bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

title: Resetting of H3K4me2 during mammalian parental-to-zygote transition

Chong Wang[1]†, Yong Shi[1]†, Jia Guo[1], Kaiyue Hu[1], Yaqian Wang[1], Yang Li[1], Jiawei Xu[1]*

[1] the First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450000 China

† These authors contribute equally to this work

* Corresponding author: jiawxu@foxmail.com

ABSTRACT

Upon sperm and oocyte fertilization, drastic histone modification reprograming occurs during preimplantation development. However, the global H3K4me2 landscape and its dynamics reprogramming in this period remains largely unknown. Herein we investigate the erasure and re-establishment of H3K4me2 in mouse GV, MII and embryos using an improved approach called Cleavage Under Targets and Release Using Nuclease (CUT&RUN) for histone modification with high-throughput sequencing. We find H3K4me2 extensively exists as a non-canonical pattern in mouse GV oocytes and early embryos. Interestingly, H3K4me2 is erased in MII oocyte and re-established in late 2-cell stage. Importantly, mouse embryos acquire widespread H3K4me2 in CpG-rich and hypomethylated regulatory regions in 4-cell stage, as well as 8-cell stage, but drastic changes happen upon ICM, these CpG-rich H3K4me2 regulatory regions then resolve to either active or repressed states. In summary, our study not only unveil that H3K4me2 transition from parental to zygote, but also provide a H3K4me2 profile during early embryo development, and this will enhance our comprehension to epigenetic reprogramming during mouse early development and in vitro fertilization.

Highlight

1. H3K4me2 present differential patterns in sperm and oocytes, as no-canonical pattern in GV oocytes but canonical of sperm in mouse.
2. GV H3K4me2 peaks were almost erased in MII oocytes, and reconstructed accompanying ZGA occurred, the first landscape of H3K4me2 during early mammalian development was provided.
3. Non-canonical H3K4me2 in mammalian embryos and pervasive H3K4me2 in CpG-rich regulatory regions in 4 and 8-cell embryos.
4. CpG-rich H3K4me2 regulatory regions resolve to either active or repressed states in ICM.

INTRODUCTION

Drastic Epigenome reprogramming occurred during early mammalian embryo development, epigenetic marks, such as DNA methylation, histone modifications have been revealed play important roles in the epigenome reprogramming process[1,2]. The epigenome reprogramming during preimplantation embryo development erases the oocyte and sperm's epigenetic memory and allows the zygote to establish a new epigenome landscape which is relevant for early development and is essential for maternal zygotic transition (MZT) and differentiation[1,3]. Deficiencies in reconstructing of the epigenomes can result in developmental defects, such as embryonic arrest[3]. Histone modifications are fundamental epigenetic regulators that control gene expression and many key cellular processes[4-6]. Different types of histone modifications deposited in specific genomic regions and perform distinct functions[3]. Most histones modifications largely accumulated around TSSs in the embryos, this feature is shared by somatic cells[7-9]. this feature is

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

commonly found in many species and cell lines[10-12], so this distribution pattern called as the 'canonical' pattern[13]. It had been reported that highly pervasive H3K27me3 is found in regions depleted of transcription and DNA methylation in oocytes, while extensive loss of promoter H3K27me3 at *Hox* and other developmental genes upon fertilization accompanied by global erasure of sperm H3K27me3 but inheritance of distal H3K27me3 from oocytes, both H3K4me3/H3K27me3 bivalent promoter marks are restored at developmental genes[14].

The K-residues at N-terminus of histone H3 exhibit mono-(H3K4me1), di-(H3K4me2), or tri-methylated (H3K4me3) catalyzed by a family of proteins that contain a SET domain, the methylation of different number of methyl groups added represent different chromatin status[15-17]. The different number and kinds of histone modification marks associated with different gene expression states[18], it is well known that H3K4me3 usually enriched nearby transcription start sites(TSS) region associated with the activation of nearby genes[19], and H3K4me1 as an global marker of enhancers, both the regions enriched H3K4me1 and H3K27ac identified as super enhancers[20-22]. It is reported that H3K4me1/H3K4me2/H3K4me3 is enriched on both sides of the TSS, play diverse and important roles in regulating gene expression in mammalians[23,24]. Previous studies revealed that the H3K4me2 distribution pattern is like the H3K4me3 in plant[25,26]. While how the H3K4me2 reprogramming in mammalians early development is still unknown.

H3K4me2 has been proved in coordinating the recovery of protein biosynthesis and homeostasis following DNA damage required for developmental growth and longevity[27]. Unlike in animals, H3K4me2 enrichment is negatively correlated with gene transcript level in plant, H3K4me2 in rice and Arabidopsis specifically distributes over the gene body region and functions as a repressive epigenetic mark[28]. H3K4me2 was found to be a key epigenetic factor in chicken PGC cell lineage determination and narrow domain activated PGC-specific expression genes and signaling pathways[29]. H3K4me2 can be acquired as early as the embryonic stem cell stage, and its distribution is dynamic and progressive throughout mouse brain differentiation and tissue development[30]. Furthermore, it is reported that 90% of the transcription factor binding regions (TFBRs) overlap with the H3K4me2-enriched regions, and together with the H3K27ac-enriched regions can greatly reduce false positive predictions of the TFBRs, moreover, the stronger signals of H3K4me2 enrichment more likely overlap with the TFBRs than those with weaker signals[31]. It had been reported that a subgroup of genes enriched with high levels of H3K4me2 within their gene body linked to T cell functions, different from the TSS centered enrichment of housekeeping genes[32]. And it had reported that many regions of the sperm epigenome bearing H3K4me2 are already present in spermatogonia, suggesting an early established of H3K4me2 in spermatogenesis and highly conserved[33].

In this study, we aimed to reveal how the H3K4me2 reprogramming in mammalian early development. To address this question, we optimized the Cleavage Under Target and Release Using Nuclease (CUT&RUN)[34,35] methods using H3K4me2 antibody to as few as 50 cells and revealed that unique H3K4me2 modifications pattern and epigenetic reprogramming modes during the parental-to-zygotic transition in mouse.

MATERIALS AND METHODS

Cell culture of K562

Human K562 cell line was cultured at 37℃, 5% CO_2 and saturated humidity, using 1640 culture medium (GIBCO) containing 10% fetal bovine serum and 90 % 1640 culture liquid. Human K562 cell was used for optimizing limited cell number CUT&RUN of antibody H3K4me2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cell culture of mouse ESC

Mouse embryo stem cells (ESC) were cultured on 0.1% gelatin-coated plates without irradiated mouse embryonic fibroblasts (MEFs), using standard ESC medium consisting of DMEM (GIBCO) supplemented with 15-20% fetal bovine serum (FBS, Sigma), 0.1 mM β-mercaptoethanol (Sigma), 0.1mM MEM non-essential amino acids (Cellgro), 2mM Glutamax (GIBCO), 1% nucleoside (Millipore), 1% EmbryoMax® Penicillin-Streptomycin Solution (Millipore) and 10 ug/L recombinant leukemia inhibitory factor (LIF) (Millipore). Mouse embryo stem cells was cultured at 37℃, 5% CO2 and saturated humidity.

Mouse gametes and early embryos collection

Overall, the mouse oocytes were collected from 4-5-week-old C57BL/6N females' mice (purchased from Charles River), In addition, the pre-implantation embryos were harvested from 4-5-week-old C57BL/6N females mated with DBA/2 males (8-week-old when purchased from Charles River). To be specific, the Germinal vesicle (GV) oocytes were obtained from 4-5-week-old C57BL/6N females 46-48 hours after injection of 10 IU of pregnant mare serum gonadotropin (PMSG, Solarbio). What's more, 4–5-week-old C57BL/6N females were administered 10 IU of human chorionic gonadotrophin (HCG)intraperitoneally, 46–48 h after PMSG disposed, then mated with DBA/2 males. PN5 zygotes were collected at 28-32 h post-hCG injection, Early_2-, Late_2-, 4cell, 8cell embryos, morulas and blastocysts were harvested at 32-36 h, 45-48 h,54-58 h,60-64 h, 68-75h, 84-96 h post-hCG injection. The metaphase of second meiosis (MII) oocyte were got 12-14 hours after injection of 10 IU of hCG intraperitoneally, the zona pellucida of all oocytes and embryos was removed by glutathione (GSH Solarbio).

Immunofluorescence

The samples (mouse oocytes or pre-implantation embryos) were washed twice in MII medium , 4% PFA was used as a fixation buffer (BOSTER, AR1068) to fix samples at room temperature about 60 mins. Next, samples were transferred by mouth pipette and washed by 1% BSA (Sigma, A1933) in 1 × PBS (Corning,21-040-CVR) for once. The samples were then permeabilized by 0.5% Triton X-100 (Sigma, T8787) in 1 × PBS at room temperature for 20 mins, followed by washing with 1% BSA in 1 × PBS, and blocked by 1% BSA in 1 × PBS at room temperature for 1 hour. The blocked samples were incubated by H3K4me2 antibody (SigmaMillpore, 07030) in 1% BSA in 1 × PBS diluted at ratio of 1:500, at room temperature for 2 hours or at 4°C for overnight, followed with washing by 1% BSA in 1 × PBS of 3 mins for at least 5 times. The samples were incubated with Alexa Fluor 568 conjugated Donkey anti Rabbit IgG (Invitrogen, A10042) in 1% BSA in 1 × PBS diluted at ratio of 1:200, at room temperature for 50 mins without light, followed with washing by 1% BSA in 1 × PBS of 3 mins for 5 times. Finally, the samples were stained by ProLong@ Gold Antifade Reagent with DAPI (Cell Signaling, 8961S), using 4-Chamber Glass Bottom Dish(100 per Case), 35mm Dish with 20mm Bottom Well (Cellvis, D35C4-20-1-N), putting the samples in room temperature for about 24 h in a dark environment. Images were acquired on Zeiss 700 confocal microscope. Raw images were projected along the Z-axis and processed using ZEN software (Zeiss) with brightness and contrast adjusted (The same parameters were used for each channel of all samples).

Zona pellucida removal and ICM/TE isolation

Before CUT&RUN experiment, to guarantee the effect of permeation, the zona pellucida of all oocytes and embryos was chemically removed by the glutathione (GSH Solarbio), GSH were added dropwise to oocytes or embryos by mouth pipette which in MII medium (Sigma), the zona

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pellucida were removed after about 10-30 seconds. Then the sample were washed twice in PBS (Corning 21-040-CVR), and then washed once with wash buffer. The inner cell mass (ICM)s or TEs of day 4-5 blastocysts were isolated mechanically. Before directly addressed to CUT&RUN, the separated cells were washed twice with PBS and once with wash buffer, then resuspended by wash buffer.

CUT&RUN library construction and next-generation sequencing

CUT&RUN was addressed following the previously published protocol[34,35] with modifications in antibody binding , adaptor ligation and DNA purification steps. To be specific, washing buffer and binding buffer were balanced to room temperature for about 30min in advance. Samples (mouse zygotes, embryos or ESCs with zona pellucida removed) were transferred into 0.2mL conventional, non-low-binding PCR tube (Axygen). The samples were resuspended by 60μL washing buffer (HEPES-KOH, pH = 7.5, 20mM; NaCl, 150mM; Spermidine, 0.5mM and with Roche complete protease inhibitor.). Concanavalin-coated magnetic beads (Polyscience, 86057) were gently washed thrice or four times and resuspended by binding buffer (HEPES-KOH, pH = 7.9, 20mM; KCl, 10mM; CaCl2, 1mM; MnCl2, 1mM) then adding Concanavalin-coated magnetic beads 10-15μL for each sample, The samples were then incubated at room temperature (about 25°C) at 450 rpm for 20 mins on Thermomixer (Eppendorf). Putting the samples on magnetic stand to facilitate removing supernatant, then resuspended by 50μL antibody buffer (washing buffer plus digitonin 1000:1, EDTA, pH = 8.0, 2mM) with H3K4me2 antibody (Sigma Millpore 07030) diluted at ratio of 1:100. The samples were incubated at room temperature (23-26℃) for 2h at 450 rpm on Thermomixer, or incubated at 4°C for overnight(≥16h) at 450 rpm on Thermomixer.

Next, the samples were held at the magnetic stand and washed by 190μL dig washing buffer (washing buffer plus digitonin, freshly pre-heated, 0.005%~0.01%, tested for each batch) for once and 100μL for a second time. The samples were then resuspended by 50μL dig washing buffer with PA/G-MNase (to a final concentration of 700ng/mL), and incubated at room temperature (23-26℃) on Thermomixer for 1h, or incubated at 4°C on Thermomixer for 3 hours at 450 rpm. The samples were washed by 190μL dig washing buffer for twice and 100μL for a third time at the magnetic stand. Then adding 100μL dig washing buffer and balanced on ice for 3 mins. Targeted cleavage was performed on ice by adding 2μL 100mM $CaCl_2$ about 30 mins, and reaction was stopped by adding 100μL 2 × stop buffer (NaCl, 340mM; EDTA, pH = 8.0, 20mM; EGTA, pH = 8.0, 4mM; RNase A, 50μg/mL; glycogen, 100μg/mL; plus digitonin, preheated, 0.005%~0.01%, tested for each batch) and fully vortexed. The samples were then incubated at 37°C of 30 mins for fragment releasing. The total samples or supernatants were digested by adding 10ng carrier RNA (Qiagen, 1068337),1μL 20% SDS and 2μL Proteinase K (Roche), and the reaction was addressed at 56°C for 45 mins and 72°C for 20 mins in PCR machine. DNA was purified by phenol chloroform followed by ethanol purification, 200ul Phenol chloroform was added, mixed and centrifuged at 20℃ for 16000g for 5min.Then adding 200ul chloroform to each sample. After mixing, the samples were centrifuged at 20℃ for 16000g for 5min and transferred to LO-binding 1.5ml tube, each sample was added 2ul glycogen ,20ul NaAC ,500ul 100% ethanol (precooled in minus 20℃), mixed and centrifuged for 30min in a refrigerator at -20℃, or incubated at -20℃ for overnight. At the same time, the centrifuge was started for pre-cooling at 4℃, and 100% alcohol was treated in a refrigerator at -20℃ for 1 hour.

