Trajectory Mapping of the Early *Drosophila* Germline Reveals Controls of Zygotic Activation and Sex Differentiation

Yi-Ru Li, Hsiao Wen Lai, Hsiao Han Huang, Hsing-Chun Chen, Sebastian D. Fugmann, Shu Yuan Yang

Supplemental tables (Tables S1, S2, S4, S5 available as separate Excel files)

Table S1. Maternal and zygotic germline markers.

Table S2. GO enrichment analysis of soma-positive zygotic germline genes.

Table S3. Early germline specific genes that encode RNA binding proteins or transcription factors.

Maternal	
RNA binding proteins:	*nos*
	osk
	bru1
Transcription factors/Zinc finger proteins:	*cad*
	ebd2
	CG4404
	dl
	CG9925
	tll
Zygotic	
RNA binding proteins:	*CG32706*
	CG6999
Transcription factors/Zinc finger proteins:	*CG34031*
	CG14036
	CG3919
	CG14135

Table S4. Top 25 male and female marker genes in the early embryonic germline.

Table S5. Top 2000 maternally- and zygotically-biased germline genes from the unsexed dataset.

Supplemental figures and figure legends

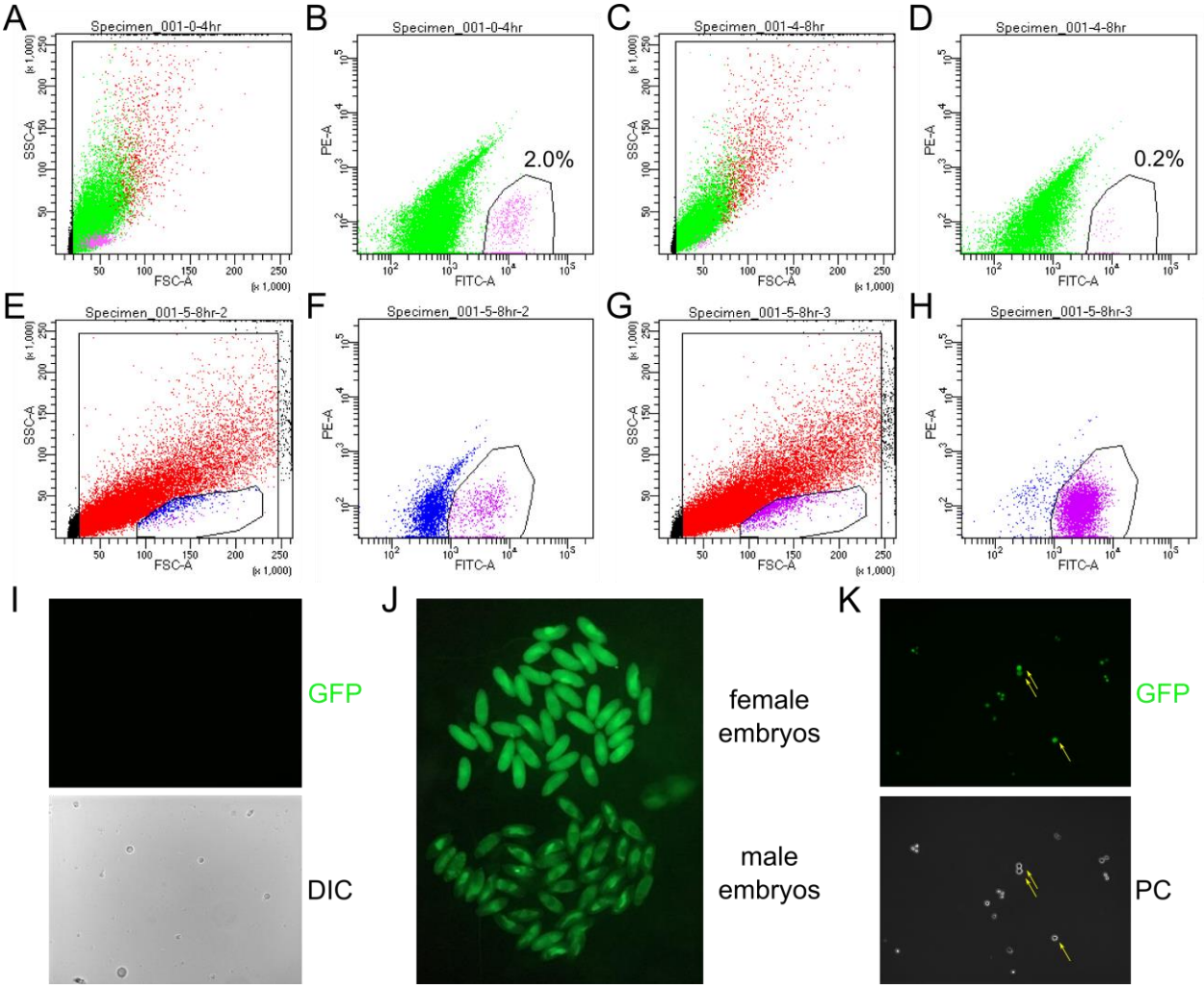

Figure S1. FACS schemes for the collection of single germ cells for scRNA-seq. A-B, FACS plots for unsexed 0-4 h germ cells. **C-D**, FACS plots for unsexed 4-8 germ cells. **A**, **C**, Gating for live cells which plots FSC-SSC. **B**, **D**, The gate for the magenta cells indicate the GFP+ cells that were sorted and their percentages. X and Y axes indicate the FITC and PE channels, respectively. **E**-**F**, FACS plots for female 5-8 h germ cells. **G**-**H**, FACS plots for male 5-8 h germ cells. **E**, **G**, The small encircled populations on the bottom of the plots are enriched for germ cells, thus this live gate was used for the FITC-PE plots. **F**, **H**, FITC-PE plots with magenta cells being the GFP+ cells sorted. The axes are the same as in **B**, **D**. **I**, Fluorescent microscopy profile of cells sorted as