SUPPLEMENTARY MATERIAL

Setdb1 is required for germ line development and silencing of H3K9me3 marked endogenous retroviruses in primordial germ cells

Included:

Supplementary Figure Legends

Supplementary Methods

Supplementary References

Supplementary Figures

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1: Validation of PGC isolation strategy and ERV expression profiling. (A) Strategy for purification of male and female E13.5 PGCs based on the expression of the SSEA-1 cell surface marker. Subsequent downstream analyses conducted on 1000-3000 PGCs from individual embryos are also shown. **(B)** Expression (RPKM) levels of all ENSEMBL annotated genes was determined by RNA-sequencing of 10^3 purified E13.5 PGCs (SSEA-1+) and soma (SSEA-1-) isolated from the same male embryo. Cell-type specific genes were identified in PGCs (green) and in soma (red) by applying both Z-score (> 2) and fold-difference (> 5) thresholds. **(C)** Expression levels of all annotated ERV1, ERVK, MaLR and ERVL elements present at >100 copies in the C57Bl/6 genome, as measured by RPKM values, are shown for the same male E13.5 PGCs and soma.

Supplementary Figure 2: Mating strategy, **experimental scheme and genotyping. (A)** Mating scheme to obtain germline-specific *Setdb1* deficient mice. The datasets generated for each genotype and sex are outlined below. **(B)** Genotyping of E13.5 embryos. Representative PCR analysis of genomic DNA isolated from the tails of embryos from a single litter is shown. *Zfy:* present in males only (chrY), *Xist*: present in males and females (chrX).

Supplementary figure 3: Genome-wide enrichment of H3K9me3 versus H3K27me3 in E13.5 PGCs. (A) Genome-wide correlation (50,000 random 1kb bins) of H3K9me3 (left panel) and H3K27me3 (right panel) between male and female E13.5 PGCs. **(B)** Enrichment of H3K9me3 and H3K27me3 in female E13.5 PGCs genome-wide (50,000 random 1 kb bins, left panel), around transcription start sites (TSSs, +/- 1kb), at gametic differentially methylated regions (DMRs), at the 1kb region in the 5' flanks of ERVs or the 1kb region in the 5' flanks of LINEs. **(C)** The percentages of the mappable genome marked by H3K9me3 and H3K27me3 in female E13.5 PGCs is shown. Enriched regions were determined by MACS2, with a p-value of 0.05.

Supplementary figure 4: Co-occurrence of H3K27me3 and H3K9me3 at TSSs and ERVs. (A) Genome browser screenshots depicting repressed genes marked by H3K9me3 and H3K27me3 around their transcription start sites. **(B)** Genome-wide relationship between H3K9me3 enrichment, H3K27me3 enrichment and DNA methylation in male Setdb1 E13.5 PGCs. The yellow gradient in the right panel depicts the relative H3K9me3 enrichment level for all data points shown. * E13.5 DNA methylation data (PBAT WT) is from Kobayashi et al. (Kobayashi et al. 2013). **(C)** Genome browser screenshots illustrating regions marked by H3K9me3, H3K27me3 and DNA methylation in male E13.5 PGCs. * E13.5 DNA methylation data (PBAT WT) is from Kobayashi et al. (Kobayashi et al. 2013). **(D)** Enrichment of H3K9me3 and H3K27me3 versus mean percentage of DNA methylation at LINEs and LTR ERVs (present at >100

copies in the BL6 genome) in female E13.5 PGCs. ERVs are sorted alphabetically according to ERV1, ERVK, ERVL, Gypsy, and MaLR classes along the X axis. Class III (ERVL and MaLR) elements are generally depleted of all three marks. **(E)** Genome browser screenshot showing the presence of residual DNA methylation and the enrichment of H3K9me3 and H3K27me3 at the 5' end of an L1Md_T element in male (blue) and female (red) E13.5 PGCs. * E13.5 DNA methylation data (PBAT WT) is from Kobayashi et al. (Kobayashi et al. 2013).

Supplementary Figure 5: Trivalent silencing marks at ERVs in E13.5 PGCs. **(A)** Correlation between H3K9me3 and H3K27me3 enrichment at individual IAP (top) or ETn (bottom) elements in male (blue) and female (red) E13.5 PGCs. RPKM values were calculated from uniquely mapped reads only. **(B)** Relationship between element age (mappability) and enrichment of silencing histone marks at all retroelements (>100 copies) in male E13.5 PGCs. Black line: H3K9me3, red line: H3K27me3 and shaded area: Mean % DNA methylation (data from Kobayashi et al. 2013). Black shading: mappability (ENCODE, 100mers). **(C)** Relative enrichment of H3K9me3 (in PGCs and ESCs) and DNA methylation (in PGCs and blastocysts) in the flanks of IAP (top) and MERVL (bottom) subfamilies. PGC (Kobayashi et al. 2013) and blastocyst (Kobayashi et al. 2012) PBAT DNA methylation datasets were published previously. **(D)** Correlation between H3K9me3 and expression of individual ETn (left) and MLV

(right) elements in male (blue) and female (red) E13.5 PGCs. RPKM values were calculated from uniquely mapped reads only.

Supplementary Figure 6: Isolation and characterization of germline specific *Setdb1* KO E13.5 PGCs. (A) FACS plots illustrating the number of PGCs in E13.5, as measured by SSEA-1 staining, from single gonads isolated from male and female *Setdb1* KO and HET littermates. % SSEA-1 positive cells is indicated. **(B)** Map of the *Setdb1* locus revealing the locations of the *lox*P sites relative to the genic exons and *in silico* PCR analysis of RNA-seq data for E13.5 PGCs depicting *Setdb1* deletion efficiency using the *Tnap*-Cre strategy. Two *lox*P sites are situated in intronic regions resulting in the deletion of exons 15/ 16 (top). Bar graph shows relative transcript levels at deleted exons (15 and 16) and control exons (6, 19). Relative coverage for the same exons for PBAT data from male HET and KO E13.5 PGCs is also shown (inset). The average percentage decrease of coverage over exons 15 and 16 is indicated above the deleted exon bars.

Supplementary Figure 7: H3K9me3 and H3K27me3 depletion in *Setdb1* KO E13.5 PGCs. (A) Genome-wide H3K27me3 profiles in HET versus *Setdb1* KO E13.5 PGCs isolated from male littermates. Yellow gradient depicts the relative H3K9me3 enrichment level in HET PGCs. **(B)** qPCR validation of H3K9me3 enrichment at IAP and MERVL ERVs in matched Setdb1 HET and KO male and PGCs. Error bars show standard deviation of 3 biological replicates. **(C)**

Relationship between H3K9me3 enrichment levels (RPKM) at LINE1, ERVK and ERV1 subfamilies in E13.5 HET female PGCs and the percentage change in H3K9me3 enrichment upon *Setdb1* deletion. Unmarked ERVs (not shown) did show a small H3K9me3 level increase in Setdb1 KO PGCs, but remained below the background threshold (0.5 RPKM). **(D)** Relationship between H3K27me3 enrichment levels (RPKM) at LINE1, ERVK and ERV1 subfamilies in E13.5 HET male PGCs and the percentage change in H3K27me3 enrichment upon *Setdb1* deletion. Unmarked ERVs (not shown) did show a small H3K27me3 level increase in Setdb1 KO PGCs, but remained below the background threshold (0.3 RPKM). **(E)** Relationship between H3K27me3 enrichment levels (RPKM) at LINE1, ERVK and ERV1 subfamilies in E13.5 HET female PGCs and the percentage change in H3K27me3 enrichment upon *Setdb1* deletion. Unmarked ERVs (not shown) did show a small H3K27me3 level increase in Setdb1 KO PGCs, but remained below the background threshold (0.3 RPKM). **(F)** Relative expression of histone methyltransferases specific for H3K9 or H3K27, *Trim28* and readers of H3K9 methylation in male (left) and female (right) HET and *Setdb1* KO E13.5 PGCs.

