**SUPPLEMENTAL INFORMATION**

Maintenance of CTCF- and transcription factor-mediated interactions from gametes to the early mouse embryo

Yoon Hee Jung, Hannah B. Gold, Isaac Kremsky, M. Jordan Rowley, Kanchana Punyawai, Alyx Buonanotte, Xiaowen Lyu, Brianna J. Bixler, Anthony W. S. Chan, and Victor G. Corces

## SUPPLEMENTAL FIGURES

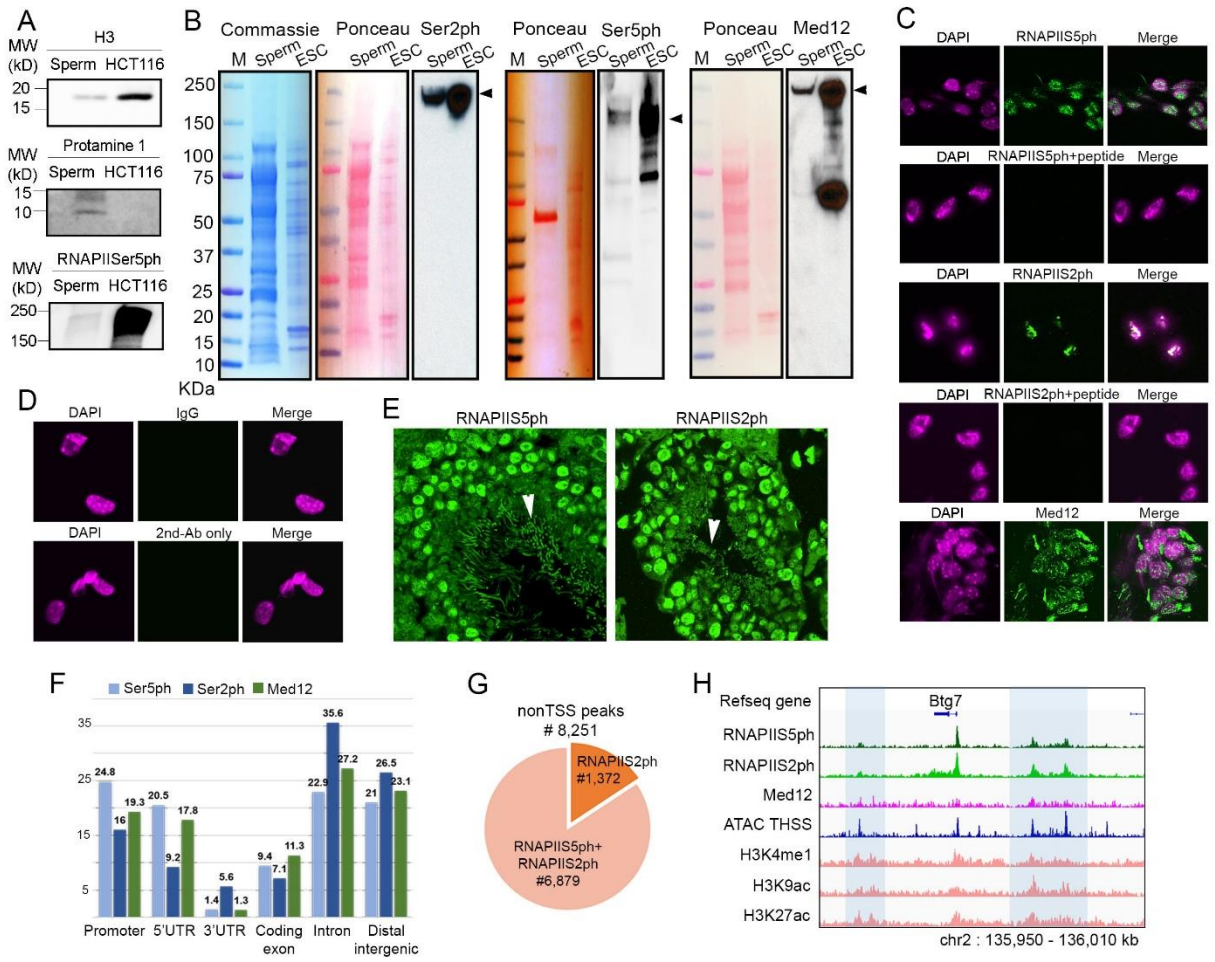



**Figure S1. RNAPII and Med12 are present in sperm. Related to Figure 1**

(A) Quality control of sperm preparations. Western blot analyses showing the presence of much lower levels of histone H3 and RNAPIISer5ph in sperm compared to somatic HCT116 cells, and the presence of Protamine 1 in sperm but not HCT116 cells. An equal number of sperm and HCT116 cells was loaded in each lane.

(B) Western blot analysis showing the presence of RNAPIISer2ph, RNAPIISer5ph and Med12 in mouse sperm. Sperm lanes contain 6-fold more genome equivalents than lanes with ESCs.