GFP- indicating that those cells indeed do not express GFP. Top image displays the GFP channel whereas lower image displays the DIC channel of the same field and focus. **J**, Female and male 5-8 h embryos can be separately based on their *Sxl-GFP* expression under a fluorescent stereoscope. The embryos in the upper half express GFP highly and are females; those in the lower half are males. **K**, Microscopy profile of cells sorted from female *vas-GFP, Sxl-GFP* embryos as GFP+. Some cells exhibit the characteristic size and shape of embryonic germ cells (yellow arrows) while many others appear to be somatic cells. Top image displays the GFP channel whereas lower panel displays the phase contrast image of the same field and focus.

A

	Number of cells	Mean reads per cell	Median genes per cell
Unsexed	3,810	33,487	3,166
Female	11,001	22,241	2,045
Male	7,222	29,231	3,482
Male 25% reduced	7,113	22,259	3,186

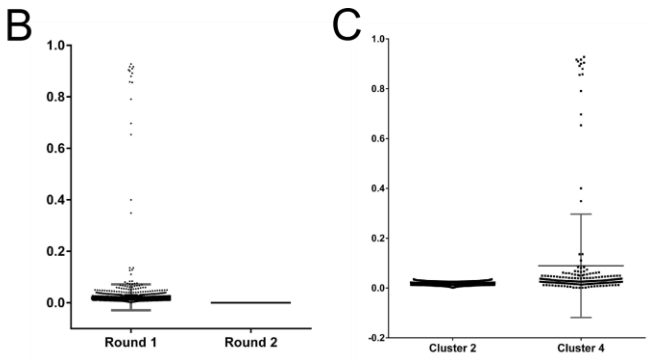

Figure S2. Quality control of the unsexed, male, and female scRNA-seq samples. A, Quality control numbers of various sequencing samples we analyzed including the unsexed sample, the female sample, the male sample, and the male sample reduced randomly by 25%. The numbers of cells detected in each sample as well as the mean reads and genes per cell detected are listed. **B**, The proportion of reads in each cell that mapped to mitochondrial genes in unsexed (round 1) and sexed (round 2) datasets. Each dot depicts one cell in our databases. There are no reads from the sexed dataset that mapped to mitochondrial genes. **C**, The proportion of reads in each cell of the indicated cluster in the unsexed dataset that mapped to mitochondrial genes. Cluster 2 is the main cluster of germ cells after the bifurcation point; cluster 4 is the small cluster positioned very close to cluster 2 after dimension reduction.

