PNAS

Supporting Information for

Sperm-inherited H3K27me3 epialleles are transmitted transgenerationally in cis

Kiyomi Raye Kaneshiro*[1,2], Thea A. Egelhofer*[1], Andreas Rechtsteiner[1], Chad Cockrum[1,3], and Susan Strome[1]

[1] Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.

[2] Present address: Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, CA 94945, USA.

[3] Present address: IDEXX Laboratories, Inc., 1 IDEXX Drive, Westbrook, ME 04092, USA.

* These authors contributed equally

Correspondence and requests for materials should be addressed to S.S. (email: sstrome@ucsc.edu)

This PDF file includes:

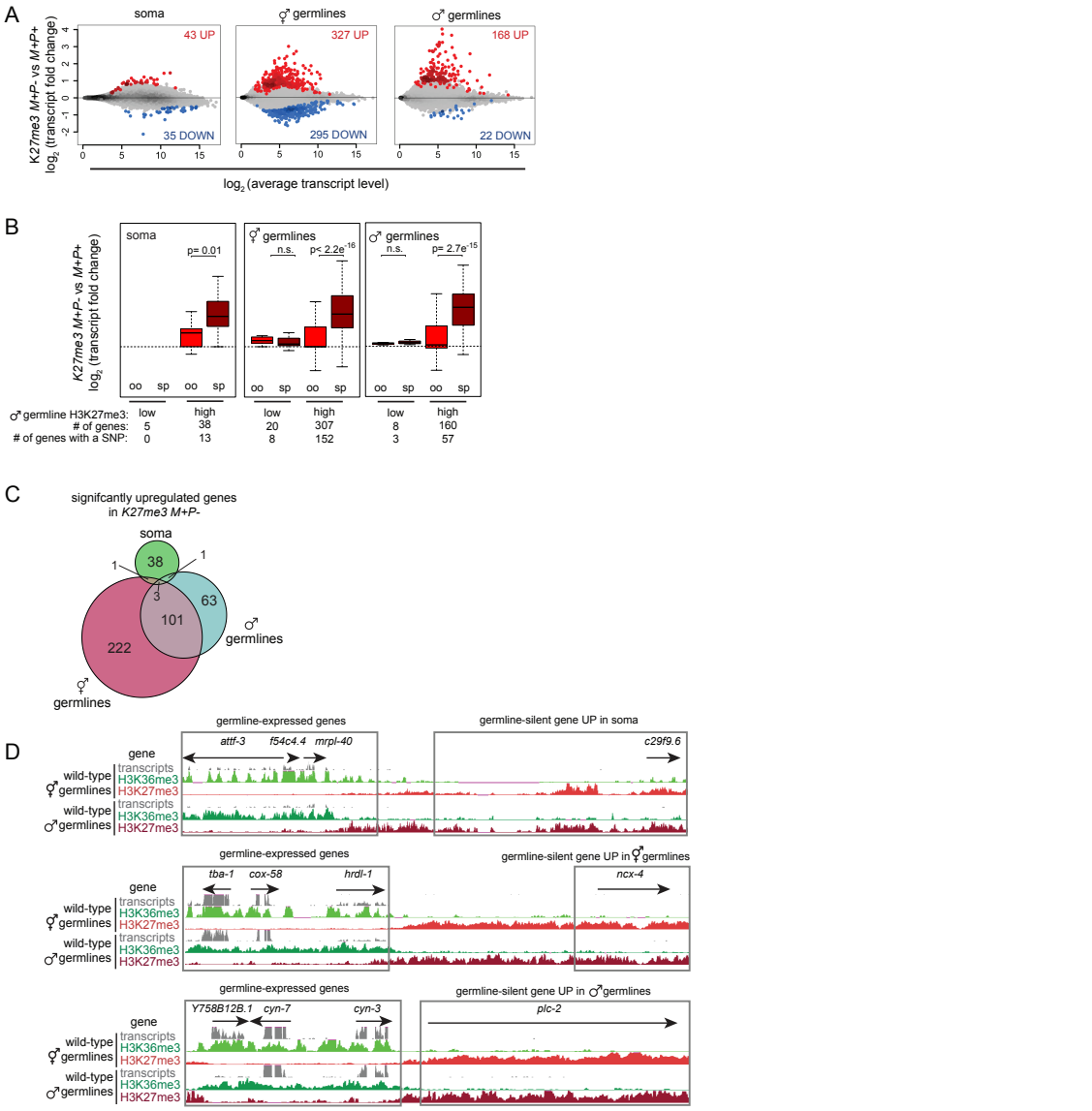

Figure S1. *K27me3 M+P-* worms misexpress genes in somatic and germline tissues. (A) MA plots of *K27me3 M+P-* vs *K27me3 M+P+* in different tissue contexts. Significantly upregulated genes (red) and downregulated genes (blue) ($p<0.05$). (B) Boxplots of oocyte (oo) and sperm (sp) allele-specific changes in transcript levels for significantly upregulated genes with low or high H3K27me3 marking in wild-type male germlines and containing a SNP (for comparison to boxplots of 300 top upregulated genes shown in Fig. 1D). Boxplots show the median (horizontal line), the 25th and 75th percentiles (boxes), and whiskers extending to the most extreme data point which is no more than 1.5 times the interquartile range. p-values were calculated using Student's t test (two-tailed). (C) Venn diagram showing overlap of significantly upregulated genes ($p<0.05$) in each of the three F1 tissues (for comparison to Venn diagram of top 300 upregulated genes in Fig. 2A). (D) UCSC genome browser images showing transcripts (grey; scale 0-200), H3K36me3 (green; scale 0-50), and H3K27me3 (red; scale 0-30) CUT&RUN sequencing reads in 'wild-type' (*K27me3 M+P+*) hermaphrodite germlines (top three tracks) and 'wild-type' (*him-8*) male germlines (lower three tracks) for a top upregulated gene in each *K27me3 M+P-* tissue context.

