

Figure S1

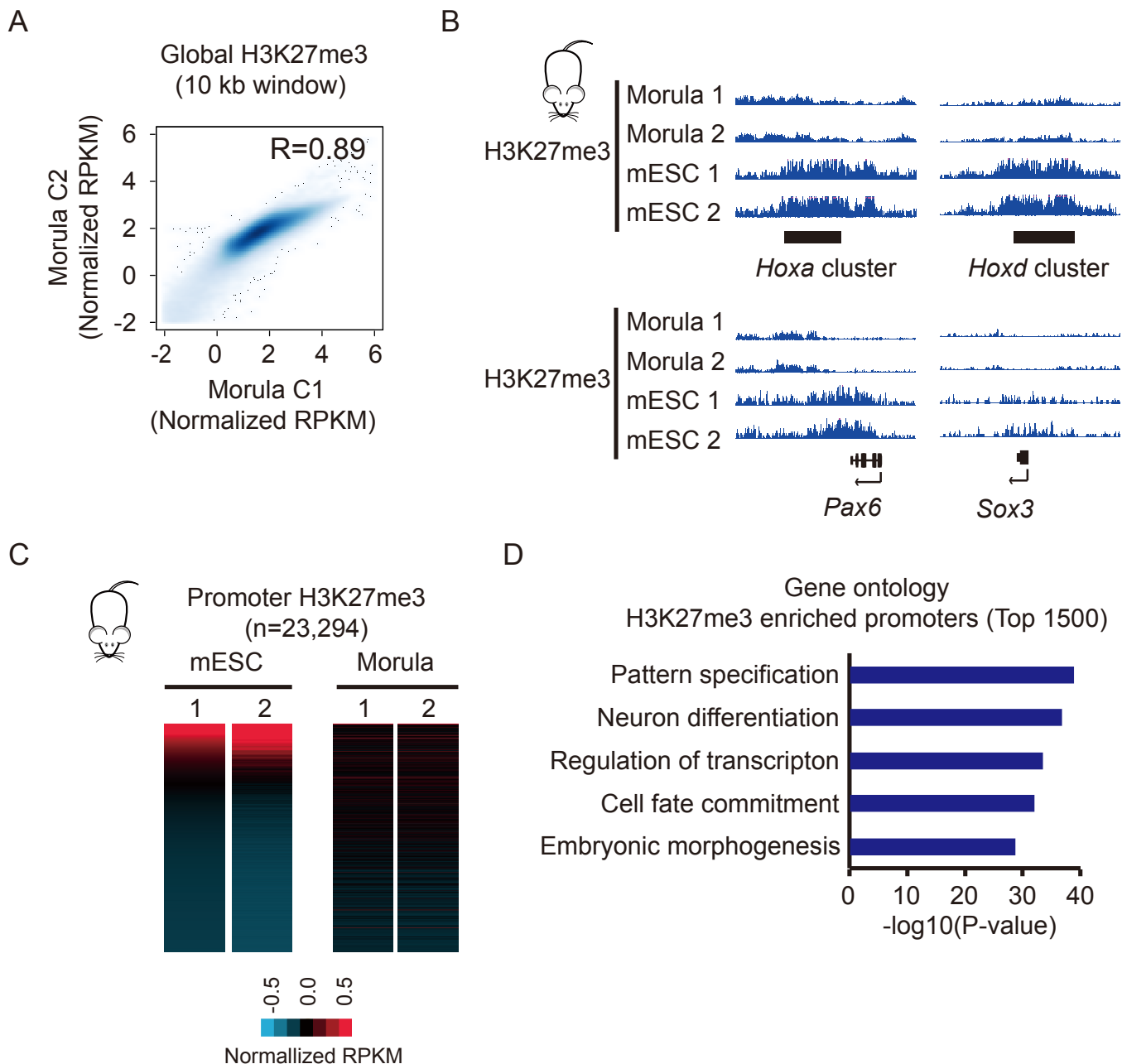


Figure S1. Comparison of H3K27me3 CUT&RUN replicates in human morulae

(A) Scatter plot showing genome-wide correlation (10 kb window) between two H3K27me3 CUT&RUN replicates. RPKM was normalized by log transform.