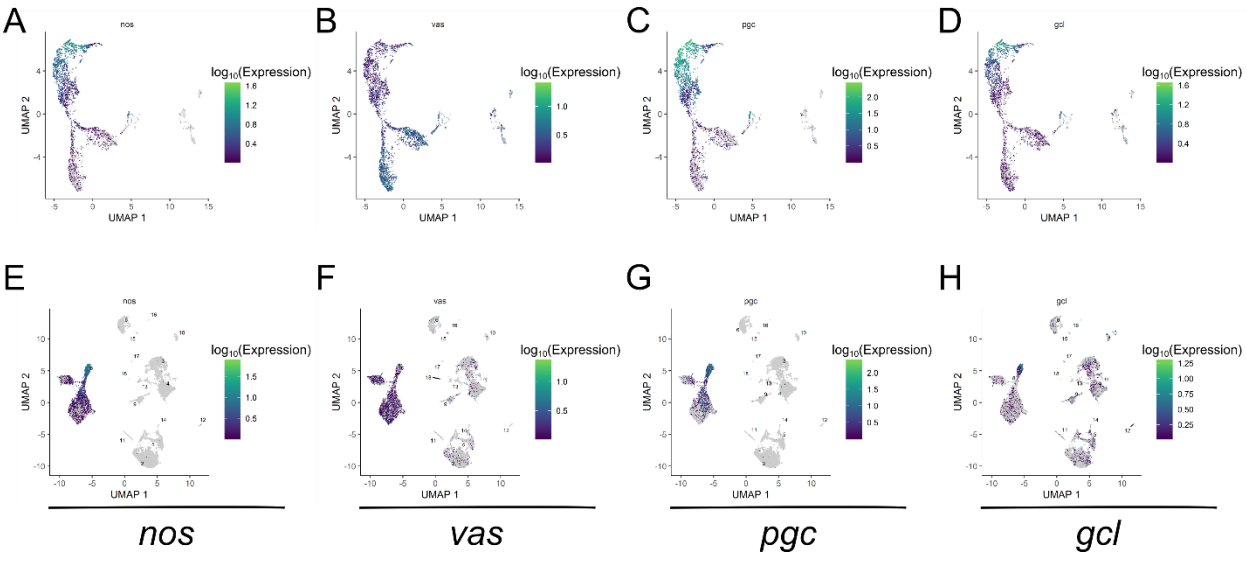

Figure S3. Expression profiles of known germline markers. A-D, Unsexed sample. **E-H**, Sexed sample. **A**, **E**, *nos* expression. **B**, **F**, *vas* expression. **C**, **G**, *pgc* expression. **D**, **H**, *gcl* expression. Color codes for expression levels are to the right of each plot.