Supplementary Figure 8: Effect of germline *Setdb1* deletion on DNA methylation and gene expression in E13.5 PGCs. (A) Histogram depicting genome-wide mean % DNA methylation (1kb bins) in male *Setdb1* KO and Het littermates. **(B)** Expression levels (RPKM) of genes involved in DNA methylation homeostasis in male (left) and female (right) HET and *Setdb1* KO E13.5 PGCs,

as determined by RNA-seq. Error bars show standard deviation between independent experiments. **(C)** Top panel shows expression levels (RPKM) in HET and *Setdb1* KO E13.5 PGCs of top 10 (male) or 12 (female) genes which are upregulated in male (left) and female (right) E13.5 PGCs (E13.5 > E12.5), as reported by Jameson et al. (Jameson et al. 2012). Bottom panel shows expression levels (RPKM) of top 10 (male) or 8 (female) genes that are down-regulated in E13.5 PGCs (E12.5 > E13.5). **(D)** Relationship between H3K9me3, H3K27me3 and mean % DNA methylation in *Setdb1* KO and Het littermates genome-wide (1kb bins), at unmarked genomic regions, or regions marked with H3K9me3 alone, H3K27me3 alone or both H3K9me3 and H3K27me3 (MACS, p-value 0.05). For % DNA methylation values, only CpGs with > 5X coverage were considered. Note that the increase in DNA methylation in the *Setdb1* KO is lowest at H3K9me3 alone and dual marked regions. **(E)** Relationship between H3K27me3 in control (HET) E13.5 PGCs and the change (Δ) in the % of methylation in these HET PGCs relative to PGCs isolated from their Setdb1 KO littermates. Data from 50,000 random regions (1kb bins) is shown. DNA methylation values were determined by PBAT.

Supplementary Figure 9: ERV de-repression in *Setdb1* KO E13.5 PGCs. (A) RNA-seq coverage (multi-match reads included) of the 19 most highly expressed ERV subfamilies (present at > 100 copies in the C57BL/6 genome) in *Setdb1* KO (light bars) and HET (dark bars) E13.5 PGCs isolated from male (blue) and female (red) littermates. Fold-change (Setdb1 KO/HET) is indicated next to the

bars. Note that specific ERVs show sex-specific expression patterns. **(B)** Comparison of the level of ERV derepression (subfamilies > 100 copies) in E13.5 *Setdb1* KO PGCs, as measured by Z-score values, from two independent experiments conducted on male (left) and female (right) Setdb1 KO and control (HET) littermates.

Supplementary Figure 10: Relationship between H3K9me3, H3K27me3 and expression of individual retroelements in HET versus *Setdb1* KO PGCs. **(A)** Relationship between H3K9me3 loss and derepression of individual MERVL (left) or MLV (right) elements in male and female HET versus *Setdb1* KO E13.5 PGCs, as measured by Z-score values. Multi-match reads were excluded from this analysis). **(B)** Relationship between H3K27me3 loss and derepression of individual IAPez (ERVK), MERVK10C (ERVK), MERVL (ERVL), ETn (ERVK), L1Md_GF (LINE1) and MLV(ERVL) elements in male and female HET versus *Setdb1* KO E13.5 PGCs.

Supplementary Figure 11: Genes upregulated in *Setdb1* KO E13.5 PGCs **(A)** Expression (RPKM) of all annotated ENSEMBL genes in HET versus *Setdb1* KO female E13.5 PGCs. 157 genes up-regulated in *Setdb1* KO PGCs (z-score KO/WT and KO/Het >1, RPKM KO/WT and KO/Het >1.5-fold) are highlighted in re(D) 31 genes down-regulated in *Setdb1* KO (z-score KO/WT and KO/Het <-1, RPKM KO/WT and KO/Het < 0.75-fold) are highlighted in green. **(B)** Gene Ontology analysis of genes upregulated in male and female Setdb1 KO PGCs.

Top 6 ontologies (> 5 genes in ontology) presented for each sex. Gray: below statistical significance. **(C)** Genome browser views, including RepeatMasker tracks, of H3K9me3 and RNA-seq coverage in male and female HET and *Setdb1* KO E13.5 PGCs. Upper panel shows the *Nfatc4* gene and the annotated ERVK10C element upstream of this locus. Lower panel shows the *Klk1b24* and *Klk1b3* genes, and a cluster of ERVs, including ERVK10C and ETnERV2 elements, upstream of these loci.

SUPPLEMENTARY METHODS

Isolation of PGCs

Isolated gonadal ridges were digested with 0.05% Trypsin at 4°C for 30min and pipetted into single cell suspensions, and cells were stained with a PE-conjugated SSEA-1 antibody (BD #560142). SSEA-1 positive (PGCs) and negative (Soma) cells from individual embryos were sorted directly into 20µL of nuclear isolation buffer (Sigma NUC-101) (ChIP-seq) or 20µL TRI reagent (Invitrogen AM9738) (RNA-seq) and snap frozen. Aliquots of 1,000 to 3,000 cells were stored for up to several months at -80°C.

Chromatin Immunoprecipitation

Chromatin was prepared from single or pooled aliquots of 3,000 cells digested with 1U/µL of micrococcal nuclease (NEB M0247) at 25°C for 5min and re-suspended in ChIP Buffer (10mM Tris-HCl PH8.0, 150mM NaCl, 2mM EDTA, 0.1% Triton X-100, 0.1% deoxycholate, proteinase inhibitor cocktail and PMSF). Chromatin was pre-cleared with protein A Dynabeads® and a 95 cell-equivalent was removed as input. The rest of the chromatin was divided into 3 aliquots (~950 cells-equivalents each), and H3K9me3 (Active Motif #39161), H3K27me3 (Diagenode pAb-069-050) and pan-H3 (Sigma I8140) antibody-beads complexes were used for immunoprecipitation (overnight at 4°C). The chromatin-antibody-beads conjugates was then washed twice with Low Salt Wash buffer (20mM

PH8.0 Tris-Cl, 0.10% SDS, 1% Triton X-100, 2mM EDTA, 150mM NaCl) and twice with High Salt Wash Buffer(20mM PH8.0 Tris-Cl, 0.10% SDS, 1% Triton X-100, 2mM EDTA,500mM NaCl), and DNA was eluted at 65°C for 2h in Hot Elution Buffer (100mM NaHCO3, 1% SDS). DNA was then extracted using phenol-chloroform (Sigma) followed by ethanol precipitation. 10% aliquots were removed for qPCR validation and the remainder for library construction.

RNA extraction and double stranded cDNA preparation

Total RNA was extracted from frozen 1,000 cells aliquots using TRIzol (Invitrogen, AM9738) according to the manufacturer's manual. RNAse inhibitor (40U per sample, Fermentas, RiboLock RNase Inhibitor, EO0381) was added to resulting RNA. Residual genomic DNA in RNA was removed by DNase I (Promega, RQ1 RNase-Free DNase, M6101) according to the manual, and ribosomal RNA was depleted using riboMinus (Invitrogen, RiboMinus Transcriptome Isolation Kit, K155002) according the low-input protocol in manufacturer's manual. First strand cDNA synthesis was carried using Superscript III (Invitrogen 18080-093) with T4 protein 32 and a combination of random 15-mers and oligo dT (NEB), followed by second strand cDNA synthesis using Klenow polymerase in the presence of RNaseH. A 10% fraction of double stranded cDNA was used for qPCR validation, and the rest were fragmented for 20 minutes (High Power mode, 30s on and 30s off for 20min) using a BioRuptor (Diagenode) for library construction.