(C) Immunofluorescence confocal microscopy showing the presence of RNAPIISer5ph, RNAPIISer2ph and Med12 in the mouse decondensed sperm. Staining with RNAPIISer5ph and RNAPIISer2ph is abolished by incubation with the same peptide used to prepare the antibodies.

(D) Immunofluorescence confocal microscopy using a non-specific IgG or the secondary antibody only does not detect any signal.

(E) Immunofluorescence confocal microscopy of testis cryosections showing the presence of RNAPIISer5ph and RNAPIISer2ph in mature sperm (white arrows) in the seminiferous tubules.

(F) Genome-wide distribution of ChIP-seq peaks from RNAPIISer5ph, RNAPIISer2ph and Med12 called by MACS.

(G) Number of distal ChIP-seq peaks of RNAPIISer5ph and RNAPIISer2ph located further than +/-500 bp from TSS sites.
(H) Track view of RNAPIISer5ph, RNAPIISer2ph and Med12 ChIP-seq data present in potential enhancers defined by enhancer-related histone modifications (highlighted by light blue stripes).

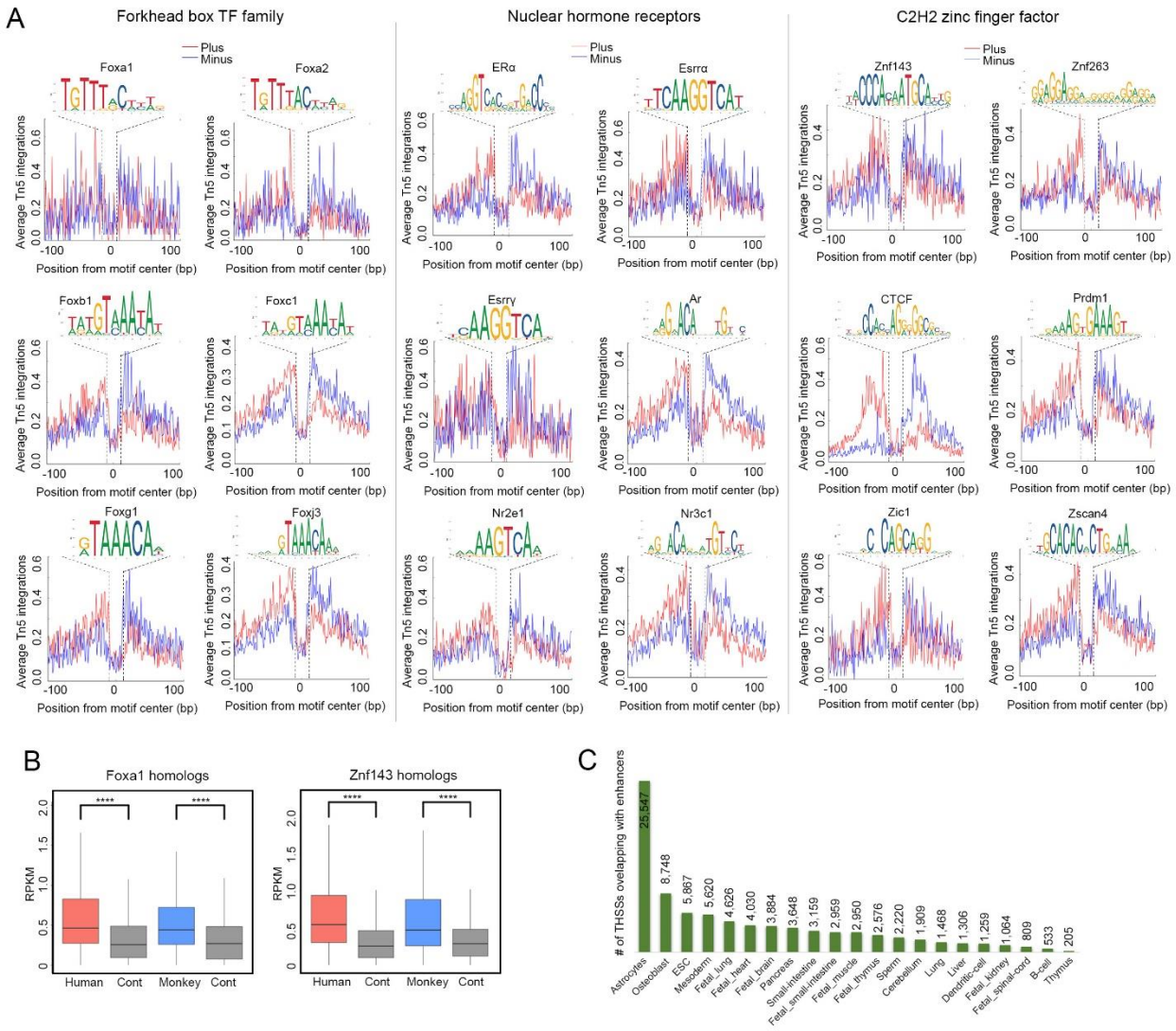



**Figure S2. Transcription factor accessible sites are present in syntenic regions of the mouse, monkey and human genomes. Related to Figure 2**

(A) Average footprint profiles of mouse omniATAC-seq distal THSSs (non TSSs) showing significant enrichment of DNA footprints for numerous transcription factors.