The samples were centrifuged at 16000g at 4℃ for 20min and remove the supernatant, 1ml

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of anhydrous ethanol was added to each sample, centrifuged at 4 degrees for 16000g for 10 min. Take out the sample to remove the supernatant and open the tube. After the liquid is naturally air-dried, Purified DNA was applied to Tru-Seq library construction using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S). For NEB kit, the DNA was resuspended in 25μL ddH2O, followed by end-repair/A-tailing with 3.5μL End Prep buffer and 1.5μL enzyme mix, the reaction was addressed at 20°C for 30 mins and 65°C for 30 mins in PCR machine. Then adding diluted 1.25μL adaptors (NEB Multiplex Oligos for Illumina, E7335S) and fully vortexed, 15μL Ligation master mix, 0.5μL Ligation enhancer, incubated at 16-20℃ for 16-20min, or at 4°C for overnight. After the ligation reaction, the samples were treated with 1.5μL USERTM enzyme, the reaction was addressed at 37°C for 15 mins in PCR machine. The samples were purified by 1×AMPure beads. The PCR procedure was performed by adding 25μL 2 × KAPA HiFi HotStart Ready Mix (KAPA biosystems, KM2602) with 2.5ul primers of NEB Oligos kit and 2.5ul Illumina universal primers, with the program of 98°C for 45s, (98°C for 15s and 60°C for 10s) with 16-17 cycles and 72°C for 1 min. The final libraries were purified by 0.9-1 × AMPure beads and applied to next-generation sequencing.

Data analysis

CUT&RUN data processing

The single-end CUT&RUN data were to mapped mm9 genome by using Bowtie2 (Version 2.2.9) [36]with the parameters –t –q –N 1 –L 25. The paired-end CUT&RUN data were mapped with the parameters: -t –q –N 1 –L 25 –X 1000 --nomixed --no-discordant. All unmapped reads, non-uniquely mapped reads and PCR duplicates were removed. For downstream analysis, we normalized the read counts by calculating the numbers of reads per kilobase of bin per million of reads sequenced (RPKM) for 100-bp bins of the genome. RPKM values were further Z-score normalized for minimizing the batch effect.

ATAC-seq data processing

The ATAC-seq reads were mapped to mm9 genome by Bowtie (version 2.2.2) [36]with parameters described previously. All unmapped reads, non-uniquely mapped reads, and PCR duplicates were removed. For downstream analysis, we normalized the read counts by calculating the RPKM for 100bp-bins of the genome. RPKM values were further normalized by Z-score normalization for minimizing batch effect.

DNA methylation data processing

we used Bismark software mapping DNA methylation reads to mm9 genome. PCR duplicates were removed. For each CpG site, the methylation level was computed as the total methylated counts divided by the total counts across all reads covering this CpG.

Analysis of the promoter H3K4me2 and H3K4me3 enrichment

Annotated promoters (TSS ± 2.5kb) of the mm9 refFlat annotation database from the UCSC genome browser were used. Whole genome Z-score normalized RPKM scores were used for calculation.

Identification of stage-specific distal ATAC-seq peaks

Stage-specific distal ATAC-seq peaks were identified as described previously. In brief, distal peaks from human embryos and ES cells were merged and then average Z-score normalized RPKM scores were calculated on them. A Shannon entropy-based method [3]was used to identify stage-specific peaks. Peaks with entropy less than 2 were selected as candidates and were further

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

selected by the following criteria: the distal peak has high enrichment at this stage (normalized RPKM > 1) and positive enrichment (normalized RPKM > 0) at no more than two additional stages.

Identification of stage-specific genes

For stage-specific genes in 8-cell embryos and ICMs, a Shannon entropy-based method [3]was used to identify stage specific-genes as previously described. Maternally expressed genes, which were defined as those expressed in GV or MII oocytes (FPKM >= 1), were excluded from this analysis to avoid the confounding effects of lingering maternal RNAs.

Clustering analysis

The k-means clustering of gene expression levels at various stages was conducted using Cluster 3.0 with the parameters –g 7 (Euclidean distance). Heat maps were generated using Java Treeview 3.0. The hierarchical clustering analysis for H3K4me3 at various stages was conducted using an R package (ape) based on the Pearson correlation between each pair of stage. The distance between two stages was calculated as: 1 − Pearson correlation.

Identification of oocyte PMDs

Briefly, average DNA methylation levels for 10-kb bins of the genome were calculated and the numbers of CpG dyads in each 10-kb bin were also counted. Bins with averaged DNA methylation level lower than 0.6 and more than 20 CpG dyads covered were selected and further merged into PMDs by BEDtools (Version v2.25.0)[37]. Promoter regions (±2.5kb) were removed from PMDs.

Identification of H3K4me2 peaks

H3K4me2 peaks were called using MACS2 (Version 2.0.9)[38] with the parameters --broad --nomodel --nolambda -g mm 1.87e9/ -g hs 2.7e9 and noisy peaks with weak signals (summed RPKM < 0) were removed from further analyses.

Statistical analysis

All statistical analyses in this study were briefly described in the text and performed using R (v3.6.1) (http://CRAN.R-project.org/).

Result

Genome-wide profiling of H3K4me2 in mouse oocytes and early embryos

Epigenome underwent drastic reprogramming during early development. To ask how the H3K4me2 inherited from sperm and oocyte to the zygotic during early development, we sought to investigate H3K4me2 in mouse gametes and preimplantation embryos using CUT&RUN methods. We optimized CUT&RUN protocol (CUT&RUN; Methods) using H3K4me2 antibody, at last, we acquired high-quality data using as few as 50 cells, most of the peaks could be retrieved using as few as 50 cells (Figure S1A, S1B). Then we collected GV oocytes, MII, PN5, early 2-cell, late 2-cell, 4-cell, 8-cell, and inner cell masses (ICMs) of mouse, and mouse embryonic stem cells (mESC) as control. We profiled H3K4me2 in oocytes and early embryos using CUT&RUN and immunofluorescence (IF). All replicates' results showed highly consistent (Figure 1A, Figure S1C). We identified 57455, 25073, 110517, 126318, 62575, and 69154 H3K4me2 peaks in the GV oocytes, late 2-cell embryos, 4-cell embryos, 8-cell embryos, ICM and mESC, respectively, the H3K4me2 was erased from GV to MII during oocytes development, paternal H3K4me2 was also erased after fertilization, and may be re-established from late 2-cell stage (Figure 1A-B). In addition, we demonstrated the H3K4me2, ATAC-seq, mRNA expression of several developmental key genes (Figure 1C). Moreover, we found H3K4me2 positively correlated with H3K4me3 and

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

chromatin accessibility in early embryos (Figure S1D-E). To our interest, the H3K4me2 genome-wide pattern of 4-cell embryos show strong association as 8-cell embryos: H3K4me2 of 4 and 8-cell embryos were in one cluster, and the correlation is high (R=0.93) (Figure S1F-G) .

Distinct H3K4me2 pattern between sperm and oocyte

We sought to systematically characterize the epigenome of mouse gametes, we included previous published DNA methylation[39], H3K4me3[40]and ATAC-seq data of early embryos[41], as well as H3K4me2 data[42] of sperm for analyses. It is reported that non-canonical pattern of H3K4me3, enriched broad domains in both promoters and distal regions, are present in mouse oocytes as broad domains in partially methylated domains (PMDs) [43,44]. Interestingly, H3K4me2 present as non-canonical pattern in GV oocytes and sperm in mouse (Figure 2A-B, 2E-F). Considering mature sperm and oocytes are in a state of inactive transcription [45,46], while it is reported that strong ChIP-Seq and CUT&RUN signal enrichment of this marker were found which is related to transcription[47]. These results raise the possibility that H3K4me2 may be deposited during gametogenesis and retained in mature gametes. To further examine this surmise, we perform K-means cluster analysis for H3K4me2 in gametes. Interestingly, by comparing promoter H3K4me2 at the sperm and GV stages, we found that promoters can be divided into four groups (Figure 2A). The first group (sperm-specific H3K4me2) are involved fertilization and germ cell development and these markers are erased at other stages lately (Figure 2A, 2C). On the contrary, the second group (GV-specific H3K4me2) are enriched for mitosis and gamete generation which are known to function in GV generation during oocyte development and they are reestablished in the lately stages (Figure 2A, 2C). The third group (sperm-GV share H3K4me2) related to mRNA processing and translation, the signal of these regions underwent erase (MII, zygote and early 2-cell stage) and rebuild (from late 2-cell stage) during development (Figure 2A, 2C). The fourth group have low H3K4me2 level at all stage (Figure 2A).

The H3K4me2 enrichment in these promoters in consistent with the dynamics changing of H3K4me3 (Figure 2A). In particularly, promoters with high CpG density are preferentially marked by H3K4me2 in sperm-GV share group (Figure 2A). Next, we further investigated H3K4me2 marked distal region in gametes. Sperm specific group are involved gamete generation and GV specific distal H3K4me2 showed functions in cell differentiation and mitotic cell cycle , those were both erased after fertilization and not re-constructed during embryos development (Figure 2B,2D). While H3K4me3 signals did not occupied the sperm specific/GV specific H3K4me2 distal regions (Figure 2B). The GV-sperm share distal H3K4me2 regions is also marked by H3K4me3, those genes participate in system development, RNA processing and regulation of gene expression (Figure 2B, 2D). Compared to GV, sperm H3K4me2 enriched in full methylated promoter but sperm distal H3K4me2 enriched in hypomethylated region the same as GV (Figure S2A-D). Therefore, these results reveal differential H3K4me2 modification patterns between sperm and oocytes.

Resetting H3K4me2 during pre-implantation embryo

Drastic epigenetic reprograming and chromosome remodeling associated with H3K4me3, H3K27me3, H3K36me3 and H3K27ac during early mammalian development[44,48,49], our previous data revealed epigenome reboot in human early development[50]. Next, we asked how parental histone mark H3K4me2 at regulatory elements reprogrammed after fertilization. To our surprise, the GV and sperm H3K4me2 peaks were almost erased in MII oocytes, and reconstructed

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

accompanying zygotic genome activation (ZGA) (Figure 3A-B). This result is further confirmed by IF analysis, showed the re-establishment of H3K4me2 may start from late 2-cell stage (Figure 1B, Figure 3B). Coincidently, The H3K4me2 methyltransferase KDM1A is up-regulated at MII stage (Figure S3A). While the CUT&RUN data which showed slightly H3K4me2 peaks (relate to other stages) re-appear at late-2-cell stage (Figure 3A). We concluded that the IF was not able to recognize the newly established weak H3K4me2 signal in the late 2-cell stage. Correlated with the activation of Prdm9 and Kmt2c at 4-cell stage (Figure S3A), H3K4me2 is enough to be detected at 4/8 cell stage (Figure 3A-B) using CUT&RUN and IF. 4/8 cell H3K4me2 targets only partially overlap with those in sperm or oocytes and these overlap targets mainly enriched in development-associated pathway or process (Figure 3C). Additionally, gametes specific H3K4me2 targets involved in gametes development and generation whereas 4/8cell specific targets were associated with cell communication and metabolic process (Figure 3C). These data indicated that the resetting of H3K4me2 involves distinct sets of genes that closely related to embryos development from mouse gametes to early embryos. In particularly, about 1% 4/8 cell H3K4me2 occurs in hypomethylated region as early as late 2-cell stage (Figure S3B-D). It implies that the resetting of H3K4me2 start from late 2-cell stage but largely re-establish at 4/8 cell stage.

Non-canonical H3K4me2 in mammalian embryos and pervasive H3K4me2 in CpG-rich regulatory regions in mouse 4 and 8-cell embryos

In GV oocytes and post ZGA stages, H3K4me2 enriched in both canonical gene promoters and partially methylated domains（PMDs）(Figure 4A-D). Our results showed that the H3K4me2 may be re-established in late 2-cell stage, and strong promoter H3K4me2 enrichment re-establishment is already finished in the 4-cell stage and maintained in 8-cell stage, furthermore, the H3K4me2 signals occupied broader regions than those at other stages (Figure 1A). We found about 65% of these gene promoters occupied by H3K4me2 in 4-cell,8-cell stage and ICM, and those genes play roles in RNA metabolic and RNA splice (Figure 4E). We then combined previous H3K4me3 and chromosome accessibility data, and we found these regions overlapped with high CG density regions, H3K4me3 enrichment and chromatin accessibility regions (Figure 4E), it shows that H3K4me2 prefer to occupy hypomethylation and high CpG density regions (Figure S4A-C).

We further asked how the stage specific distal peaks reprogrammed, about 35% of the genes promoter which associates with cell differentiation and development were not large-scale re-established H3K4me2 in 4-cell, 8-cell and blastocyst stage (Figure 4E). In particularly, we identified stage-specifically expressed genes and examined their promoter H3K4me2 (Figure S4E). We found promoters with constantly high H3K4me2 that are preferentially CpG-rich, whereas promoters with constantly low H3K4me2 are generally CpG-poor (Figure 4F, Figure S4E). Interestingly, a group of genes showed dynamic promoter H3K4me2 that correlates with gene expression (Figure S4E). These genes preferentially play roles in development, differentiation, and morphogenesis. Therefore, our data indicated that promoter H3K4me2 in early embryo development correlates with both gene activities and CpG density.

Next, we sought to investigate the distal H3K4me2, 4/8-cell embryos also showed widespread distal H3K4me2 signals (Figure 4C). We firstly identified the distal H3K4me2 regions in mouse gametes and pre-implantation embryos. These regions are relatively CpG-rich (although lower than promoters) and hypo-methylated (Figure S4B and S4C). However, compared with

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

promoters, distal H3K4me2 correlated with CpG density in a continuous and monotonic manner (Figure S4A). In sum, these data reveal widespread distal H3K4me2 in CpG-rich and hypomethylated regions in mouse 4/8-cell embryos.

Transcription regulation, CcREs priming and resolving during maternal-to-zygotic transition

Previously study found that most of the transcription factor binding regions overlap with H3K4me2-enriched regions[51]. To investigate the association between H3K4me2 and transcription regulation. We further identified stage specific distal regions and the accessibility in these distal peaks of H3K4me2 (Figure 4F). Then, we identified the binding motifs for transcription factors (TFs) enriched in these distal regions (Figure S4D). TFs in mESC and 8-cell include pluripotency factors involved system development[52,53], we found CTF is specifically enriched in 8-cell embryos. The GATA, KLF families are enriched in ICMs.

