Dynamic sex chromosome expression in *Drosophila* male germ cells

Sharvani Mahadevaraju1,10, Justin M. Fear1,10, Miriam Akeju2,10, Brian J. Galletta3, Mara M. L. S. Pinheiro4, Camila C. Avelino4, Diogo C. Cabral-de-Mello5, Katie Conlon2, Stafania Dell’Orso6, Zelalem Demere2, Kush Mansuria2, Carolina A. Mendonça4, Octavio M. Palacios-Gimenez4,8, Eli Ross2, Max Savery1, Kevin Yu2, Harold E. Smith7, Vittorio Sartorelli6, Haiwang Yang1,9, Nasser M. Rusan3, Maria D. Vibranovski4,11, Erika Matunis2,11 & Brian Oliver1,11

Given their copy number differences and unique modes of inheritance, the evolved gene content and expression of sex chromosomes is unusual. In many organisms the X and Y chromosomes are inactivated in spermatocytes, possibly as a defense mechanism against insertions into unpaired chromatin. In addition to current sex chromosomes, *Drosophila* has a small gene-poor X-chromosome relic (4th) that re-acquired autosomal status. Here we use single cell RNA-Seq on fly larvae to demonstrate that the single X and pair of 4th chromosomes are specifically inactivated in primary spermatocytes, based on measuring all genes or a set of broadly expressed genes in testis we identified. In contrast, genes on the single Y chromosome become maximally active in primary spermatocytes. Reduced X transcript levels are due to failed activation of RNA-Polymerase-II by phosphorylation of Serine 2 and 5.
O

rganisms as diverse as flies and humans have XY males and XX females. In *Drosophila*, the Y chromosome is large, heterochromatin-rich, and gene poor. It carries genes essential for fertility, but not viability.\(^1,2\). The X chromosome is present as a single copy in males. The small 4th chromosome is an ancestral X chromosome\(^3,4\), that is now distomic. Given the sex chromosome copy number differences and their unique modes of inheritance, the unusual evolved gene content and expression of sex chromosomes is widely studied\(^5\). In many organisms, including mammals\(^6, C. elegans\), and *Drosophila*, X expression is reduced in testes. Non-exclusively unique reasons for this reduction include: inactivation of unpaired chromosomes\(^9,18\), absence of upregulation of the X chromosome in the germline (canonical *Drosophila* dosage compensation)\(^11\), and sexually antagonistic selection followed by evolutionary re-localization of genes required in males from the X chromosome to the autosomes\(^8,12-14\). Some combination of these models is possible, if not likely.

Meiotic X-chromosome inactivation is well described in mammals, where premature transcriptional inactivation of the X and Y chromosome during mid-spermatogenesis is essential for fertility\(^15\). The X and Y precisely condense into a heterochromatic “XY body” within the pachytene spermatocyte nucleus. Transcriptional inactivation is thought to protect these largely dissimilar chromosomes by defending un-synapsed regions from stealthy invasive transposons that cannot be detected as foreign in the absence of a homolog\(^16\). Inactivation may also protect against unwanted recombination between the X and Y, or to repair damage created by lack of recombination\(^17\). Inactivation might also mark paternal X chromosomes for later imprinting for inactivation\(^18,19\). Some of these models are unlikely in the case of *Drosophila* males, which completely lack meiotic recombination\(^20\). Exploiting differences, and commonalities, between *Drosophila* and other models can illuminate causes of unusual sex chromosome expression patterns.

In many organisms, the potentially lethal imbalance of X-linked gene transcripts from the single X relative to the paired autosomes must be alleviated in the soma to ensure viability\(^21\). In *Drosophila* somatic cells, X chromosome dosage compensation is essential and involves large-scale transcriptional upregulation of the single X chromosome\(^22\). It is unclear if X chromosome dosage compensation exists in *Drosophila* germ cells, which do not require the core dosage compensation machinery used in the soma\(^23,24\). If dosage compensation occurs, it relies on a undiscovered mechanism. Analysis of transcripts expressed in portions of hand-dissected testes\(^25,26\) or in mutants that block spermatogenesis in mitotic stages\(^13,21,27\) suggest that *Drosophila* testes enriched in spermatogonia show upregulation of the X, consistent with dosage compensation in pre-meiotic cells. Portions of adult testes enriched in late stage spermatocytes\(^25\) show greatly reduced X chromosome expression and autosomal genes inserted on the X show dramatically reduced expression in spermatocytes, far exceeding simple failure of dosage compensation\(^23,24,28\), consistent with X inactivation in primary spermatocytes.

Given the dramatic differences in the gonads and gametes between the sexes, the optimal male and the optimal female genome will differ. For autosomes, which reside in each sex in equal dose, selection is balanced. In stark contrast, sex chromosome residency is not balanced. In a population with equal numbers of males and females, 2/3rds of X chromosomes reside in females. The X chromosome residency profile is expected to result in more opportunities for selection of alleles favoring females. The Y, of course, resides only in males and is under selection only in males. The presence of a homolog is also important. The single X and single Y chromosomes in males are under immediate selection, while only alleles with some degree of dominance are immediately selected in females. Assuming that there is at least subtle dominance\(^29\), then the X chromosome should be both feminized and demasculinized (alleles with female advantage selected for, and alleles with male advantage selected against), and the Y should be both masculinized and defeminized\(^3,11\). These patterns have been observed in *Drosophila* species, where expression from the X chromosome is reduced, and where genes required in males have evolutionarily relocated to other chromosomes\(^12\). Evolutionary arguments for sex chromosome gene content and expression present an interesting causality dilemma. In the model, antagonistic selection for female functions on the X drives removal of genes that males need for development from the X. Removal of those genes is permissive for events such as X inactivation in the male germline\(^18\). It then follows that X inactivation in the male germline would provide even more selective pressure against X genes with male-biased functions. In this work, we show that sex chromosome expression is dynamic at tissue and single cell resolutions.

**Results**

**Sex chromosome expression in different tissues.** *Drosophila* have X and Y sex chromosomes, two major autosome pairs and a pair of “dot” 4th chromosomes (Fig. 1a). The Y and 4th chromosomes are gene poor, while the remaining chromosome arms are gene rich (Fig. 1b). To examine sex-biased gene expression patterns, we focused on the distribution of male-biased gene expression across chromosomes or chromosome arms (chromosome elements) for each tissue. We measured adult gene expression (quadruplicates) in the whole body (Fig. 1c) as well as seven tissues (Fig. 1d–j): head, thorax (viscera removed), abdomen (viscera and all reproductive organs removed), viscera (including digestive and excretory organs), reproductive tract (gonads and genitalia removed), terminalia (including genitalia and analia), and gonads in females and males from two strains. We found a significant deviation from random distributions of sex-biased gene expression (we use \( p \leq 0.01 \) throughout this study) in whole body, head, thorax, viscera, reproductive tract, and gonad (\( \chi^2 \) test of independence, Supplementary Dataset 1). Sex chromosomes and former sex chromosomes are the major contributors to the non-randomness.

Overall gene expression of the sex chromosomes varied by tissue. For X-chromosomes, we observed under-expression of genes in the whole body and gonads of males from either of two wildtype strains (Fig. 1c, j), as previously reported\(^6\). We also observed reduced X-chromosome expression in heads, thorax, reproductive tract, and terminalia (Fig. 1d, e, h). These non-gonadal patterns of X chromosome expression are difficult to explain by absence of germline X chromosome dosage compensation or meiotic sex chromosome inactivation, since there are no germ cells in those tissue. By elimination, this suggests that sexual selection drives gene expression patterns of X-chromosome expression in many somatic cell types. Under-representation in gonads could be explained in full or part by the absence of dosage compensation or meiotic sex chromosome inactivation.