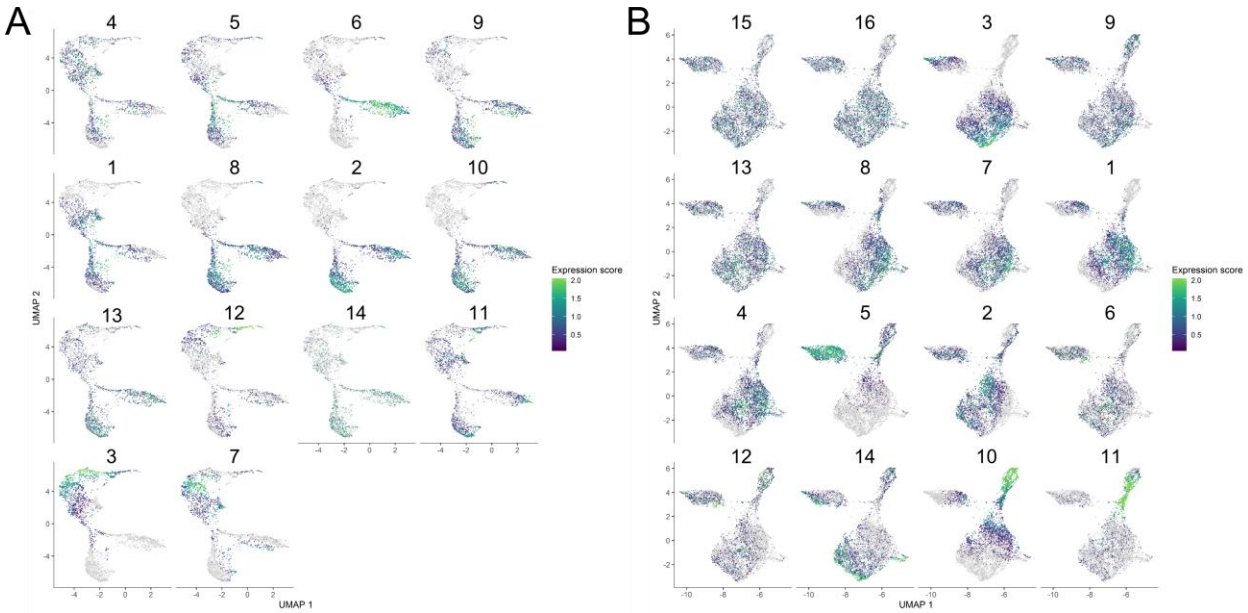

Figure S4. Expression modules for the germline clusters in the unsexed (A) and sexed (B) datasets. The module numbers are indicated on top of each individual module, and color codes for expression levels are indicated on the right.