Our recent study in human found widespread deposition of H3K4me3 at promoters, including those for developmental genes, prior to ZGA, which subsequently resolve to either active or repressed states after ZGA[50]. Such promoter priming is related to embryonic development. To ask whether this is also happened in mouse, we identified specific H3K4me2 target genes in 4/8-cell and compared them with genes which are expressed or inactive at all stages. The promoters of 4/8-cell specific group decreased in ICM (Figure 5A and S5A). “Active” genes showed strong H3K4me2, H3K4me3 and ATAC-seq signals before and after ZGA (Figure 5A and S5A). We then asked if such priming-resolving epigenetic transition also occurs in the distal regions. Distal H3K4me2 sites are enriched on regulatory elements, as 75.4% (compared to 16.2% of random) of them overlap with candidate cis-Regulatory Elements (CcREs) identified by ENCODE (Figure 5B).

The enhancers are DNA sequences that increase promoter activity and thus the frequency of gene transcription. The enhancer region is usually modified by H3K27ac catalyzed by P300 /CBP and H3K4me1 histone 2 catalyzed by MLL3/4[54], while the H3K4me3 modification level is low, which is the opposite of the promoter region. We compared histone modifications to H3K4me2 and H3K27ac and found no significant signal overlap between the two histone modifications at specific developmental stages in the distal region. Distal 4/8-cell H3K4me2 marked CcREs show high enrichment of chromatin accessibility (Figure 5C, S5B). Interestingly, after 4/8-cell, one class (“active”) remain accessible and are associated with basic biological processes (GREAT analysis). By contrast, the other class (Figure 5, S5B, “poised”) are associated with H3K27me3 in ICM and somatic and located preferentially near developmental genes. These data indicate that the widespread *de novo* H3K4me2 was re-established following ZGA occurred, and H3K4me2 may be globally reset to a permissive state before they resolve after 4/8-cell (Figure 5D).

DISCUSSION

Histone modifications are fundamental epigenetic marks, the reprogramming of histone marks play important roles in early development[9]. Here in our study, we profiled the H3K4me2 landscape in mESC, oocyte and early embryos using modified low-input sensitivity CUT&RUN method. We found the H3K4me2 presented differential patterns in sperm and oocytes, exhibited non-canonical pattern in GV oocytes while canonical of sperm in mouse[33], furthermore, the GV H3K4me2 peaks were almost erased in MII oocytes, and reconstructed accompanying ZGA occurred, our results indicated that maternal H3K4me2 were lost during oocyte maturation. The

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

non-canonical H3K4me2 in mammalian embryos and pervasive H3K4me2 in CpG-rich regulatory regions in 4 and 8-cell embryos and in ICM. Hence, we provided the first landscape and revealed the potential function of H3K4me2 during early mammalian development.

How the paternal epigenetic inheritance is important for development and disease in offspring, the drastic epigenome reprogramming occurred and associated with key process during early mammalian development[1,3,9,14,39-41,55-59]. It is reported that genome-wide H3K4me3 exhibited a non-canonical pattern (ncH3K4me3) in full-grown and mature oocytes, means it exists as broad peaks at promoters and lots of distal loci[40,44,48], here in our study, we found that H3K4me2 was a non-canonical pattern in GV oocyte, but was erased in MII oocyte, indicating that H3K4me2 play an important role during oocyte maturation. The H3K4me3 was inherited in pre-implantation embryos and was erased in the late two-cell embryos following canonical H3K4me3 starts to be established[40,44], while the H3K4me2 mainly starts to be re-established in late 2-cell stage, and it exists as non-canonical pattern in the following stages. Emerging data reveal that in many species, including mouse, zebrafish and human, the parental epigenome undergoes dramatic reprogramming after terminally differentiated gamete fertilization followed by subsequent re-establishment of the embryo epigenome, leading to epigenetic ‘‘rebooting’’[9]. We found H3K4me2 were erased in MII oocyte, but reserved in sperm[33], while the paternal H3K4me2 was erased upon fertilization, and start to be re-established in late 2-cell stage, indicating the H3K4me2 participated the embryo epigenome rebooting process.

H3K4me3 is known as an active gene expression marker, mainly marks the gene promoter region. H3K4me1 and H3K4me2 mark both enhancer and promoter, were reported to have multiple functions for gene activation according to different enzymes and complexes[22,27,32,60-69]. It is reported that distinct pattern of H3K4me2 enrichment, spanning from the transcriptional start site (TSS) to the gene body, and marks a subset of tissue-specific genes, which indicated unique H3K4me2 profile mark tissue-specific gene regulation[30,32]. Our results showed that non-canonical H3K4me2 in mammalian embryos and pervasive H3K4me2 in CpG-rich regulatory regions in mouse 4 and 8-cell embryos. We next interrogate whether the H3K4me2 associated active or repressed genes in embryo development, and our data revealed that promoter H3K4me2 in early mouse embryo development correlates with both gene activities and high CpG density, as expected, the promoter H3K4me2 is consistent with H3K4me3.

Histone modification and chromatin accessibility profiles are usually used to identify transcription factor binding regions and motifs, we previously found large scale chromatin accessibility in pre-ZGA human embryos, and these regions enriched lots of transcription factors[3]. It is reported about average 90% of the transcription factor binding regions overlap with the H3K4me2-enriched regions and the stronger signals of H3K4me2 enrichment, the more likely overlap with the transcription factor binding regions [31]. Interesting, here in our study, we identified the binding motifs for transcription factors enriched in these distal regions, and CCAAT transcription factor (CTF) is specifically enriched in 8-cell embryos. And the GATA, KLF families are enriched in ICMs, GATA is critical regulators of primitive endoderm (PE), which is part of ICM [70], and KLF factors support the naive pluripotency[71]. Thus, these results showed stage-specific distal H3K4me2 involve transcription circuitry in mammalian embryonic development.

The state of chromatin influences DNA replication, DNA repair and gene expression. It has reported that accessible chromatin occupies regulatory sequences, such as promoters and

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

enhancers[72]. These elements interact with cell type-specific transcription factors to start transcriptional programs that determinate cell fate[73]. Early animal embryos undergo extensive histone modification reprogramming and chromatin remodeling before terminally differentiated gametes develop into pluripotent cells. Previous study found that extensive overlaps of accessible promoters and enhancers at 4-cell stage in human[50]. This phenomenon indicated crosstalk function between histone modification and chromatin state. In our study, most of the H3K4me2 re-established in chromatin accessible regions including promoters and distal regions at 4-cell stage. So, we infer that the H3K4me2 establishment may play an important shaping role in chromatin state. Therefore, our results furthermore support the crosstalk between accessible chromatin and histone modification. But how the chromatin state influences H3K4me2 deposition or the opposite need further investigation.

Dynamic histone modifications play important roles during mammalian Maternal-to-zygotic transition (MZT), ~22% of the oocyte genome associated with broad H3K4me3 are anti-correlated with DNA methylation and then become confined to transcriptional-start-site (TSS) regions in 2-cell embryos, concomitant with the onset of major ZGA[48], while H3K4me2 was erased in MII oocytes, the sperm H3K4me2 was erased after fertilization. The same as H3K4me3, the embryos' H3K4me2 was also re-established accompanied with the onset of major ZGA. The ncH3K4me3 in oocytes overlaps almost exclusively with partially methylated DNA domains[40], the H3K4me2 was also enriched in both canonical gene promoters and partially methylated domains (PMDs).

Taken together, our investigation provided a genome-wide landscape of H3K4me2 modification in mammalian oocytes and pre-implantation embryos, facilitating further exploration of the epigenetic regulation network in mammalian early development.

DATA AVAILABILITY

All raw sequencing data generated during this work have been submitted to the Genome Sequence Archive in National Genomics Data Center (China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences)

H3K4me2 CUT&RUN data of Human K562 cell line

http://bigd.big.ac.cn/gsa-human

H3K4me2 CUT&RUN data of mouse gametes and early embryos

https://ngdc.cncb.ac.cn/gsa/

FUNDING AND ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China (2019YFA0802200 and 2019YFA0110900 to Jiawei Xu), the National Natural Science Foundation of China (31870817 and 32170819 to Jiawei Xu), the Scientific and Technological Innovation Talent Project of Universities of Henan Province (20HASTIT045 to Jiawei Xu), the Central Plains-Youth Talents Program (ZYQR201810155 to Jiawei Xu).

Author contributions

Conceptualization: JWX

Methodology: CW

Experiments performing: CW, SQY, XQX, KYH, YQW, YL, SLS

Data NGS (next generation sequencing): CW

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bioinformatics analysis: YS, JWX,CW

Supervision: JWX
Writing—original draft: JWX, CW, YS
Writing—review & editing: JWX, CW, YS

Competing interests

The authors declare no competing interests.

REFERENCES

1 Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. Science 293, 1089-1093, doi:10.1126/science.1063443 (2001).

2 Burton, A. & Torres-Padilla, M. E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat Rev Mol Cell Biol 15, 723-734, doi:10.1038/nrm3885 (2014).

3 Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature 557, 256-260, doi:10.1038/s41586-018-0080-8 (2018).

4 Stillman, B. Histone Modifications: Insights into Their Influence on Gene Expression. Cell 175, 6-9, doi:10.1016/j.cell.2018.08.032 (2018).

5 Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. & Zhao, Y. SnapShot: histone modifications. Cell 159, 458-458.e451, doi:10.1016/j.cell.2014.09.037 (2014).

6 Peterson, C. L. & Laniel, M. A. Histones and histone modifications. Curr Biol 14, R546-551, doi:10.1016/j.cub.2004.07.007 (2004).

7 Hammoud, S. et al. Distinctive chromatin in human sperm packages genes for embryo development. Nature 460, 473-478, doi:10.1038/nature08162 (2009).

8 Hammoud, S. et al. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell stem cell 15, 239-253, doi:10.1016/j.stem.2014.04.006 (2014).

9 Xia, W. & Xie, W. Rebooting the Epigenomes during Mammalian Early Embryogenesis. Stem Cell Reports 15, 1158-1175, doi:10.1016/j.stemcr.2020.09.005 (2020).

10 Goldberg, A. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140, 678-691, doi:10.1016/j.cell.2010.01.003 (2010).

11 Jin, C. et al. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature genetics 41, 941-945, doi:10.1038/ng.409 (2009).

12 Mito, Y., Henikoff, J. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. Nature genetics 37, 1090-1097, doi:10.1038/ng1637 (2005).

13 Ishiuchi, T. et al. Reprogramming of the histone H3.3 landscape in the early mouse embryo. Nature structural & molecular biology 28, 38-49, doi:10.1038/s41594-020-00521-1 (2021).

14 Zheng, H. et al. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell 63, 1066-1079, doi:10.1016/j.molcel.2016.08.032 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

15 Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol 20, 259-266, doi:10.1038/nsmb.2470 (2013).

16 Soares, L. M. et al. Determinants of Histone H3K4 Methylation Patterns. Mol Cell 68, 773-785 e776, doi:10.1016/j.molcel.2017.10.013 (2017).

17 Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol 8, 307-318, doi:10.1038/nrm2143 (2007).

18 Eissenberg, J. C. & Shilatifard, A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol 339, 240-249, doi:10.1016/j.ydbio.2009.08.017 (2010).

19 Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell 25, 15-30, doi:10.1016/j.molcel.2006.12.014 (2007).

20 Rickels, R. et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat Genet 49, 1647-1653, doi:10.1038/ng.3965 (2017).

21 Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. Nat Struct Mol Biol 21, 210-219, doi:10.1038/nsmb.2784 (2014).

22 Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

23 Hu, D. et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol Cell Biol 33, 4745-4754, doi:10.1128/mcb.01181-13 (2013).

24 Liu, T. et al. Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res 21, 227-236, doi:10.1101/gr.115519.110 (2011).

25 Du, Z. et al. Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in Oryza sativa L. Japonica. Mol Plant 6, 1463-1472, doi:10.1093/mp/sst018 (2013).

26 van Dijk, K. et al. Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana. BMC Plant Biol 10, 238, doi:10.1186/1471-2229-10-238 (2010).

27 Wang, S., Meyer, D. H. & Schumacher, B. H3K4me2 regulates the recovery of protein biosynthesis and homeostasis following DNA damage. Nat Struct Mol Biol 27, 1165-1177, doi:10.1038/s41594-020-00513-1 (2020).

28 Liu, Y. et al. H3K4me2 functions as a repressive epigenetic mark in plants. Epigenetics & chromatin 12, 40, doi:10.1186/s13072-019-0285-6 (2019).

29 Zhang, C. et al. Narrow H3K4me2 is required for chicken PGC formation. J Cell Physiol 236, 1391-1400, doi:10.1002/jcp.29945 (2021).

30 Zhang, J., Parvin, J. & Huang, K. Redistribution of H3K4me2 on neural tissue specific genes during mouse brain development. BMC Genomics 13 Suppl 8, S5, doi:10.1186/1471-2164-13-S8-S5 (2012).

31 Wang, Y., Li, X. & Hu, H. H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics 103, 222-228, doi:10.1016/j.ygeno.2014.02.002 (2014).

32 Pekowska, A., Benoukraf, T., Ferrier, P. & Spicuglia, S. A unique H3K4me2 profile

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

marks tissue-specific gene regulation. Genome Res 20, 1493-1502, doi:10.1101/gr.109389.110 (2010).

33 Lambrot, R., Siklenka, K., Lafleur, C. & Kimmins, S. The genomic distribution of histone H3K4me2 in spermatogonia is highly conserved in spermdagger. Biol Reprod 100, 1661-1672, doi:10.1093/biolre/ioz055 (2019).

34 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc 13, 1006-1019, doi:10.1038/nprot.2018.015 (2018).

35 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6, doi:10.7554/eLife.21856 (2017).

36 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359, doi:10.1038/nmeth.1923 (2012).

37 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

38 Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

39 Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. Cell 157, 979-991, doi:10.1016/j.cell.2014.04.017 (2014).