(B) A genome browser view showing H3K27me3 enrichment at *Hoxa*, *Hoxd* clusters, *Pax6*, and *Sox3* in both mouse morulae and mouse embryonic stem cells (mESCs).

(C) Heatmap showing H3K27me3 signal at developmental gene promoters (TSS \pm 2.5 kb) in mESCs, but not in mouse morulae. RPKM was subjected to Z-score normalization.

(D) DAVID Gene ontology analysis showing the enriched GO terms of the top 1500 promoter-H3K27me3-enriched genes.

Figure S2

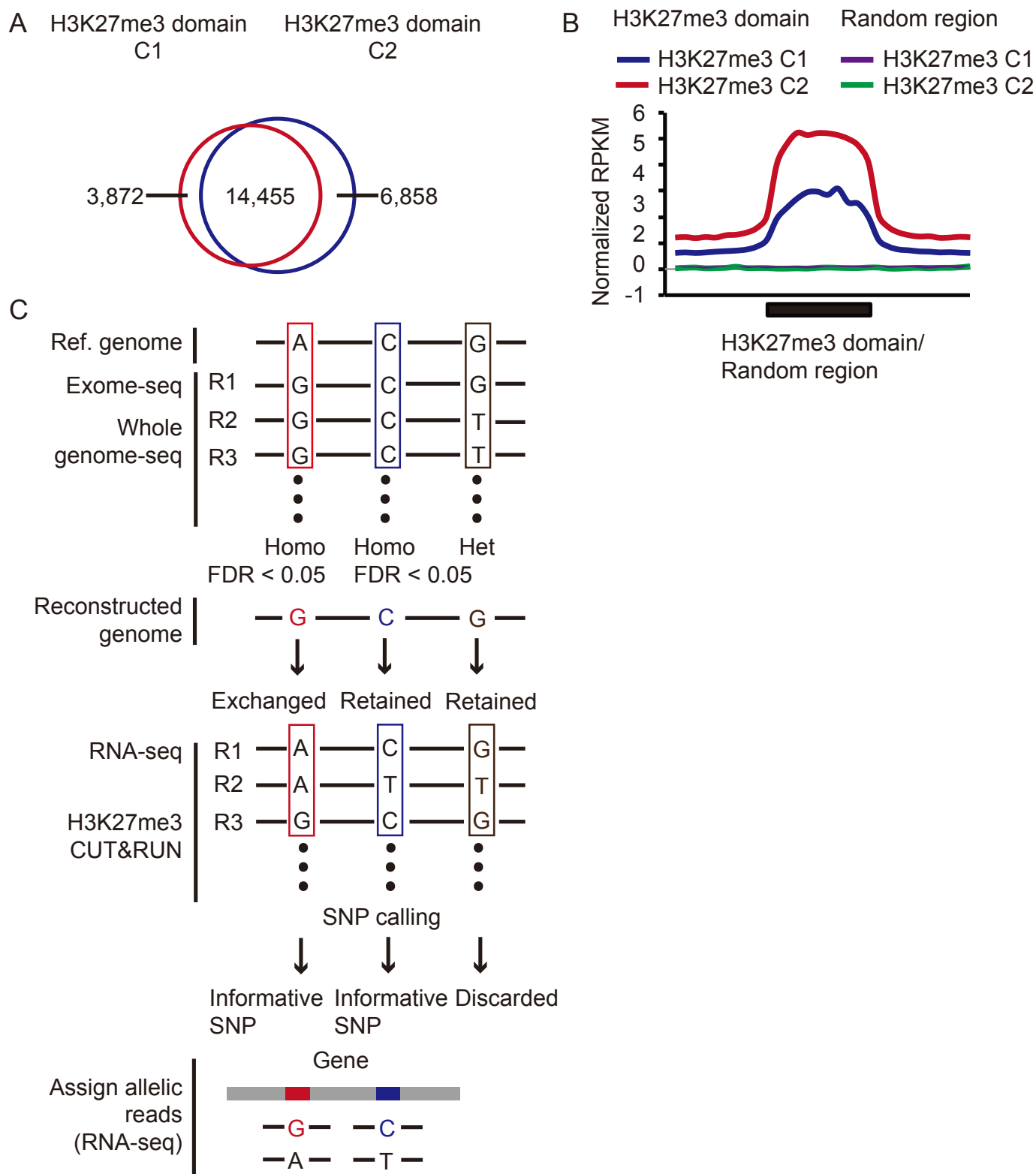


Figure S2. Identification of H3K27me3 domains in the human morula and SNP calling strategy

(A) Venn diagram showing overlaps of H3K27me3 domains between two H3K27me3 CUT&RUN data sets of human morulae. Only the common domains were used for downstream analysis.