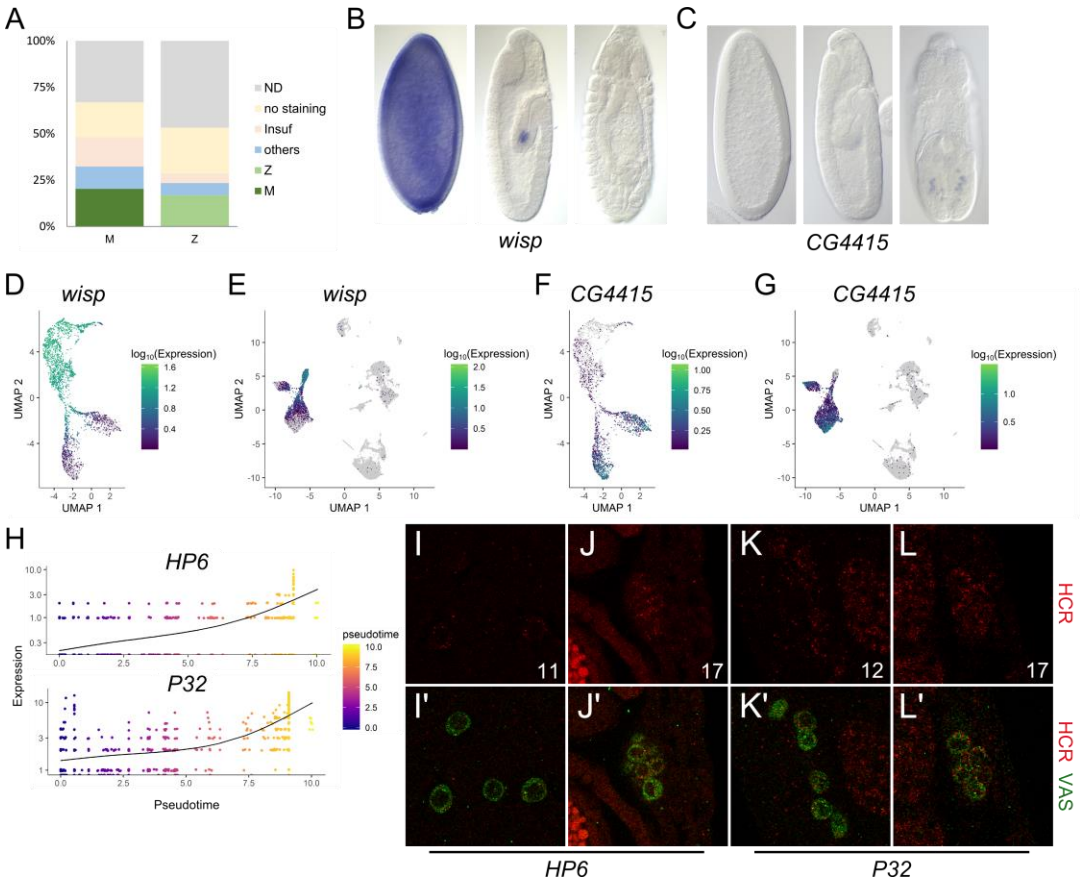

Figure S5. Expression validation of the scRNA-seq results. A, Expression patterns of the all germline marker genes in the BDGP database. The left bar is for maternally-deposited, germline-specific genes (M). The bar on the right is for zygotically-activated, germline-specific genes (Z). Dark green indicates highest expression in stages 4-6 whereas light green indicates germline-specific expression that begins in stages 7-10 and increases in later stages. Blue indicates that the BDGP in situ pattern was mainly outside of germ cells. Pink indicates insufficient data at BDGP to make a proper determination. Yellow indicates no staining. Gray indicates the fractions that have not been profiled in the database. **B**-**C**, *In situ* images from the BDGP database of progressive embryonic stages from left to right for *wisp* (**B**) and *CG4415* (**C**), from the "maternally-deposited" and "zygotically-activated" categories. **D**-**E**, Expression profiles of *wisp* in the unsexed (**D**) and sexed (**E**) datasets. **F**-**G**, Expression profiles for *CG4415* in the unsexed (**F**) and sexed (**G**) datasets. **H**, Expression levels of *HP6* (upper graph) and *P32* (lower graph) of every cell in the common stem and male branch in the germline cluster of the unsexed sample is graphed along pseudotime. The black lines indicate the trends of expression along pseudotime. The X- and Y-axes plot pseudotime and expression levels, respectively, and the pseudotime color codes are indicated to the right. **I**-**L**, Fluorescent in situ HCR images for *HP6* (**I**-**J**) and *P32* (**K**-**L**) showing expression at stage 17 (**J, L**) and the stages we could detect the first signs of expression (stage 11 for *HP6*, panel **I**, and 12 for *P32*, panel **K**). **I**-**L** display the HCR channel alone (red) while **I'**-**L'** display merged images with α-Vasa staining (green) to mark the germ cells.

