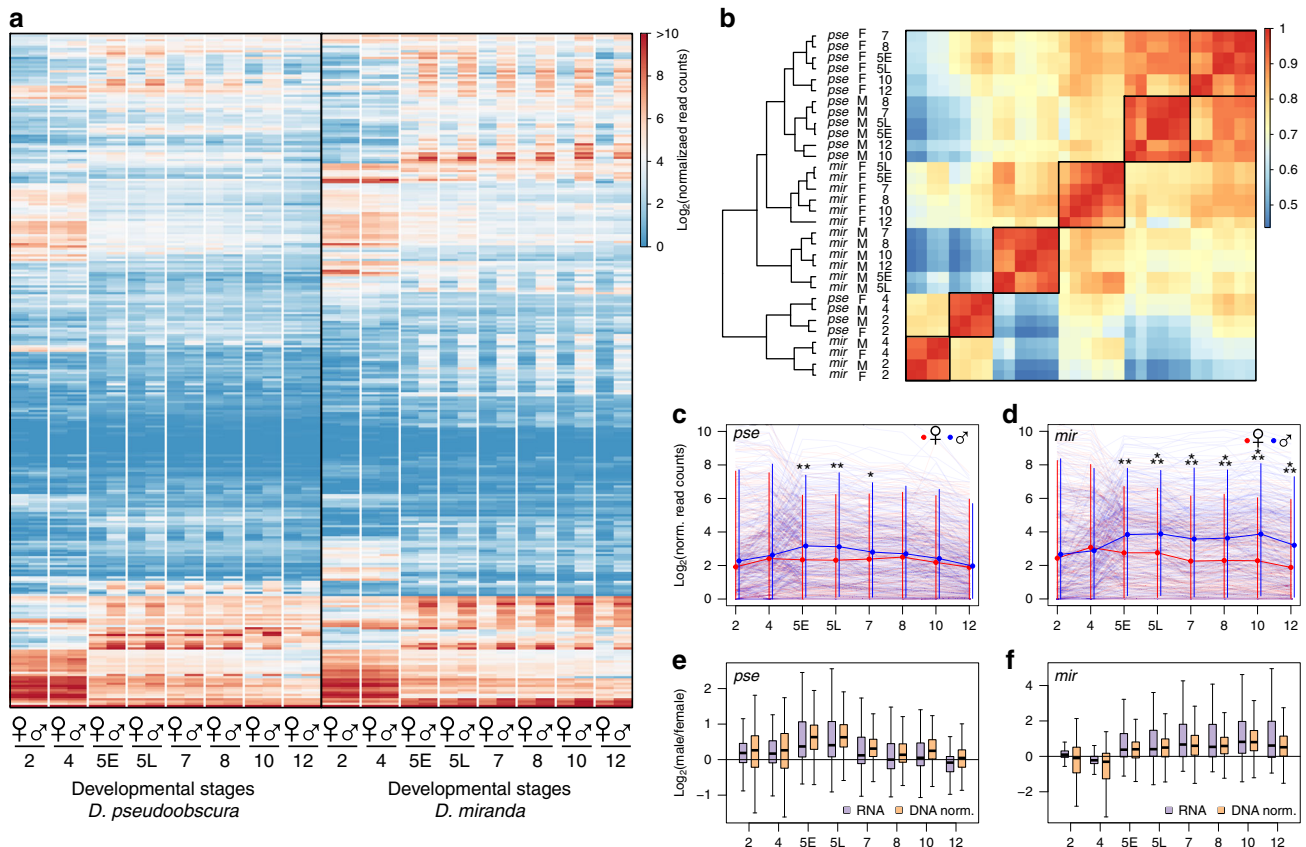


Fig. 2 Sex-specific transcriptional regime of transposable elements across early embryogenesis. **a** Heatmap showing the normalized read counts mapping to TEs from developmentally staged and sexed embryos (average of three embryos) from *D. pseudoobscura* and *D. miranda*. The list of TE names can be found in Supplementary Fig. 1. For heatmap in FPKM, see Supplementary Fig. 2. **b** Heatmap depicting the correlation of TE transcript abundance between species, sex, and developmental stages. Samples were ordered by hierarchical clustering. **c, d** TE transcript abundance across developmental stages, and sexes in *D. pseudoobscura* and *D. miranda*. Light lines depict TE transcript abundance for individual TEs, and dark lines depict the median with 95% confidence intervals demarcated by vertical lines. *, **, and *** denote significant differences between males and females with *p* values of <0.01, <0.0001, and <0.000001, respectively (two-tailed Wilcoxon rank-sum test, *n* = 303 consensus TEs from the repeat index, no multiple testing correction). **e, f** Distribution of the fold differences in TE transcript abundance (purple) between sexes in log₂ scale. RNA-seq read counts at TEs are further normalized by their DNA-seq read counts to account for copy number, and the fold differences between sexes are depicted (orange); *n* = 303 consensus TEs from the repeat index. In the boxplots, whiskers delineate the minima and maxima of the distributions, the boxes demarcate the 25th and 75th percentiles, and the centers mark the medians. Source data are provided as a Source data file.

c, d and Supplementary Fig. 3); no such differences are seen at autosomal genes (Supplementary Fig. 4). As zygotic expression increases, TEs are activated more highly in males than in females (Fig. 2a). *D. pseudoobscura* males have significantly higher TE expression than females immediately after the transition; at late stage 5 where the difference is greatest, males have on average 1.68-fold higher TE expression than females (Fig. 2c). TE expression in males then gradually decreases throughout development, resulting in similar expression levels between the sexes at stage 12 (Fig. 2a, c, e). This suggests that efficient silencing mechanisms are established by then, in both sexes (see below). In *D. miranda*, TE transcripts are generally more abundant than in *D. pseudoobscura* (Fig. 2a, c, d), and males similarly show significantly higher TE expression than females. However, elevated TE expression is maintained in *D. miranda* males throughout early development (Fig. 2a, d). At stages 10 and 12, *D. miranda* males have on average more than twice the TE expression of females (2.37 and 2.04-fold, respectively; Fig. 2d, f), suggesting that TEs may be evading silencing in *D. miranda* males. Notably, elevated transcript abundance in males is not simply due to the higher copy number of repeats; after normalizing the TE read counts with their copy numbers in

males and females, higher transcript abundance in males persists in both species (Fig. 2e, f).