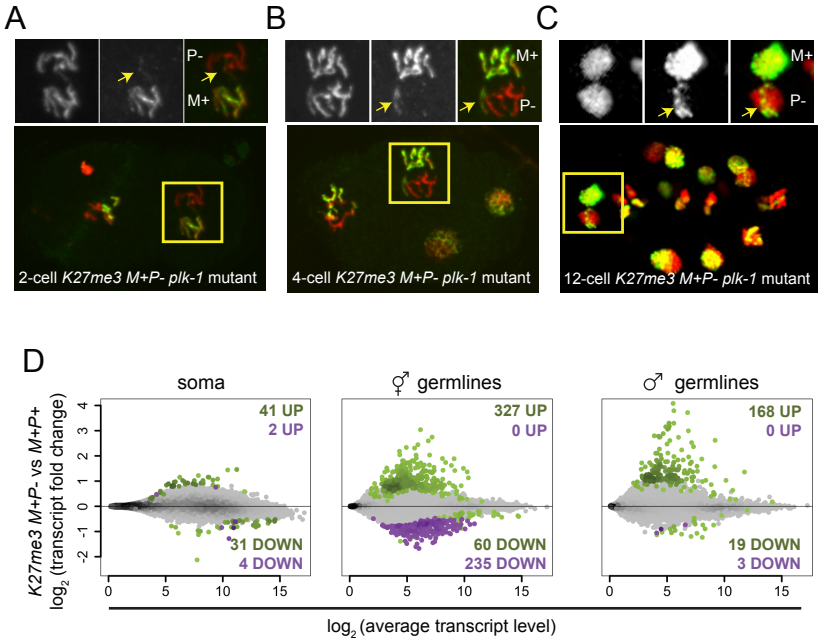

Figure S2. In *K27me3 M+P-* embryos, the sperm-inherited X chromosome (X^{sp}) acquires H3K27me3 during embryogenesis, preventing its upregulation in *K27me3 M+P-* larvae and adult tissues. (A-C) Immunofluorescence images of (A) 2-cell, (B) 4-cell, and (C) 12-cell *K27me3 M+P-* embryos in a *plk-1* background, which keeps sperm- and oocyte-inherited chromosomes in separate nuclei (1, 2). Red is DNA (DAPI) or chromatin stain (H3K4me3), green is H3K27me3. The upper panels are zoomed images of the region in the yellow box. A single H3K27me3(+) chromosome among the sperm-inherited chromosomes (yellow arrow) was detected at the 2-cell stage and got progressively brighter in later-stage embryos. (D) MA plots of *K27me3 M+P-* vs *K27me3 M+P+* in different tissue contexts. Significantly misregulated X-linked genes (purple) and autosomal genes (green) ($p<0.05$).

Extended Methods

***C. elegans* strains.** All strains are in the N2 background except CB4856 and SS1292.

N2 wild-type isolate from Bristol, England
CB4856 wild-type isolate from Hawaii
CB1489 *him-8(e1489) IV*
SS1292 *fem-2(b245ts) III* - mutation CRISPR-engineered into CB4856
DH0245 *fem-2(b245ts) III*
SS1167 *mes-3(bn35) I/ hT2-GFP (I;III); fem-2(b245ts) III/ hT2-GFP (I;III)*
SS1099 *mes-3(bn35) I/ hT2-GFP (I;III); him-8(e1489) IV*
SS1495 *glh-1[sams24[glh-1::GFP::3xflag]) mes-3(bn199) I/ hT2-GFP (I;III); fem-2(b245ts) III/ hT2-GFP (I;III)*
OCF65 *plk-1(or683) III; fog-1(e2121) unc-11(e47) I/ hT2 [qIs48] (I:III); itIs37[pAA64: pie-1p::mCHERRY::his-58 + unc-119(+)]*

RNA sequencing. Tubes containing germlines or L1s in 300 uL TRizol were subjected to three freeze-thaw cycles. 1 uL linear polyacrylamide (LPA) was added to each sample, vortexed 10 seconds, then transferred to Phase Lock Gel-Heavy 2 mL tubes. 60 uL chloroform was added to each sample, shaken vigorously for 15 seconds, then spun at 12,000 xg for 10 minutes. The aqueous phase was transferred to a clean tube, and 0.75 volume of isopropanol was added. After overnight precipitation at -20˚C, the samples were spun at 13,000 xg for 25 minutes, and the supernatant was removed. The pellet was washed five times with 500 