40 Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 537, 553-557, doi:10.1038/nature19361 (2016).

41 Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature 534, 652-657, doi:10.1038/nature18606 (2016).

42 Siklenka, K. et al. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science (New York, N.Y.) 350, aab2006, doi:10.1126/science.aab2006 (2015).

43 Hanna, C. W. et al. MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nature structural & molecular biology 25, 73-82, doi:10.1038/s41594-017-0013-5 (2018).

44 Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 537, 558-562, doi:10.1038/nature19362 (2016).

45 Li, L., Lu, X. & Dean, J. The maternal to zygotic transition in mammals. Molecular aspects of medicine 34, 919-938, doi:10.1016/j.mam.2013.01.003 (2013).

46 Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. Annual review of cell and developmental biology 30, 581-613, doi:10.1146/annurev-cellbio-100913-013027 (2014).

47 Liu, Y.-W. et al. LincRNAFEZF1-AS1 represses p21 expression to promote gastric cancer proliferation through LSD1-Mediated H3K4me2 demethylation. Molecular cancer 16, 39, doi:10.1186/s12943-017-0588-9 (2017).

48 Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature 537, 548-552, doi:10.1038/nature19360 (2016).

49 Xu, Q. et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nat Genet 51, 844-856, doi:10.1038/s41588-019-0398-7 (2019).

50 Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. Science 365, 353-360, doi:10.1126/science.aaw5118 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

51 Wang, Y., Li, X. & Hu, H. H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics 103, 222-228, doi:10.1016/j.ygeno.2014.02.002 (2014).

52 Dominguez, M. H., Ayoub, A. E. & Rakic, P. POU-III transcription factors (Brn1, Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex. Cerebral cortex (New York, N.Y. : 1991) 23, 2632-2643, doi:10.1093/cercor/bhs252 (2013).

53 Sarkar, A. & Hochedlinger, K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell stem cell 12, 15-30, doi:10.1016/j.stem.2012.12.007 (2013).

54 Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. Annual review of biochemistry 89, 213-234, doi:10.1146/annurev-biochem-011420-095916 (2020).

55 Xavier, M. J., Roman, S. D., Aitken, R. J. & Nixon, B. Transgenerational inheritance: how impacts to the epigenetic and genetic information of parents affect offspring health. Hum Reprod Update 25, 518-540, doi:10.1093/humupd/dmz017 (2019).

56 Chen, Q., Yan, W. & Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. Nat Rev Genet 17, 733-743, doi:10.1038/nrg.2016.106 (2016).

57 Zhang, B. et al. Widespread Enhancer Dememorization and Promoter Priming during Parental-to-Zygotic Transition. Molecular cell 72, doi:10.1016/j.molcel.2018.10.017 (2018).

58 Chen, Z., Djekidel, M. & Zhang, Y. J. N. g. Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. 53, 551-563, doi:10.1038/s41588-021-00821-2 (2021).

59 Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. 365, 353-360, doi:10.1126/science.aaw5118 (2019).

60 Agarwal, S. et al. KDM1A maintains genome-wide homeostasis of transcriptional enhancers. doi:10.1101/gr.234559.118 (2021).

61 Ye, R., Cao, C. & Xue, Y. J. E. i. b. Enhancer RNA: biogenesis, function, and regulation. 64, 883-894, doi:10.1042/ebc20200014 (2020).

62 Kang, Y., Kim, Y., Kang, J. & Kim, A. J. F. j. o. p. o. t. F. o. A. S. f. E. B. Histone H3K4me1 and H3K27ac play roles in nucleosome eviction and eRNA transcription, respectively, at enhancers. 35, e21781, doi:10.1096/fj.202100488R (2021).

63 Kang, H. et al. Dynamic regulation of histone modifications and long-range chromosomal interactions during postmitotic transcriptional reactivation. 34, 913-930, doi:10.1101/gad.335794.119 (2020).

64 Ishihara, H. et al. Primed histone demethylation regulates shoot regenerative competency. 10, 1786, doi:10.1038/s41467-019-09386-5 (2019).

65 Murphy, P., Wu, S., James, C., Wike, C. & Cairns, B. J. C. Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. 172, 993-1006.e1013, doi:10.1016/j.cell.2018.01.022 (2018).

66 Yan, J. et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. 28, 204-220, doi:10.1038/cr.2018.1 (2018).

67 Soldi, M. et al. Chromatin proteomics reveals novel combinatorial histone modification signatures that mark distinct subpopulations of macrophage enhancers. 45, 12195-12213,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

doi:10.1093/nar/gkx821 (2017).

68 Dorighi, K. et al. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. 66, 568-576.e564, doi:10.1016/j.molcel.2017.04.018 (2017).

69 Kaikkonen, M. et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. 51, 310-325, doi:10.1016/j.molcel.2013.07.010 (2013).

70 Morrisey, E. E. et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & development 12, 3579-3590 (1998).

71 Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nature reviews. Molecular cell biology 17, 155-169, doi:10.1038/nrm.2015.28 (2016).

72 Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. Annual review of biochemistry 57, 159-197 (1988).

73 Rivera, C. M. & Ren, B. Mapping human epigenomes. Cell 155, 39-55, doi:10.1016/j.cell.2013.09.011 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIG LEGENDS

Fig. 1. Mapping H3K4me2 in mouse (C57BL6/N) gametes and pre-implantation embryos.

(A) The IGV view showing the global landscape of H3K4me2 signals in mouse sperm, GV oocytes, MII oocytes, pre-implantation embryos. H3K4me2 data in sperm is from ENCODE (Lambrot R et al). ICM were isolated from blastocyst.

(B) Immunofluorescence of H3K4me2 in mouse(C57BL6/N) GV oocytes, MII oocytes, PN5 zygote, early 2-cell, late 2-cell, 4-cell, 8-cell, morula and blastocyst, scale bar is also shown.

(C) H3K4me2 enrichment near typical genes and their expression levels. H3K4me2 signal, ATAC-seq and Gene expression are shown.

Fig. S1. Validation of H3K4me2 CUT&RUN data in Human K562 cell line, oocytes and early embryos.

(A) The IGV views showing H3K4me2 distributions by CUT&RUN using various numbers of Human K562 cell line. The ENCODE references are added for comparison. GEO accession: GSM733651 .

(B) The Pearson correlation coefficients showing the comparison between CUT&RUN of H3K4me2 using various numbers of Human K562 cells and conventional ChIP–seq data from ENCODE.

(C) Scatter plots showing the correlations of biological replicates (n=2) of H3K4me2 signals (2kb-bin, whole genome) by CUT&RUN for each stage of mouse oocytes and early embryos. The Pearson correlation coefficients are also shown.

(D) Scatter plots comparing the H3K4me2 signals with ATAC-seq (Assay for Transposase-Accessible Chromatin with high throughput sequencing) at promoters or distal accessible regions in mouse oocytes and early embryos. Spearman correlation coefficients are also shown.

(E) Scatter plots comparing the H3K4me2 signals with H3K4me3 signals in mouse sperm, oocytes and early embryos. Spearman correlation coefficients are also shown.

(F) Hierarchical clustering of H3K4me2 mouse oocytes and early embryos. Pearson correlation was used to measure distances.

(G) Scatter plots comparing the H3K4me2 signals between 4-cell and 8-cell in mouse embryos. Spearman correlation coefficients are also shown.

Fig. 2. Comparison of the H3K4me2 between sperm and oocyte.

(A-B) Heatmaps showing H3K4me2 and H3K4me3 expressed genes between sperm and GV oocyte in promoter and distal region among different stages. CpG densities are also mapped and shown.

(C) The enriched GO terms of H3K4me2 marked genes in GV and sperm specific or shared genes, P value is also calculated and shown.

(D) The GREAT analysis result for distal peaks in GV and sperm specific or shared genes is shown. P value is also calculated and shown.

(E-F) The IGV views showing H3K4me2 expressions in GV and sperm specific or shared genes at representative promoters (left, shaded) or distal accessible regions (right, shaded).

Fig. S2. H3K4me2 in mouse gametes.

(A) Heatmaps showing H3K4me2 signals at all promoter regions (TSS ± 2.5kb) in mouse gametes. DNA methylation (TSS ± 2.5kb) is also mapped. DNA methlation data cited in GEO dataset: GSE56697.

(B) The box plots showing DNA methylation level at all promoter regions in mouse gametes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Random peaks are shown as a control. DNA methlation data cited in GEO dataset: GSE56697.

(C) Heatmaps showing H3K4me2 signals at all distal regions (peak ± 2kb) in mouse gametes. DNA methylation (peak ± 2kb) is also mapped. DNA methlation data cited in GEO dataset: GSE56697.

(D) The box plots showing DNA methylation level at all distal regions in mouse gametes. Random peaks are shown as a control. DNA methlation data cited in GEO dataset: GSE56697.

Fig. 3. Resetting H3K4me2 during mouse pre-implantation development.

(A) H3K4me2 peaks number in mouse embryo stem cell(mESC), mouse gametes and pre-implantation embryos was calculated and shown.

(B) The IGV view showing H3K4me2 signals in mESC, mouse gametes and pre-implantation embryos, Ncc1 gene sites was shown.

(C) Venn diagrams showing the overlaps of H3K4me2 marked genes between sperm/GV oocytes, separately (left) or jointly (right), and 4-cell. The enriched GO terms for shared or non-overlapping genes are also shown.

Fig. S3. H3K4me2 in late 2-cell embryos.

(A) Heatmaps showing the expression of H3K4me2 related regulators in mouse (based on published RNA-seq data[40]).

(B) The box plots showing DNA methylation level at all promoter regions and distal region in mouse 2-cell stage. Random peaks are shown as a control.

(C-D) Venn diagrams showing De novo 4-cell H3K4me2 promoter and distal peaks number.

Fig. 4. Pervasive H3K4me2 in CpG-rich regulatory regions in mouse embryos.

(A-B) H3K4me2 and H3K4me3 enrichment near TSS (TSS ± 5kb) and PMD (3× PMD) in mouse gametes, pre-implantation embryos and mESC is shown.

(C) The IGV view showing promoter H3K4me2 signals and widespread distal H3K4me2 signals in mouse oocytes and early embryos.

(D) The IGV view showing H3K4me2 signals and DNA methylation distributions in mouse oocytes and early embryos at PMD (partially methylation district) region.

(E) Heat maps showing H3K4me2, H3K4me3 enrichment at the promoters in all stage shared and 4-cell/8-cell specific embryos. ATAC-seq, promoter CpG density and GO analysis results are also shown.

(F) Heatmaps showing stage-specifically expressed distal peaks of H3K4me2 among 8-cell, ICM and mESC. ATAC-seq is also shown.

Fig. S4. H3K4me2 in mouse 4/8-cell embryos.

(A) Scatter plots comparing the promoter and distal H3K4me2 signals with CpG densities in mouse gametes, 4-cell, 8-cell, ICM and mESC. The Spearman correlation coefficients are also shown.

(B) The box plots showing the CpG density and promoter/distal H3K4me2 peaks. Random peaks are shown as a control.

(C) The box plot showing DNA methylation in promoter and distal 4-cell H3K4me2 peaks. Random peaks are shown as a control.

(D) TF motifs identified from active enhancers in mouse early embryos (8-cell, ICM) and mESC. showing motif enrichment p value < 1e-20 at least at one stage were included. Circle size showing TF enrichment and the expression of the TF is color coded.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(E) Heatmaps showing stage-specifically expressed genes further classified by their H3K4me2 states (high, dynamic, and low). RNA-seq, H3K4me3 state, CpG densities and GO analysis results are also mapped and shown.

Fig. 5. Chromatin states at specific stages and H3K4me2 mode chart during mouse maternal-to-zygotic transition.

(A) Heatmaps showing H3K4me2, H3K4me3 and ATAC-seq, "All active" and "All inactive" promoters and corresponding genes expression. CpG densities are also mapped and shown.

(B) The bar chart showing H3K4me2 distal peaks overlap with CcREs 4/8-cell stage. Random peaks are shown as a control.

(C) Heat maps showing the chromatin states of CcREs overlapping with H3K4me2 and H3K27ac distal peaks. Clustering was conducted using ATAC-seq, H3K4me2 in 4-cell, 8-cell and ICM stages. The GREAT analysis result for each cluster is shown (right), Mouse histone modified H3K27ac data refer to GEO database : GSE72784.

(D) Dynamic reprogramming of H3K4me2 during mouse parental-to-zygotic transition. A schematic model showing "H3K4me2 resetting" during parental-to-zygotic transition. After fertilization, the parental H3K4me2 is almost globally erased. De novo H3K4me2 promoters and distal regions in CpG-rich and DNA hypomethylation regions were re-established at late 2-cell stage, possibly as a priming chromatin state. This leads to resolution of primed promoters to "active" promoters with H3K4me3, or "poised". A similar transition occurs for putative enhancers.

Fig. S5. Chromatin states at specific stages and distal CcREs.

(A) Box plots showing H3K4me2, H3K4me3 and ATAC-seq signals, "All active" and "All inactive" promoters in mouse gametes, embryos and mESC. CpG density for each group is also shown (right).

(B) The box plot showing the ATAC-seq, H3K4me2 and H3K27ac signals at putative active, poised or random enhancers in mouse gametes, early embryos or mESC, Mouse histone modified H3K27ac data refer to GEO database : GSE72784.

Supplementary Table

The samples applyed in this study are listed in the table below.