Y-linked TEs drive elevated TE expression in males. The presence of the Y and neo-Y chromosomes in males substantially increases the repeat content of the cell¹⁰. Elevated TE expression in males can either result from misregulation of Y-linked TEs or global reduction in repeat suppression¹⁰. Given the repetitive nature of TEs, it is typically not possible to pinpoint the specific genomic copy from which a TE transcript originates, especially for highly active families with large numbers of identical insertions. Instead, we identified TE families preferentially located on the Y chromosome based on their relative abundance in males vs. females (twofold or higher mapping of DNA-seq reads); this resulted in 20 and 79 Y-enriched TEs in *D. pseudoobscura* and *D. miranda*, respectively (out of a total of 303 TEs; Fig. 3a, b). In both species, Y-enriched TEs are significantly more highly expressed in males after the MZ transition compared to females, and the magnitude of male bias is also significantly higher than for the remaining TEs (Fig. 3c–f). These results are consistent with misregulation of Y-linked TEs driving elevated repeat

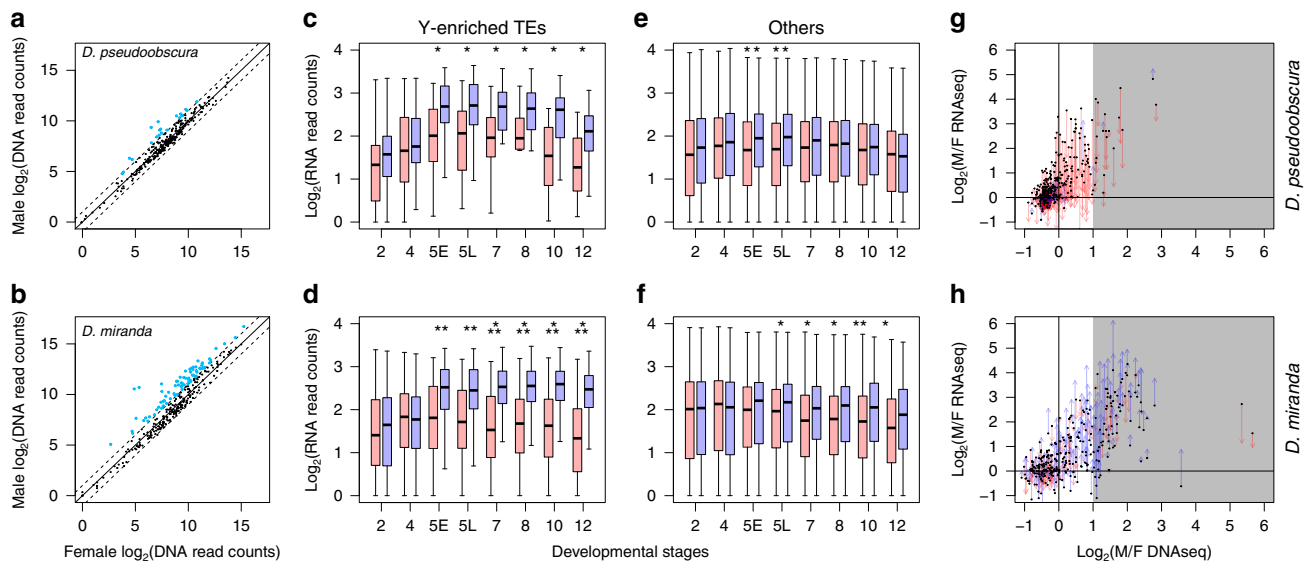


Fig. 3 Y-linked TEs drive male-biased expression in both *D. pseudoobscura* and *D. miranda*. **a, b** Regression between female and male DNA-seq read counts mapping to TEs. TEs with more than twice the abundance in males are deemed Y-enriched (blue). Solid diagonal delineate the identity line, dotted lines delineate ± 2 -fold differences. **c, f** Boxplots depict the transcript abundances of either the Y-enriched (**c, d**) or remaining (**e, f**) TEs between the sexes (female in red and male in blue) and across developmental stages. *, **, and *** denote significant differences between males and females with p value of <0.05 , <0.0005 , and <0.000005 , respectively (two-tailed Wilcoxon rank-sum test, no multiple testing correction; $n = 20, 79, 283$, and 224 consensus TEs for **c, d, e**, and **f**, respectively). In the boxplots, whiskers delineate the minima and maxima of the distributions, the boxes demarcate the 25th and 75th percentiles, and the centers mark the medians. **g, h** Fold difference between female and male DNA-seq read counts plotted against that from RNA-seq read counts at stage 5. Gray background demarcate cutoff for Y-enriched TEs. Fold expression differences between sexes for each TE are shown at developmental stages 5 (dots) and 12 (arrowheads). The color of the vertical line denotes whether the expression of a TE becomes less (red) or more (blue) male biased across the development. Source data are provided as a Source data file.