uL freshly prepared 75% ethanol. The pellet was dried for 1 minute, resuspended in 13 uL RNase-free water, and incubated 5 minutes at 55˚C to facilitate dissolution. RNA concentration was measured (Qubit Quant-IT RNA Assay Kit) and RNA quality assessed (Agilent RNA Nano Kit). Total RNA was polyA selected (NEBNext Poly(A) mRNA Magnetic Isolation Module, NEB #E7490), and libraries were prepared (NEBNext Ultra RNA Library Prep Kit for Illumina, NEB #E7530). Sequencing libraries prepared from L1 larvae were sequenced on the Illumina NextSeq platform to acquire paired-end 75 bp reads. All other libraries were sequenced on the Illumina NovaSeq platform to acquire paired-end 50 bp reads.

Analysis of RNA sequencing data. All reads were mapped to the Bristol (N2) genome version WS220. Reads longer than 50 bp were trimmed to 50 bp. For RPKM and RPSM calculations, 50 bp reads were mapped to the WS220 genome twice: once allowing 1 mismatch (1mm) and a second time allowing 0 mismatches (0mm). For all reads, RPKM was calculated by first scaling total reads to 1 million then dividing by transcript length in kb (union of all exons). For SNP reads, RPSM (SNP reads per million) was calculated by scaling total SNP reads to 1 million then dividing by the number of SNPs per transcript. For differential expression analysis, reads were mapped using TopHat2 (v2.1.1) (3) with Bowtie2 (2.3.5.1) for paired-end (PE) reads using -g 1 (1 max multihit) with a mate inner distance set to 200, a mate standard deviation of 50, and a segment length of 25. Reads per transcript or gene body were counted using HTSeq (4) and a gtf file specifying transcript or gene body (union of all exons) or SNP locations per transcript or gene body. Significantly misexpressed genes were determined using the differential analysis program DESeq2 (v1.22.2) (5) from 50 bp reads (trimmed if necessary). Read counts from replicated samples from the same tissue (hermaphrodite germlines, male germlines, and L1s) were imported into DESeq2, and differentially expressed genes (DEGs) were determined using the contrast function for the various comparisons reported. Genes with a multiple hypothesis adjusted p-value < 0.05 were considered