(B) Average distribution of H3K27me3 signals (average enrichment, calculated by Z-score normalized RPKM) across the H3K27me3 domains and random control regions, respectively.

(C) Schematics for identification of parental-specific SNPs and PEGs using WES, WGS, CUT&RUN, and/or RNA-seq datasets. Briefly, maternal genomes were reconstructed based on the homozygous SNPs identified using WES or WGS (See Supplementary methods). SNPs for H3K27me3 CUT&RUN and RNA-seq data were identified based on maternal genome.

Figure S3

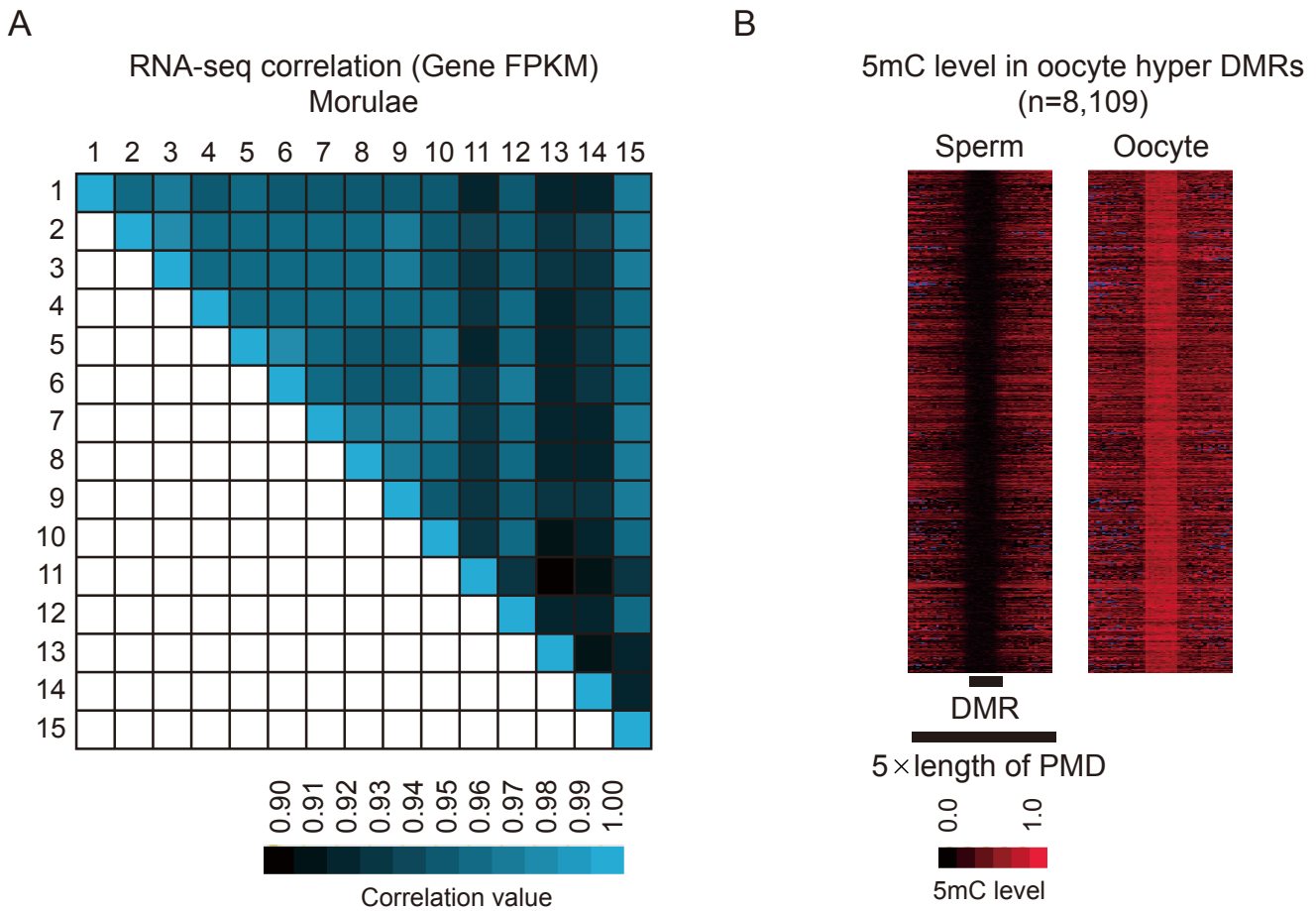


Figure S3. Transcriptome reproducibility and oocyte hyper-DMRs

(A) A heatmap showing Pearson correlation of gene expression levels for single morula RNA-seq data sets.

(B) Heatmap showing sperm and oocyte DNA methylation level surrounding oocyte hyper-DMRs identified in this study.