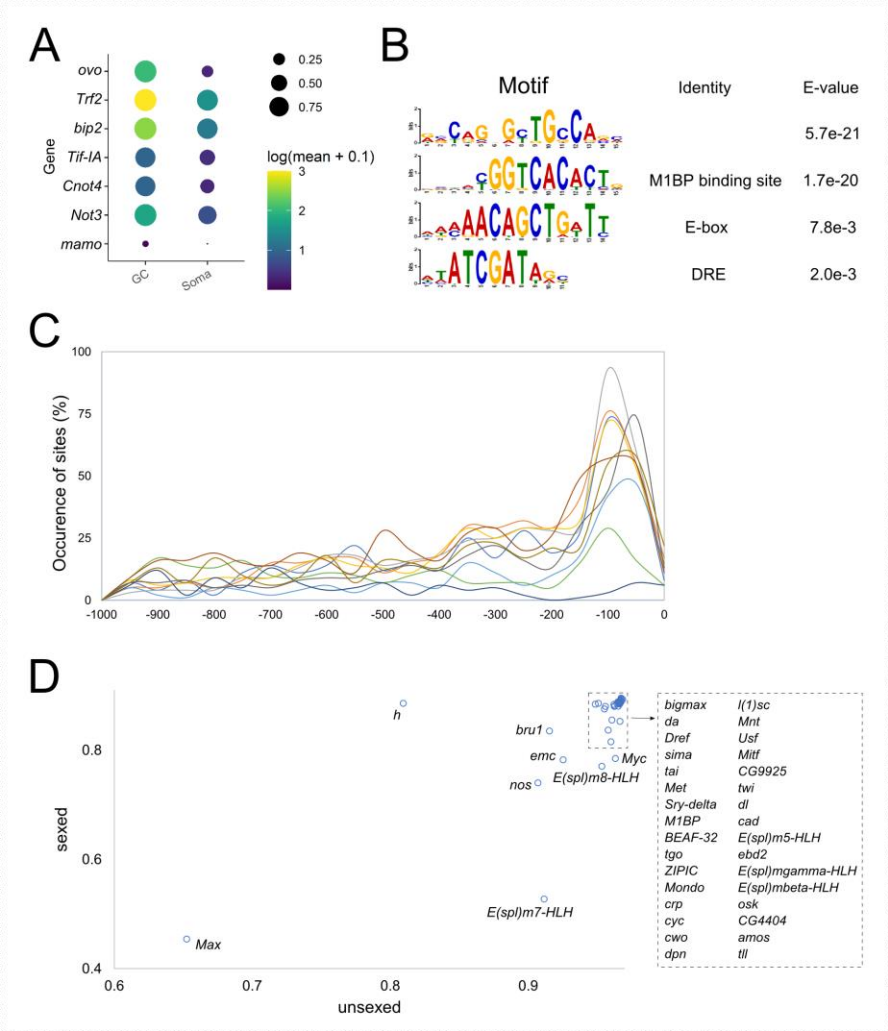

Figure S6. Analyses of expression regulation of zygotic germline genes. A, Average expression of various TFs previously reported to be germline-specific in embryogenesis in the germline and somatic cells of our sexed dataset. The color of the dots indicate mean expression levels as referenced by the color code to the right whereas the size of the dots indicate the percentage of cells within each group that express the designated genes. **B,** Enriched motifs discovered *de novo* in the promoter regions of zygotic germline genes and their likely identity based on comparisons with consensus sequences of known TFs. DRE is the binding site of Dref, a factor whose binding site was found to be enriched in **Fig. 3A**. **C**, Distribution of TF sites enriched in the promoter regions of zygotic germline genes. Sites of TFs identified from both the enrichment analysis and *de novo* discovery are included. Each color line represents a different enriched motif. The x-axis indicates the position in relations to the transcription start site (+1) whereas the y-axis denotes the percentage of each motif that occurred at a given position. **D**, Median of the pairwise correlation values of a given TF- or RNA-binding protein-encoding gene to all zygotic germline genes included in the GRN analysis using the unsexed (x-axis) and sexed (y-axis) datasets. The names of the genes residing in the dotted box in the top right corner of the plot is listed to the right in decreasing order based on their median values from the unsexed dataset from top to bottom, left to right.

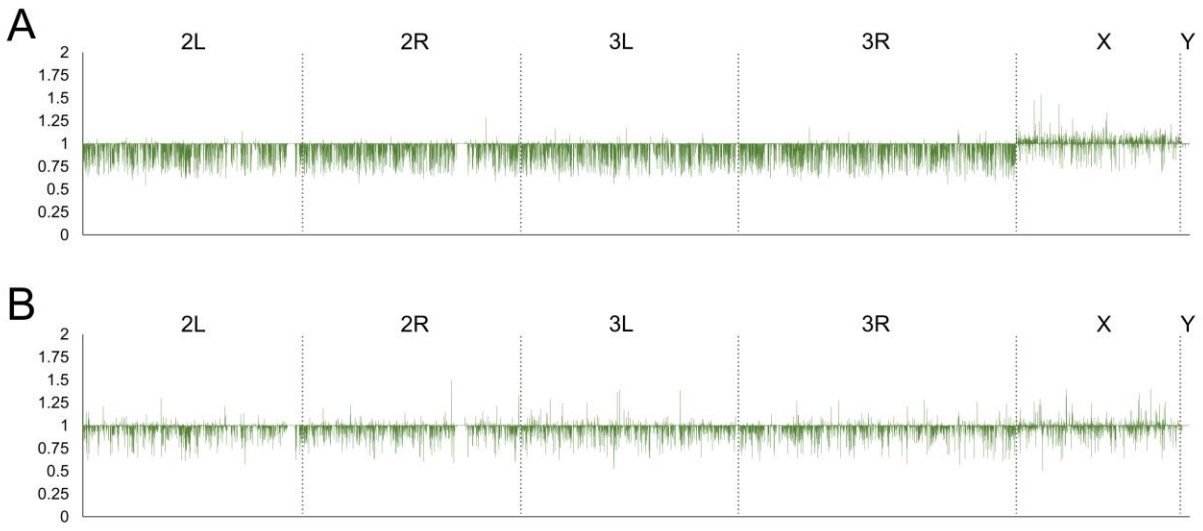

Figure S7. Expression differences between the female and male germline in the original sexed dataset. **A**, Female-to-male expression ratios in the germline cluster for genes across all chromosomes from the original sexed dataset. The X-axis plots the positions on individual chromosomes as indicated and the Y-axis plots the fold differences of expression of female over male cells. **B**, Female-to-male expression ratio of all genes in somatic cells included in the original sexed sample. Axes are same as in (**A**).