ChIP and RNA sequencing analysis

For analysis of relative ChIP or RNA enrichment at unique loci, unique reads with a MapQ > 5 (uniquely mapped) were utilized. Multi-aligned reads were included for calculating the relative ChIP enrichment at agglomerated transposable elements (TEs). As they are not properly mapped in the B6 genome (Fennelly et al. 1996; F U Reuss 1996) IAPEy, IAPLTR3, RLTR5 and MURVY elements were not included in our analyses. Normalization of relative ChIP enrichment was calculated as Reads Per Kilobase per Million mapped reads (RPKM) (Mortazavi et al. 2008; Pepke et al. 2009). For pair wise sample comparisons, a Z-score was calculated:

$Z\text{-}\ score = (RPKM_A\ -\ RPKM_B)/\sqrt{(RPKM_A + RPKM_B)}$, where $RPKM_A$ and $RPKM_B$ are RPKMs in the region of interest of A and B samples, respectively. For RNA-seq analysis, RPKM values were calculated for exonic reads only. To calculate the proportion of the genome covered by H3K9me3 and H3K27me3 in E13.5 PGCs, peaks were called using MACS2 (Zhang et al. 2008; Feng et al. 2012) with a p-value of 0.05. False peaks detected in our pan-H3 ChIP-seq library were subtracted. Gene Ontology (GO) analysis was conducted on InnateDB, with Benjamini Hochberg correction of GO enrichment p-values.

SUPPLEMENTARY REFERENCES

Reuss FU, Frankel WN, Moriwaki K, Shiroishi T, Coffin JM. 1996. Genetics of intracisternal-A-particle-related envelope-encoding proviral elements in mice. *Journal of Virology* **70**: 6450.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**: 1728–1740.

Fennelly J, Harper K, Laval S, Wright E, Plumb M. 1996. Co-amplification to tail-to-tail copies of MuRVY and IAPE retroviral genomes on the Mus musculus Y chromosome. *Mamm Genome* **7**: 31–36.

Jameson SA, Natarajan A, Cool J, DeFalco T, Maatouk DM, Mork L, Munger SC, Capel B. 2012. Temporal Transcriptional Profiling of Somatic and Germ Cells Reveals Biased Lineage Priming of Sexual Fate in the Fetal Mouse Gonad ed. G.S. Barsh. *PLoS Genetics* **8**: e1002575.

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, et al. 2012. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks ed. W. Reik. *PLoS Genetics* **8**: e1002440.

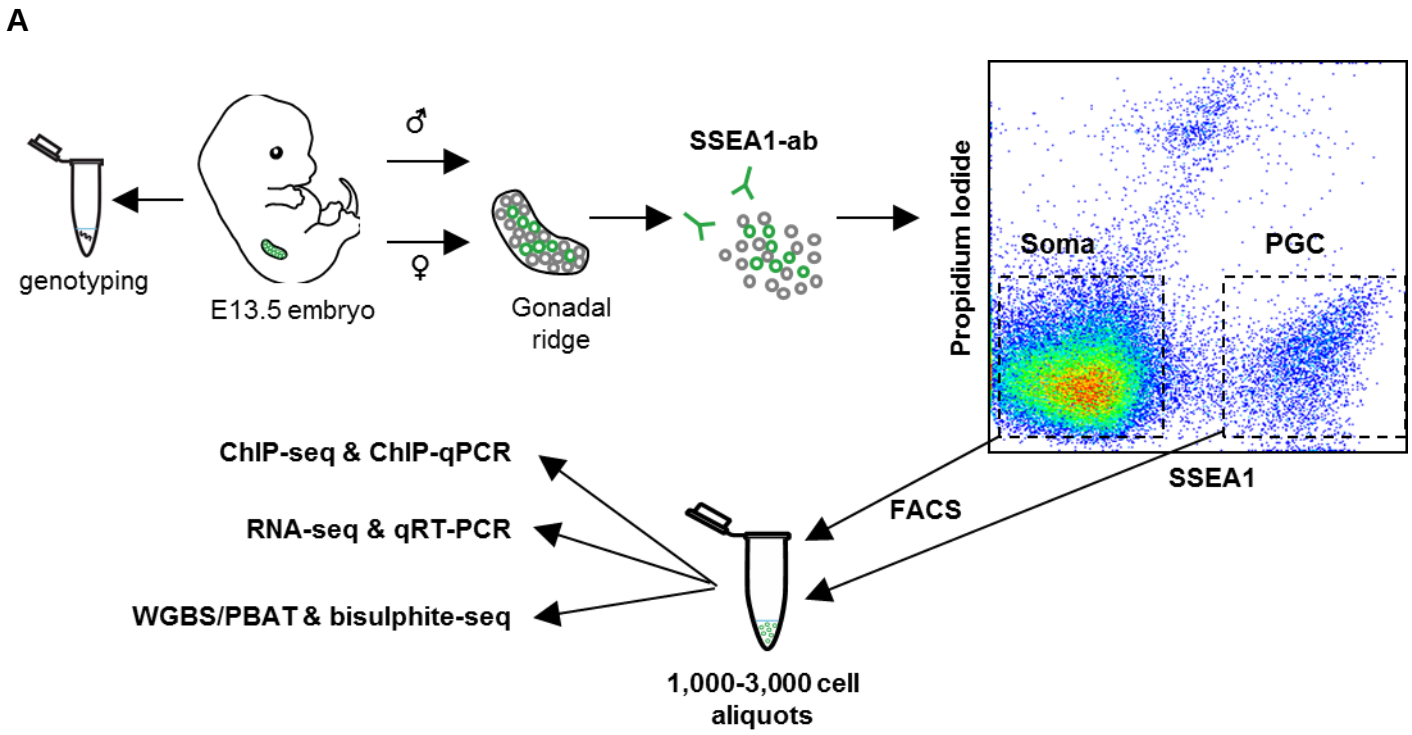
Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, et al. 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research* **23**: 616–627.

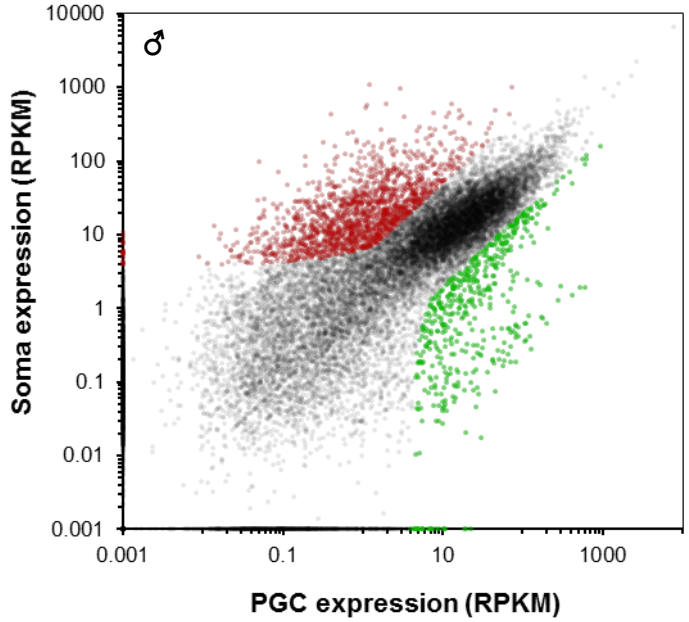
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621–628.

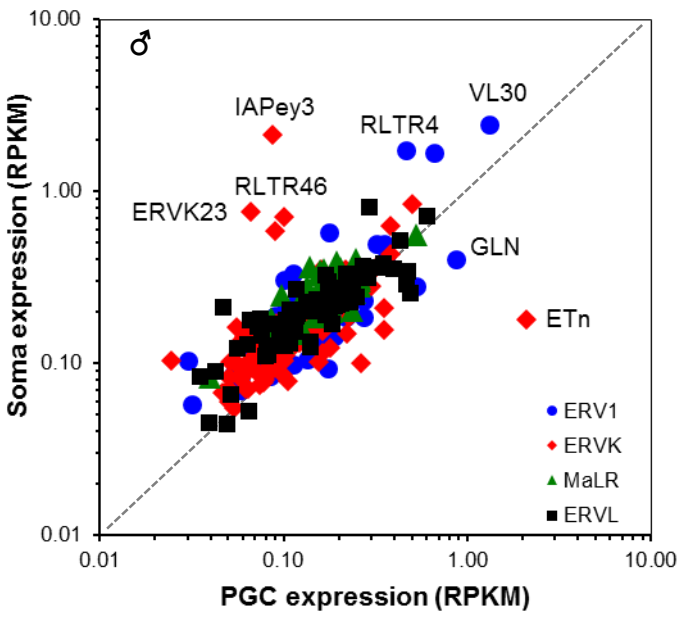
Pepke S, Wold B, Mortazavi A. 2009. Computation for ChIP-seq and RNA-seq studies. *Nature Methods* **6**: S22–S32.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**: R137.