(B) Boxplots showing the RPKM of ATAC-seq THSSs in human and monkey sperm at regions orthologous to distal accessible regions in mouse containing Foxa1 or Znf143 motifs.

(C) Overlap of human ATAC-seq THSSs (non TSS) with known tissues enhancers.

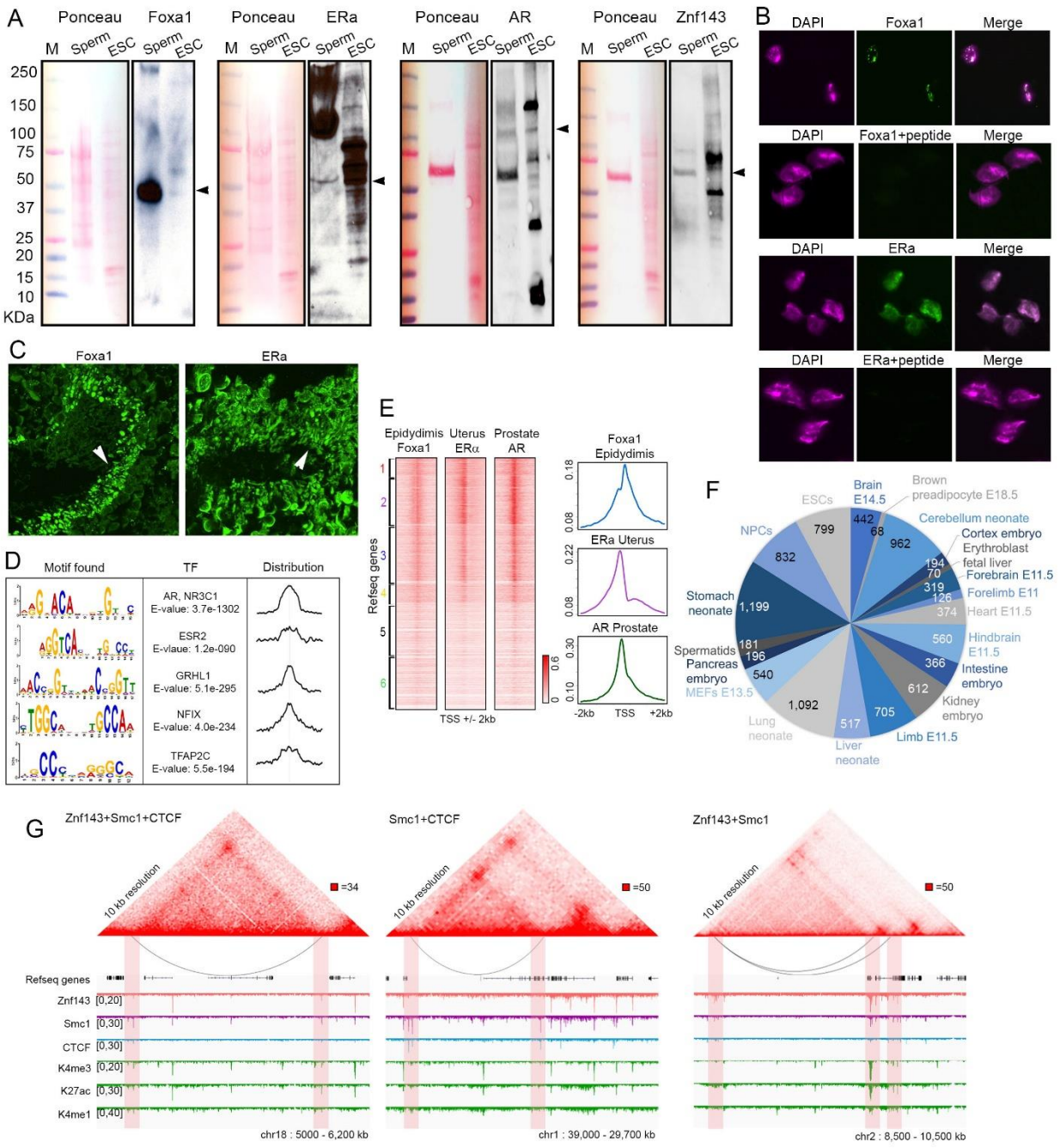



**Figure S3. Foxa1, ERα, an AR are present in sperm. Related to Figure 3**

(A) Western analysis showing the presence of Foxa1, ERα, AR, and Znf143 in mouse sperm.

(B) Immunofluorescence confocal microscopy detecting Foxa1 and ERα in mouse decondensed sperm. Staining with Foxa1 and ERα is abolished by incubation with the same peptides used to prepare the antibodies.

(C) Immunofluorescence confocal microscopy of testis cryosections showing the presence of RNAPIISer5ph and RNAPIISer2ph in mature sperm (white arrows) in the seminiferous tubules.