Males with no Y chromosome are viable, but sterile and the Y chromosome is known to be expressed in spermatocytes\(^30\). However, the tissue-specific Y chromosome gene expression pattern is poorly described. We report that Y-chromosome gene expression was detectable only in whole males and gonads (Fig. 1c, j).

The 4th chromosome showed a decrease in gene expression in male whole bodies, reproductive tracts, and gonads (Fig. 1c, h, j). As a former X chromosome, 4th chromosome expression in the gonads mirrored the X-chromosome underrepresentation of male-biased gene expression. Additionally, and unlike the X chromosome, 4th chromosomes are present in two copies in
males. Because there are two copies of the 4th chromosome genes, under-representation of expression cannot be explained by the absence of dosage compensation. These data again suggest that there has been sexual selection of sex chromosomes, but only gonads show sex-biased expression of the \textit{Drosophila} X, Y, and 4th chromosomes.

Given the dynamic expression of the sex chromosomes by tissue, it is reasonable to assume that there could be differences by cell type and stage within the gonad. Examination of X-linked gene expression from hand-dissected anterior, middle, and posterior regions of adult testes shows a dynamic pattern of X chromosome expression that decreases as germ cells mature\textsuperscript{25}.
We decided to use single cell RNA sequencing (scRNA-seq) for a higher resolution picture. We did not want to use read depth to sequence transcriptionally inactive secondary spermatocytes and sperm, so we selected Drosophila third instar larval (L3) testis for our experiments. They contain abundant germ cells, including the critical transition from mitotic spermatogonia to meiotic primary spermatocytes. Prior to using L3 testes for single cell analysis, we asked whether the patterns of chromosome-specific gene expression known to occur in adult gonads are also present at this earlier stage of gonadogenesis. We performed RNA-Seq on eight biological replicates each of cleaned 3rd instar larval testes and ovaries (Supplementary Fig. 1) to compare with adult tissues. The overall profiles, including X, Y and chromosome expression are similar in both adult and larval gonads (Fig. 1j, k).

**Single cell resolution gene expression in the testis.** Determining which cells of the testis contribute to unusual patterns of sex chromosome expression is challenging because of the cellular diversity and dynamic nature of gene expression in the gonad. Drosophila spermatogenesis (Fig. 2a) begins at the apical end of the testis where germline stem cells divide asymmetrically, generating mitotically active differentiating daughters called gonialblasts that produce clusters of interconnected spermatogonia (G). After their fourth mitotic division, spermatogonia quickly enter pre-meiotic S-phase, becoming early primary spermatocytes (E1°) that are initially morphologically indistinguishable from spermatogonia. As early primary spermatocytes traverse an extended G2 phase, they undergo a burst of transcriptional activity that is accompanied by a 25-fold increase in volume, becoming large middle primary spermatocytes (M1°) and then slightly smaller late primary spermatocytes (L1°). Fully mature late primary spermatocytes progress through two rapid meiotic divisions, becoming spermatids. Pairs of quiescent somatic cyst cells (C1–C4) envelope each gonialblast and descendants, differentiating alongside the germ cells they support. Cyst cells are a source of intercellular signals and act as a permeability barrier. The entire L3 testis is encapsulated by an epithelial monolayer of pigment cell precursors (P), which become adult pigment cells and are required for joining the gonad with the reproductive tract. The basal end of the L3 testis contains terminal epithelial precursor cells (T) which will ultimately contribute to maturation in the final steps of spermiogenesis.

To capture the transcriptional profiles of the cell types in the Drosophila L3 testis, we dissected staged male third instar larvae, enzymatically removing the associated fat body before dissociating 20–40 testes to a single cell suspension for scRNA-seq ("Methods", Supplementary Fig. 1a–d). We piloted dissociation conditions using fly strains expressing GFP in subsets of cells (Supplementary Fig. 1g, h) coupled with dye exclusion (Supplementary Fig. 1i) (70–75% viability, n = 1000/replicate), ensuring consistent retrieval of single cell suspensions of viable germline and somatic cells. We then performed scRNA-seq on wild-type L3 testes. We identified 18,965 single cells across three biological replicates based on the intersection of cells from cell range court and DropletUtils emptyDrops (Supplementary Dataset 2). Potential cell doublets were detected using scrublet and removed. Based on preliminary cluster analysis using the 2000 most variably expressed genes, we set the perplexity threshold in Seurat to 0.3. This yielded ten clusters, each potentially representing a distinct cell type or state with each of three biological replicates contributing to the clusters (Spearman expression rank correlation ≥ 0.93, p < 0.01, Fig. 2b). While the replicates show the absence of large effects, systematic errors in reporting the biology were still possible. If we captured the majority of the cells and cell types, then we should observe a strong correlation between RNA-seq from the whole organ and the total of all single cells. Indeed, the correlation between replicates bulk RNA-Seq from whole L3 testes and sum of single cells from L3 testes was significant (Fig. 2c; Supplementary Dataset 3), indicating that major cell types are well represented in our scRNA-seq dataset. Gene Set Enrichment Analysis GSEA on the ten clusters indicated that genes with male-biased expression in whole gonads are also enriched in germ cells relative to somatic cells in the L3 gonad (compare red and blue plots, Fig. 2d). Thus, the germine is the major contributor to the pattern of sex-biased expression of the X, Y, and 4th chromosomes.

To align gene expression patterns with cell types, we first identified genes with preferential expression in particular cell types, then compared the expression profiles of cells in each of the ten groups with that of all remaining cells in an iterative process for each cell cluster (Fig. 2e, Supplementary Dataset 2). To determine what the cell types were, we projected genes with well-known expression patterns, carefully re-curated from published images (Supplementary Dataset 4), onto the cell level expression profiles. We also used transgenic protein-trap reporter genes to identify expression patterns for genes with enriched expression in each cell type (Fig. 2f–k, Supplementary Fig. 3, Supplementary Dataset 4). This allowed us to annotate all ten clusters. 91% of 79 genes from the literature or reporter expression patterns overlapped cell types identified by scRNA-seq. 46% were expressed in the same or developmentally preceding cell types, and an additional 45% were expressed in the same cell lineage. The 7% of genes showing discordant assay-dependent expression patterns were protein-trap reporters showing either weak or widespread expression, both of which hinder assignment to specific cell types (Supplementary Fig. 3, Supplementary Dataset 4). These data indicate that we have identified many of the major cell types in the L3 testis. We did not identify previously unidentified cell types and failed to unambiguously identify cells in the stem cell niche.