Supplementary Figure 1

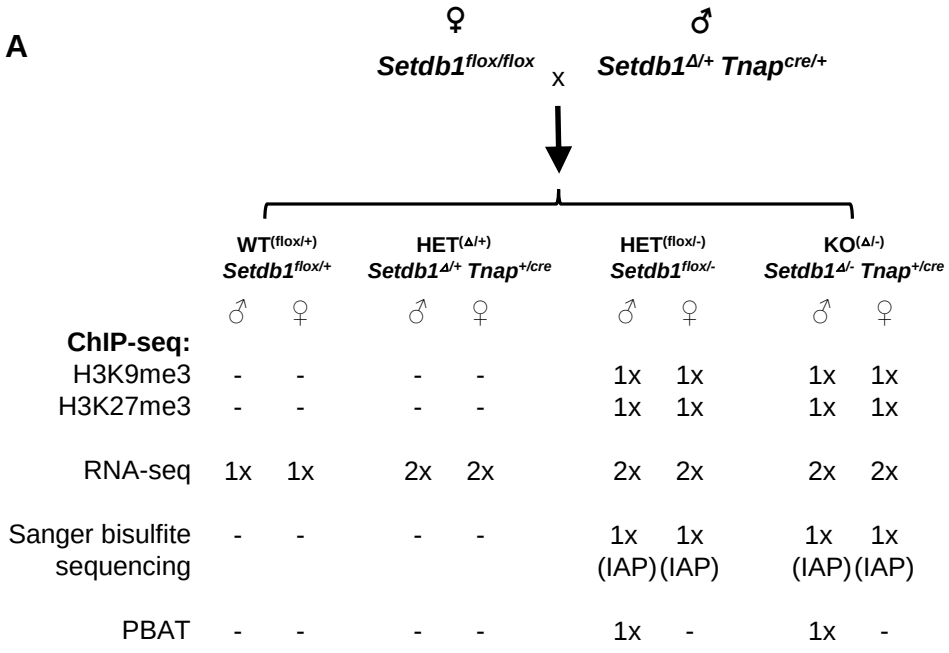




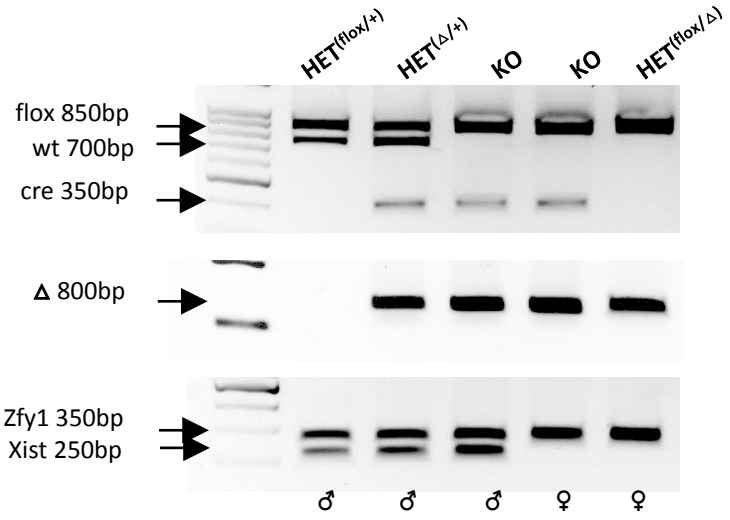


Supplementary Figure 2

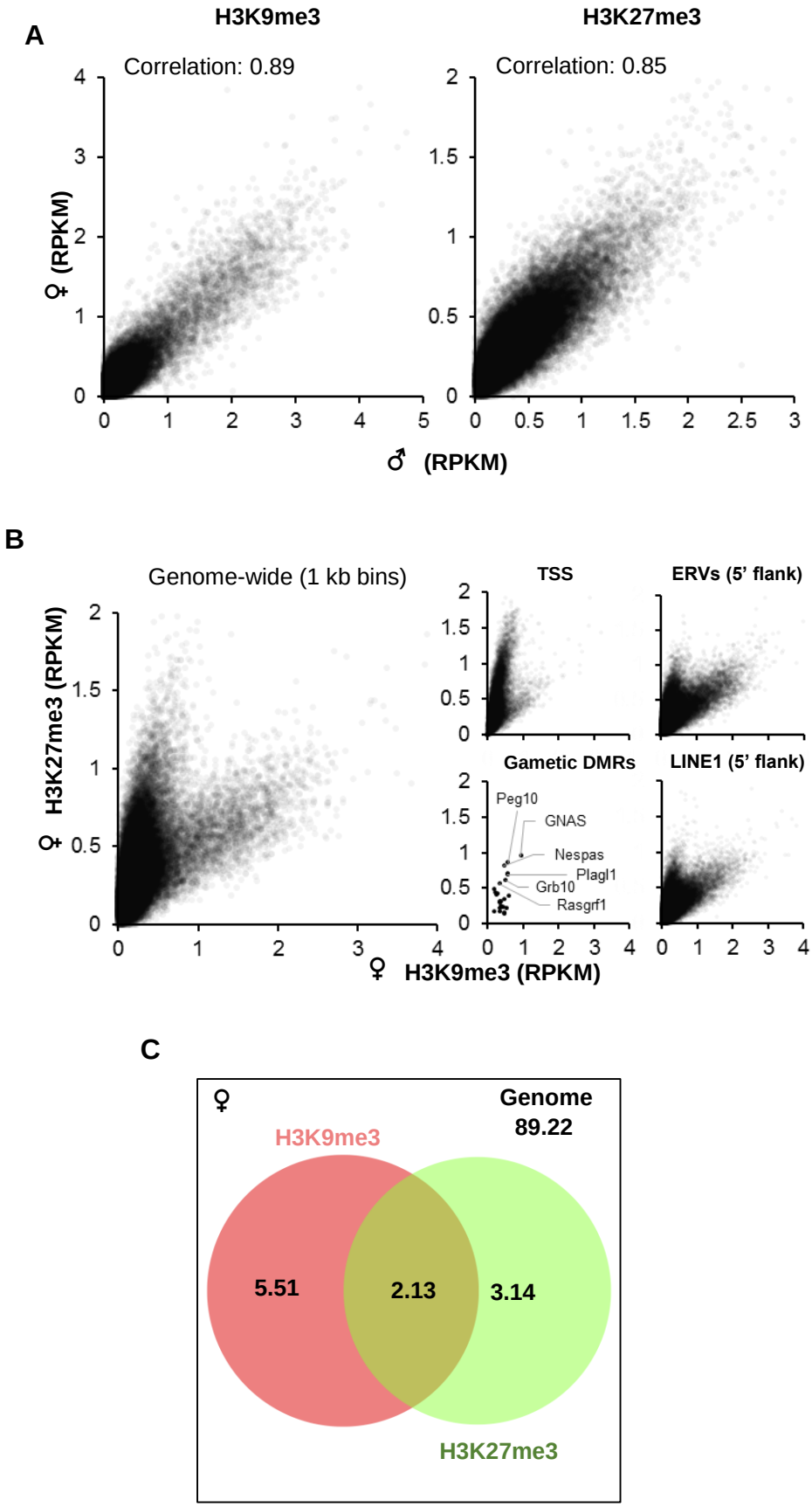
A



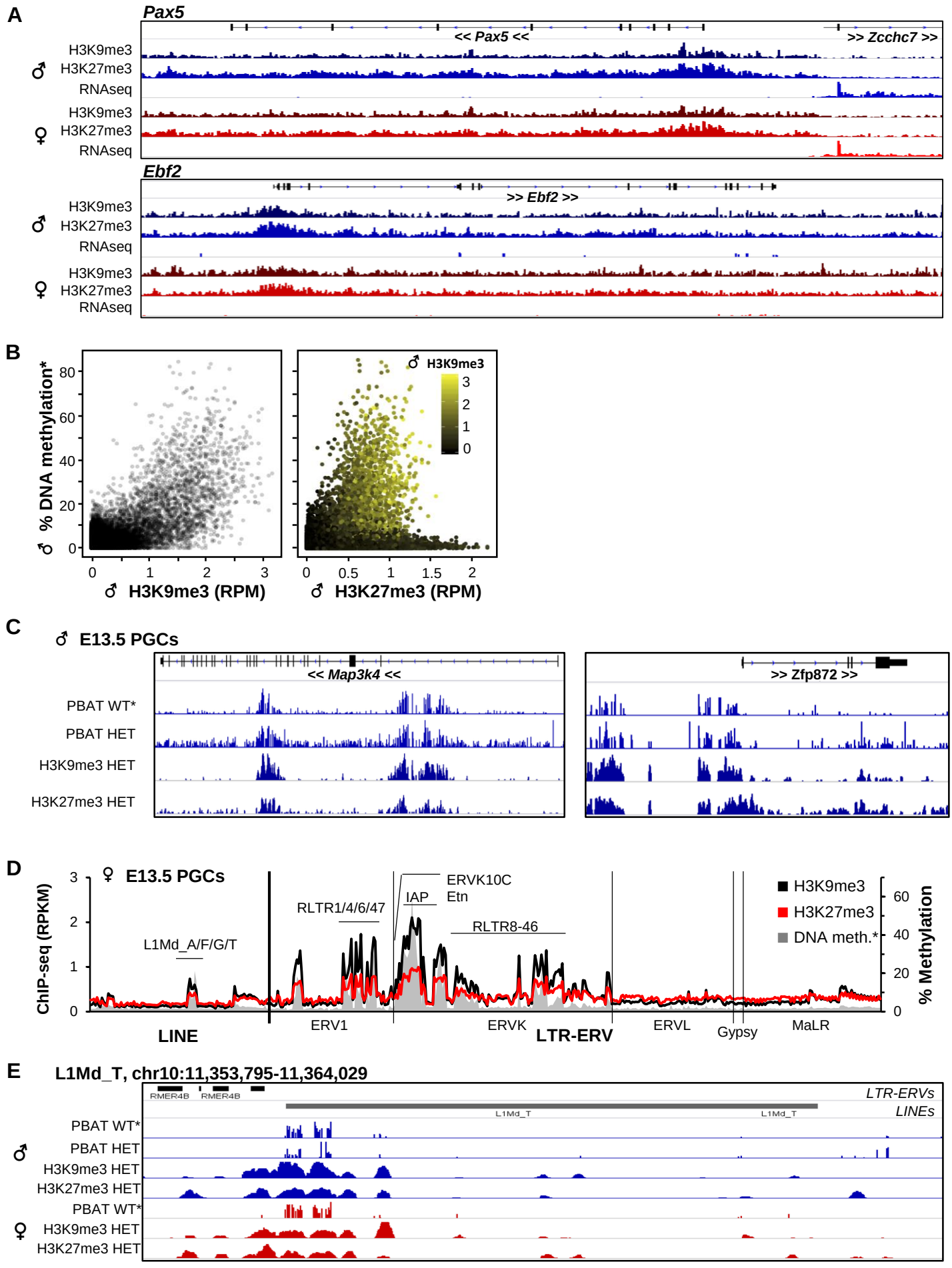
B



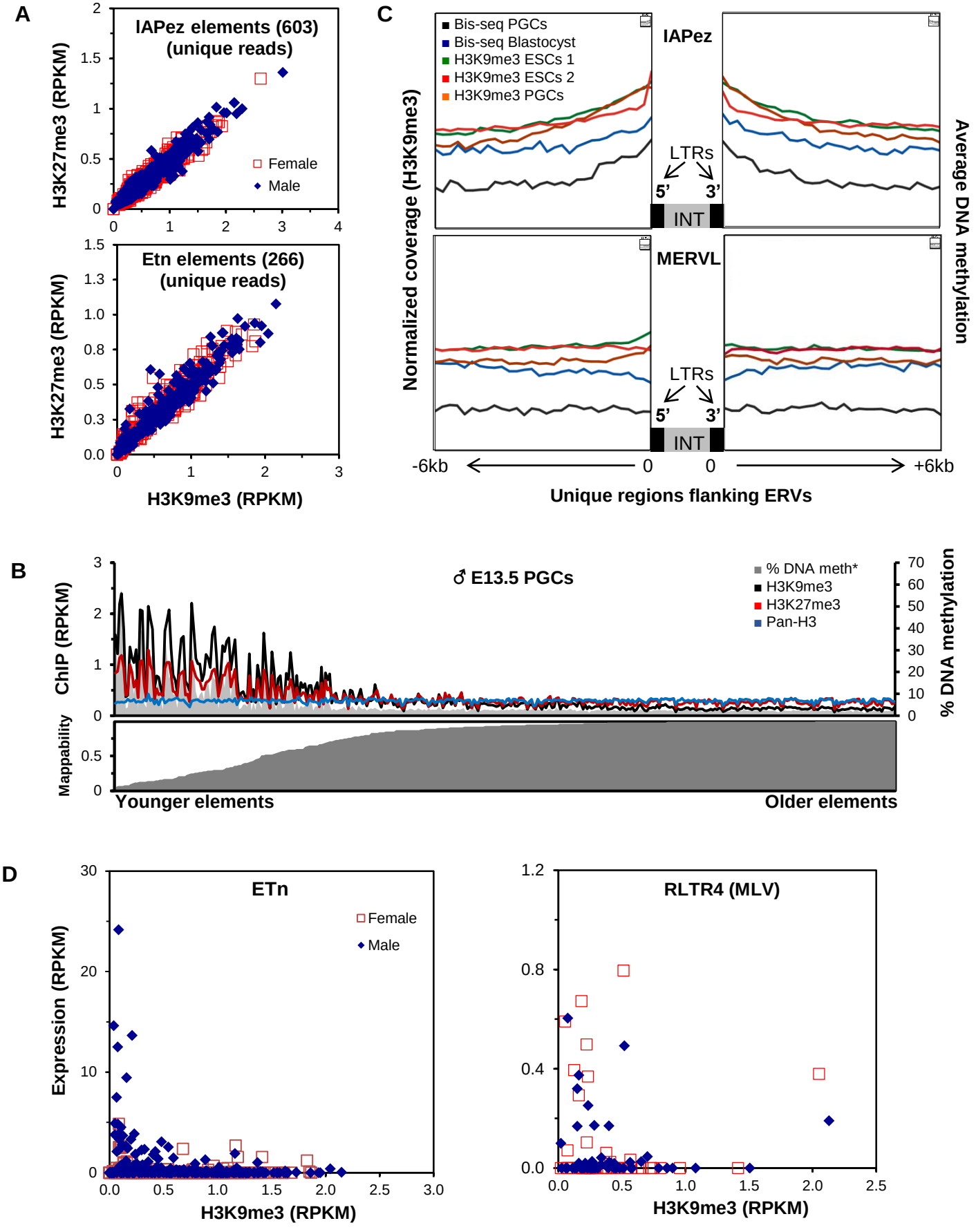
Supplementary Figure 3



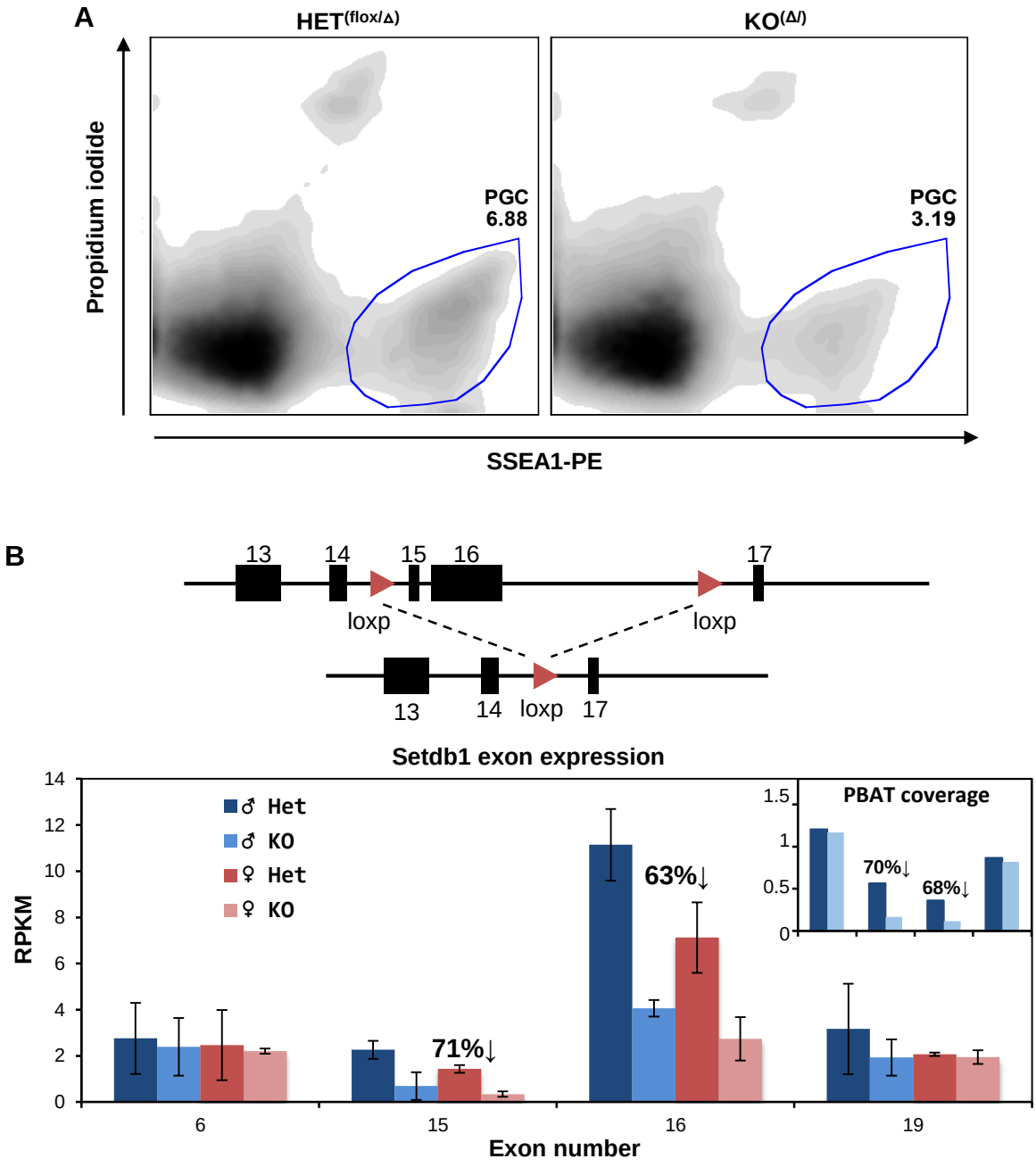
Supplementary Figure 4



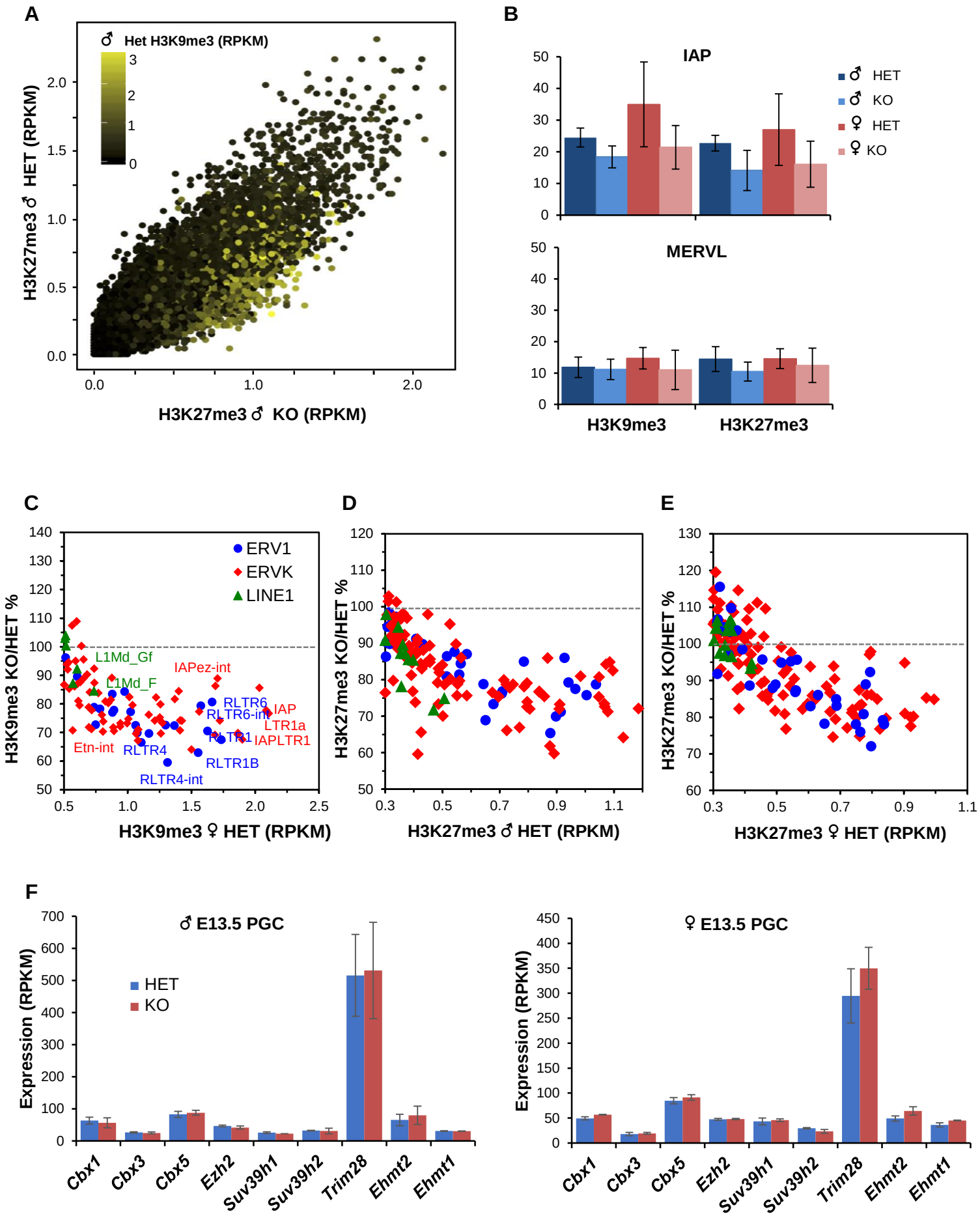
Supplementary Figure 5



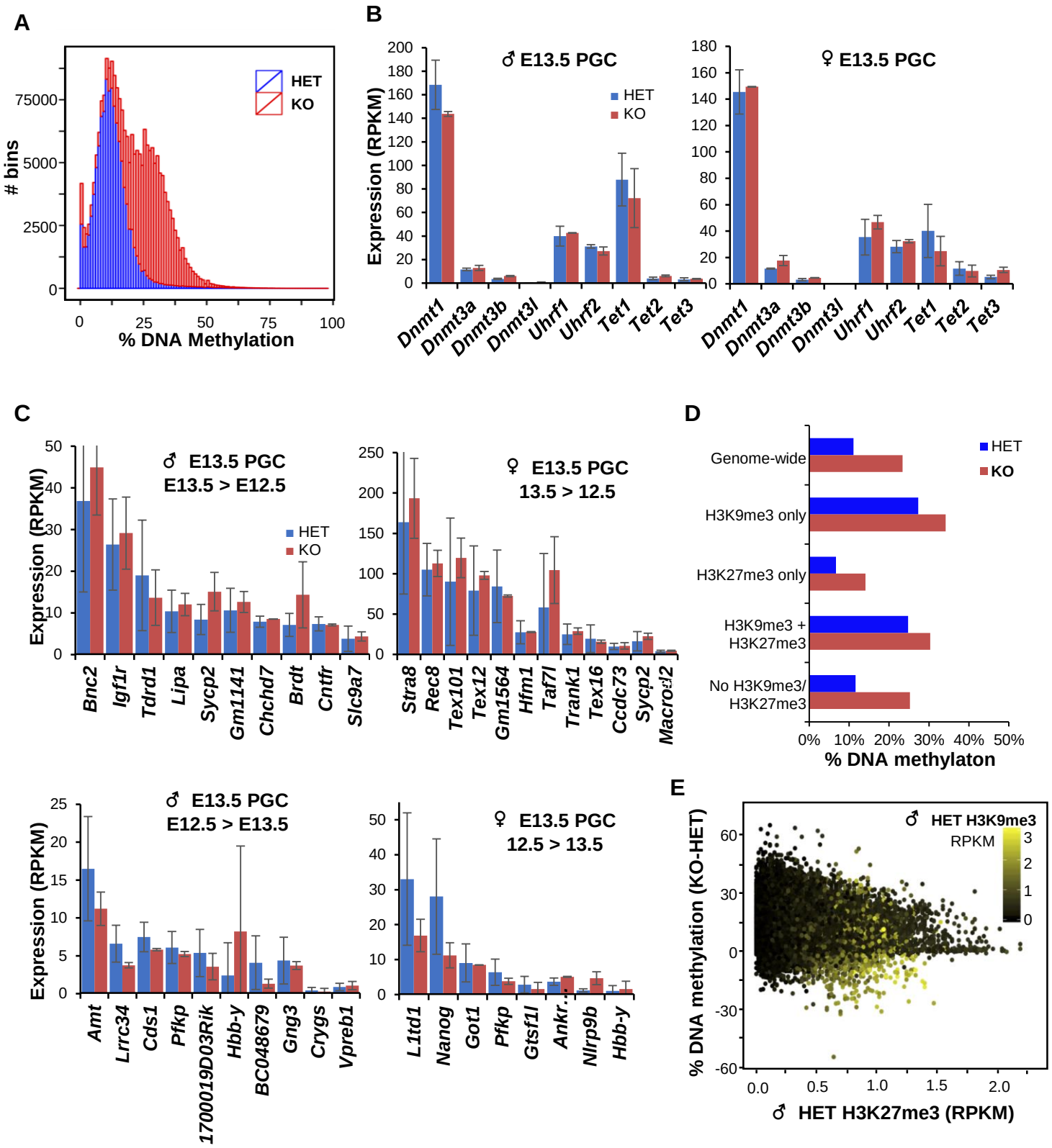
Supplementary Figure 6



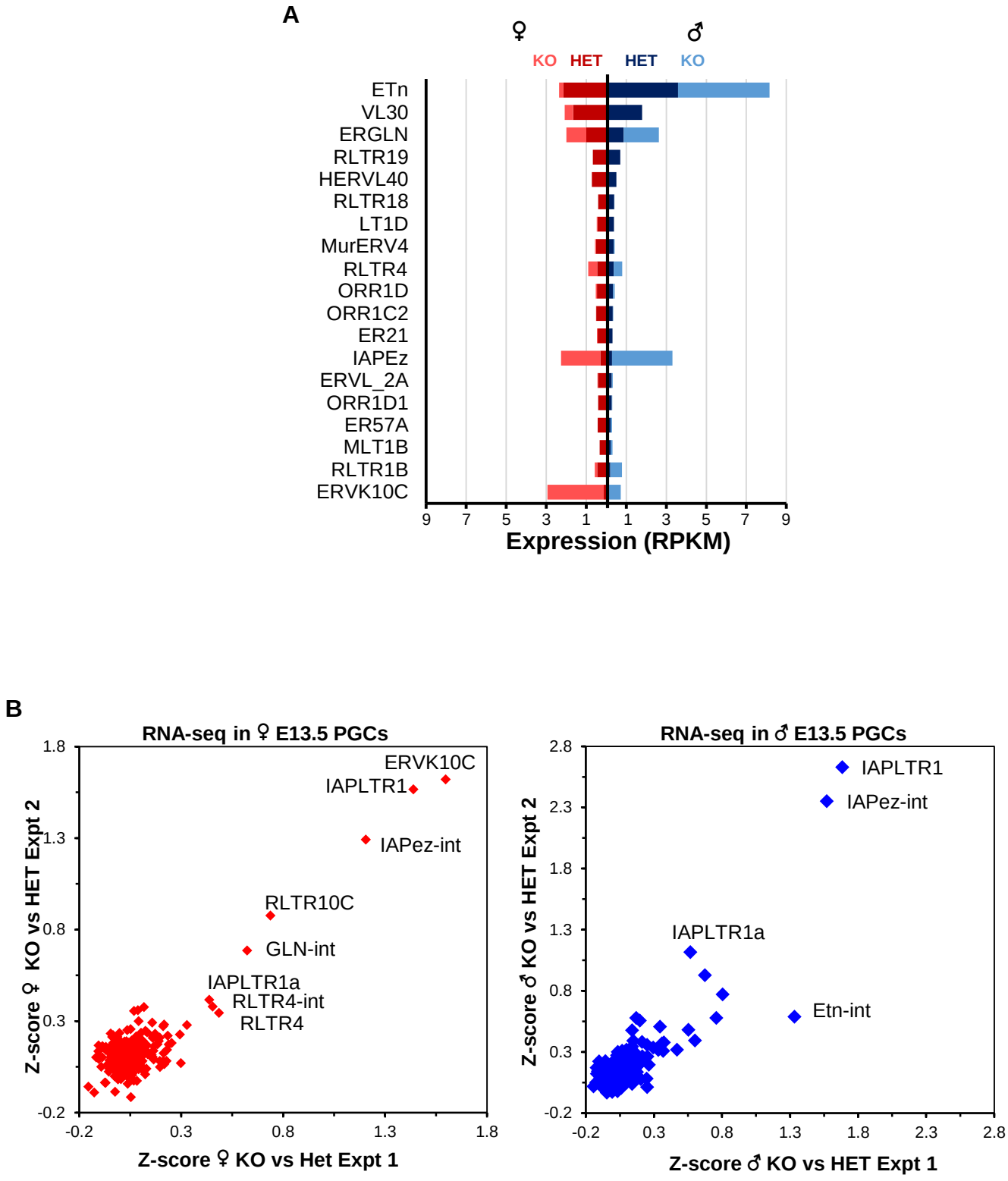
Supplementary Figure 7



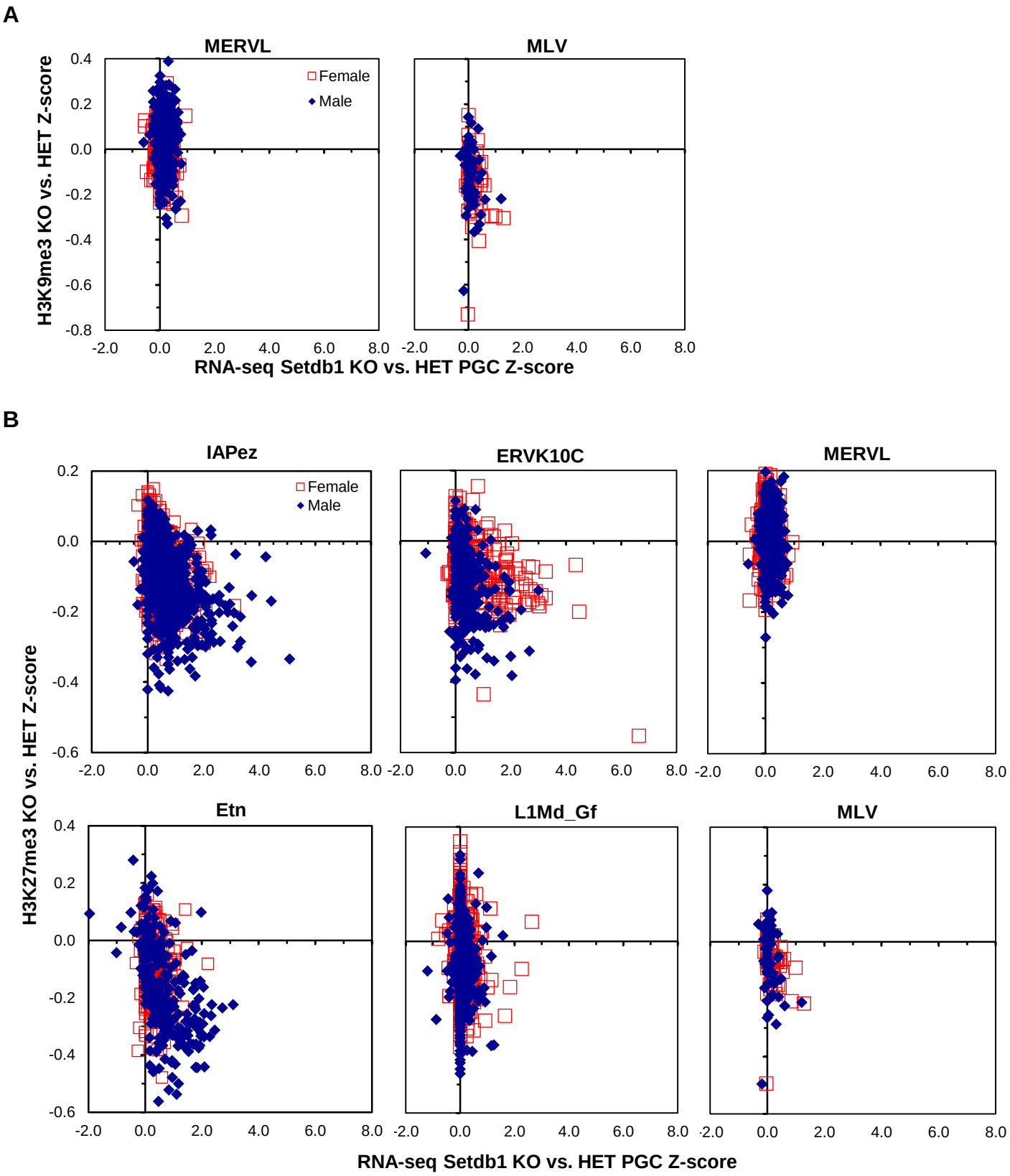