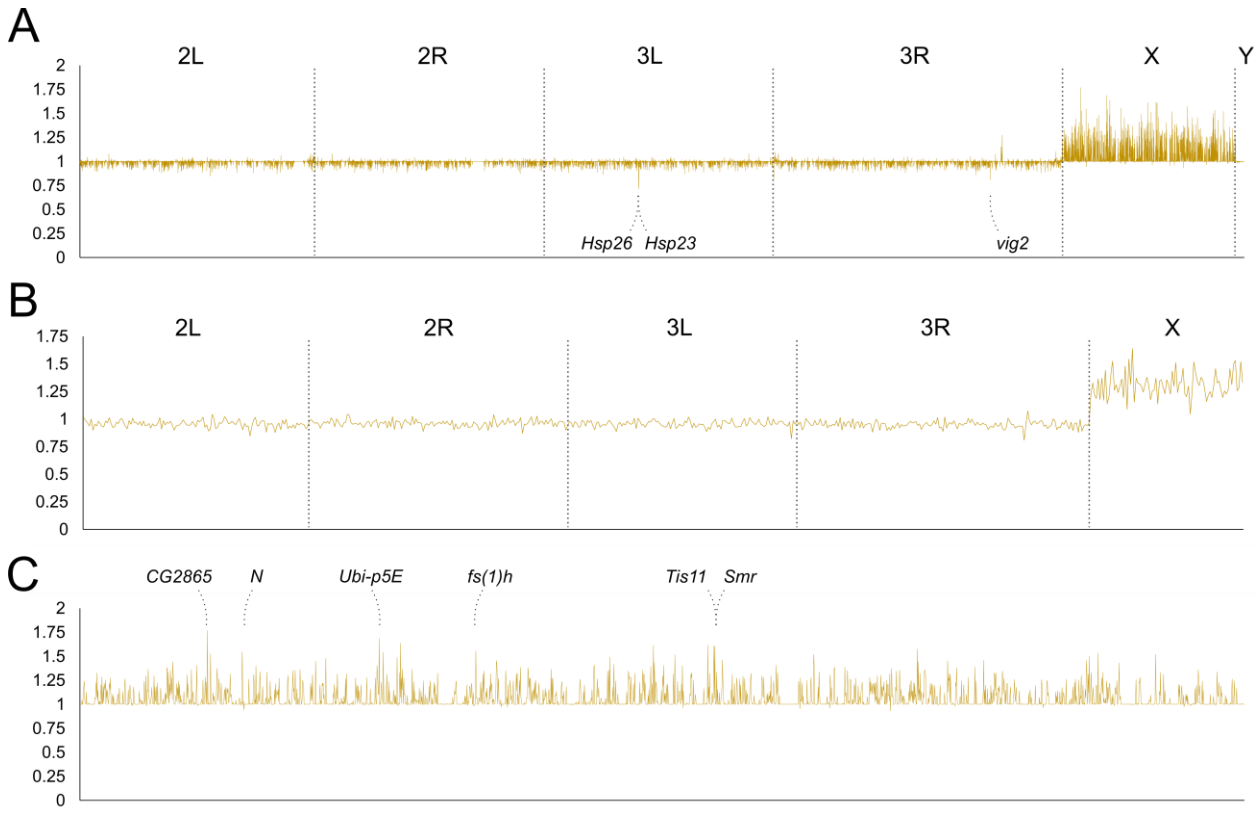

Figure S8. Expression differences between the female and male germline in the unsexed dataset. A, Female-to-male expression ratios for genes across all chromosomes from the unsexed dataset. The axes in **A**-**C** are the same: the X-axis plots the positions on individual chromosomes as indicated and the Y-axis plots the fold differences of expression of female germ cells over male germ cells. **B**, Expression differences between the two sexes for the 647 zygotically activated germline genes from the unsexed dataset. **C**, Sex expression differences from the unsexed dataset highlighted for the X-chromosome. The identities of the highest peaks are indicated.