Spermatogonia (G) were characterized by the biased expression of 624 genes (Fig. 2e #1) including: known early germ cell markers argonaute 3 (AGO3), aubergine (aub), and vasa (vas) the early differentiation signal bag-of-marbles (bam), the mitotic cell marker p53 (Fig. 2f), the chromatin proteins...
Dek\textsuperscript{47} and enhancer of yellow 3 (e(y))\textsuperscript{3}\textsuperscript{48} (Fig. 2e #1). While expression of these spermatogonial genes also extended to spermatocytes, we could readily distinguish spermatogonia from spermatocytes by their lack of spermatocyte-specific gene expression. The earliest primary spermatocytes (E1\textsuperscript{o}) were characterized by high expression of 235 genes including three markers of the transition from spermatogonia to spermatocyte: always early (aly), spermatocyte arrest (sa)\textsuperscript{49}, and testis-specific expressed bromodomain containing protein-1 (tbrd-1)\textsuperscript{50} (Fig. 2e #3). Sa is a testis-specific TATA-binding protein associated factor (tTAF), Aly serves as a testis-specific meiotic arrest complex (tMAC) member\textsuperscript{49}, and tbrd-1 is associated with both transcriptional complexes\textsuperscript{51}. In addition, we show that ADD domain-containing protein 1 (Add1), which encodes a heterochromatin associated protein that interacts with Heterochromatin Protein 1 (HP1) to maintain heterochromatin\textsuperscript{52,53}, has enriched expression in early primary spermatocytes in the scRNA-seq data. The Add1 protein accumulated throughout spermatocyte stages, consistent with translational control and/or protein stability over several days of germ cell development (Fig. 2g). The tTAFS and tMACs...
Fig. 2 Identification and annotation of larval testis cell types from scRNA-seq and comparison to bulk testis RNA-Seq. a Illustration of the 3rd instar larval testis, showing germline cell types spermatogonia (G); early- (EP); middle- (EP); and late- (LP) primary spermatocytes; four groups of somatic cysts cells (C1–C4), terminal epithelium precursors (T) and pigment cells (P). These cell types are from germline or soma linages (dashed boxes). See abbreviation and color code key used throughout this manuscript. b Consistency of cell type expression profiles. Uniform Manifold Approximation and Projection (UMAP) for each scRNA-seq replicate. Each cell position in multi-dimensional space projected onto the two-dimensional UMAP space and clustered using k-nearest neighbors. c Correlation between whole testis and single cell profiles. Density scatterplot (high density, red and low density: blue); regression line and Spearman rank correlation, ρ, of gene expression from the sum of all cells from the larval testis scRNA-seq (x-axis) versus bulk larval testis RNA-Seq (y-axis). Transcripts Per Million reads ~1 (TPM ~1). d Distribution of genes showing testis-biased expression in bulk RNA-Seq using Gene Set Enrichment Scores (y-axis) among cell types (x-axis) defined by L3 testis scRNA-seq. Each bar is a cell type, so sample size is the number of cells in that cell type: G = 1065, E1 = 1267, M1 = 4570, L1 = 2667, C1 = 4064, C2 = 1230, C3 = 910, C4 = 945, T = 1379, P = 868. Boxplot lines (box = interquartile range (IQR), notch = 95% confidence interval of median, whiskers = ±1.5×IQR). e Patterns of differential gene expression among putative cell types. Only differentially expressed genes are shown (two-sided Wilcoxon rank-sum test, FDR corrected P ≤ 0.01). Gene expression is summarized (Z-score; low: blue, high: yellow) for each differentially expressed gene (rows) for all cell type-by-replicate (columns). Genes are ordered and grouped (horizontal dashed lines) manually into 16 classes (numbered) based on differential gene expression. Each class is annotated with the number of genes (parentheses) and the key genes used to identify the cell types (Supplementary Dataset 2). f–k Immunofluorescence images showing protein expression of representative genes from the protein trap reporters used for cell type annotation. f p53, g Add1; h bol i rdo j Ngf3, and k Piezo. Each panel (counter clockwise from top left) consists of color (Fas III, 1B1: red, DAPI: blue, and GFP: green) and grayscale (GFP) images of protein trap expression, normalized gene expression values (normalized to maximum expression) plotted by cell type (Each bar is a cell type, so sample size is the number of cells in that cell type: G = 1065, E1 = 1267, M1 = 4570, L1 = 2667, C1 = 4064, C2 = 1230, C3 = 910, C4 = 945, T = 1379, P = 868. Error bars are bootstrapped 95% confidence intervals around the mean), and scRNA-seq gene expression projected onto the UMAP (Z-score; low: blue; high: yellow; see key).

Dynamic and cell type specific sex chromosome expression. We looked at the dynamics of sex chromosome gene expression in germ cells in addition to all the other cell types from the single cell dataset (Fig. 3, Supplementary Dataset 5). The somatic cell X chromosome expression relative to autosomes hovered near 1.0 despite the 2-fold dose difference, a pattern consistent with the known canonical X chromosome dosage compensation mechanism in somatic cells, expected to increase expression from the single X22. This dosage compensation mechanism does not exist in germ cells. Nevertheless, spermatogonia showed similar levels of X chromosome expression relative to autosomes, providing evidence for non-canonical dosage compensation of the X chromosome in pre-meiotic germ cells. There was a significant decrease in expression of the X chromosome in early, middle, and late primary spermatocytes (M1° and L1°) relative to either spermatogonia or somatic cells. This decrease in X expression approached 2-fold which could be due to either a loss of dosage compensation in germ cells as they mature into primary spermatocytes, or to the gain of meiotic X-chromosome inactivation during the transition from mitotic spermatogonia to meiotic spermatocytes. At least some genes on mammalian X chromosomes are also overexpressed prior to inactivation65, raising the possibility that this dynamic transition in X chromosome expression in the male germline is conserved.

Expression of 4th chromosome genes parallels what was seen for the X (Fig. 3b). The expression ratio of the two 4th chromosomes relative to the two sets of major autosomes hovered near 1. There was a significant decrease in relative 4th chromosome expression in middle and late primary spermatocytes (M1° and L1°) compared to expression in either spermatogonia or somatic cells. Since we can rule out failed dosage compensation as a cause of 4th chromosome decreased expression, there must be a gain of inactivation during the developmental transition from mitotic spermatogonia to meiotic spermatocytes. This X chromosome like behavior may reflect the evolutionary history of the 4th chromosome; specifically, that the 4th retained X-inactivation after reacquiring autosomal status.

Finally, the entire pattern of Y chromosome gene expression is inverted relative to that seen for the X and 4th chromosomes (Fig. 3c). We observed poor expression of the Y in somatic cells and spermatogonia, and increased expression in primary spermatocytes (E1°, M1°, and L1°), concomitant with decreased expression of the X and 4th chromosomes. This occurs from expression of a 42 transcriptionally active Y-linked genes (Supplementary Dataset 2), consistent with the diffuse chromatin and Y-loops originally identified by cytology of primary spermatocytes66,67. The decrease in X and 4th chromosome
expression in middle and late primary spermatocytes does not reflect an overall decrease in total gene expression compared to somatic lineages, as genome-wide expression was robust in all cell types (Fig. 3d). Since the single X and the two 4th chromosomes showed reduced expression, while the single Y showed ongoing expression, these data demonstrate that there is no simple rule for dose-dependent sex chromosome gene expression in *Drosophila* primary spermatocytes. Genes on chromosomes with no homolog can go up (Y) or down (X), and the 4th appears to retain X-like reduced expression despite having a homolog. These differences in spermatocyte expression fully account for the differences in X, Y and 4th chromosome expression observed in bulk RNA-seq of larval testis.

Since genes with high expression in the testis are not uniformly distributed in the genome8,13, it was possible that the reduced expression of the X and 4th chromosomes was due to the absence of genes highly expressed in spermatocytes rather than a chromosome-wide reduction in expression due to a more global inactivation. A way to avoid this potential confounding effect, is to explore the expression of widely expressed “housekeeping” genes. We explored three data-driven methods to determine X and 4th chromosome expression of genes with housekeeping functions. In the first two methods, we used low tissue-specificity genes based on Tau and Tissue Specificity Score (TSPS) using our data68,69. The third method was a more granular low cell-type specificity metric within the scRNA-seq experiments (CTSP). Specifically, a set of widely expressed genes expressed in ≥33% of all cells. These methods reduced the expressed gene set numbers to varying degrees, with CTSP being the most stringent (Table 1).