significant. The top 300 upregulated genes per F1 tissue context were selected based on having a positive fold change and the lowest p-values in *K27me3 M+P-* samples relative to their *K27me3 M+P+* controls. We defined 6,664 germline-silent genes as genes for which the RPKM of transcript reads was 0 in *K27me3 M+P+* germlines. We defined 5,931 germline-expressed genes as genes for which the RPKM of transcript reads was > 5 in *K27me3 M+P+* germlines. To identify genes that failed to reestablish H3K27me3 in *K27me3 M+P-* hermaphrodite germlines, genes were first selected that were H3K27me3-enriched (H3K27me3 gene body z score > 0.2) in *K27me3 M+P+* hermaphrodite germlines. This set of genes was split into 2 categories: UP (transcriptionally upregulated) in *K27me3 M+P-* vs *M+P+* germlines ($\log_2$ fold change > 0 and adjusted p-value < 0.05) and not UP (not transcriptionally upregulated) (all other H3K27me3-enriched genes). This yielded 310 and 9,318 H3K27me3-enriched UP and not UP genes, respectively. To perform gene set enrichment analysis on the top 300 upregulated genes in different F1 tissue contexts, the hypergeometric test (phyper in R) was performed using previously identified tissue-unique gene sets, hermaphrodite germline- and male germline-specific gene sets, and a germline-specific gene set.

SNP assessment. The annotated single nucleotide polymorphisms (SNPs) of the Hawaiian CB4856 genome compared to the Bristol N2 WS220 genome were downloaded from WormBase. To determine which annotated SNPs reliably distinguish between Hawaii and Bristol reads, 50 bp CUT&RUN-seq and RNA-seq reads from wild-type Hawaii and Bristol embryos and germlines, respectively, were mapped to the WS220 Bristol genome. Mapped reads that overlapped with 1 of the 78,780 gene-associated SNPs, either intronic or exonic, were filtered and assessed for expected mappability (TopHat2) to the Bristol genome (WS220) when mapping parameters allowed 1 mismatch *(mm1)* versus 0 mismatches *(mm0).* For each SNP i, we calculated a score $bz_i = (bmm1_i – bmm0_i)/(bmm1_i + bmm0_i)$ *from* SNP-overlapping reads from Bristol libraries and $hz_i = (hmm1_i – hmm0_i)/(hmm1_i + hmm0_i)$ for SNP-overlapping reads from Hawaii libraries where $mm1_i$ and $mm0_i$ represents a read overlapping SNP i when mapping allowed 1 mismatch or 0 mismatches, respectively. We expect SNP-overlapping reads from Bristol samples to map to the Bristol genome whether allowing 1 mismatch or 0 mismatches. Therefore, we expect bzi to be close to 0. In contrast, SNP-overlapping reads from Hawaii samples mapped to the Bristol genome are expected to only map when allowing 1 mismatch. Therefore, we expect hzi to be close to 1. SNPs for which $|bz_i| > 0.25$ and $|hzi| < 0.75$ were kept for use in assessing SNP-overlapping reads that emanate from Hawaii (oocyte-inherited) versus Bristol (sperm-inherited) genomes in hybrid worms. Of the 78,780 gene-associated (exonic and intronic) SNPs, 52,360 SNPs that spanned 12,834 genes remained. Of these SNPs, 17,793 are in exons specifically and span 7,950 genes. We could therefore use SNPs in our hybrid experiments to analyze the gamete-specific origin of CUT&RUN-seq reads from 12,834 genes and RNA-seq reads from 7,950 genes. The genome locations of these SNPs were used to filter mapped reads (intersectBed) that overlap these SNPs for gamete-specific analysis of hybrid worm libraries. This analysis assessed the mappability of SNPs individually and allowed us to identify more usable SNPs than we identified previously. To increase the opportunity to capture a sequencing read that contained a SNP, we mapped our paired-end reads singly (R1 and R2 reads mapped separately). R1 and R2 reads mapped when allowing 0 mismatches were collated in R (*mm0*). Similarly, R1 and R2 read counts mapped when allowing 1 mismatch were collated in R (*mm1*). Collated read counts for *mm0* files were Bristol in origin and thus emanated from the sperm-inherited genome. Read counts that were Hawaii in origin were calculated by subtracting the *mm0* read counts from *mm1* read counts. Total SNP-overlapping reads (Bristol + Hawaii reads)