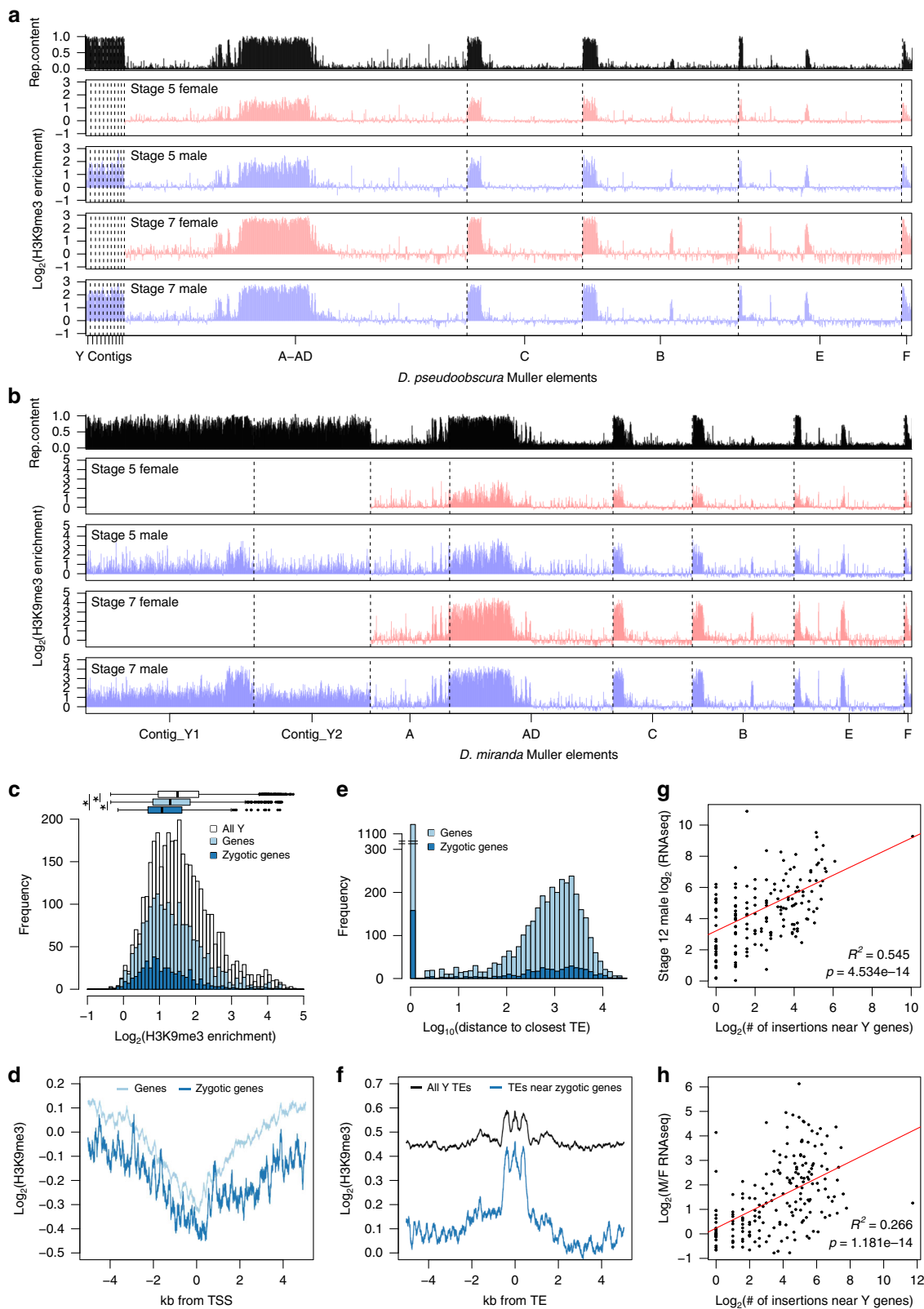
expression in males. TEs not classified as Y-biased nonetheless show moderately male-biased expression, suggesting that global TE regulation may also be affected in males (Fig. 3e–f). The extent of male-biased expression of Y-enriched TEs decreases through development in *D. pseudoobscura* (Fig. 3g), as their transcript abundance declines in both females and males (Fig. 3c, e, g). Only a small set of Y-enriched TEs do not drop to the same levels as in females, suggesting that some Y-linked copies may not be fully silenced (Supplementary Fig. 4). In contrast, a large number of Y-enriched TEs maintain their high expression level throughout development in male *D. miranda* (Fig. 3d and Supplementary Fig. 3); in fact, they become more male biased with time (Fig. 3h), largely due to decreased TE expression in females (Fig. 3d, f). Thus, Y-linked TEs appear to be poorly suppressed in *D. miranda* causing persistently elevated expression in males.

Reduced levels of heterochromatin at TEs on the *D. miranda* neo-Y. To determine if incomplete epigenetic silencing may drive elevated expression of Y-linked TEs after the MZ transition, we characterized genome-wide enrichment profiles of the repressive chromatin mark H3K9me3 in single male and female embryos at stages 5 and 7 (Fig. 4a, b). These two time points reflect the transition of initiation of heterochromatin (at the onset of cellularization of the blastoderm in early stage 5) toward maturation into a stable, repressive chromatin compartment (gastrulation of the embryo at stage 7)^{18,20,21}. Overall, H3K9me3 is enriched at repeat-rich regions in both species at both stages, including the Y/neo-Y chromosomes and the pericentromeres (Fig. 4a, b). As expected, H3K9me3 enrichment is higher at stage 7, consistent with the progressive establishment and spreading of heterochromatin after the MZ transition^{18,20,21}. Therefore, this suggests that decreasing male-biased TE expression across the development in *D. pseudoobscura* likely results from increased heterochromatic suppression, and efficient silencing of Y-linked TEs

(Fig. 4a). In contrast, the neo-Y chromosome of *D. miranda* does not appear to become fully heterochromatinized despite containing tens of thousands of TEs (Fig. 4b). Chromosome-wide H3K9me3 enrichment levels on the neo-Y are markedly less than that of pericentric heterochromatin at the X and autosomes, at both developmental stages, even though their repeat content is similar (Fig. 4b and Supplementary Fig. 5).