Figure S4

A

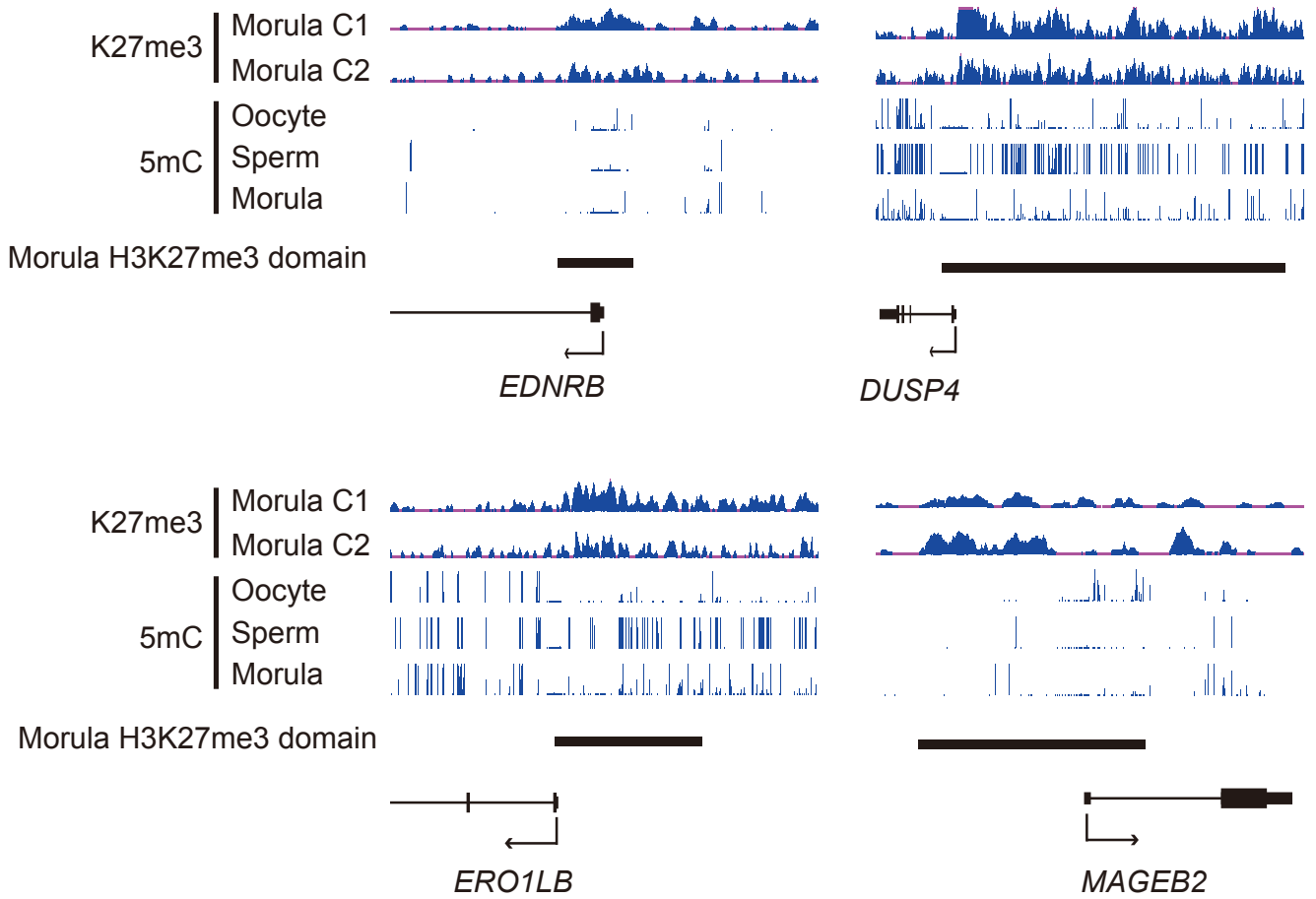


Figure S4. Examples of H3K27m3 associated PEGs

Genome browser view showing H3K27me3 enrichment in morula and DNA methylation in oocyte, sperm and morula at the promoters of PEGs, which overlapped with morula H3K27me3 domains.

**Maternal-biased H3K27me3 correlates with paternal-specific gene expression
in the human morula**

Wenhao Zhang, Zhiyuan Chen, Qiangzong Yin, Dan Zhang,
Catherine Racowsky, and Yi Zhang

Supplementary Methods

H3K27me3 CUT&RUN

Human morula H3K27me3 CUT&RUN libraries were prepared as previously described (Skene et al. 2018) with a few modifications. Briefly, morulae (7 from one couple, 8 from another) were combined in a 1.5-ml tube containing 50 μ l wash buffer [20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM Spermidine (Sigma-Aldrich), and 1 \times protease inhibitor cocktails (Sigma-Aldrich)]. The embryos were then captured with BioMagPlus Concanavalin A (Polysciences) and incubated with a rabbit anti-H3K27me3 antibody (C15410069, Diagenode) for 16 hours at 4°C in antibody incubation buffer [20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1 \times protease inhibitor cocktails (EDTA-free tablet, Roche), 2 mM EDTA, and 0.02% Digitonin (Life technologies)]. After unbound antibodies were washed away, protein A-MNase (pA-MN, a gift from Dr. Steven Henikoff) was added at a 1:280 ratio (500 ng/ml) and incubated for 3 hours at 4°C. After washing, CaCl₂ was added