Table1 CUT&RUN experimental sample information

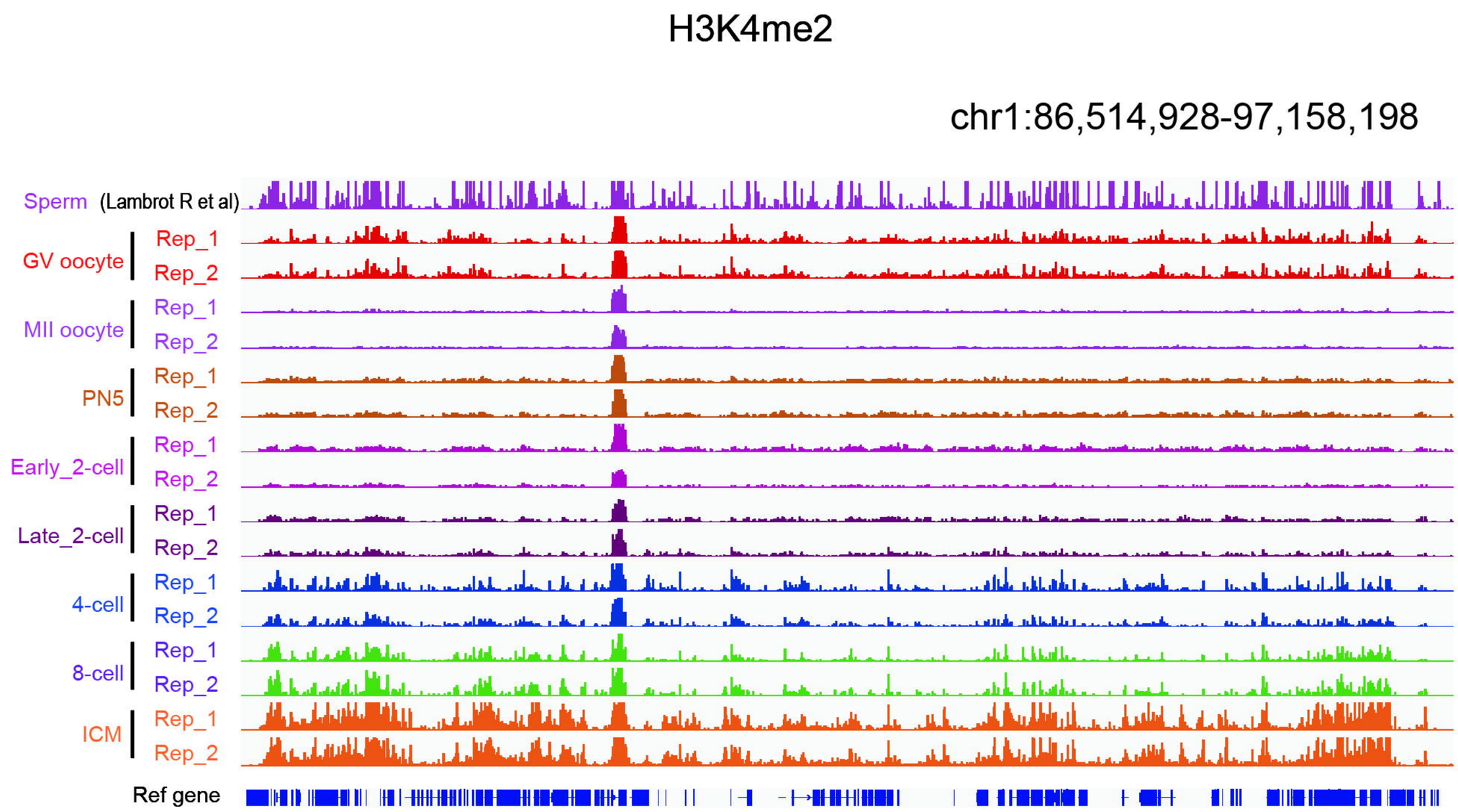
Sample name	Experiment	Replicate	Sample number	Cell number
K562 cell line	H3K4me2CUT&RUN	Rep1	50	50
K562 cell line	H3K4me2CUT&RUN	Rep2	50	50
K562 cell line	H3K4me2CUT&RUN	Rep1	100	100
K562 cell line	H3K4me2CUT&RUN	Rep2	100	100
K562 cell line	H3K4me2CUT&RUN	Rep1	200	200
K562 cell line	H3K4me2CUT&RUN	Rep2	200	200
K562 cell line	H3K4me2CUT&RUN	Rep1	500	500
K562 cell line	H3K4me2CUT&RUN	Rep2	500	500
K562 cell line	H3K4me2CUT&RUN	Rep1	5000	5000
K562 cell line	H3K4me2CUT&RUN	Rep2	5000	5000
K562 cell line	H3K4me2CUT&RUN	Rep1	10000	10000
K562 cell line	H3K4me2CUT&RUN	Rep2	10000	10000
MES cell line	H3K4me2CUT&RUN	Rep1	50	50
MES cell line	H3K4me2CUT&RUN	Rep2	50	50
MES cell line	H3K4me2CUT&RUN	Rep1	100	100
MES cell line	H3K4me2CUT&RUN	Rep2	100	100

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

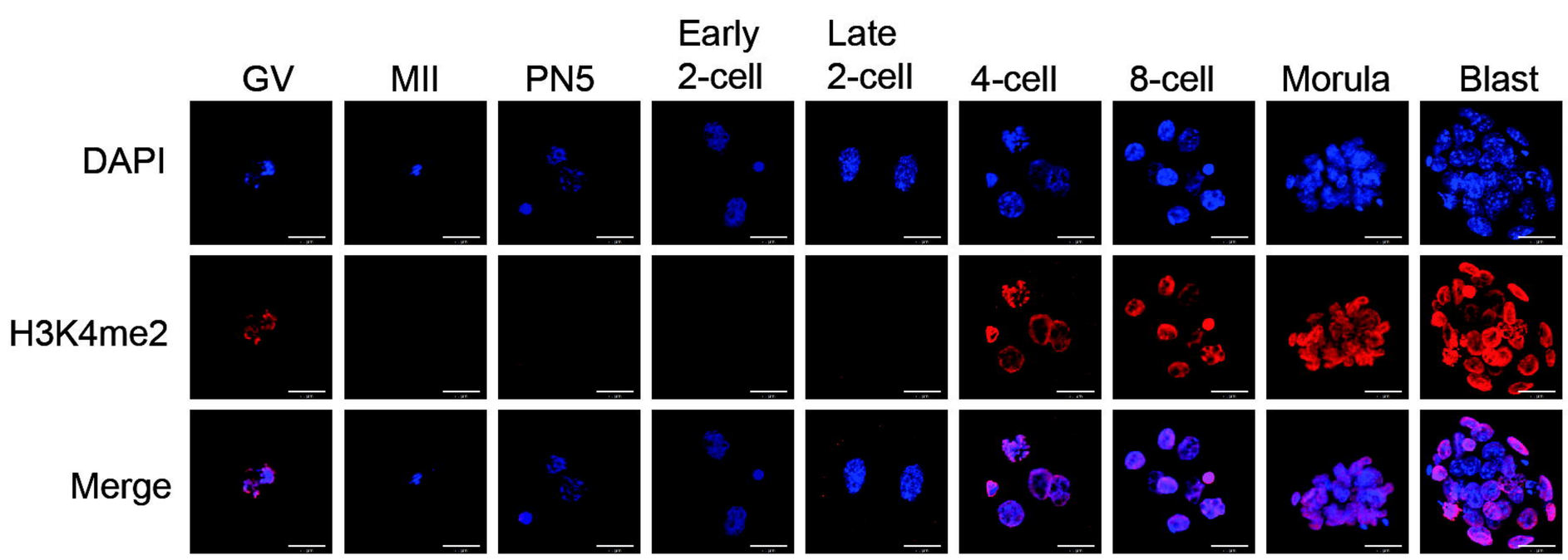
Sample name	Experiment	Replicate	Sample number	Cell number
MES cell line	H3K4me2CUT&RUN	Rep1	200	200
MES cell line	H3K4me2CUT&RUN	Rep2	200	200
MES cell line	H3K4me2CUT&RUN	Rep1	500	500
MES cell line	H3K4me2CUT&RUN	Rep2	500	500
MES cell line	H3K4me2CUT&RUN	Rep1	5000	5000
MES cell line	H3K4me2CUT&RUN	Rep2	5000	5000
MES cell line	H3K4me2CUT&RUN	Rep1	10000	10000
MES cell line	H3K4me2CUT&RUN	Rep2	10000	10000
GV oocyte	H3K4me2CUT&RUN	Rep1	190	190
GV oocyte	H3K4me2CUT&RUN	Rep2	190	190
MII oocyte	H3K4me2CUT&RUN	Rep1	165	165
MII oocyte	H3K4me2CUT&RUN	Rep2	165	165
PN5 zygote	H3K4me2CUT&RUN	Rep1	250	250
PN5 zygote	H3K4me2CUT&RUN	Rep2	250	250
Early_2cell	H3K4me2CUT&RUN	Rep1	86	172
Early_2cell	H3K4me2CUT&RUN	Rep2	86	172
Late_2cell	H3K4me2CUT&RUN	Rep1	98	196
Late_2cell	H3K4me2CUT&RUN	Rep2	98	196
4cell	H3K4me2CUT&RUN	Rep1	65	260
4cell	H3K4me2CUT&RUN	Rep2	65	260
8cell	H3K4me2CUT&RUN	Rep1	25	200
8cell	H3K4me2CUT&RUN	Rep2	25	200
ICM	H3K4me2CUT&RUN	Rep1	14	≈420
ICM	H3K4me2CUT&RUN	Rep2	14	≈420

Fig. 1. Mapping H3K4me2 in mouse gametes and pre-implantation embryos.