The Y chromosome was expressed in an exquisitely tissue-specific matter and has no widely expressed genes using any metric. To determine if the functions of these three reduced gene sets are consistent with generic gene function, we systematically analyzed Gene ontology (GO) enrichment for all three subsets of genes (Fig. 4a; Supplementary Dataset 5). There are differences in function in the three gene sets. For example, in the Molecular ontology, all three CTSP gene sets, while regulators (which are less likely to be generic) were more enriched in the Tau gene set. In the Biological ontology, all three sets were enriched for protein metabolism, consistent with “housekeeping”, but the tissue-level Tau and TSPS gene sets were enriched for genes with development and female gamete functions, which is not commonly thought to be generic. Housekeeping genes are often highly expressed. All the reduced gene sets had higher median expression than all expressed genes,

### Table 1 Chromosome element genes and expression.

| Chromosome element | Annotated^a | Expressed^b | CTSP^c | Tau^d | TSPS^e |
|--------------------|-------------|-------------|--------|-------|--------|
| X                  | 2675        | 2207        | 65     | 336   | 1130   |
| Y                  | 113         | 42          | 0      | 0     | 0      |
| 2L                 | 3501        | 2767        | 126    | 367   | 1252   |
| 2R                 | 3628        | 2861        | 138    | 406   | 1413   |
| 3L                 | 3466        | 2867        | 101    | 371   | 1347   |
| 3R                 | 4201        | 3467        | 145    | 498   | 1673   |
| 4                  | 111         | 102         | 3      | 7     | 54     |

^aGenes annotated in Flybase 6.26, including non-coding genes.  
^bGenes detectably expressed in L3 scRNA-seq experiments.  
^cGenes expressed in ≥33% of cells in L3 scRNA-seq experiments.  
^dGenes expressed across multiple adult tissues (tissue specificity score r ≤ 0.5).  
^eGenes expressed across multiple adult tissues (tissue specificity score TSPS ≤ 1).
but elevated expression was most pronounced in the CTSP gene set (Fig. 4b–e). Additionally, the CTSP gene set showed greater uniformity in expression levels across cell types. Based on these results, we concluded that the CTSP gene set was the best subset for exploring expression of “housekeeping” genes.

We then used the reduced gene sets to examine the expression of the X and 4th chromosomes in all testis cell types. Importantly, when we examined relative expression, all three reduced gene sets showed significantly reduced X/A expression in germline cells (Fig. 4f–h). However, Tau and TPS gene sets showed reduced X/A expression in somatic cells (Fig. 4i–k).
expression in all cell types resulting in X/A ratios approaching 0.5 in both somatic and germline cells (Fig. 4f, g). At face value, failed dosage compensation might be expected to approach 0.5. These observations suggest a reason for the previous conclusion that there is no dosage compensation in male germline11 following the analysis of widely expressed genes using tissue-specificity scores. This conclusion is likely spurious, as testis somatic cells express the dosage compensation genes (Supplementary File 1) and the protein complex decorates the X in those somatic cells as occurs in X-chromosome dosage compensation in other somatic cells23. In contrast, the CTSP gene set showed reduced X chromosome expression, approaching 0.5, only in the late primary spermatocytes (Fig. 4h). Spermatogonia and somatic cells showed X/A rations approaching 1.0. Like the analysis of all expressed genes
Fig. 5 Spatial location of chromosomal territories in late primary spermatocyte nuclei. a Schematic model summarizing chromosomal spatial location in the spermatocyte nucleus and chromosome localization on chromosomes. There are two autosomal territories (A/A), X and 4/4 territory close to the nucleolus and Y chromosome. Experimental 1.688 satellite probe to detect X heterochromatin (red), AATAC satellite to detect Y heterochromatin (green), AATAT to detect 4th heterochromatin (cyan), oligo paints to detect X-euchromatin (yellow) and 2L (purple) are represented on the fly chromosomes (not to scale). b DAPI (gray) distinguishing three chromosomal territories where X and Y satellite probes clearly detects the chromosomes, respectively. Primary spermatocyte nuclei is outlined (white-dashed line) based on DAPI. Asterisk was used to stage individual cells by progression of centriole elongation (insets). c, d Spatial location of X and 4th chromosomes. e, g Representative samples showing spatial location of X and 2L by oligopaints. f, h Box plots showing the distributions of mean distances between the X (red), or 4th chromosomes (cyan), to the nucleolus (nucleolar). f, h Correlation of X and 4th chromosome territories (identical). The large sample size of cells was used to test a hypothesis independent of median distance (p ≤ 0.01). We observed a significant decrease in the median distance for the 4th chromosome (Fig. 5j). The magnitude of spermatocyte decrease was nearly as close (median distance 0.7 µm), well within the same prominent territory (Fig. 5d, f). Since the 4th occupies the same territory as the X, these chromosomes could be regulated independently, or coordinately, due to territory-level regulation.

The location of the satellite sequences identify the territories, but we were most interested in chromosome structure in gene rich euchromatic regions. Hence, we used oligopainting to examine the euchromatic portions of the X chromosome for evidence of compaction that might accompany inactivation. Oligopaint probes show that inactive X chromosomes have greater compaction (decreased volume) and increased sphericity compared to active X chromosomes in mammalian cells72,76, so we measured both these parameters. We probed similar sized euchromatic regions of the X chromosome (22.3 Mb) and the left arm of the 2nd chromosome (2L, 22.7 Mb) with oligopaints (Fig. 5e). We converted raw in situ data in Imaris to create masks (n = 23) of pixel intensities (Supplementary Movie 1) and obtained volumetric measurements of the X and 2 territories (Fig. 5g). We found that almost the probe length corrected X chromosome volume was reduced relative to chromosome 2L (Fig. 5h). This was not significant at the p ≤ 0.01 level we have used in this work (p = 0.03). However, when we corrected the data to account for 2L copy number (divided by two), the X had significantly greater volume than 2L (Fig. 5i) which was inconsistent with inactivity resulting from compaction. The assumption that volume scales with copy number is of dubious validity. The X chromosome was significantly more spherical than 2L (Fig. 5j), which was consistent with the hypothesis that compaction accompanies inactivation for regulation of X expression in Drosophila primary spermatocytes. In mammalian fibroblasts, the inactive X has a sphericity (ψ) of 0.67, while the active X has ψ = 0.5772,76. We found that the Drosophila spermatocyte X had ψ = 0.58, while 2L had ψ = 0.53. Collectively, these results do not provide strong evidence that Drosophila spermatocyte X chromosome activity is regulated by overall condensation levels.

X chromosome condensation in primary spermatocytes. Given that the X expression falls in maturing primary spermatocytes, we asked if the expression shift in primary spermatocytes correlated with precocious chromosome condensation, as has been previously reported70. To locate chromosomes, we used satellite and oligopaint probes that were validated on somatic metaphase chromosomes (Supplementary Fig. 4). The major autosomal bivalents and the X chromosome (along with the 4th chromosomes) reside in three distinct chromosome territories that abut the nuclear envelope of primary spermatocytes71 (Fig. 5a). The nuclear interior, more diffusely stained by DAPI, was occupied by the large transcriptionally active Y chromosome66,67,72 (Fig. 5a). The euchromatic portions of the X chromosome are the rDNA clusters. Only the Y-linked rDNAs are addressed transcriptional status of chromosome territories in spermatocytes. Given the rDNA repeat size and the rDNA copy number, the Y chromosome is the rDNA clusters. Only the Y-linked rDNAs are regulated by overall condensation levels.

Inactive RNA polymerase II mechanism. We more directly addressed transcriptional status of chromosome territories in spermatocytes by determining the phosphorylation status of the regulatory C-Terminal Domain (CTD) of RNA polymerase II (Pol-II). The transcription cycle begins when Pol-II binds the promoter. Serine 5 phosphorylation (pSer5-CTD) causes RNA Pol-II to initiate transcription, and subsequent Serine 2 phosphorylation (pSer2-CTD) induces transcriptional elongation (Fig. 6a). In the whole testis, we observed pSer2-CTD in spermatogonia and in the two autosomal territories in spermatocytes (Fig. 6b). In spermatocytes, we observed clear pSer2-CTD in two autosomal territories, but the X and 4th territory (identified by nucleolar proximity) was poorly stained as previously reported78 and pSer5-CTD localization was similar to pSer2-CTD pattern (Fig. 3), the parsimonious explanation is that spermatogonia show dosage compensation and spermatocytes show inactivation or reduced X chromosome compensation.