(D) Analysis of transcription factor motifs by MEME-ChIP found at Foxa1 peaks.

(E) Distribution of Foxa1, ERα, and AR at TSSs in different tissues of adult mice.

(F) Distal Foxa1 sites in mouse sperm are present at enhancers active in different tissues of the embryo during mouse development.
(G) Juicebox browser views of sperm Hi-C showing significant interaction mediated by Znf143, Smc1 and/or CTCF present at anchors. Lower panels depict arc views of significant interactions defined by FitHi-C from sperm Hi-C. IGV views of the distribution of Znf143, Smc1, CTCF, H3K4me3, H3K27ac, and H3K9ac ChiP-seq are also shown.

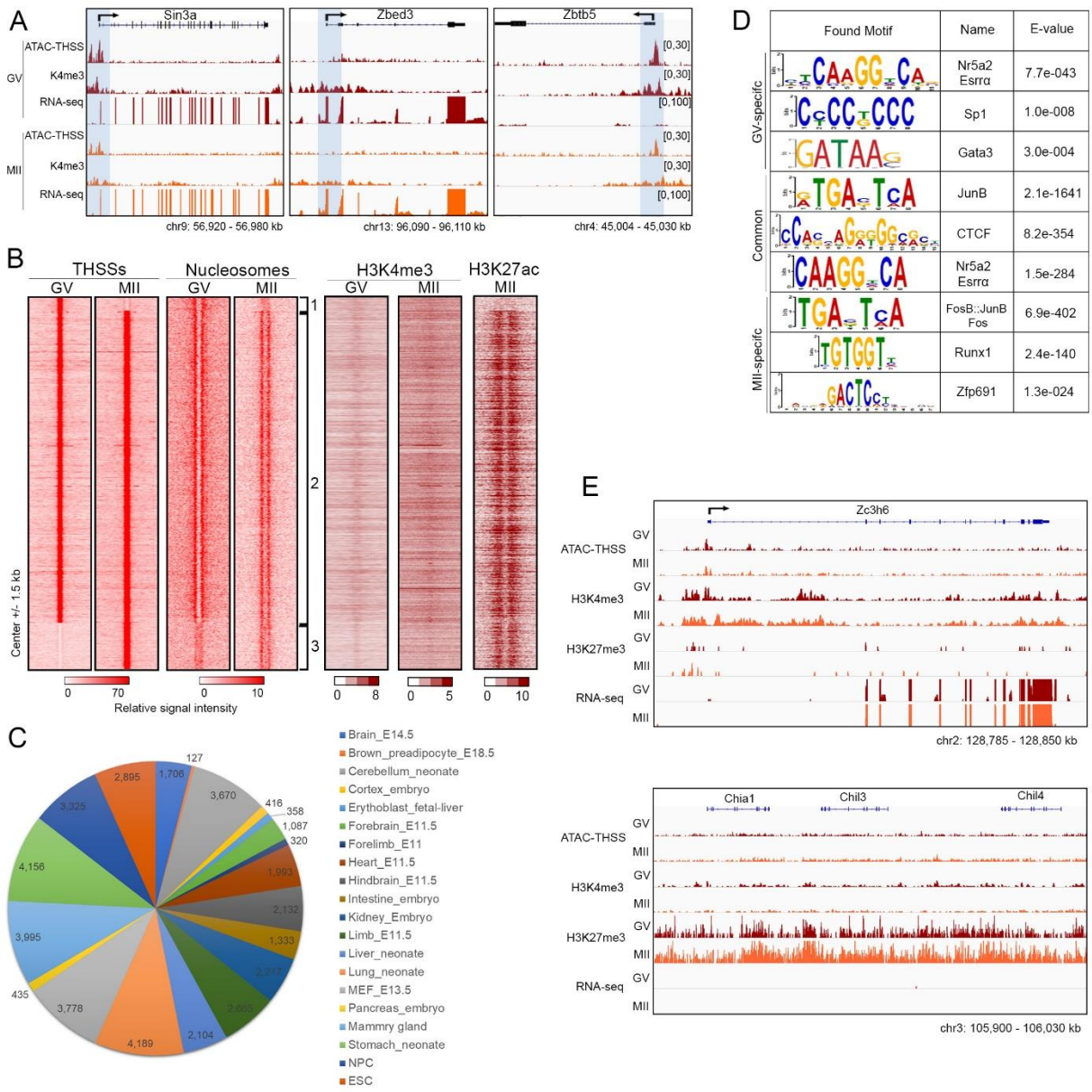



**Figure S4. Chromatin accessibility in mouse oocytes. Related to Figure 4**

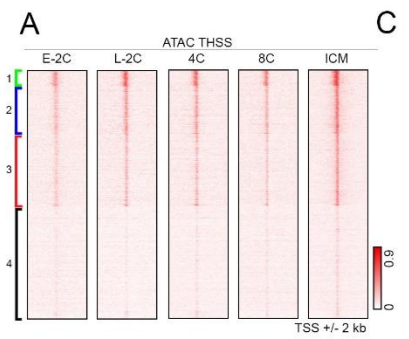
(A) Track view of ATAC-seq THSSs, H3K4me3, and RNA-seq at specific genes in GV and MII oocytes. TSSs are highlighted.