Transcription of neo-Y genes impedes heterochromatin formation at TEs. The genome architecture of the neo-Y differs from pericentromeres; despite consisting mostly of repetitive DNA, the neo-Y still contains thousands of functional genes with 6448 genes annotated in the current assembly^{14,22}. Active transcription at neo-Y genes may impede heterochromatin formation, creating islands of euchromatin across the neo-Y²³. Indeed, while heterochromatic Y chromosomes do not form polytene chromosomes in *Drosophila*²⁴, segments of the neo-Y chromosome maintain euchromatin-like banding patterns in polytene spreads interspersed with under-replicated heterochromatin²⁵. TEs may therefore exploit these euchromatic environments to maintain elevated activities. Supporting this hypothesis, segments of the neo-Y adjacent to active genes are less heterochromatic: neo-Y windows overlapping annotated genes have significantly lower H3K9me3 enrichment compared to all neo-Y windows (1.2-fold lower; Wilcoxon rank-sum test $p < 2.2e-16$; Fig. 4c), and windows with zygotically expressed Y-linked genes have even less H3K9me3 enrichment (Wilcoxon rank-sum test $p = 4.689e-07$; Fig. 4c). H3K9me3 levels are depleted near the transcription start sites of both Y-linked and zygotically expressed Y-linked genes, and gradually increase distally (Fig. 4d).

To determine whether TEs profit from reduced silencing near Y-linked genes, we evaluated the distribution of genes and TE insertions on the neo-Y. Y-linked genes are on average only 512 bp away from the closest TEs and 25.4% have at least one



insertion within the gene (and 27.6% of zygotically expressed Y genes), presumably in the introns and UTRs (Fig. 4e). In contrast, genes and TEs are significantly farther apart on autosomes, with an average distance of 4127 bp and 7.7% of autosomal genes have internal TE insertions (Wilcoxon rank-sum test $p < 2.2e-16$; Supplementary Fig. 6). Despite elevated H3K9me3 enrichment, TEs near Y-linked zygotically expressed genes are less enriched

than TEs across the entire chromosome (Fig. 4f). Consistently, TEs are expressed more highly in males if they have a larger number of insertions around (± 5 Kb) zygotically expressed neo-Y genes ($p = 4.53e-14$; Fig. 4g), and these TEs show higher male bias in their expression ($p = 4.87e-05$; Fig. 4h). Therefore, TEs neighboring transcribed neo-Y-linked genes are poorly suppressed and likely a main source of the persistently elevated

Fig. 4 The heterochromatin landscape on the neo-Y. **a, b** *D. pseudoobscura* and *D. miranda* H3K9me3 enrichment between sexes and across stages 5 and 7. The repeat content of the chromosomes are depicted at the top (in black). Chromosomes and contigs are separated by vertical dotted lines. For H3K9me3 enrichment estimation using only uniquely mapped reads, see Supplementary Fig. 5. **c** Distribution of H3K9me3 enrichment in 50 kb windows on the neo-Y (white, $n = 3642$ windows). Windows overlapping annotated genes and zygotically expressed genes are in light ($n = 1904$ windows) and dark blue ($n = 480$ windows), respectively. * denotes significant difference in H3K9me3 enrichments between windows with p value < 0.00001 (two-tailed Wilcoxon rank-sum test, no multiple testing correction; $p < 2.2 \times 10^{-16}$ in all vs. genes, $p < 2.2 \times 10^{-16}$ in all vs. zygotic genes, and $p = 4.69 \times 10^{-7}$ in genes vs. zygotic genes). In the boxplots, whiskers demarcate the minima and maxima of the distributions, the boxes delineate the 25th and 75th percentiles, and the centers mark the medians. **d** Median H3K9me3 enrichment ± 5 Kb around the transcription start site of all (light blue) and zygotically expressed (dark blue) genes on the neo-Y. **e** Distance between genes and the nearest TE on the neo-Y. 0 indicates the TE is within a gene. Light blue represents TE insertions near all neo-Y genes ($n = 4401$ insertions), and dark blue represents TE insertions near zygotically expressed neo-Y genes ($n = 573$ insertions). **f** Median H3K9me3 enrichment around all annotated TEs and those within ± 5 Kb of a gene on the neo-Y. **g, h** The number of times TEs are found to have inserted near zygotically expressed genes on the neo-Y (within ± 5 Kb) is plotted against the TE transcript abundance (**g**), and the fold difference of TE transcript abundance between males and females (**h**) at stage 12. Line of best fit is in red; Pearson's R of the regressions and the significance of the correlations are labeled. Only TEs found on the neo-Y chromosome are plotted ($n = 198$). Source data are provided as a Source data file.

expression in *D. miranda* males. Notably, while TEs near zygotically expressed genes show reduced H3K9me3 levels, they are nevertheless enriched for heterochromatin, suggesting some extent of epigenetic silencing of TEs. The deposition and spreading of silencing heterochromatin at TEs near euchromatic genes is deleterious, and euchromatic TE insertions are under purifying selection in *Drosophila*²⁶. The close proximity of thousands of genes and TEs on the neo-Y therefore results in an epigenetic conflict between silencing TEs via heterochromatin formation while maintaining the activity of functionally important genes during the development.

Increased rates of TE insertions in *D. miranda* males. Elevated TE expression could lead to increased rates of TE insertions. To test if differences in TE activity result in sex-specific differences in TE movement, we identified novel TE insertions in male and female embryos by deep sequencing, leveraging the input DNA reads from our single-embryo ChIP-seq data. Insertions were defined by paired-end reads, where one read mapped uniquely to the genome and the other to a TE^{27,28} (Fig. 5a). To avoid capturing chimeric reads, we required that both the 5' and 3' junctions of insertions were identified (Fig. 5a, b), and TEs found in more than one sample are discarded to ensure only novel insertions are counted. As any de novo insertion is likely found in a tiny fraction of cells, our approach is a highly conservative estimate of the number of total insertion events (Supplementary Data 3). To ensure that our approach is robust to artifacts, we tested it using exon sequences instead of repeats; indeed, no more than four “insertions” are called across any library (and a median of 0 “insertions” per library; Supplementary Data 4).