A

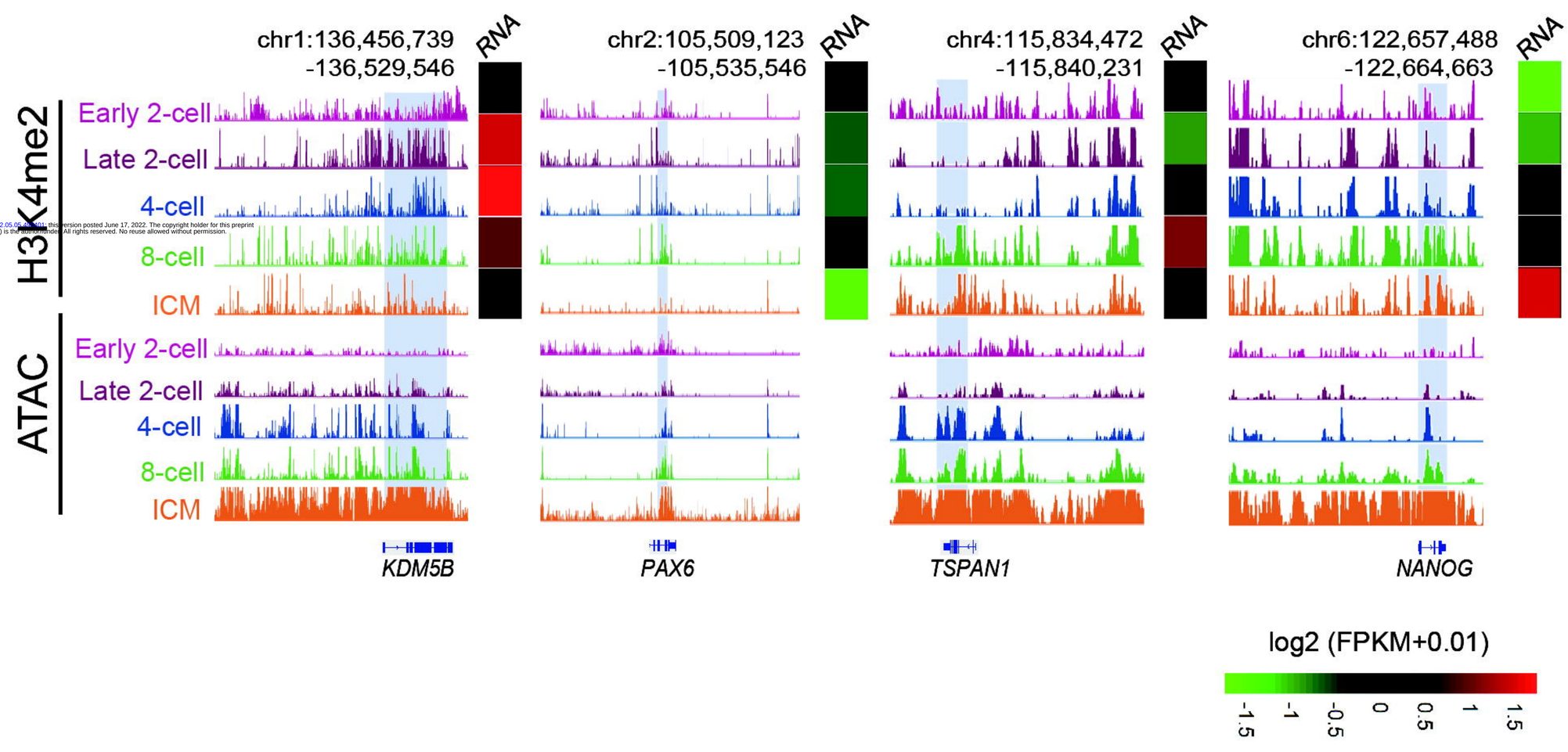


B



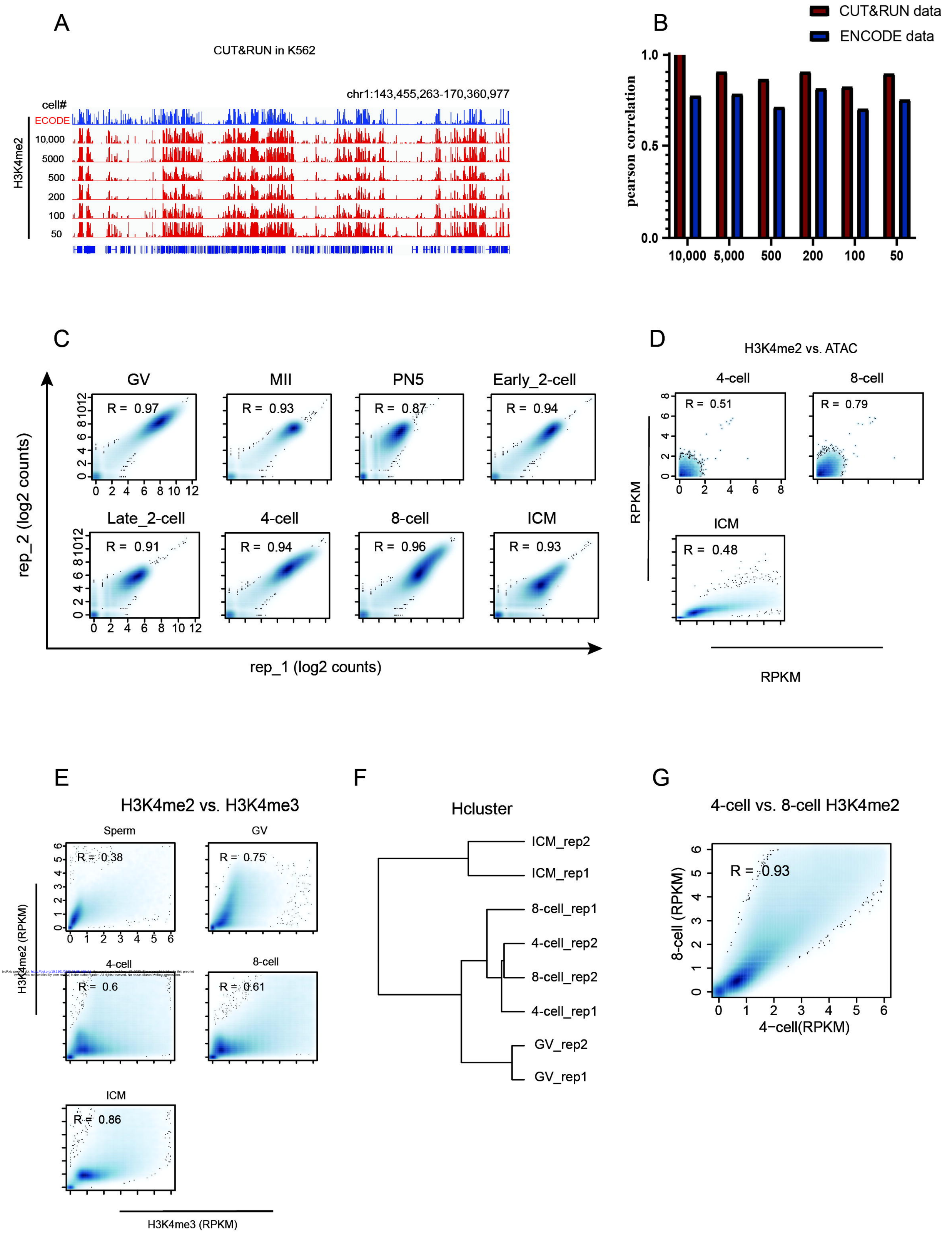
*scale bar=20μm

C



bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.491491 this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

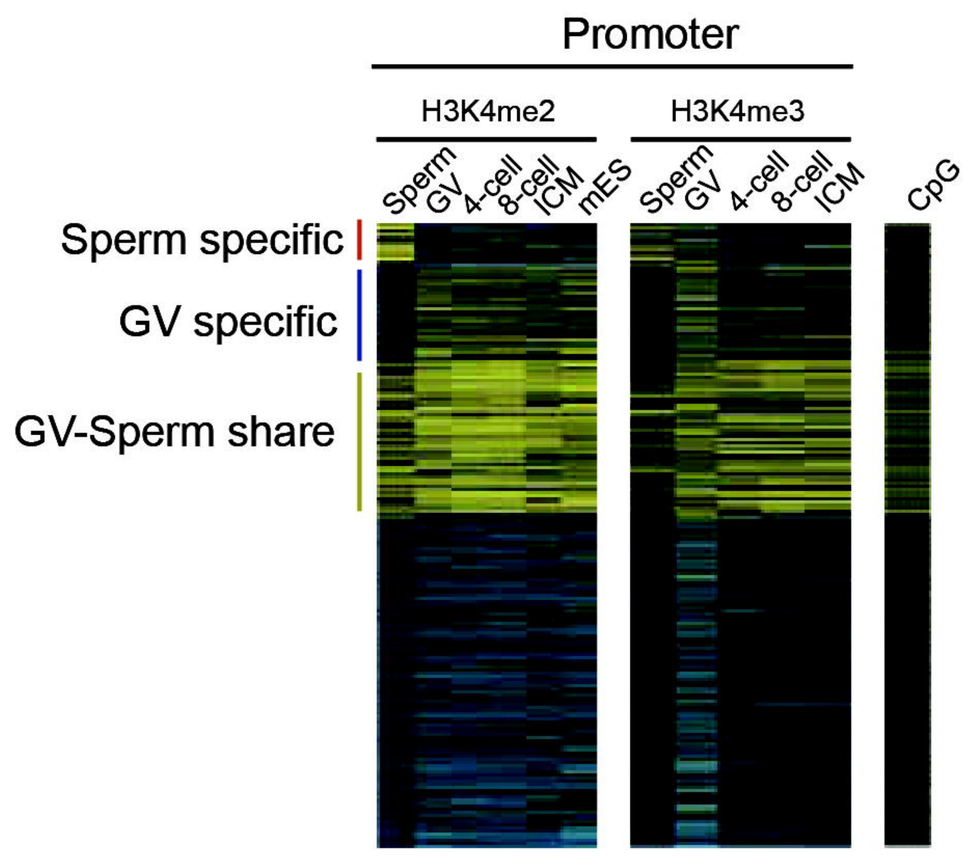
Fig. S1. Validation of CUT&RUN data



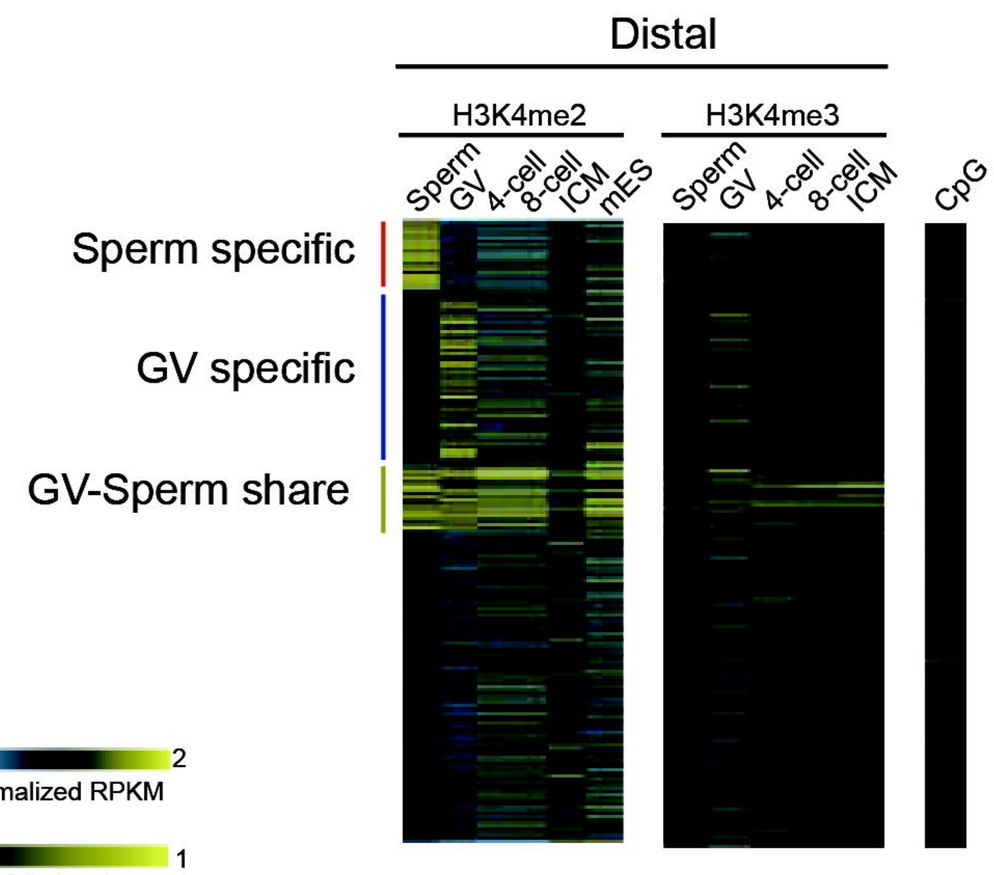
bioRxiv preprint doi: https://doi.org/10.1101/2023.05.05.499491 this version posted June 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

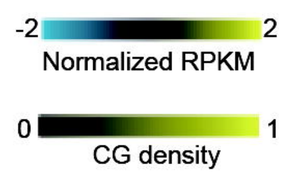
Fig. 2. Comparison of H3K4me2 between sperm and oocyte

A



B





C

GO enrichment of H3K4me2 marked genes

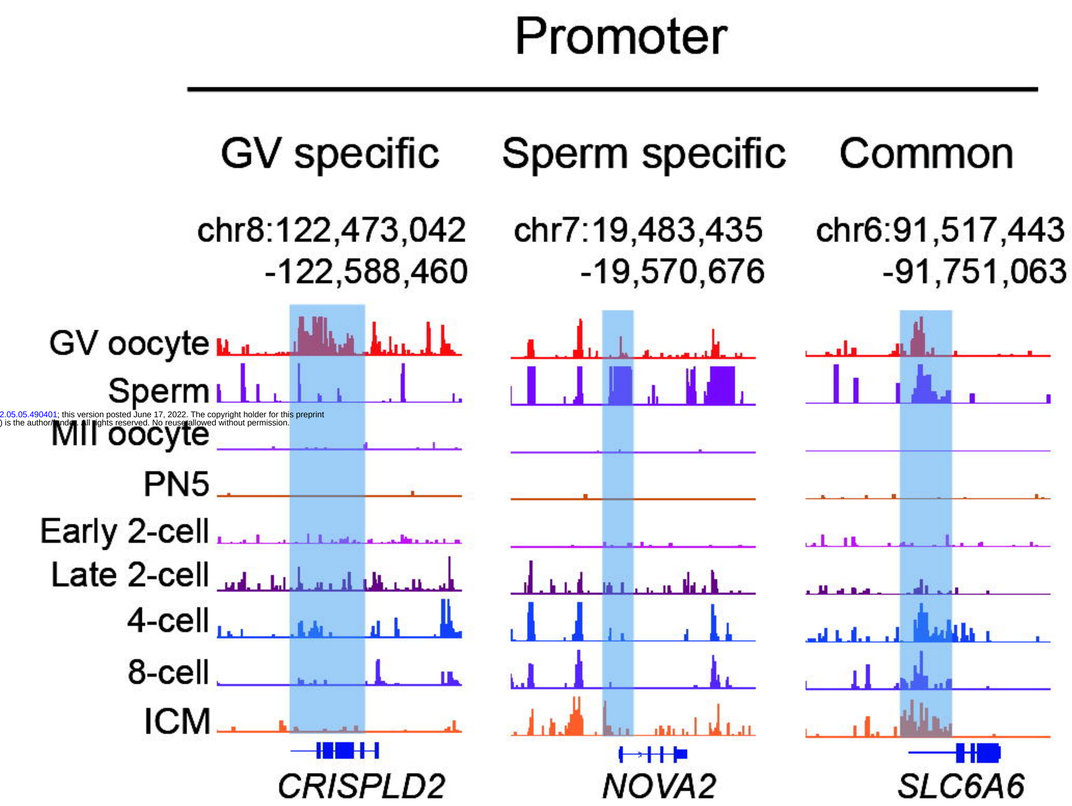
Sperm-spcific	Fertilization *(P = 0.021)*
	Germ cell development *(P = 0.018)*
GV-spcific	Mitosis *(P = 0.031)*
	Gamete generation *(P = 0.041)*
Share	mRNA processincg *(P = 0.035)*
	Mitotic cell cycle *(P = 0.044)*
	Regulation of translation *(P < 0.001)*

D

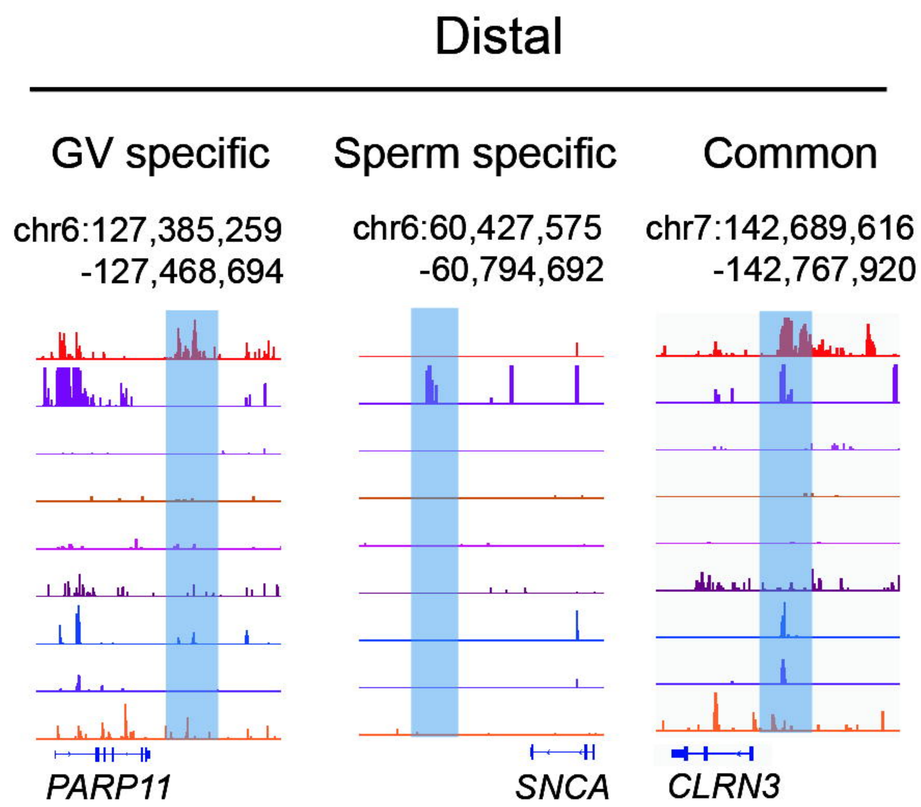
GREAT analysis terms of distal peaks

Sperm-spcific	Spermatogenesis *(P = 0.013)*
	Gamete generation *(P = 0.025)*
GV-spcific	Cell differentiation *(P < 0.001)*
	Mitotic cell cycle *(P < 0.001)*
Share	System development *(P = 0.040)*
	RNA processing *(P < 0.001)*
	Regulation of gene expression *(P = 0.002)*