were scaled to 1 million. SNP-overlapping reads per SNP per million SNP reads (RPSM) was calculated for each gene by dividing the total SNP-overlapping reads per gene by the number of SNPs per gene. $\log_2$ (fold change) was calculated by log transforming the fold change of RPSM plus a 0.5 pseudocount of M+P- vs M+P+ samples (L2FC = $\log_2$ ((Pm RPSM + 0.5)/(Pp RPSM + 0.5)). To calculate SNP-specific z scores, log transformed RPSM GPM ($\log_2$ (RPSM + 1)) were scaled using the R package 'scale'. To calculate the change in gene body H3K27me3 or H3K36me3 in *K27me3 M+P-* vs *M+P+,* the z score for *K27me3 M+P+* was subtracted from the z score for *K27me3 M+P-.*

Analysis of the paternal X chromosome in *K27me3 M+P-* embryos in a *plk-1* background for Fig. S2. *plk-1; fog-1 unc-11* females (*K27me3 M+*) were mated with *mes-3* M-Z-; *him-8* males (*K27me3* P-) and incubated overnight at 23˚C. Embryos were dissected from mated females in egg buffer (25 mM HEPES pH 7.4, 118 mM NaCl, 48 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$) with 1 mM levamisole on polylysine-coated slides. Embryos were flattened between a slide and coverslip by wicking away buffer. Slides were frozen in liquid nitrogen, coverslips were removed, then slides were fixed in ice-cold methanol (10 min) and acetone (10 min) and allowed to air dry for up to 3 days. Slides were blocked (1.5% ovalbumin, 1.5% BSA in PBS) for 25 minutes, then incubated with primary antibody (1:3000 original stock solution of rabbit anti-H3K27me3 (C36B11 Cell Signaling MAb#9733 lot#C36B11) overnight at 4˚C. Slides were washed 3× in PBST (0.1% tween in PBS pH 7.2), blocked for 5 minutes, incubated with secondary antibody and DAPI at room temperature for 2 hours, and washed 3× in PBST. Coverslips were mounted on slides with gelutol. Slides were allowed to dry for up to 3 days before being imaged. Stacks of optical sections were acquired on a Solamere spinning disk confocal system. Stacks were collapsed into maximum intensity projections using Micro-Manager. Single-channel and merge images were generated in Photoshop.

SI References

1. M. M. Rahman, *et al.*, *Caenorhabditis elegans* polo-like kinase PLK-1 is required for merging parental genomes into a single nucleus. *Mol. Biol. Cell* **26**, 4718–4735 (2015).

2. K. R. Kaneshiro, A. Rechtsteiner, S. Strome, Sperm-inherited H3K27me3 impacts offspring transcription and development in C. elegans. *Nat. Commun.* **10**, 1271 (2019).

3. D. Kim, *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

4. S. Anders, P. T. Pyl, W. Huber, HTSeq – A Python framework to work with high-throughput sequencing data. *Bioinformatics.* **31**, 166-169 (2015).

5. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).