We similarly examined the expression of the reduced gene sets for the 4th chromosome. Genes with low Tau were over-represented on the 4th chromosome, especially in M11 germ cells and C1 somatic cells, resulting in an exaggerated over-expression relative to the major autosomes across all cell types (Fig. 4h), while low TSPS and CTSP resulted in significantly lower relative expression of the 4th chromosomes only in spermatocytes (Fig. 4i, j). The magnitude of spermatocyte decrease was magnified when we used the CTSP gene set, but overall the 4/A ratios were near 1.0 (Fig. 4j). The large sample size of cells resulted in tightly centered distributions, but note that the number of genes contributing the 4th chromosome measurements was small (Table 1). To briefly summarize, we observed a decrease in X and 4th chromosome expression with all genes (Fig. 3) and with reduced gene sets (Fig. 4), suggesting a chromosome-wide change in gene expression in spermatocytes, and not simply a reduced number of X-linked and 4-linked genes with male-biased expression.
We also observe localization of total Pol-II, but very little pSer2-CTD on the X territory when we co-stained with X euchromatic oligopaints (Supplementary Fig. 4), consistent with the idea that reduced levels of X chromosome transcripts may be due to decreased active transcription. To obtain more quantitative information on transcriptional status of the X chromosome, we determined the ratio of pSer2-CTD and pSer5-CTD to total Pol-II within a given territory compared to the autosomes in late primary spermatocytes (Fig. 6c, d). We found a >2-fold decrease in pSer2-CTD (log₂ = −1.4) and a >2-fold decrease in pSer5-CTD (log₂ = −0.6) on the X relative to the autosomes in late primary spermatocytes (Fig. 6c, f). These data indicate that the decline in X chromosome transcripts seen by scRNA-seq is due to a block in the transcriptional cycle regulated by CTD tail phosphorylation. Given that the 4th is also in this territory, it may be subject to the same fate.

**Discussion**

Our data clearly support a dynamic model, where X chromosomes are expressed at a higher rate in spermatogonia than one would expect based on DNA copy number alone, supporting the idea of X chromosome dosage compensation in the pre-meiotic male germline. This initial upregulation of X chromosome expression is followed by a dramatic decrease. We suggest that lower expression of the X in early meiosis is not due to the

![Fig. 6 Transcriptional events in late primary spermatocyte nuclei.](image)

a. Labelled schematic of RNA Pol-II preinitiation, initiation, and elongation at a promoter. Optical sections of an immuno-stained L3 testis (b) and spermatocytes (c, d). Serine 2 phosphorylated Pol-II localization (green), the nuclear envelope marker Lamin (white), and the DNA stain DAPI (red). We indicate spermatogonia and spermatocytes, separated by a line (dashed, white). Three/five visible pigment cells are indicated. c, d From left to right: DAPI grayscale, Phospho-CTD grayscale, total Pol-II grayscale, color merge. Primary spermatocyte nuclei are outlined (white-dashed line) based on DAPI. Insets in the merged images are anti-Asterless staining (grayscale) used to stage by centriole length, a measure of spermatocyte age. The ratio of e Serine 2 phosphorylated Pol-II (261 territories scored) or f Serine 5 phosphorylated Pol-II to total Pol-II pixel intensity on the X chromosome and autosome territories (N = 148 territories). We corrected for background using non-territory nuclear staining in the same nucleus. Significance (p ≤ 0.01, two-sided Welch’s t-test) (asterisk).
absence of X-chromosome dosage compensation in the germ-line\textsuperscript{11}, but to an even more extreme reduction in gene expression due to silencing\textsuperscript{10,29}. While the canonical dosage compensation pathway, acting to up-regulate X expression, is absent in male germ cells\textsuperscript{13,24}, there is also evidence for non-canonical dosage compensation in testis\textsuperscript{21,27}. The mechanism of germline dosage compensation in Drosophila is unknown. Our data provide an important additional argument against failed dosage compensation in the male germline as a cause of reduced X chromosome expression in spermatocytes, based on the fact that the 4th chromosome undergoes a similar dramatic decrease in transcript levels, despite being present in two copies. Silencing of genes in spermatocytes is independent of copy number.

Mechanistically, the reduced expression of the X and 4th chromosomes in spermatocytes correlates with the failure to activate RNA Pol-II. The Y chromosome is concomitantly active. This begs the question, why? This expression pattern could be due to the simple absence of genes expressed in spermatocytes on the X and 4th chromosomes and the presence of genes that must be expressed from the Y chromosome (Fig. 7a). If there are few genes expressed, there will be little active Pol-II ipso facto. However, expression of “housekeeping” genes suggests a chromosome-wide decrease in X and 4th expression. A pure gene content model predicts that genes newly arriving on the X with would be expressed, as evolutionary modification of regulation takes time. In fact, the autosomal ocnus gene is precisely expressed in spermatocytes, but shows extremely reduced reporter activity when inserted onto the X\textsuperscript{28}. This is consistent with a model where the X is a generally unfavorable environment for spermatocytes gene expression, due to either chromosome- or territory-level repression (Fig. 7b, c). This is reminiscent of meiotic sex chromosome inactivation in mammals, where X chromosome expression is high in spermatogonia, followed by X inactivation associated with a distinct organelle like XY body\textsuperscript{15}. The inactivation of both the X and Y chromosomes in mammals may be a special case of a more general inactivation of unpaired chromosome regions in a genomic defense model\textsuperscript{16}. Lack of homology could signal intruding transposable elements seeking to hijack the germline for vertical transmission to the next generation. Active recognition and silencing would be useful to the host organism. We observed two violations of the prediction that unpaired chromosomes are silenced in primary spermatocytes. Specifically, the 4th chromosome would be active, and the Y would be inactive in the simplest versions of this model. However, the 4th has retained its X chromosome-like silencing despite having two copies and the Y is maximally expressed in spermatocytes despite having a single copy. One way to achieve this would be the creation of a repressed territory occupied by both the X and 4th chromosomes (Fig. 7c). The evolutionarily retained inactivation of the 4th could be due to this localization, perhaps originally triggered by monosomy in ancestral species. It is also possible that the non-recombining 4th chromosome\textsuperscript{4} is not recognized as having a homolog. The single Y is highly diffuse and very little of it is in this repressed territory. However, allele-specific expression of the Y-linked rRNA genes drive the activity of the nucleolus\textsuperscript{73}, so at least part of the Y is expressed while in a repressed territory. It is possible that Y-linked genes, including the rDNA cluster, required for spermatogenesis escape inactivation as occurs for a subset of X linked genes on inactive X chromosomes in mammals\textsuperscript{80}. X to 2nd or 3rd chromosome translocations result in breakpoint-independent dominant male sterility, whereas X to 4th do not\textsuperscript{74}. Spreading repression or activation along a chromosome element, or relocation of parts of elements to novel territories might result in such a phenotype. Experiments to test these models will help us understand the evolution of sex chromosome expression in flies, and probably many other species.

Fig. 7 Models for chromosome element expression in spermatocytes. Non-mutually exclusive models. a Gene-level regulation. Movement (green dashed arrow) of genes with male-biased expression (blue ovals) off the X due to deleterious changes resulting from antagonistic selection in females or extinction of genes with male-biased expression on the X, or female-biased expression on the Y (strike through) result in a feminized X and a masculinized Y. X and 4th (red), Y (green), and major autosome (black) chromosomes (elongated bars) shown. b Chromosome-level regulation. This model relies on chromosome-specific regulation, including X-chromosome inactivation. Gene expression depends on chromosome elements that are active (green chromosomes) or repressed (red). Centromeres (Black circles). c Territory-level regulation. In this model the X and 4th chromosomes are restricted to a repressed address (red lozenge) within the nucleus (large circle). This could be the result of a barrier preventing activation of Pol-II or active repression within the territory. The 2nd and 3rd chromosome territories (green lozenges) and the diffuse Y (green lines) are expressed normally. The active nucleolus (n, green circle) is embedded within the repressed domain, but depends on expression of Y rDNA, not X rDNA.