Supplementary Figure 8



Supplementary Figure 9

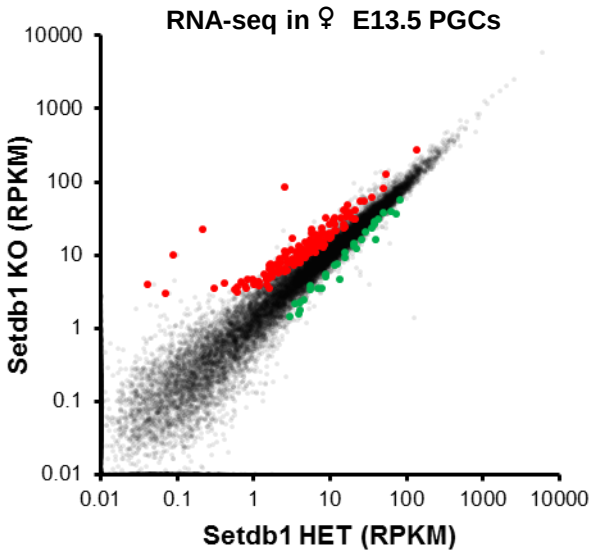


Supplementary Figure 10



Supplementary Figure 11

A



B

Genes upregulated in ♂ KO PGCs

GO term Name	GO term Id	p-value (corrected)	Gene Symbols
nucleic acid binding (MF)	GO:0003676	0.007	D1Pas1 ; Gm10351 ; Gm13145 ; Gm13154 ; Gm13242 ; Gm3376 ; Gm5698 ; Gm5699 ; Gm6712 ; Gm7221 ; Hltf ; Zbtb39 ; Zfp560 ; Zfp600 ; Zfp618 ; Zfp936 ;
iron ion binding (MF)	GO:0005506	0.014	Alox12e ; Alox15 ; Cyp17a1 ; Cyp2b23 ; Ogfod2 ; Plod2 ;