We identified a total of 1054 and 8191 insertions across 37 *D. pseudoobscura* and 42 *D. miranda* single-embryo libraries (Supplementary Data 3). The majority of these insertions are likely somatic as the number of somatic cells are at least three orders of magnitude more numerous than the pole (germ) cells at these embryonic stages. As expected, the number of TE insertions identified from the embryos is strongly dependent on the sequencing depth (Fig. 5c, e). For each of the two species, we fitted an ANOVA model with median coverage, developmental stage, and sex as independent variables, and number of insertions as the response variable. While library coverage has the strongest effect in both species, we find a significant effect of sex in *D. miranda*, but not in *D. pseudoobscura* (Supplementary Table 1). In addition, developmental stage also shows a strong effect in both species reflecting the increase in TE activities through the MZ transition, and the increasing number of cells where insertions can occur. After removing the effect of library coverage, we observed significantly more TE insertions in males than in females for stages 5 and 7 embryos in *D. miranda*, but not

D. pseudoobscura (Fig. 5d, f and Supplementary Fig. 7). Increased rates of TE insertions in *D. miranda* males are also observed when considering only autosomal insertions (Supplementary Fig. 8). The magnitude of difference is greater in stage 7 embryos, likely due to the increased amount of time and cells in which de novo mutations can occur (Fig. 5f). Female and male embryos at stage 4 have similarly low numbers of insertions (Fig. 5f), consistent with the absence of male-biased TE expression prior to the MZ transition. As expected, de novo TE insertions resulted from repeats that show higher expression in early embryos²⁹, and the number of insertions summed across all embryos of the same sex and stage is significantly correlated with their transcript abundance (Fig. 5g). Altogether, these results reveal that elevated TE expression in *D. miranda* males is associated with higher rates of TE insertions in males compared to females. Interestingly, we find that insertions are nonrandomly distributed across chromosomes. In both males and females, there are significantly fewer autosomal insertions than expected based on the chromosomal sizes (Fisher's exact test, $p < 0.0001$; Fig. 5f). TE insertions are significantly overrepresented on Muller-AD in females and the neo-Y in males (FET, $p < 0.0001$ for both; Fig. 5f), and either transposition bias or selection could contribute to the nonrandom spatial distribution of TEs³⁰. The elevated rate of somatic TE insertions in *D. miranda* males is expected to impose a deleterious fitness cost unique to males, and may contribute to the female-biased sex ratio found in this species³¹, and shorter lifespan of *D. miranda* males compared to females³². In addition, if insertions rates are also higher in the male germline, the species as a whole is expected to have a higher mutational TE burden. Concordantly, the *D. miranda* genome overall has higher repeat content than its close relative *D. pseudoobscura*, even outside of its neo-sex chromosomes¹².

Model for neo-Y toxicity and TE accumulation. Y chromosomes have evolved repeatedly in different species from a pair of ordinary autosomes. Y evolution is typically characterized by progressive gene loss, an accumulation of repetitive DNA, and heterochromatin formation³³. Here, we show that nascent Y chromosomes can form a “genomic liability” for males, especially if epigenomic defense mechanisms are compromised. This occurs in the early development when the zygotic genome is reprogrammed to create a set of totipotent cells capable of generating a new organism. Heterochromatin loss also occurs in old individuals, and Y chromosomes can contribute to faster male aging in *Drosophila*³⁴. We show that incomplete silencing of Y-linked TEs in early development results in a surge of repeat expression in males, resulting in more somatic TE insertions. Repeat-rich Y chromosomes that still contain functional genes create a dilemma, as actively transcribed euchromatin antagonizes heterochromatin assembly²³. Thus, competition