Methods
Sample preparation. We collected gonads from the immobile third instar (L3) larvae just prior to the final prepsalp defecation and prior to evasion of the anterior spiracle (Supplementary Fig. 1a). Staging was aided by staining gut contents using 0.5 mg/ml Sulforhodamine B yeast paste fed to L3 larvae (clear foregut,
full midgut (hindgut) (Supplementary Fig. 1g–h). The

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-20897-y | www.nature.com/naturecommunications

1× PBX; incubated with secondary antibodies (1:200 Alexa Fluor 488

...developmental analysis of the Astler (Asl) centriole marker. We created X and 2L oligopaint probe channel masks (n = 23 for both X and 2L) of the pixel intensities in late primary spermatocyte cells. We measured volume and sphericity (shape) of the masks and performed Paired t-tests.

Immunostaining. For immunostaining formaldehyde fixed tissues, we used slightly different conditions depending on the antibodies used and location. For protein traps: we fixed in 4% formaldehyde in PBS for 20 min; blocked in 1× PBX and then incubated in the secondary antibodies (1:20 mouse or 1:100 guinea pig) at 4 °C overnight; washed twice in 1× PBX, once in 1× PBS; and mounted onto a microscope slide in Vectashield. Images were acquired using Zeiss LSM 800 laser scanning confocal microscopy with 63X/1.4 NA oil immersion objective and images (Supplementary Fig. 2) were acquired using Nikon Eclipse Ti with a 100×/1.4 NA oil immersion objective, a spinning disc confocal microscope with rhodamine (Roche). The preparations were counterstained using DAPI and Vectashield. Images from metaphase spreads and adult spermatocytes were obtained using a Zeiss Axiohot 2 microscope equipped with bright field and epifluorescence optics.

For fluorescence in situ hybridization using oligopaints, we fixed testes in 5% formaldehyde washed in 1X PBX and then 0.3 M NaCl, 30 mM NaCitrate, 1 Tween 20 (2X SSC), then successively in 20, 40, and 50% formamide in 2X SSC. Tissue were pre-fixed at 37 °C for 4 h, 92 °C for 3 min, and finally in 2x SSC at room temperature for 10 min. Probes labeled with biotin-16-DUTP were detected using avidin-FITC conjugate (Fisher Scientific) and probes labeled with digoxigenin-11-DUTP were detected using anti-digoxigenin rhodamine (Roche). The preparations were counterstained using DAPI and mounted in Vectashield. Images from metaphase spreads and adult spermatocytes were obtained using a Zeiss Axioshot 2 microscope equipped with bright field and epifluorescence optics.

In situ hybridization. To detect deuteromycin on X, Y and 4Chromosomes, we synthesized probes (1.688 satellite of X chromosome and AATAC repeats for 4th and AATAC repeats for Y chromosomes) using D. melanogaster genomic DNA as a template. The PCR was performed using 50 ng of genomic DNA and Taq Platinum DNA polymerase following the manufacturer’s instructions. Probes were PCR-labeled with digoxigenin-11-DUTP or biotin-14-DUTP during their synthesis at the 5’ end. To detect the euchromatin, we synthesized oligopaint probes using 5’ fluorophore labeled and 5’ phosphorylated PCR primers, amplified with the following cycles: 98 °C for 5 min; 40 cycles of 30°C for 30 s, 52°C for 30 s and 72°C for 15 s, with a final extension step at 72°C for 5 min. The PCR product was purified using Zymo spin columns (D4031) and then digested using lambda exonuclease for 30 min at 37°C and 10 min at 75°C. The digested probe products were precipitated using ethanol and quantified using a nanodrop.

We stained mitotic chromosomes to assess probe hybridization (Supplementary Fig. 3). For mitotic metaphase chromosomes, we immersed dissected brains in 0.05% Colchicine in PBS for 20 min, hypotonic in tap water for 15 min, and fixed in 3:1 (v/v) ethanol-acetic acid for 15 min, transferred to 60% acetic acid and squashed on a slide. Fluorescence in situ hybridization for heterochromatin probes was performed according to Pinkel et al.9 using mitotic chromosome spreading or spermatocytes from adult testes. Post-hybridization washes were performed as follows: two times in 2x SSC at 42°C for 5 min, two times in 0.1x SSC at 42°C for 5 min, one time in 2x SSC at 42°C for 5 min and finally in 2x SSC at RTT temperature for 10 min. Probes labeled with biotin-16-DUTP were detected using avidin-FITC conjugate (Fisher Scientific) and probes labeled with digoxigenin-11-DUTP were detected using anti-digoxigenin rhodamine (Roche). The preparations were counterstained using DAPI and mounted in Vectashield. Images from metaphase spreads and adult spermatocytes were obtained using a Zeiss Axioshot 2 microscope equipped with bright field and epifluorescence optics.

RNA-Seq—library preparation. For bulk sequencing, we added fatbody-free gonads to RNeasy 96 kit 350 µl PL RT buffer mixed two times for 30 s, froze in a dry ice-ethanol bath and stored at -80°C. RNA was extracted using RNeasy 96 kit following the manufacturer’s protocol. We quantified RNA using the Quant-iT RiboGreen quantification kit and used 200 ng of total RNA for library preparation.9 We spiked in External RNA Control Consortium (ERCC) spike-in pools 78A and 78B (transcribed from a certified standard reference) prior to using 50 ng. For multiplexing, we used eight different TruSeq v2 barcoded adapters. We sequenced stranded multiplexed libraries using a single-ended 50 bp strategy and generated RNA-Seq profiles of biological quadruplicates on a HiSeq Illumina 2500.

For single-cell sequencing, we loaded dissociated cells (6k for Replicates 1 and 2, 7k for Replicates 3 and 4; for replicate 3) onto 10 cm dishes (at 30°C) for 2.5 h and then co-cultured with 100 µg ml−1 doxycycline for 2 days. Chromosome spreads were performed according to the single-cell library preparation following the user guide for Single Cell 3’ Reagent Kits v2. We sequenced libraries with Quant-iT PicoGreen and confirmed 300–500 bp insertions...
sizes on a TapeStation 2200. We generated scRNA-seq profiles in biological triplicate pools on the 10X Chromium System and sequenced (Read1 = 26 bp, Read2 = 98 bp and Read3 = 8–10 bp) on a HiSeq Illumina 2500.

**RNA-Seq—data processing.** For the bulk RNA-Seq, we demultiplexed and converted Binary Base Call (BCL) files to FASTQ using bcl2fastq (v2.17.1.14, Illumina, San Diego, CA, USA). We processed FASTQ files through custom RNA-Seq workflow (commit: 0069ec8572, Supplementary Dataset 6). Briefly, reads are trimmed of Illumina adapters using cutadapt with default parameter except for --q 20 and --minimum-length=25. We mapped with Hisat2 with default parameters except --max-intronlen 300,000 and --known-splice-site-infile using annotated splice sites (FlyBase r6-26). We removed multi-mapping reads and low-quality alignments using samtools view -q 20. Gene expression and intergenic expression (FlyBase r6-26) is quantified in a strand-specific manner using FeatureCount from the subread package with the -a 2 option. The workflow outputs a number of quality control metrics from FastQC, FastQC Screen, Picard Mark Duplicates, and Picard CollectPerfMetrics (Supplementary Dataset 6). In addition, we quantified intergenic expression and ERCC spike-in expression using FeatureCount. See Supplemental Dataset 1 for visualization of all analyzed scRNA-Seq data and all code is available at the Zenodo repository: https://doi.org/10.5281/zenodo.3973174. Bulk RNA-Seq data is available at the Gene Expression Omnibus (GEO) GSE115478 (laval testes) and GSE115511 (larval ovaries). Bulk adult testes and ovaries are from GSE995745, Supplementary Dataset 3.