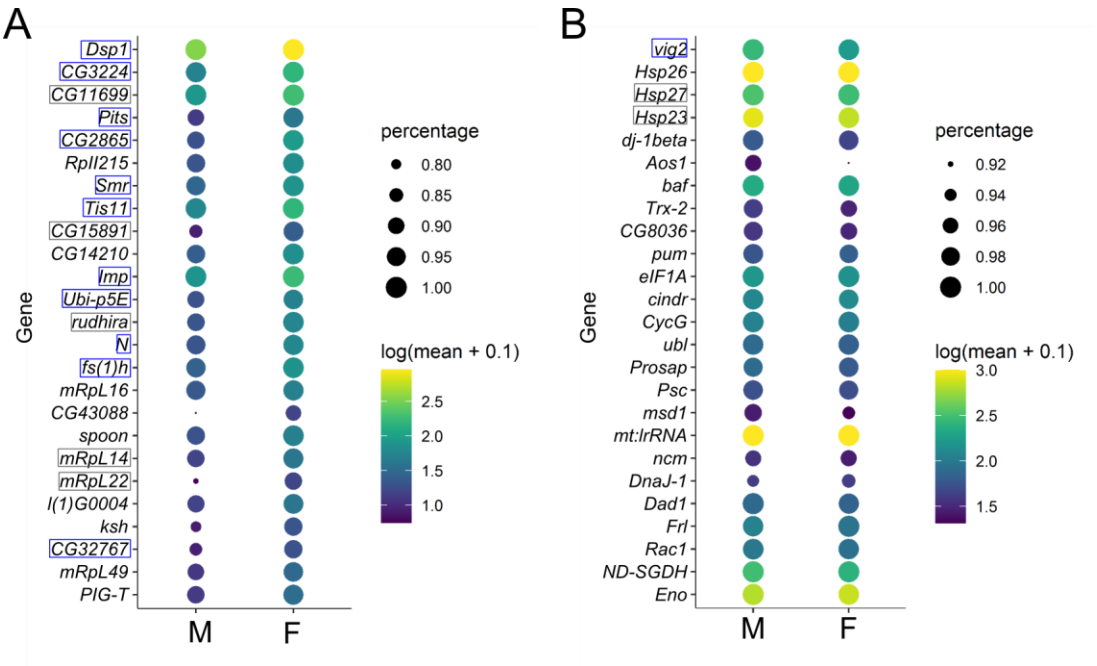

Figure S9. Top sex markers from the unsexed dataset. A, Expression profiles of the top 25 female markers in the male and female germline clusters after zygotic activation are plotted with the colors reflecting mean expression as indicated on the right; the size of the dots reflects percentages of cells in the clusters with detectable expression of each gene. The gene names boxed in gray or blue are those also determined as top female markers in the sexed dataset; the ones predicted to be transcription or chromatin-associated factors are marked with blue boxes. **B**, Expression patterns of the top 25 male markers plotted in the same way as in (**A**).

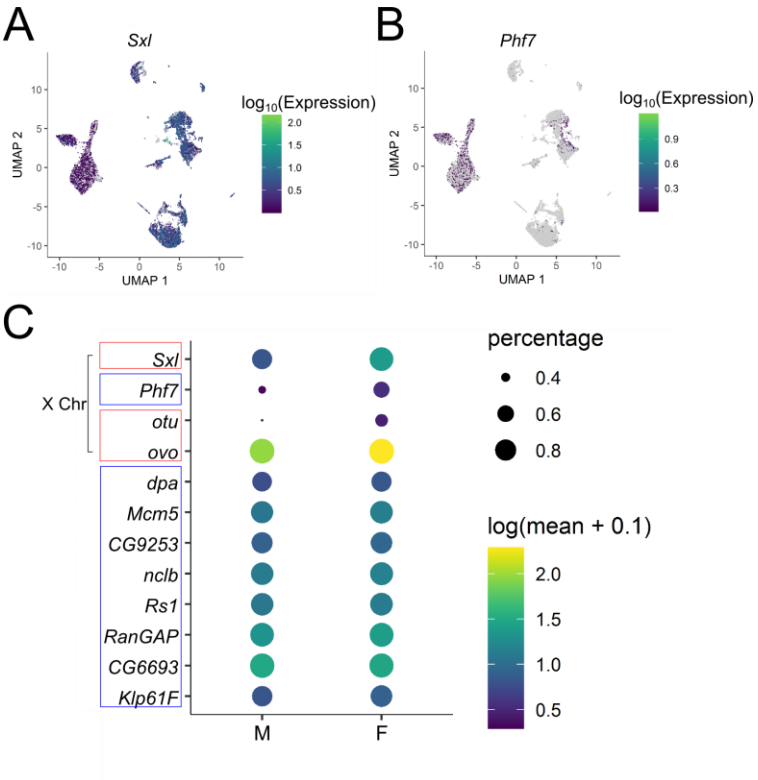

Figure S10. Expression profiles of known sex-biased genes in the embryonic germline in the sexed dataset. A-**B**, Expression patterns of *Sxl* (**A**) and *Phf7* (**B**). Color codes of expression levels are indicated to the right. **C**, Comparisons of expression levels in the male vs. female cells after the bifurcation point in the sexed dataset of various genes previously reported to exhibit sex-biased expression patterns in the embryonic germline. The color of the dots indicates mean expression levels as referenced by the color code to the right whereas the size of the dots indicates the percentage of cells within each group that express the designated genes. The genes reported to exhibit male and female-biased expression are boxed in blue and red, respectively. The first four genes on the top are located on the X chromosome.