(B) Heatmaps showing H3K4me3 ChIP-seq in GV and MII oocytes, and H3K27ac ChIP-seq in MII oocytes at distal ATAC-seq THSSs present in GV and MII oocytes. Distribution of nucleosomes around the same THSSs are also shown.

(C) Overlap between cluster 2 THSSs shown in Figure S4B and enhancers characterized in various embryonic tissues.

(D) TF motifs identified at distal ATAC-seq THSSs (non TSSs) of GV and MII oocytes at each of the clusters shown in Figure 4B.

(E) Track views of ATAC-seq THSSs, H3K4me3, H3K27me3, and RNA-seq of GV and MII oocytes at specific examples corresponding to genes in the two clusters shown in Figure 4J.




C

Coincidence of promoter accessibility in sperm, oocyte, and preimplantation embryo

| | | Sperm (FPKM>1) | Sperm (FPKM<1) | Oocyte (FPKM>1) | Oocyte (FPKM<1) |
|---|---|---|---|---|---|
| ATAC-seq | Sperm | 1 | 1 | 0.907105 | 0.723634 |
| | MIIoocyte | 0.834682 | 0.835094 | 1 | 1 |
| DNase-seq | PN3_Pat | 0.792678 | 0.656945 | 0.801954 | 0.611202 |
| | PN3_Mat | 0.732614 | 0.737348 | 0.748135 | 0.698497 |
| | PN5_Pat | 0.697393 | 0.771559 | 0.710746 | 0.729508 |
| | PN5_Mat | 0.776334 | 0.721583 | 0.795293 | 0.68429 |
| | Early-2C | 0.504535 | 0.826738 | 0.519893 | 0.806148 |
| | 4C | 0.678434 | 0.802144 | 0.692274 | 0.75929 |
| | 8C | 0.874561 | 0.681066 | 0.87913 | 0.613934 |
| | Morula | 0.927597 | 0.600189 | 0.928242 | 0.530738 |
| ATAC-seq | Early-2C | 0.497589 | 0.847076 | 0.51492 | 0.827732 |
| | Late-2C | 0.631364 | 0.833675 | 0.653375 | 0.805464 |
| | 4C | 0.665604 | 0.783383 | 0.691741 | 0.763661 |
| | ICM | 0.666258 | 0.901151 | 0.692362 | 0.865574 |

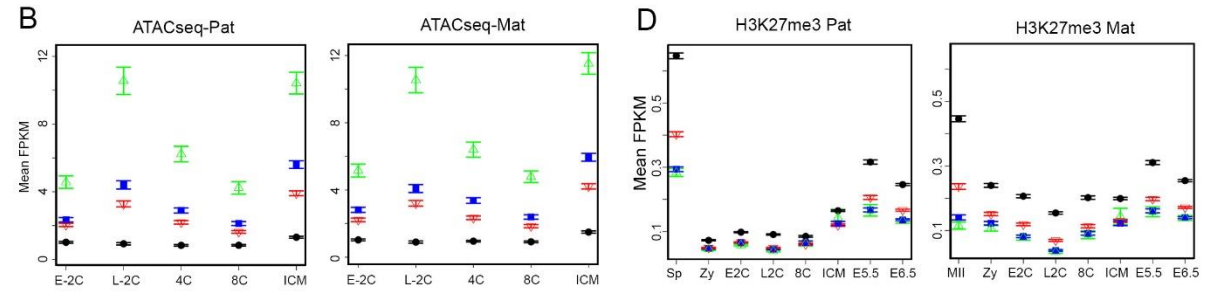



**Figure S5. Promoter accessibility in mouse preimplantation embryos. Related to Figure 5**

(A) Heatmap showing ATAC-seq THSS signal at TSSs in preimplantation embryos at the same clusters of promoters shown in Figure 5B.

(B) Plots showing average allele-specific signal of ATAC-seq THSSs in early embryos at the same clusters of promoters shown in Figure 5B.

(C) Table showing the coincidence of promoter accessibility between sperm, oocytes, and preimplantation embryos.

(D) Plots showing average allele-specific signal of H3K27me3 ChIP-seq at the TSS clusters defined in Figure 5A. The paternal signal is shown on the left side of each panel, and maternal signal on the right. E5.5 and E6.5 signify embryonic stages 5.5 and 6.5 days after fertilization, respectively. The number of TSSs present in each cluster is as follows: Cluster 1, 594; Cluster 2, 1809; Cluster 3, 2792; Cluster 4, 4525. Other symbols as in Figure 5F.