E

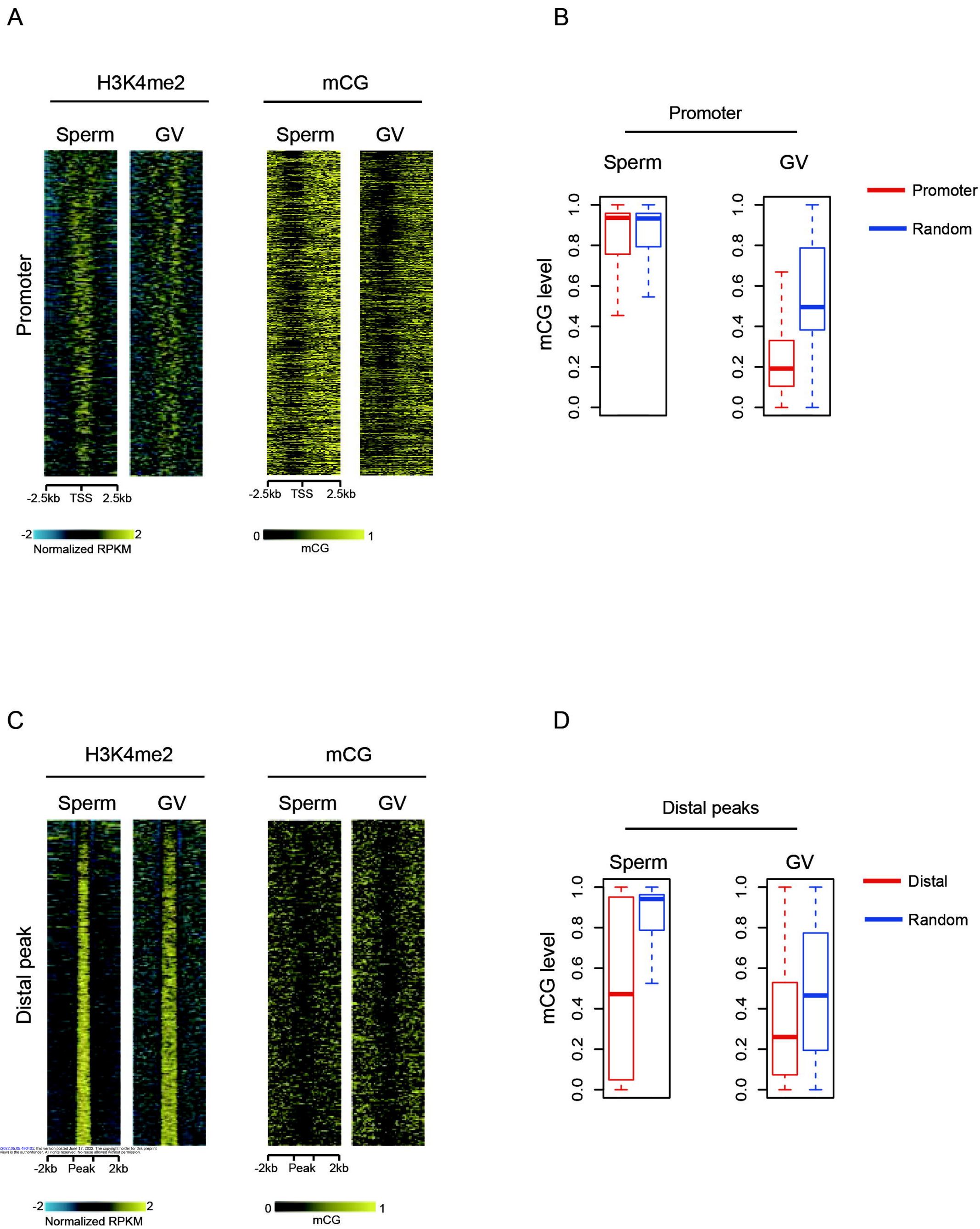


F



bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

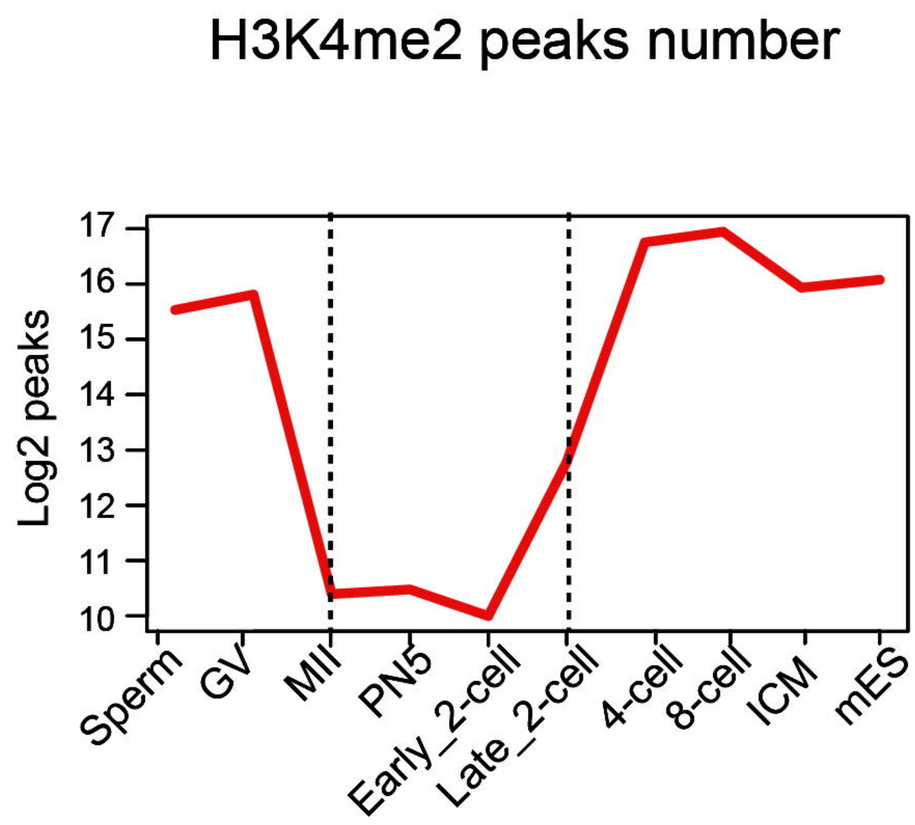
Fig. S2. H3K4me2 in mouse gametes



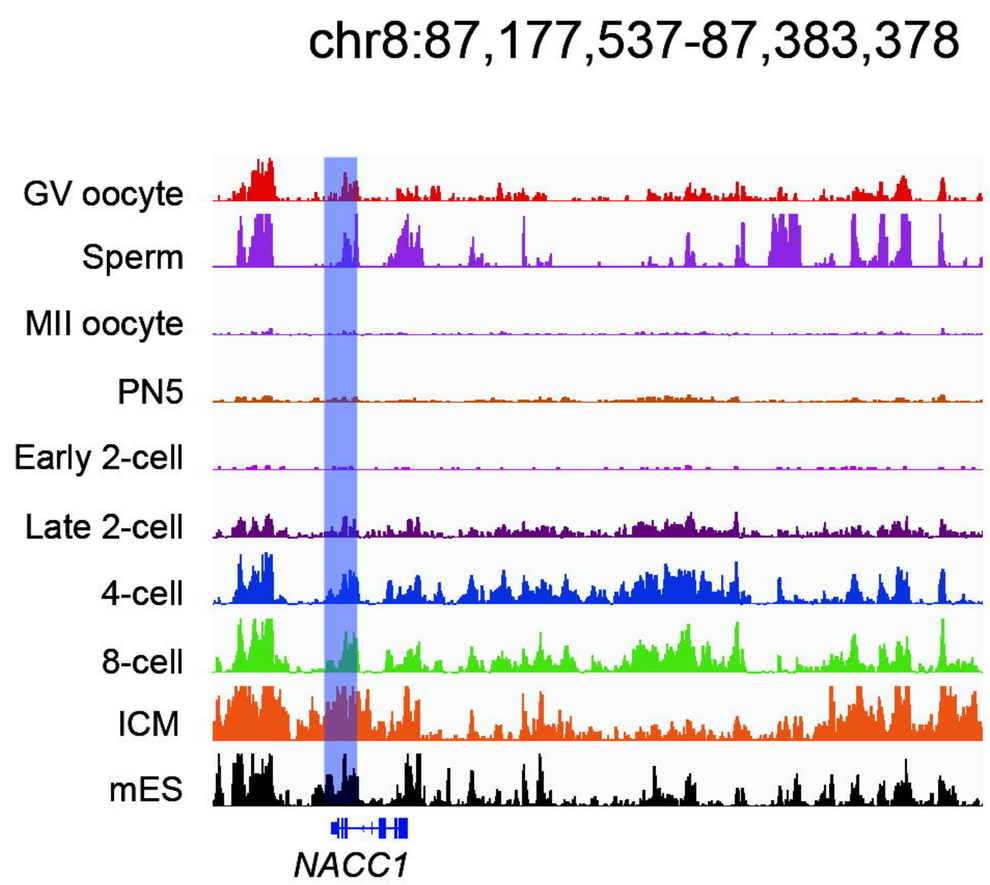
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 3. Resetting H3K4me2 during mouse pre-implantation development