to a final concentration of 2 mM to activate pA-MN, and the reaction was allowed for 20 min at 4°C and then stopped by adding 1/10 volume of 10 \times STOP buffer [1,700 mM NaCl, 100 mM EDTA, 20 mM EGTA, 250 μ g/mL RNase A (Invitrogen) and 250 μ g/mL glycogen (Invitrogen)]. The protein-DNA complexes were released by 10 min incubation at 37°C followed by centrifugation for 5 min at 16,000g at 4°C. After transferring the supernatant to a new Lo-bound tube, 1/100 volume of 10% SDS and 1/80 volume of 25 mg/ml Proteinase K (Life technologies) were added and incubated at 55°C for at least 1 hour. DNA was then precipitated by phenol/chloroform/isoamylalcohol followed by ethanol precipitation with glycogen, and then dissolved in water.

Sequencing libraries were prepared using a NEBNext Ultra II DNA library preparation kit for Illumina (New England Biolabs) according to the manufacturer's instructions with a few modifications. Briefly, end repair was conducted at 20°C for 30 min followed by dA-tailing at 65°C for 30 min. After adaptor ligation at 20°C for 30 min, the DNA fragments were purified by 1.8 \times volume of SPRIselect beads (Beckman Coulter) followed by 17 cycles of PCR amplification with NEBNext Ultra II Q5 Master Mix (NEB). The PCR products were cleaned up with 0.9 \times volume of SPRIselect beads and quantified using Qubit dsDNA HS assay kit (Agilent Technologies). The libraries were sequenced on a HiSeq2500 with paired-end 100bp reads (Illumina).