For scRNA-Seq, we converted BCL files to FASTQ using cellranger mkfastq (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger, Supplementary dataset 6). We demultiplexed pooled, aligned reads, and quantified gene-level expression using cellranger count with default parameters. For alignment and quantification, we used the D. melanogaster genome assembly (dm6) with gene annotations from FlyBase r6.26. Larval testes scRNA-seq data is available at the GEO GSE125947.

To quantify how well scRNA-Seq captured the overall expression patterns of the testis, we calculated a set of four metrics for each of four replicates and normalized scRNA-seq and bulk RNA-seq libraries using Transcripts Per Million Reads plus 1 (TPM +1). We calculated pairwise Spearman rank correlation coefficients among all samples (Supplementary Dataset 3).

**Sex-biased gene expression.** For sex-biased gene expression, we used publicly available adult RNA-Seq counts from w1118 and Oregon-R (GEO: GSE99574). Briefly, counts tables from w1118 and Oregon-R were downloaded for each tissue. Gene identifiers were mapped to current FlyBase FBgn using the provided annotation files and FlyBase’s secondary ID. We estimated differential gene expression by comparing males and females using DESeq2 for each adult tissue and for L3 bulk RNA-Seq generated in this study.

**scRNA-seq cell selection, clustering, and cell type annotation.** Prior to the downstream analysis, we removed cells ID’s that were likely to be empty and identified cell IDs that likely contain two or more cells known as multiplets (Supplementary Dataset 2). There are two distinct classes of multiplets: homotypic and heterotypic. Homotypic multiplets contain 2 or more cells from the same cell-type. Homotypic multiplets tend to have high UMI counts and are found at the top of the distribution. However, setting an upper UMI threshold may exclude genes of high RNA content cell-types. To identify an optimized upper UMI threshold, we calculated the X:A,A, and 4,4:A,A ratios for each cell by taking the total X-linked (or 4th-linked) reads divided by the total autosomal-linked reads of the UMI distribution however, setting an upper UMI threshold may bias against homotypic multiplets that were putative homotypic doublets. Heterotypic multiplets contain 2 or more cell types from different cell-types. This leads to a mixture of expression profiles causing cells to look like intermediate cell-types, which are identifiable using in silico mixing cell-types. We removed 410 cell ID’s that behave like the mixture.

We combined all three scRNA-Seq replicates into one dataset using single-cell integration and clustered cells with K-Nearest Neighbors using the 2000 most variably expressed genes. To determine cluster boundaries, we iteratively tested mixing of cell-types. We removed 410 cell ID’s that behave like the mixture. For scRNA-seq downstream analysis. We know that the transcriptional profile of the testis is unique and expressed genes are not equally distributed among chromosome elements. Therefore, we wanted to interrogate a multiple gene sets to ensure our observations are unbiased. We created three “housekeeping” gene sets, two based tissue-specific expression in adult tissues and two gene sets based on scRNA-seq gene expression from L3 testis. Using the adult tissue panel described in “Sex-biased gene expression”, we compared gene counts across all adult tissues using two tissue-specificity estimates Tau and TSPS. We consider genes with r ≤ 0.5 or TSPS ≤ 1 as not tissue specific (i.e., “housekeeping”). We consider genes expressed anywhere in our scRNA-seq data (expressed) or gene expressed in at least 33% of cells (low cell type specificity, CTSP) which are effectively genes expressed across multiple cell types. Expression by chromosome arms is complicated as gene expression of an individual cell is sparse in scRNA-Seq, and simple aggregation of cell counts is not appropriate because missingness is not random. Therefore, we used a cell level permutation test approach for downstream analysis. Briefly, we calculate each the measure of interest (see below) for each individual cell using all gene sets (see above). For pairwise comparisons, we compare the distributions of cells from two cell types using Mann–Whitney U test. We then permute cell type labels and repeat the Mann–Whitney U test comparisons for 10,000 iterations. We calculate the overall p-value based on the number of times the observed test was more extreme than the original observed test. For mean comparisons, we compare the distributions of cells from a given cell type to a random sample, of the same size, from all cells in the experiment using the Mann–Whitney U test for 10,000 iterations. We calculated the overall p-value as the average number of times the observed distribution was significantly different than the permuted distribution. We calculate 3 different metrics for each cell used with the described permutation test approach. Normalized Gene Expression use the Log10 Transcripts Per Million (TPM) UMI counts. Male-biased Enrichment Scores use Gene Set Enrichment Analysis (GSEA) with TPM UMI counts and L3 bulk testis-based expressed genes to estimate an cell level enrichment score. Finally, we compared X chromosome degeneration. We consider genes expressed anywhere in the X chromosome arms. We used the Log10 Transcripts Per Million (TPM) UMI counts. Male-biased Enrichment Scores use Gene Set Enrichment Analysis (GSEA) with TPM UMI counts and L3 bulk testis-based expressed genes to estimate an cell level enrichment score. Finally, we compared X chromosome degeneration. We consider genes expressed anywhere in the X chromosome arms. We used the Log10 Transcripts Per Million (TPM) UMI counts.

**Code availability.** All raw sequencing data has been deposited and is publicly available through the NCBI Gene Expression Omnibus under accession codes “GSE125947”, “GSE115511”, and “GSE1147487”. The projections of the expression of each Drosophila gene onto the UMAP can be accessed at the NIH Fighare (https://doi.org/10.35092/yhcz.11945746). All data repository details and links are listed in the FlyBase Author Reagent format (Supplementary Dataset 6). All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. Source data are provided with this paper in the Supplementary Data files.

**References.**

1. Bachfor, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14, 113–124 (2013).
2. Carvalho, A. B., Dobos, B. A., Vibration, M. D. & Clark, A. G. Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc. Natl Acad. Sci. USA* 122, 13232–13237 (2001).

3. VICOSO, B. & Bachtrog, D. Re-analysis of the larval testis data on meiotic sex chromosome inactivation. *Nature* 44, 322–325 (2013).

4. Riddle, N. C. & Elgin, S. C. R. The Drosophila dot chromosome: where genes flourish amidst repeats. *Genetics* 201, 757–772 (2018).

5. vicosa, B. & Chibon, R. X chromosome: unusual patterns and processes. *Nat. Rev. Genet.* 7, 645–653 (2006).

6. khil, P. P., Smirnova, N. A., Romanenko, P. J. & Camerini-Otero, R. D. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat. Genet.* 36, 642–646 (2004).

7. Kelly, W. G. et al. X-chromosome silencing in the germline of C. elegans. *Development* 121, 492 (2002).

8. Parisi, M. et al. Patterning of genes on the Drosophila X chromosome showing male-biased expression. *Science* 299, 697–700 (2003).

9. Lee, J. T. Sex chromosome inactivation: the importance of pairing. *Curr. Biol.* 15, R249–R252 (2005).

10. Vibration, M. D. Meiotic sex chromosome inactivation in Drosophila. *J. Genet.* 2, 104–117 (2014).

11. Meiklejohn, C. D., Lande, E. L., Cook, J. M., Kingan, S. B. & Presgraves, D. G. Sex chromosome-specific regulation in the Drosophila male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biol.* 9, e1001126 (2011).

12. Bao, L. E., Thomson, E. K. & Long, M. Retraced new genes out of the X chromosome. *Genome Res.* 12, 1854–1859 (2002).

13. Sturgill, D., Zhang, Y., Parisi, M. & Oliver, B. Dermalization of X chromosomes in the Drosophila genus. *Nature* 450, 238–241 (2007).

14. Vibration, M. D., Zhang, Y. & Long, M. General gene off the X chromosome in the Drosophila genus. *Genome Res.* 19, 897–903 (2009).