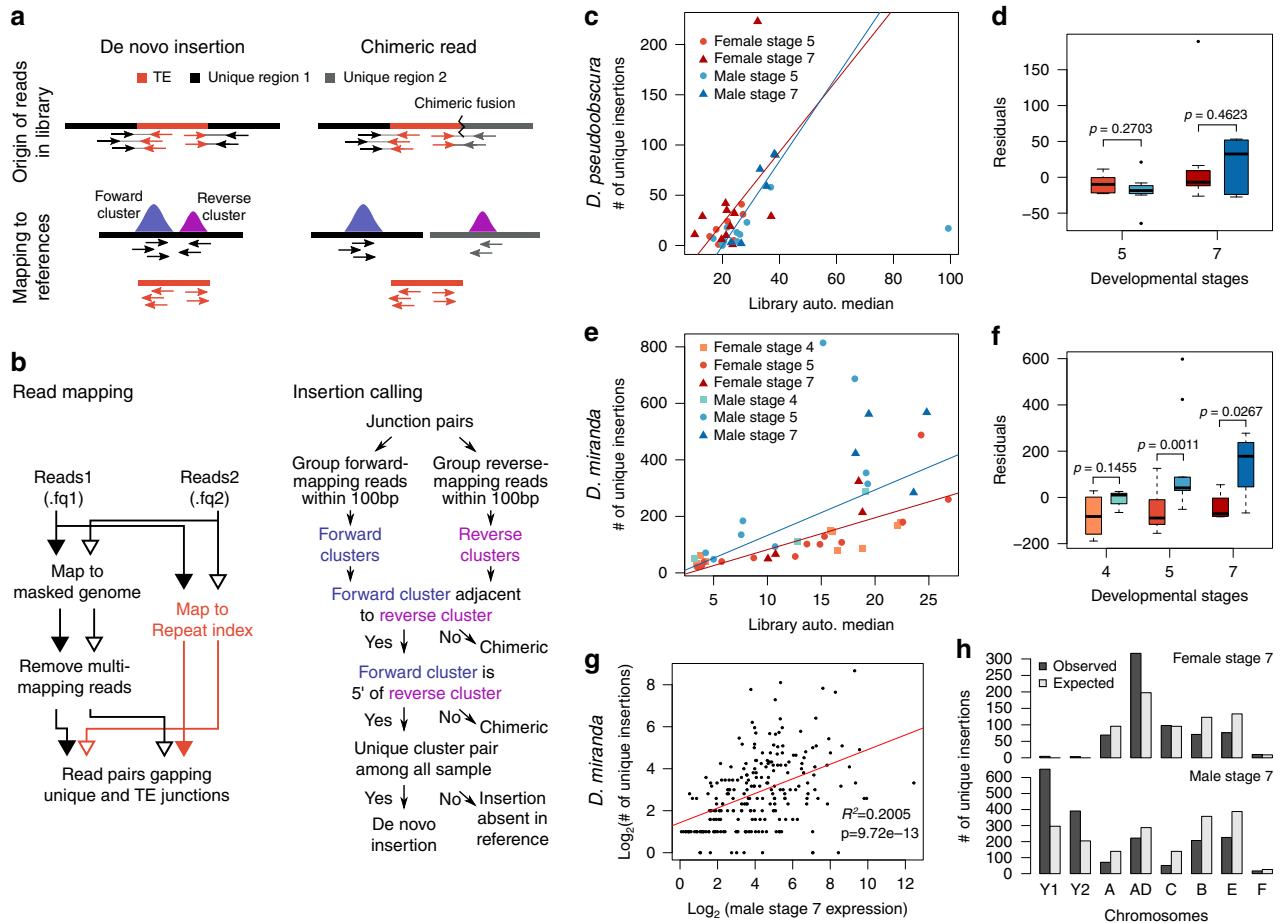


Fig. 5 *D. miranda* males show more de novo TE insertions than females. **a** Schematic diagram of rationale behind TE insertion identification. **b** Computational pipeline implementing the strategy in **a**. The left panel outlines the mapping strategy to identify mate pairs at insertion junctions. The right panel outlines the filtering strategy to identify de novo TE insertions. See Supplementary Data 3 for summary of insertions filtered and identified for each library. **c** The numbers of de novo TE insertions identified in *D. pseudoobscura* embryos are plotted against the autosomal median coverages. Lines of best fit are shown for the male (blues) and female (reds) samples, separately. **d** The effect of library size is removed by regression of TE insertion numbers against library coverage and taking the residuals. Boxplot depicts the residuals of the regression binned by sex and developmental stages. *P* values are labeled above the comparisons (two-tailed Wilcoxon rank-sum test, no multiple testing correction). In the boxplots, whiskers delineate the minima and maxima of the distributions, the box delineates the 25th and 75th percentiles, and the centers mark the medians. **e, f** Same as **c** and **d**, but for *D. miranda* embryos. **g** The number of de novo insertions across all male *D. miranda* stage 7 embryos is plotted against the transcript abundance. Line of best fit is in red and R^2 of the regression and the significance of the correlation are shown. Only TEs with de novo insertions are plotted. **h** The observed and expected distribution of insertions across the chromosomes are depicted in females (top) and males (bottom). Source data are provided as a Source data file.

between the opposing mechanisms of heterochromatin formation and genic transcription likely explains the incomplete silencing of TEs on the transcriptionally active neo-Y in *D. miranda*. While the accumulation of repetitive elements on the Y chromosome appears to be near universal during sex chromosome evolution, our study reveals an unappreciated aspect of this process (Fig. 6). The conflict between genic expression and TE silencing on a nascent Y creates a “toxic environment”, elevating the mutational burden in the male genome. This discord is maximized on Y chromosomes of intermediate evolutionary age that still contain an appreciable number of genes, but also a high repeat density. Resolution of this conflict may select for the adaptive degeneration of remaining protein-coding genes on the Y, and further repeat accumulation to strengthen epigenetic silencing, thereby reducing the toxicity of the Y. Epigenetic conflicts therefore represent a novel mechanism driving the degeneration of the Y chromosome.

Methods

Fly strains. We used *D. pseudoobscura* strain SS-R2 and *D. miranda* strain MSH22 kept at 18 °C (the preferred temperature for these species and the same temperature

they were reared when the RNA-seq data and ChIP-seq data were generated) and *D. melanogaster* strain Oregon-R kept at 25 °C.

RNA-seq data analysis. We used published RNA-seq data¹⁹ to analyze sex-specific repeat expression during embryogenesis. For read counts at genes, pair-end reads were mapped to the *D. pseudoobscura*¹² or *D. miranda*¹⁴ genomes using bwa mem (v0.7.15)³⁵ on default settings, and sorted with samtools (v1.5)³⁶. We then used featureCounts (v1.6.2) from the Subread package³⁷ to determine the number of reads mapping to annotated genes of the two genomes. For counts at TEs, we mapped reads with bwa mem to the TE library specific to the *D. pseudoobscura* species subgroup generated by ref. 11. Reads mapping to each repeat entry were then tallied from the sam files. We also used bowtie2 (v2.3.4.1), which generated similar patterns in sex-biased expression, indicating that our results are robust to mapping strategies (Supplementary Fig. 9). We normalized the gene and TE read counts by the median read counts at autosomal genes to avoid the large contribution of sex-specific expression from the sex chromosome, especially the neo-Y. After normalization, one pseudocount is added to each gene. All hierarchical clustering and correlation procedures were conducted on the log₂-transformed read counts. For sample information and mapping statistics, see Supplementary Data 1.

Chromatin immunoprecipitation and sequencing. We performed ChIP-seq using a protocol adapted from ref. 38. *D. pseudoobscura* data were newly collected, and *D. miranda* data were downloaded from the SRA under BioProject PRJNA601450.

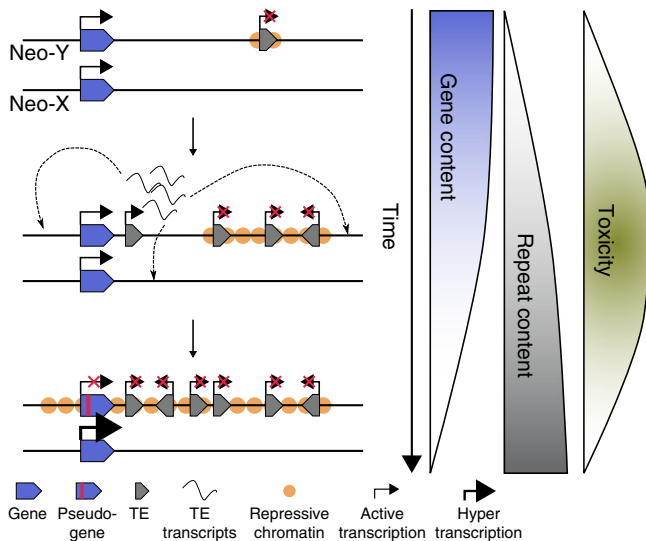


Fig. 6 Toxic Y chromosome and adaptive Y degeneration. Epigenetic conflicts between host-specific genes (blue) and selfish genetic elements (gray) are maximized on Y chromosomes of intermediate age (maximum toxicity). Active transcription of Y-linked genes impedes heterochromatin formation at repeats that accumulate on degenerating Y chromosomes and their mobilization can create a mutational burden in males (middle). Selective pressure to efficiently silence a toxic Y chromosome and dosage compensation of the X homologs could drive adaptive degeneration of Y genes.