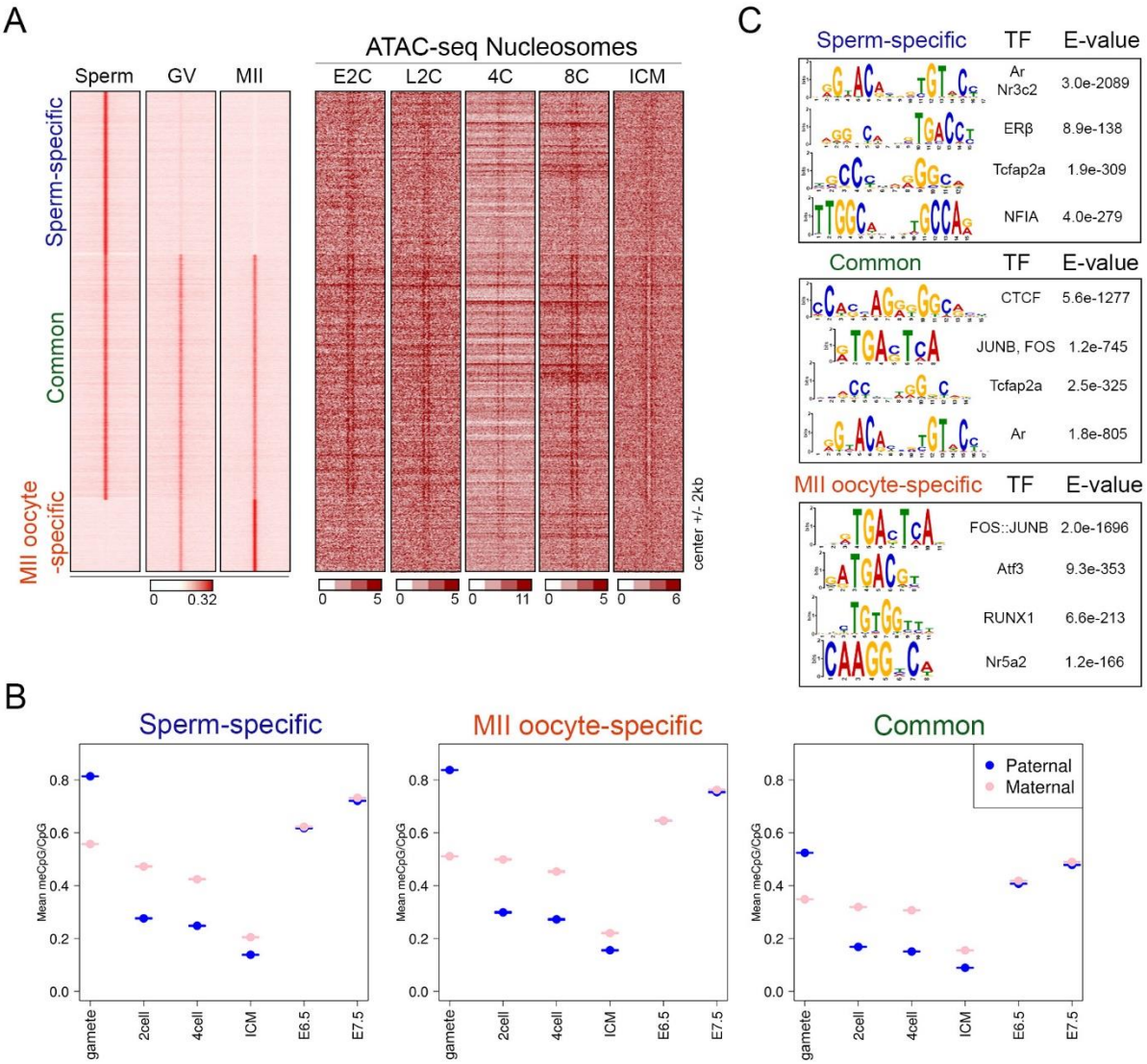



**Figure S6. Enhancer accessibility in mouse pre-implantation embryos. Related to Figure 6**

(A) Heatmaps showing ATAC-seq nucleosome signals in preimplantation embryos at distal THSSs found in germ cells. Sites are the same as in Figure 6B.

(B) TF motifs found in ATAC-seq distal peaks in sperm, MII oocytes, or both.

(C) Analysis of allele-specific DNA methylation at sperm-specific, oocyte-specific, and sperm-oocyte common regions shown in Figure 6B