15. Vibration, M. A. Meiotic silencing in mammals. *Annu. Rev. Genet.* 39, 395–412 (2015).

16. Huynh, K. D. & Lee, J. T. X chromosome inactivation: a hypothesis linking ontogeny and phylogeny. *Nat. Rev. Genet.* 6, 410–418 (2005).

17. Lu, L.-Y. & Yu, X. Double-strand break repair on sex chromosomes: challenges during male meiotic prophase. *Cell Cycle* 14, 516–525 (2015).

18. Wu, C. T.-C. & Xu, E. Y. Sexual antagonism and X-inactivation—the SAXI hypothesis. *Trends Genet.* 19, 243–247 (2003).

19. Handel, M. A. The XY body: a specialized meiotic chromatin domain. *Exp. Cell Res.* 296, 57–63 (2004).

20. McKee, B. D., Yan, R. & Tsai, J.-H. Meiosis in male Drosophila. *Spermatogenesis* 2, 167–184 (2012).

21. Deng, X. et al. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat. Genet.* 43, 1179–1185 (2011).

22. Kuroda, M. I., Hifflker, A. & Lucchesi, J. C. Dosage compensation in Drosophila—a model for the coordinate regulation of transcription. *Genetics* 201, 435–450 (2018).

23. Rastelli, L. & Kuroda, M. I. An analysis of maleless and histone H4 acetylation in Drosophila melanogaster spermatogenesis. *Mech. Dev.* 71, 107–117 (1998).

24. Oliver, B. Genetic control of germline sexual dimorphism in Drosophila. *Int. Rev. Cytol.* 219, 1–60 (2002).

25. Vibronovski, M. D., Lopes, H. F., Karr, T. L. & Long, M. Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet.* 5, e1000731 (2009).

26. Vibronovski, M. D. et al. Analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the Drosophila X chromosome. *BMC Biol.* 10, 42 (2012), author reply 50.

27. Gupta, V. et al. Global analysis of X chromosome dosage compensation. *J. Biol.* 3, 5 (2006).

28. Hense, V., Baines, J. F. & Parsch, J. X chromosome inactivation during Drosophila spermatogenesis. *PLoS Biol.* 5, e273 (2007).

29. Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38, 735–742 (1984).

30. Hess, O. & Meyer, G. F. Genetic activities of the Y chromosome in Drosophila during spermatogenesis. *Adv. Genet.* 14, 171–223 (1968).

31. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* 34, 1145–1160 (2016).

32. Hennig, W. Spermatogenesis in Drosophila. *Int. J. Dev. Biol.* 40, 167–176 (1996).

33. Fuller, M. T. Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. *Semina. Cell Dev. Biol.* 9, 433–444 (1998).

34. Greenspan, L. J., da Caevas, M. & Matunis, E. Genetics of gonadal stem cell renewal. *Annu. Rev. Cell Dev. Biol.* 31, 291–315 (2015).

35. Vicosa, B. et al. Aneuploidy, translocation and benign gonial cell neoplasms in the germline to restrict proliferation during Drosophila spermatogenesis. *Development* 124, 4361–4371 (1997).
65. Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. An abundance of X-linked genes expressed in spermatogonia. Nat. Genet. 27, 423–426 (2001).
66. Redhouse, J. L., Morezoniacci, J. & White, R. A. H. Co-transcriptional architecture in a Y loop in Drosophila melanogaster. Chromosoma 120, 399–407 (2011).
67. Fingerhut, J. M., Moran, J. V. & Yamashita, Y. M. Satellite DNA-containing gigantic introns in a unique gene expression program during Drosophila spermatogenesis. PLoS Genet. 15, e1008028 (2019).
68. Ravasi, T. et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, 744–752 (2010).
69. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21, 650–659 (2005).
70. Cooper, K. W. Normal spermatogenesis in Drosophila. in Biology of Drosophila melanogaster. 1:1–61 (1965).
71. Thomas, S. E. & McKee, B. D. Analysis of chromosome dynamics and chromosomal proteins in Drosophila spermatocytes. Methods Mol. Biol. 558, 217–234 (2009).
72. Tsai, J.-H., Yan, R. & McKee, B. D. Homolog pairing and sister chromatid cohesion in heterochromatin in Drosophila male meiosis I. Chromosoma 120, 335–351 (2011).
73. Reugels, A. M., Kurek, R., Lammermann, U. & Bünemann, H. Mega-introns in Drosophila melanogaster. Genes Dev. 20, 2922–2936 (2006).
74. El-Sharnouby, S., Redhouse, J. & White, R. A. H. Genome-wide and cell-specific epigenetic analysis challenges the role of polycistron in Drosophila Cooperstown. PLoS Genet. 9, e1003812 (2013).
75. Parsch, J., Meiklejohn, C. D., Hauschreck-Jungen, E., Hunziker, P. & Hartl, D. L. Molecular evolution of the ocnus and janus genes in the Drosophila melanogaster species subgroup. Mol. Biol. Evol. 18, 801–811 (2001).
76. Carrel, I. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404 (2005).
77. Hsieh, T. & Brutlag, D. Sequence and sequence variation within the 16S rRNA gene in E. coli. Nucleic Acids Res. 41, e108 (2013).
78. Anim pelvis, S. & Andrews, S. FastQC: a tool for multi-genome mapping and quality control. Fir000Res. 7, 1338 (2018).
79. Wang, Y. et al. Re-annotation of eight Drosophila genomes. Life Sci. Alliance 1, e201800156 (2018).
80. Stuart, T. et al. Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21 (2019).
81. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. Bioinformatics 19, 562–578 (2018).

Acknowledgements

This work utilized the computational resources of the NIH HPC Biowulf cluster (http://biowulf.nih.gov), NCBI and FlyBase databases, and the Bloomington Drosophila Stock Center. We thank Bob Johnson for oligopaint probe libraries, Gustavo Kuhn for providing the X chromosome satellite probe, Eduardo Gorah for total RNA Pol-II antibodies, and Veronic Souza for checking the 4th chromosome localization in our experiments. This research was supported in part by the Intramural Research Program of the NIH, the National Institute of Diabetes and Digestive and Kidney Diseases (DK015600) and the National Heart Lung and Blood Institute (HL06126), the National Institute of General Medical Sciences (GM126752 and GM12107), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD052937) the São Paulo Research Foundation (FAPESP 2015/20844-4, CEPID 2013/00826-1, PT 2015/16661-1, MS 2017/26609-2, MS 2017/14923-4, DD 2019/15212-0 and DD 2019/14788-5).

Author contributions

S.M., J.M.F., B.O., M.D.V. and E.M. conceptualized the project; S.M. and J.M.F. developed single cell methodology; J.M.F. developed code; M.A., M.M.L.S.P., C.C.A., D.C.C. M., K.C., Z.D., K.M., C.A.M., O.M.P.G., E.R., M.S. and K.Y. validated results; J.M.F., M.A., B.J.G., C.C.A., and C.A.M. performed formal analysis; S.M., B.A., B.J.G., M.M.L.S. P., C.C.A., D.C.C.M., K.C., S.D., Z.D., K.M., C.A.M., O.M.P.G., E.R., M.S., K.Y., H.Y., H.E.S. and V.S. performed experiments; J.M.F. curated metadata; S.M., J.M.F., B.O., M.D.V. and E.M. wrote the original draft; S.M., J.M.F., M.A., C.C.A., B.O. and M.D.V. prepared figures; S.M., H.E.S., V.S. N.R., B.O., M.D.V. and E.M. directed the work; S.M. and B.O. administered the project; N.M.R., B.O., M.D.V., M.A. and E.M. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-021-20897-y.

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

Peer review information Nature Communications thanks Todd Nystul and the other anonymous, reviewer(s) 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/.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021