Briefly, single embryos were homogenized with a pipette tip, and chromatin was digested for 7.5 min at 21 °C using micrococcal nuclease (MNase; New England Biolabs). We spiked DNA from *D. pseudoobscura* single embryos with DNA from *D. melanogaster* stage 7 (gastrulation) embryos so that each sample had 20% spike (*D. melanogaster*) DNA (i.e., one *D. melanogaster* embryo was used for four *D. pseudoobscura* embryos). We set aside 10% of each sample as input and incubated the remaining chromatin with Dynabeads Protein G (Invitrogen) for 2–6 h. The H3K9me3 antibody (Diagenode, 1.65 µg/ul) was incubated for >3 h with Dynabeads Protein G to bind the antibody to the beads, before adding it to the chromatin (0.25 µl per embryo) for overnight incubation. The chromatin–antibody–bead complexes were washed first with low-salt buffer, then with high-salt buffer. DNA was eluted from the chromatin–antibody–bead complexes by shaking at 65 °C for 1–1.5 h. We extracted DNA from our ChIP samples and from our input using a phenol/chloroform/isoamyl alcohol mixture, then cleaned the DNA with Agencourt AmpureXP beads before library preparation. Libraries were prepared using the ThruPLEX DNA-seq kit (Rubicon) followed by two rounds of cleaning with AmpureXP beads. A total of 100 bp paired-end sequencing of our samples was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

ChIP-seq data processing and normalization. ChIP-seq libraries for *D. miranda* were downloaded from the SRA under BioProject PRJNA601450. Like the *D. pseudoobscura* libraries they were spiked with *D. melanogaster* embryos. To distinguish the spike-in reads, pair-end reads were aligned using bwa mem to a concatenated reference combining the *D. melanogaster* genome (r6.12), with either the *D. pseudoobscura* or *D. miranda* genomes. Since bwa mem by default only reports the best alignment, spike-in reads are then extracted as those mapping to the *D. melanogaster* contigs. We used bedtool's genomeCoverageBed³⁹ to obtain per base coverage of the sample, and spike-in reads after sorting with samtools sort. For library and mapping information see Supplementary Data 2.

For the genome-wide coverage H3K9me3 enrichment in both spike-in and actual samples, enrichment at 50 kb window is calculated as:

$$\frac{\text{No. of ChIP reads}/\text{Median autosomal coverage of ChIP library}}{\text{No. of input reads}/\text{Median autosomal coverage of input library}}$$

For spike-in normalization, we first generated an enrichment “reference” by averaging the enrichment of all spike-ins. Since all spike-ins should have the same enrichment, systematic differences between spike-ins are predominantly the result of different antibody pulldown efficiencies during library preparation. We used a quantile normalization procedure, matching the distribution of the enrichment values to that of the reference. The extent of change for each enrichment value after the normalization is then applied to windows of the same enrichment value in the actual sample. In effect, if windows with enrichment of 1.5 in the spike-in is

increased to 2 after quantile normalization, windows with enrichment of 1.5 in the actual sample will be increased to 2. For details on the procedure, see Supplementary Fig. 10.

H3K9me3 enrichment analysis. We used the R package IRanges⁴⁰ to infer overlaps and distances between windows with elevated H3K9me3, genes, and TEs. H3K9me3 enrichment around genes were inferred by extracting the per bp enrichment 5000 bp upstream and downstream of the TSS of the relevant genes; the median enrichment at each position is plotted. For TEs, enrichments were extracted 5000 bp upstream and downstream of the midpoint of the relevant TE annotation.

TE annotation and repeat masking. We annotated the genomes of the two species and masked the repeats with the repeat library using RepeatMasker⁴¹ (v3.3) with the following command:

```
RepeatMasker -norna -nolow -dir output.directory -gff -u -lib TE.library.fasta genome.fa
```

Identification of de novo TE insertions. Pair-end reads of inputs from the ChIP-seq experiments are mapped as single ends to their respective repeat-masked genomes and the repeat library. We first identified read pairs where one maps uniquely to the genome and the other maps only to the TE library, thus identifying pairs that flank insertion junctions. Reads of the same mapping orientation (forward or reverse strand mapping) that map <100 bp of each other in the genome are deemed to be capturing the same junction. Because fusion of DNA fragments during library preparation can generate chimeric reads, which can create such TE-to-unique junctions, we conservatively estimated insertions by further requiring that both the 5' and 3' junctions of the insertions must be identified. Thus, an insertion is only called if a forward mapping junction at the 5' is followed by a reverse mapping junction <100 bp away at the 3'. Note, since TEs can insert in either direction, we do not stipulate any directionality to reads mapping within the TEs. For true de novo insertions in the embryos, we removed all insertions that are found within < 50 bp of any other insertion found across all samples.

With each of the two species, we used the following ANOVA model in R:

$$\text{No. of insertions} \sim \text{coverage} + \text{sex} + \text{developmental stage},$$

with coverage being the median autosomal coverage. For ANOVA summary statistics, see Supplementary Table 1. To remove the effect of sequence depth, we used the linear regression model:

$$\text{No. of insertions} \sim \text{coverage}.$$

The residuals are then used to compare the difference in the amount of insertions due to sex and developmental stage. We calculated the expected number of TE insertions per chromosome by assuming uniform insertion rates; we multiplied the number of observed insertions genome wide by the size of the chromosomes (in the genome assembly) proportional to the diploid genome of females and males.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All the sequencing data have been posted on GenBank under BioProject PRJNA625074. All processed data files have been posted on Dryad: <https://doi.org/10.6078/D1B12G>. Source data are provided with this paper.