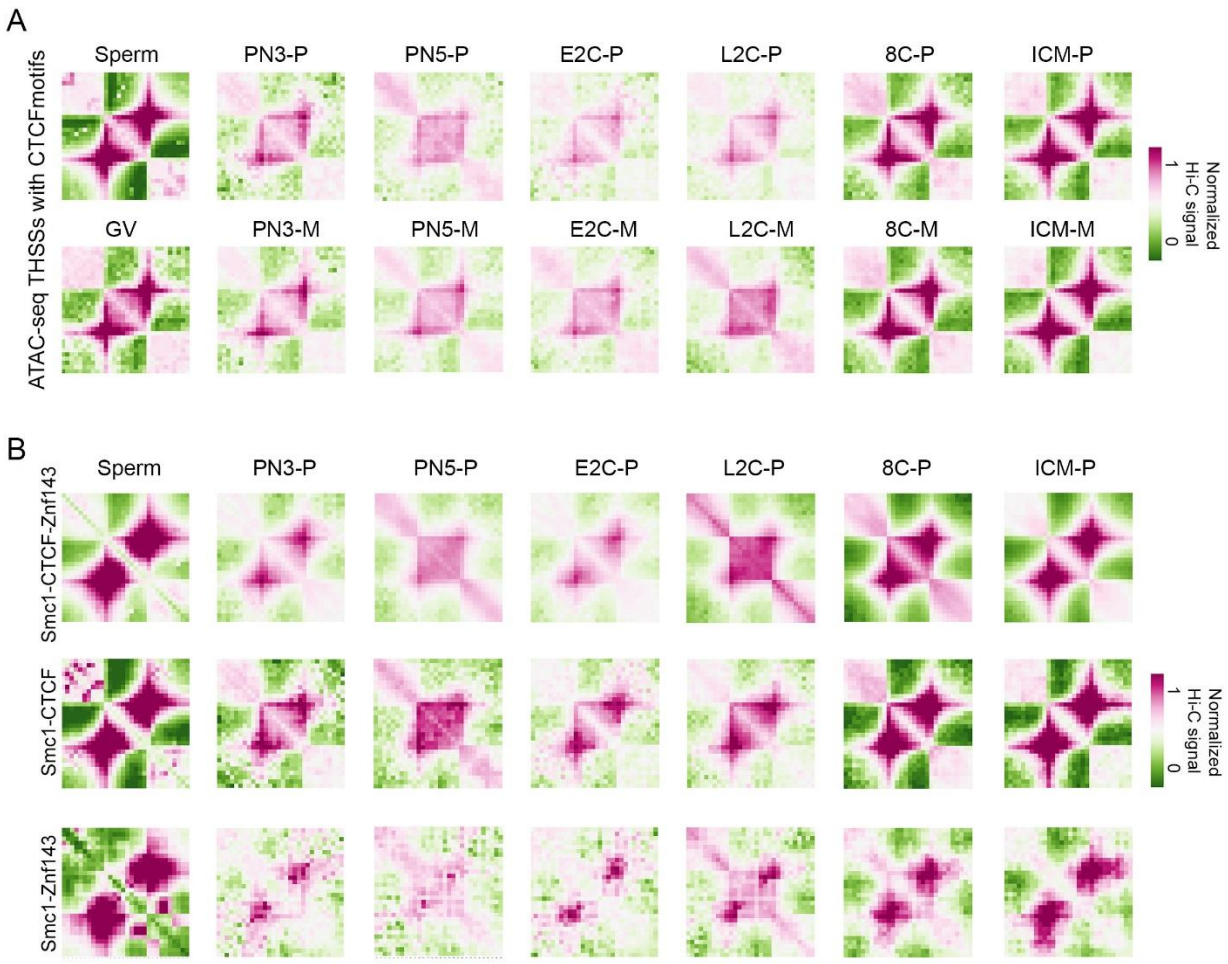



**Figure S7. Maintenance of CTCF-mediated loops in gametes and preimplantation embryos. Related to Figure 7**

(A) Metaplot showing median, distance-normalized Hi-C signal determined by FitHi-C in sperm, GV oocytes, and maternal or paternal chromosomes from preimplantation embryos at THSSs common to sperm, GV, and MII oocytes containing the CTCF motif.

(B) Metaplot showing median, distance-normalized Hi-C signal determined by FitHi-C in sperm, GV oocytes, and maternal or paternal chromosomes from preimplantation embryos at THSSs common to sperm, GV, and MII oocytes containing Smc1, CTCF, and/or Znf143.