cellular component (CC)	GO:0005575	0.017	4933403G14Rik ; AI481877 ; Acap3 ; Atxn1l ; C130073F10Rik ; Cdh19 ; Clca4 ; Cox7b2 ; Crispld1 ; Cyp2b23 ; Dnahc6 ; Dph5 ; Efhc2 ; Gm10351 ; Gm10491 ; Gm10760 ; Gm1110 ; Gm12794 ; Gm13145 ; Gm13154 ; Gm13242 ; Gm5698 ; Gm5699 ; Gm7221 ; Gm7579 ; Gm9112 ; Gramd1c ; Klk1b24 ; Lysmd2 ; Mup6 ; Ogfod2 ; Pilrb2 ; Sectm1a ; Tex13 ; Tex15 ; Trim52 ; Ttc18 ; Usp9y ; Zbtb39 ; Zfp560 ; Zfp600 ; Zfp936 ;
lipoxygenase activity (MF)	GO:0016165	0.020	Alox12e ; Alox15 ;
biological process (BP)	GO:0008150	0.052	1700029P11Rik ; 4933403G14Rik ; 9430020K01Rik ; AI481877 ; Acap3 ; Atxn1l ; C130073F10Rik ; Ccnb3 ; Cdh19 ; Clca4 ; Cox7b2 ; Crispld1 ; Cyp2b23 ; Dnahc6 ; Dph5 ; Efhc2 ; Fndc4 ; Gm10351 ; Gm10491 ; Gm10760 ; Gm1110 ; Gm12794 ; Gm13145 ; Gm13154 ; Gm13242 ; Gm5698 ; Gm5699 ; Gm7221 ; Gm7579 ; Gm9112 ; Gramd1c ; Lysmd2 ; Mup6 ; Pilrb2 ; Tex13 ; Trim52 ; Ttc18 ; Zbtb39 ; Zfp560 ; Zfp600 ; Zfp618 ; Zfp936 ; Zscan4d ;
male gonad development (BP)	GO:0008584	0.053	Cyp17a1 ; Nkx3-1 ; Ren1 ;

Genes upregulated in ♀ KO PGCs

GO term Name	GO term Id	p-value (corrected)	Gene Symbols
extracellular matrix (CC)	GO:0031012; GO:0005578	0.00099559	Adamts9 ; Col14a1 ; Egfl6 ; Emilin1 ; Htra1 ; Lama5 ; Ltbp1 ; Pxdn ; Serpinf1 ; Wnt7a ;
definitive hemopoiesis (BP)	GO:0060216	0.00625453	Hoxb3 ; Hoxb4 ; Meis1 ;
urogenital system development (BP)	GO:0001655	0.01158553	Pax2 ; Pax8 ; Srd5a1 ;
embryonic skeletal system morphogenesis (BP)	GO:0048704	0.01305687	Hoxb3 ; Hoxb4 ; Hoxb5 ; Hoxb7 ; Sox11 ;
identical protein binding (BP)	GO:0042802	0.02523512	Amotl1 ; Cldn3 ; Emilin1 ; Grb7 ; Grik5 ; Inhba ; Kctd17 ; Sh3gl3 ; Trim8 ;
signal transduction (BP)	GO:0007165	0.02684925	Angptl6 ; Arhgap33 ; Arhgap39 ; Arnt2 ; Chrnb4 ; Gm266 ; Grb7 ; Nfkb2 ; Plce1 ; Plxnb1 ; Rac3 ; Rrad ; Unc5d ; Wnt7a ;

C