Code availability

Scripts used for normalization and insertion calls can be found on KW's github page <https://github.com/weikevinhc/heterochromatin>⁴².

Received: 7 May 2020; Accepted: 24 September 2020;

Published online: 02 November 2020

References

- Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *Elife* **3**, e1003428 (2014).
- Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).
- Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696 (1982).

4. Haig, D. Transposable elements: self-seekers of the germline, team-players of the soma. *Bioessays* **38**, 1158–1166 (2016).
5. Padeken, J., Zeller, P. & Gasser, S. M. Repeat DNA in genome organization and stability. *Curr. Opin. Genet. Dev.* **31**, 12–19 (2015).
6. Hedges, D. J. & Deininger, P. L. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* **616**, 46–59 (2007).
7. Elgin, S. C. R. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5**, a017780 (2013).
8. Girton, J. R. & Johansen, K. M. Chromatin structure and the regulation of gene expression: the lessons of PEV. *Drosophila* **61**, 1–43 (2008).
9. Bourque, G. et al. Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
10. Brown, E. J., Nguyen, A. H. & Bachtrog, D. The *Drosophila* Y chromosome affects heterochromatin integrity genome-wide. *Mol. Biol. Evol.* **37**, 2808–2824 (2020).
11. Hill, T. & Betancourt, A. J. Extensive exchange of transposable elements in the *Drosophila pseudoobscura* group. *Mob. DNA* **9**, 20 (2018).
12. Bracewell, R., Chatla, K., Nalley, M. J. & Bachtrog, D. Dynamic turnover of centromeres drives karyotype evolution in *Drosophila*. *Elife* **8**, 923 (2019).
13. Bachtrog, D. & Charlesworth, B. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323–326 (2002).
14. Mahajan, S., Wei, K. H.-C., Nalley, M. J., Gibilisco, L. & Bachtrog, D. De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol.* **16**, e2006348 (2018).
15. Lécuyer, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007).
16. Pritchard, D. K. & Schubiger, G. Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes Dev.* **10**, 1131–1142 (1996).
17. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
18. Yuan, K. & O'Farrell, P. H. TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in *Drosophila* embryos. *Genes Dev.* **30**, 579–593 (2016).
19. Lott, S. E., Villalta, J. E., Zhou, Q., Bachtrog, D. & Eisen, M. B. Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. *PLoS Genet.* **10**, e1004159 (2014).
20. Vlassova, I. E., Graphodatsky, A. S., Belyaeva, E. S. & Zhimulev, I. F. Constitutive heterochromatin in early embryogenesis of *Drosophila melanogaster*. *Mol. Gen. Genet.* **229**, 316–318 (1991).
21. Lu, B. Y., Ma, J. & Eissenberg, J. C. Developmental regulation of heterochromatin-mediated gene silencing in *Drosophila*. *Development* **125**, 2223–2234 (1998).
22. Bachtrog, D., Mahajan, S. & Bracewell, R. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nat. Ecol. Evol.* **3**, 1587–1597 (2019).
23. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
24. Leach, T. J., Chotkowski, H. L., Wotring, M. G., Dilwith, R. L. & Glaser, R. L. Replication of heterochromatin and structure of polytene chromosomes. *Mol. Cell Biol.* **20**, 6308–6316 (2000).
25. Macknight, R. H. & COOPER, K. W. The synapsis of the sex chromosomes of *Drosophila miranda* in relation to their directed segregation. *Proc. Natl Acad. Sci. USA* **30**, 384–387 (1944).
26. Lee, Y. C. G. The role of piRNA-Mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. *PLoS Genet* **11**, e1005269 (2015).
27. Jiang, C., Chen, C., Huang, Z., Liu, R. & Verdier, J. ITIS, a bioinformatics tool for accurate identification of transposon insertion sites using next-generation sequencing data. *BMC Bioinformatics* **16**, 72 (2015).
28. Treiber, C. D. & Waddell, S. Resolving the prevalence of somatic transposition in *Drosophila*. *Elife* **6**, 2185 (2017).
29. Charlesworth, B. & Langley, C. H. The evolution of self-regulated transposition of transposable elements. *Genetics* **112**, 359–383 (1986).
30. Bousios, A., Nützmann, H.-W., Buck, D. & Michieletto, D. Integrating transposable elements in the 3D genome. *Mob. DNA* **11**, 8–10 (2020).
31. Dobzhansky, T. *Drosophila miranda*, a new species. *Genetics* **20**, 377–391 (1935).
32. Nguyen, A. H. & Bachtrog, D. Increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome. <https://doi.org/10.1101/2020.07.21.214528> (2020).
33. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
34. Brown, E. J., Nguyen, A. H. & Bachtrog, D. The Y chromosome may contribute to sex-specific aging in *Drosophila*. *Nat. Ecol. Evol.* **4**, 853–862 (2020).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
38. Brind'Amour, J. et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, 6033 (2015).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
40. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
41. Tempel, S. Using and understanding RepeatMasker. *Methods Mol. Biol.* **859**, 29–51 (2012).
42. Wei K. H.-C. Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *weikevinhc/heterochromatin: Heterochromatin scripts v1.0*. <https://doi.org/10.5281/zenodo.4033521> (2020).

Acknowledgements

This work was supported by NIH grants (nos. R01GM076007, R01GM101255, and R01AG057029) to D.B. Publication made possible in part by support from the Berkeley Research Impact Initiative (BRII) sponsored by the UC Berkeley Library.

Author contributions

D.B. conceived the study. D.B., L.G., and K.H.-C.W. designed the analyses. L.G. collected the samples and generated the sequence data. K.H.-C.W. developed the analytical tools and pipelines. K.H.-C.W. and L.G. analyzed the data. D.B. and K.H.-C.W. prepared the manuscript and addressed reviewer comments. D.B. acquired the funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19134-9>.

Correspondence and requests for materials should be addressed to D.B.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons

Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020