Genomic DNA extraction, whole genome and exome sequencing

DNA was extracted from the cumulus cells using the AllPrep DNA/RNA mini kit (Qiagen, 80204). WGS library for the cumulus cells was prepared using the Nextera XT Library Prep Kit (Illumina). Briefly, 1 ng DNA extracted from the cumulus cells was mixed with 10 µl TD buffer and 5 µl ATM buffer and incubated at 55°C for 5 minutes. After neutralization with ND buffer, the DNA library was subjected to PCR amplification (12 cycles) with NPM master mix by following the manufacturer's instructions. WES libraries were prepared using SureSelectXT2 Enrichment System for Illumina (Agilent Technologies) according to the manufacturer's instructions. Specifically, 100 ng genomic DNA extracted from the maternal cumulus cells were sheared into 150-200 bp using a Covaris M220 sonicator (Covaris) with microTUBE-50 (Covaris). Following end-repair and ligation, five cumulus cell libraries were indexed, pooled and amplified. To capture human exon sequences including the untranslated regions (UTRs), SureSelectXT2 Human All Exon V6+UTR (Agilent Technologies) probes were used to hybridize with the pooled DNA libraries. Following hybridization, the SureSelect-enriched libraries were purified and undergo quality check by Qubit dsDNA HS Assay (Agilent Technologies) and Bioanalyzer (Agilent Technologies). Both WGS and ES libraries were sequenced on an Illumina HiSeq 2500 (Illumina) with paired-end 100bp reads.

Single morula RNA sequencing

Single morula RNA sequencing was performed as previously described (Inoue et al. 2017). Specifically, SMARTer Ultra Low Input RNA cDNA preparation kit (Clontech) was used to perform reverse transcription and cDNA amplification (13 cycles) using the whole-embryo lysate. cDNAs (80pg) were then fragmented, adaptor-ligated and amplified (12 cycles) using Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's instructions. Paired-end 100bp sequencing was performed on a HiSeq 2500 sequencer (Illumina).

CUT&RUN and RNA-seq data processing

Reads alignment. Adaptor sequences were first trimmed by Trim Galore to remove adaptor sequences and low-quality bases (version 0.5.0) (parameters: --Illumina --paired). Following reads trimming, H3K27me3 CUT&RUN reads were aligned to the human reference genome (hg19) using Bowtie2 (version 2.3.4.1)(Langmead and Salzberg 2012) paired-end alignment mode (parameters: -N 1 -L 25 -X 2000

--no-mixed --no-discordant). RNA-seq reads were aligned to hg19 using TopHat (version 2.1.1) (Kim et al. 2013) paired-end alignment mode (parameters: --no-coverage-search).

RPKM calculation for CUT&RUN. Following alignment, multiple-aligned reads were discarded and reads duplicates were removed by MarkDuplicate from Picard Tools (version 2.18.11). The number of reads per kilobase of bin per million reads sequenced (RPKM) were calculated at 100-bin throughout the genome.

FPKM calculation for RNA-seq. Following alignment, FPKM for all genes was calculated using Cufflinks (version 2.2.1) (Trapnell et al. 2010) with NCBI RefSeq (hg19) from UCSC genome browser supplied as the annotation file.

Identification of parental-specific SNPs

Maternal genome reconstruction. To ensure the accuracy of SNP calling, Burrows-Wheeler Aligner (BWA, version 0.7.17) (Li and Durbin 2009) was used with parameter 'aln' to align the WGS and WES to hg19. Reads with mapping quality $q < 10$ and multiple-mapped reads were discarded. Read duplicates were further filtered by MarkDuplicate from Picard Tools (version 2.18.11). The remaining reads were then subjected to local realignment and base-score recalibration using GATK (version 4.0.7.0) (McKenna et al. 2010). Raw variants were called using UnifiedGenotyper tools from GATK (parameter: -stand_call_conf 0 -stand_emit_conf 0 -U ALLOW_N_CIGAR_READS -out mode EMIT_VARIANTS_ONLY). For quality filtering, variants with quality < 20 and located within first six bases of a read were discarded. Furthermore, only variants showing maternal homozygous genotypes (FDR < 0.05) with ≥ 10 reads coverage were considered as homozygous SNPs. The original nucleotides in hg19 were replaced with the homozygous SNPs to restrict a customized maternal reference genome (Figure S2C).

SNP calling from CUT&RUN data. BWA was used with parameter 'aln' to map the CUT&RUN data to hg19 (Li and Durbin 2009). The SNP calling and filtering steps were the same as for WGS and WES. Only variants with read coverage over 10 were used for assessing allelic enrichment of H3K27me3. The parental-origins of the reads that overlap SNPs were determined by corroborating the SNPs in WGS or WES data (Fig. S2C). Allele-specific read counts at SNPs were generated using samtools pileup (version 0.1.12a) (Li et al. 2009).