**Table S1. Summary of read numbers from ChIP-seq and ATAC-seq. Related to Figure 1 through Figure 6**

| Experiment Description | Sample Name | Sequenced Read Pairs | Unique Read Pairs |
|---|---|---|---|
| ATAC-Seq | MOUSE_SPERM_ATACseq_Rep1 | 79,057,603 | 59,788,683 |
| ATAC-Seq | MOUSE_SPERM_ATACseq_Rep2 | 47,601,884 | 37,974,991 |
| ATAC-Seq | MONKEY_SPERM_ATACseq_Rep1 | 73,627,563 | 55,958,510 |
| ATAC-Seq | MONKEY_SPERM_ATACseq_Rep2 | 48,112,542 | 38,240,871 |
| ATAC-Seq | HUMAN_SPERM_ATACseq_Rep1 | 72,254,780 | 57,944,270 |
| ATAC-Seq | HUMAN_SPERM_ATACseq_Rep2 | 62,610,191 | 51,103,617 |
| ATAC-Seq | MOUSE_GV_ATACseq_Rep1 | 88,101,376 | 65,903,356 |
| ATAC-Seq | MOUSE_GV_ATACseq_Rep2 | 51,208,023 | 33,339,596 |
| ATAC-Seq | MOUSE_MII_ATACseq_Rep1 | 42,029,649 | 29,369,446 |
| ATAC-Seq | MOUSE_MII_ATACseq_Rep2 | 26,794,219 | 13,333,081 |
| ChIP-Seq | CONT_SPERM_RNAPIIS5p_Rep1 | 19,776,004 | 14,795,775 |
| ChIP-Seq | CONT_SPERM_RNAPIIS5p_Rep2 | 19,962,054 | 14,079,661 |
| ChIP-Seq | CONT_SPERM_RNAPIIS2p_Rep1 | 22,807,600 | 16,966,936 |
| ChIP-Seq | CONT_SPERM_RNAPIIS2p_Rep2 | 37,838,696 | 27,279,492 |
| ChIP-Seq | CONT_SPERM_Med12_Rep1 | 25,756,433 | 18,341,206 |
| ChIP-Seq | CONT_SPERM_Med12_Rep2 | 28,756,798 | 13,678,062 |
| ChIP-Seq | CONT_SPERM_Med12_Rep3 | 31,315,280 | 12,799,313 |
| ChIP-Seq | CONT_SPERM_Znf143_Rep1 | 58,090,922 | 29,170,904 |
| ChIP-Seq | CONT_SPERM_Znf143_Rep2 | 35,058,068 | 24,289,858 |
| ChIP-Seq | CONT_SPERM_Foxa1_Rep1 | 17,963,312 | 12,496,725 |
| ChIP-Seq | CONT_SPERM_Foxa1_Rep2 | 49,226,170 | 29,676,467 |
| ChIP-Seq | CONT_SPERM_ERα_Rep1 | 36,309,378 | 22,019,797 |
| ChIP-Seq | CONT_SPERM_ERα_Rep2 | 208,437,014 | 83,887,218 |
| ChIP-Seq | CONT_SPERM_AR_Rep1 | 323,098,216 | 232,477,887 |
| ChIP-Seq | CONT_SPERM_AR_Rep2 | 313,452,079 | 233,695,275 |
| ChIP-Seq | CONT_SPERM_H3.3_Rep1 | 26,461,235 | 12,831,941 |
| ChIP-Seq | CONT_SPERM_H3.3_Rep2 | 24,755,822 | 14,881,953 |
| ChIP-Seq | CONT_SPERM_H2A.Z_Rep1 | 22,218,909 | 13,155,472 |
| ChIP-Seq | CONT_SPERM_H2A.Z_Rep2 | 45,276,027 | 28,441,383 |

**Table S2. Summary of the processing step of sequencing reads obtained in mouse sperm Hi-C experiments. Related to Figure 3.**

| | Mouse sperm_Rep1 | Mouse sperm_Rep2 |
|---|---|---|
| Sequenced Read Pairs | 1,367,611,035 | 1,225,497,862 |
| Normal Paired | 1,230,900,108 (90.00%) | 967,264,048 (78.93%) |
| Chimeric Paired | 262,077 (0.02%) | 186,465 (0.02%) |
| Chimeric Ambiguous | 251,659 (0.02%) | 178,103 (0.01%) |
| Unmapped | 136,197,191 (9.96%) | 257,869,246 (21.04%) |
| Ligation Motif Present | 307,797,257 (22.51%) | 749,901,642 (61.19%) |
| Alignable (Normal+Chimeric Paired) | 1,231,162,185 (90.02%) | 967,450,513 (78.94%) |
| Unique Reads | 519,361,739 (37.98%) | 768,053,227 (62.67%) |
| PCR Duplicates | 454,775,015 (33.25%) | 190,691,690 (15.56%) |
| Optical Duplicates | 252,611,226 (18.47%) | 4,772,911 (0.39%) |
| Library Complexity Estimate | 684,012,123 | 2,078,677,383 |
| Intra-fragment Reads | 79,509,010 (5.81% / 15.31%) | 58,973,277 (4.81% / 7.68%) |
| Below MAPQ Threshold | 179,653,777 (13.14% / 34.59%) | 225,145,347 (18.37% / 29.31%) |
| Hi-C Contacts | 260,198,952 (19.03% / 50.10%) | 483,934,603 (39.49% / 63.01%) |
| Ligation Motif Present | 59,414,349 (4.34% / 11.44%) | 276,239,478 (22.54% / 35.97%) |
| 3' Bias (Long Range) | 86% - 14% | 94% - 6% |
| Pair Type %(L-I-O-R) | 25% - 25% - 25% - 25% | 25% - 25% - 25% - 25% |
| Inter-chromosomal | 68,841,438 (5.03% / 13.26%) | 155,387,605 (12.68% / 20.23%) |
| Intra-chromosomal | 191,357,514 (13.99% / 36.84%) | 328,546,998 (26.81% / 42.78%) |
| Short Range (<20Kb) | 95,867,972 (7.01% / 18.46%) | 142,929,697 (11.66% / 18.61%) |
| Long Range (>20Kb) | 95,482,496 (6.98% / 18.38%) | 185,615,860 (15.15% / 24.17%) |