SNP calling from RNA seq data. SNPs in RNA-seq data were identified using SNPiR, a highly accurate method specifically designed for SNP calling from RNA-seq data (Piskol et al. 2013). Briefly, RNA-seq reads were aligned using BWA 'aln' to a customized reference genome containing both the reconstructed maternal hg19 genome and splicing junction sequences with a 95bp span over the adjacent junctional sites. Low quality ($q < 10$), multi-mapped, and duplicated reads were removed, and the raw variants were identified as describe above. However, as SNP calling from RNA-seq tends to yield more false positives, more stringent filtering steps were applied to remove suspicious SNPs (Piskol et al. 2013). First, variants with quality < 20 and located within the first six bases of a read were discarded. Variants in simple repeats were removed according to the Repeat-Masker annotation from UCSC genome browser. Furthermore, intronic variants were discarded if they were within 4bp of known splice junctions. To ensure the reads supporting a variant were uniquely aligned to the genome, BLAT (Kent 2002) was used to re-align all the reads to the genome. Only reads with the best hit overlap with the expected position and the second best hit having a score $< 95\%$ of the best hit were considered as uniquely aligned. Variants covered with at least 10 reads were retained. Finally, variant overlaps with all currently known RNA-editing sites (A-to-I RNA editing database) were removed.

Identification of PEGs

Allele origins of the RNA-seq reads were assigned according to the parental-specific SNPs identified by the SNPiR method. Allele-specific read counts (discordantly aligned read pairs were excluded) were summarized for each gene and the Binomial test was used to identify significant PEGs. To estimate the FDR, RNA-seq reads overlapping the informative SNPs were randomly assigned to either maternal or paternal origin to generate a randomized dataset. Then p -value was set accordingly for each sample to ensure $FDR < 0.02$. To minimize the random allelic variation, only genes with consistent paternal biased expression (paternal/maternal > 2 -fold) in at least two morulae were considered as PEGs.

Identification of PMDs

Human morula single cell DNA methylation data were obtained from ((Zhu et al. 2018). For each single cell, only CpGs covered with at least three reads were used for further analysis. The methylation level for each CpG site was calculated as the mean of all the single cells ($n=17$). PMDs in morulae were identified as described before (Lister et al. 2009). Briefly, average DNA methylation levels were calculated for every

10kb bins in the genome. Bins with the average methylation level less than 0.4 and with more than 20 CpGs covered were selected and merged into PMDs.

Identification of oocyte hyper-DMRs

Human oocyte and sperm single-cell DNA methylome data were obtained from (Zhu et al. 2018). For each single cell, only CpGs covered with at least three reads were used for further analysis. The methylation level for each CpG site was calculated as the mean of all the single cells (sperm n=21, oocyte n=28). Oocyte hyper-CpG was defined as the CpG with methylation difference between oocyte and sperm greater than 0.5. Regions with at least 10 consecutive oocyte hyper-CpGs and average methylation level less than 15% in sperm were considered as oocyte hyper-DMRs, with the distance between two consecutive oocyte hyper-CpG being less than 500bp.

Identification of H3K27me3 domains and analysis of its enrichment in promoters

To identify H3K27me3 domains, average H3K27me3 enrichment was calculated for every 5 kb bin across the genome. Bins with RPKM greater than 1.5-fold of the genome-wide average RPKM were considered as H3K27me3-enriched bins. H3K27me3-enriched bins with a distance less than 5kb were further merged into H3K27me3 domains.

For the promoter H3K27me3 enrichment analysis, the H3K27me3 CUT&RUN or ChIP-seq signals at promoters (TSS \pm 2.5kb) were computed by sum signals from each bin within the promoter and Z-score normalization was performed across all promoters in the genome.

Supplementary References

- Inoue A, Jiang L, Lu F, Suzuki T, Zhang Y. 2017. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**: 419-424.
- Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome research* **12**: 656-664.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**: R36.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315-322.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**: 1297-1303.
- Piskol R, Ramaswami G, Li JB. 2013. Reliable identification of genomic variants from RNA-seq data. *American journal of human genetics* **93**: 641-651.
- Skene PJ, Henikoff JG, Henikoff S. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature protocols* **13**: 1006-1019.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**: 511-515.
- Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, Lian Y, Fan X, Hu B, Gao Y et al. 2018. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nature genetics* **50**: 12-19.