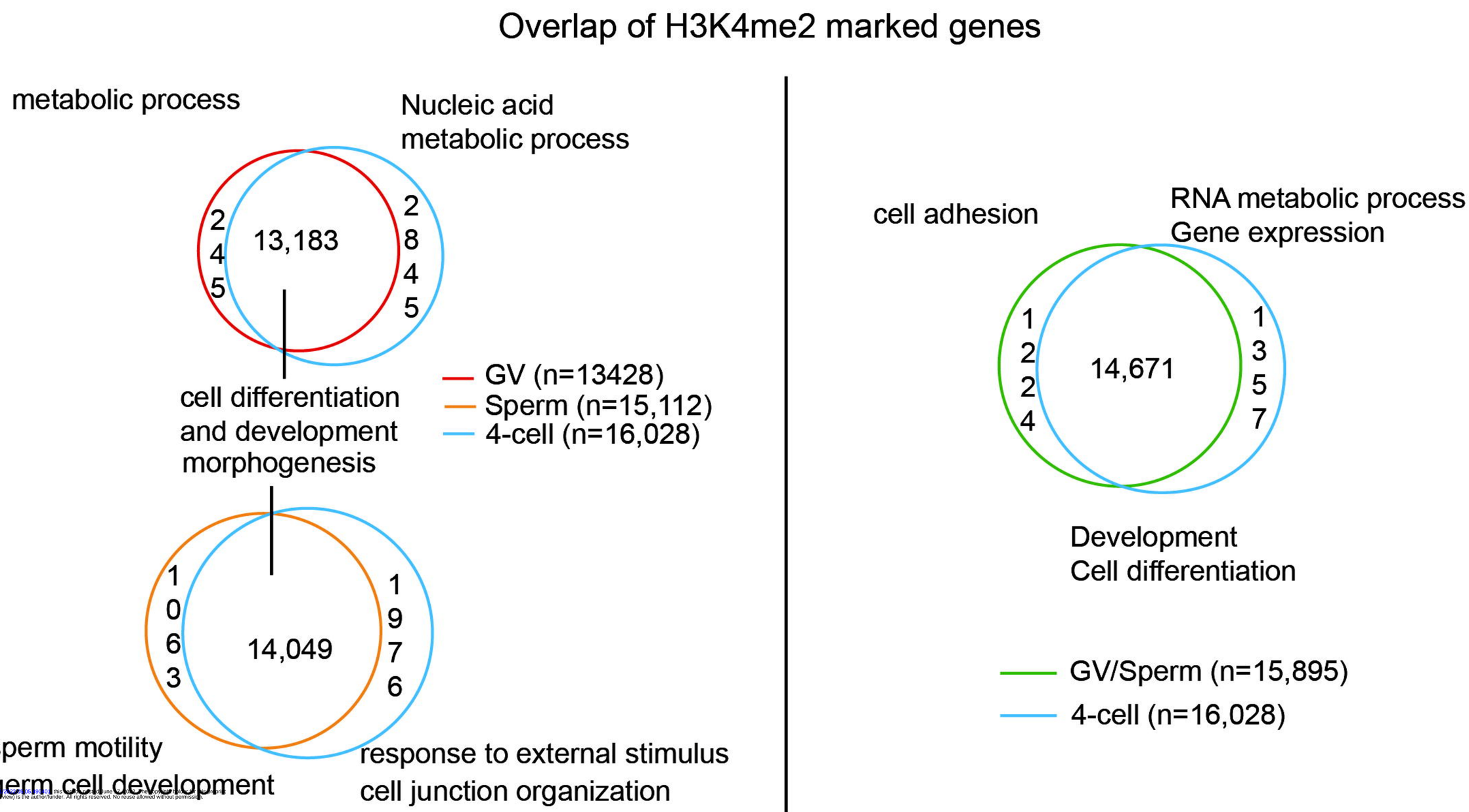
A



B



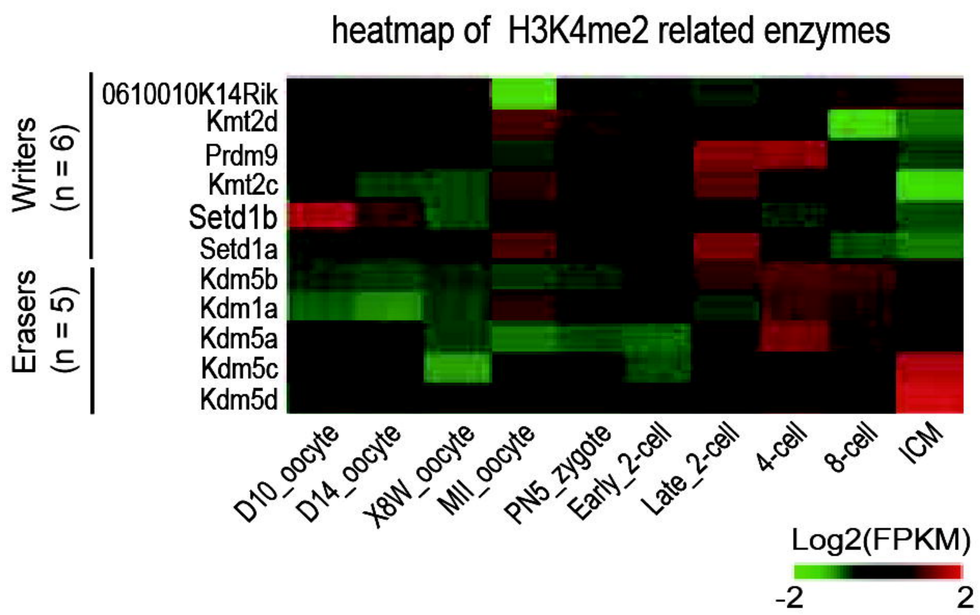
C



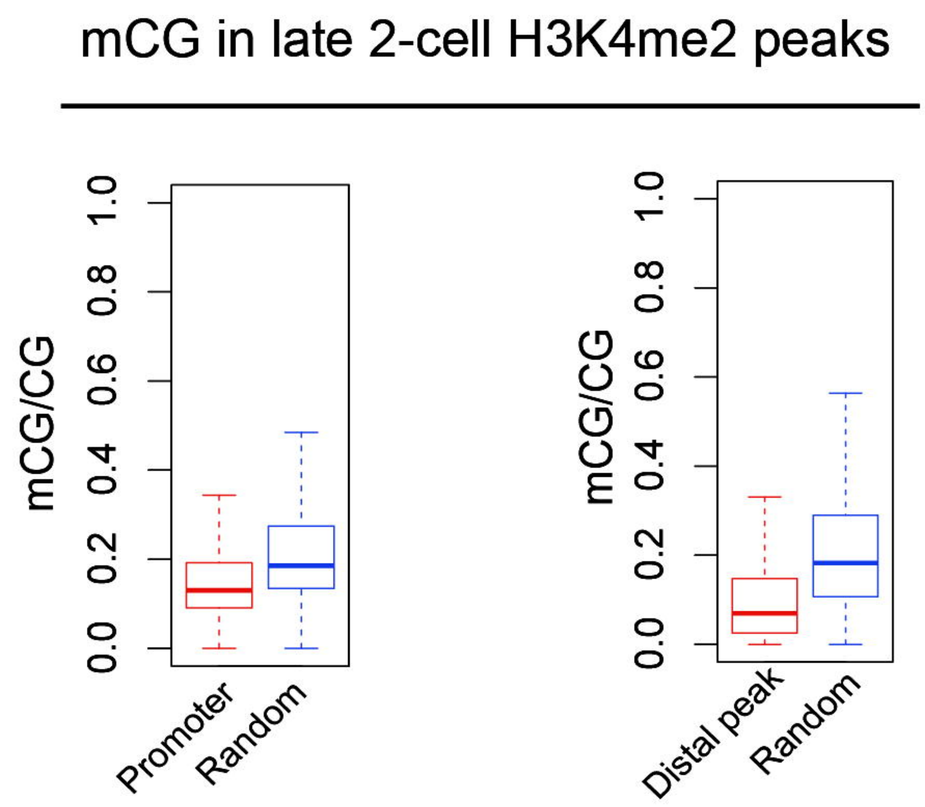
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.19.102011 this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. S3. H3K4me2 in late 2-cell embryos

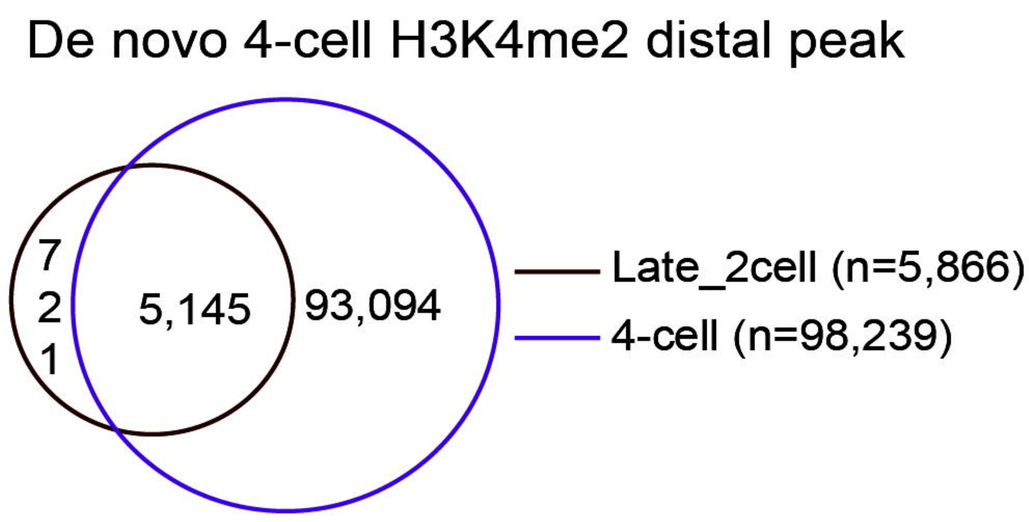
A



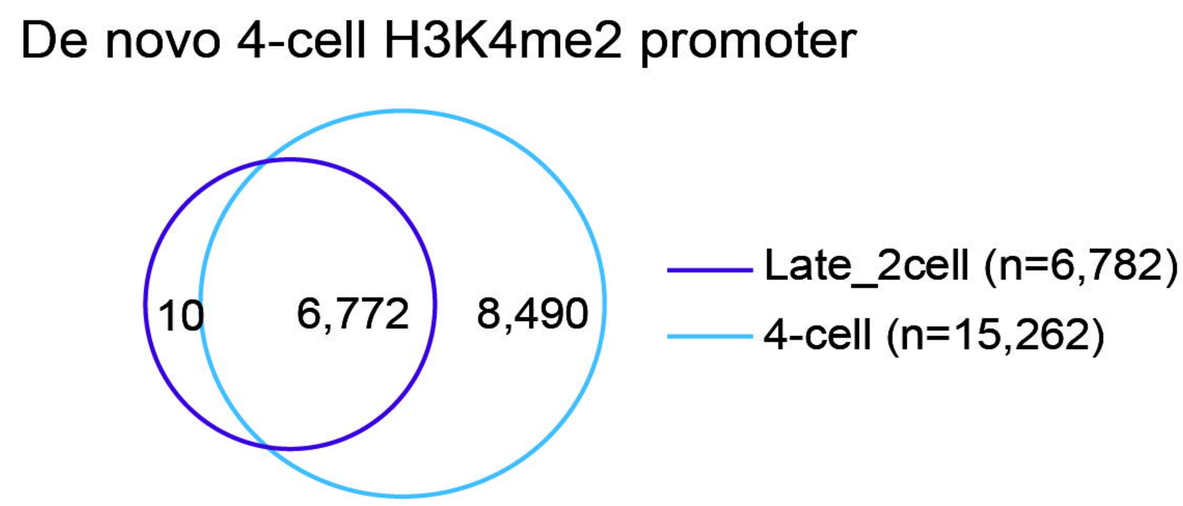
B



C



D



bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 4. Pervasive H3K4me2 in CpG-rich regulatory regions in mouse embryos

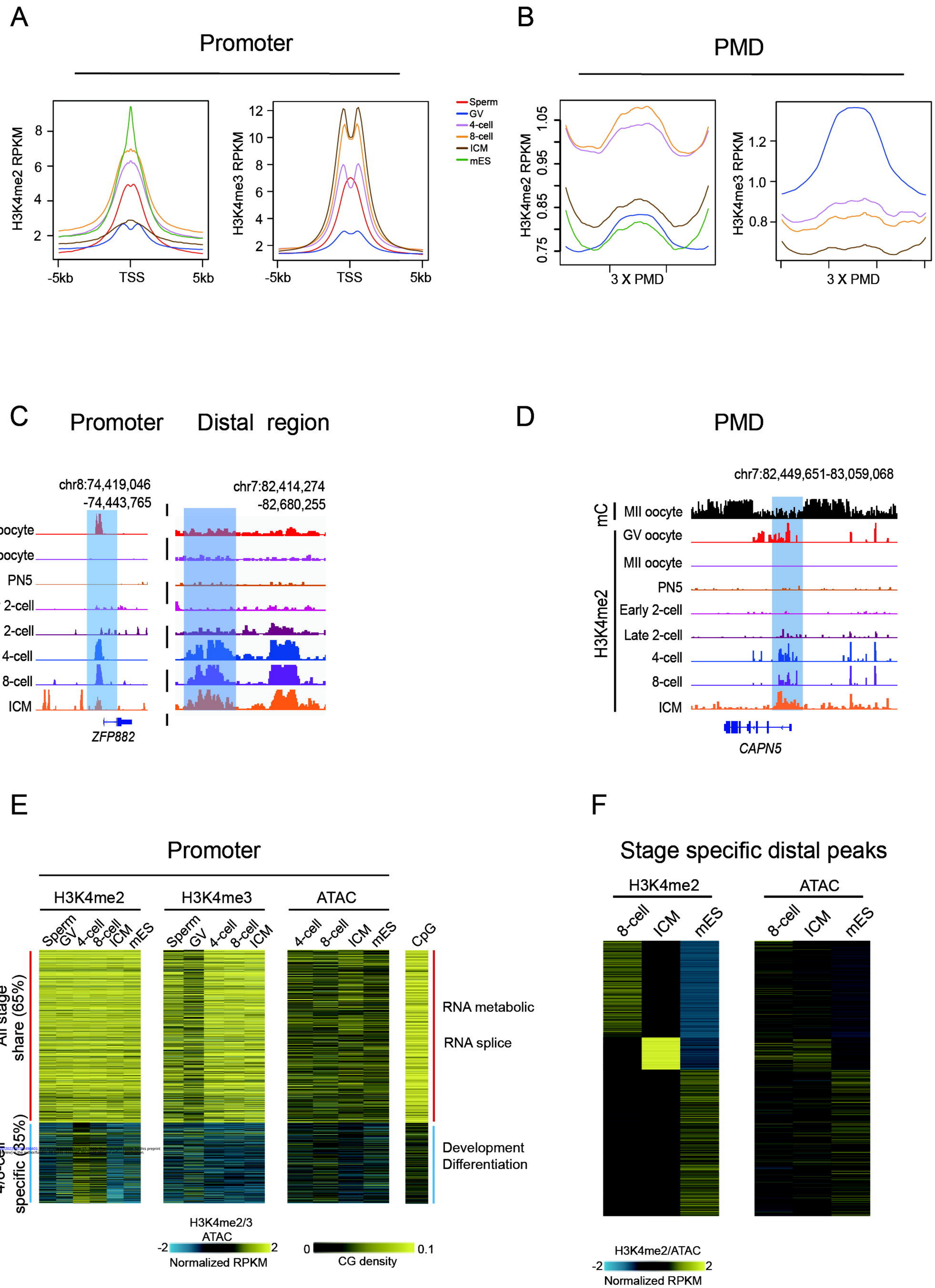
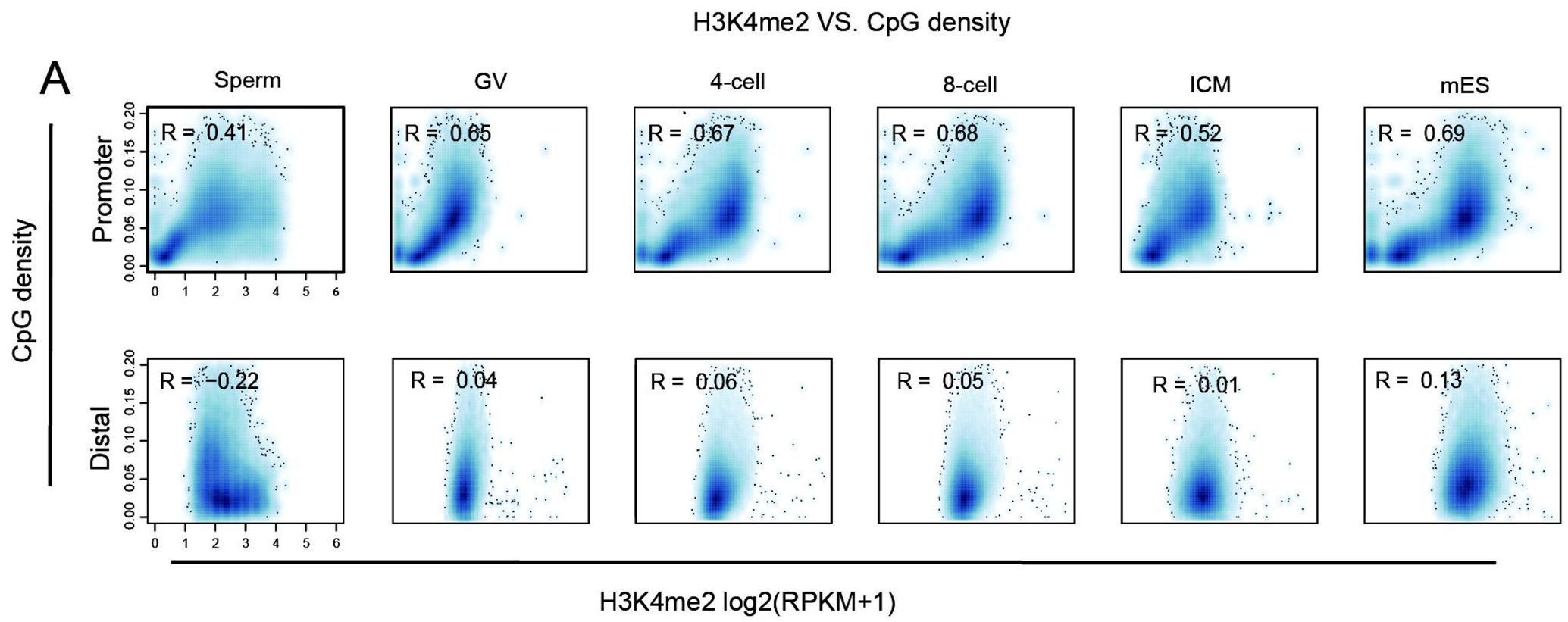
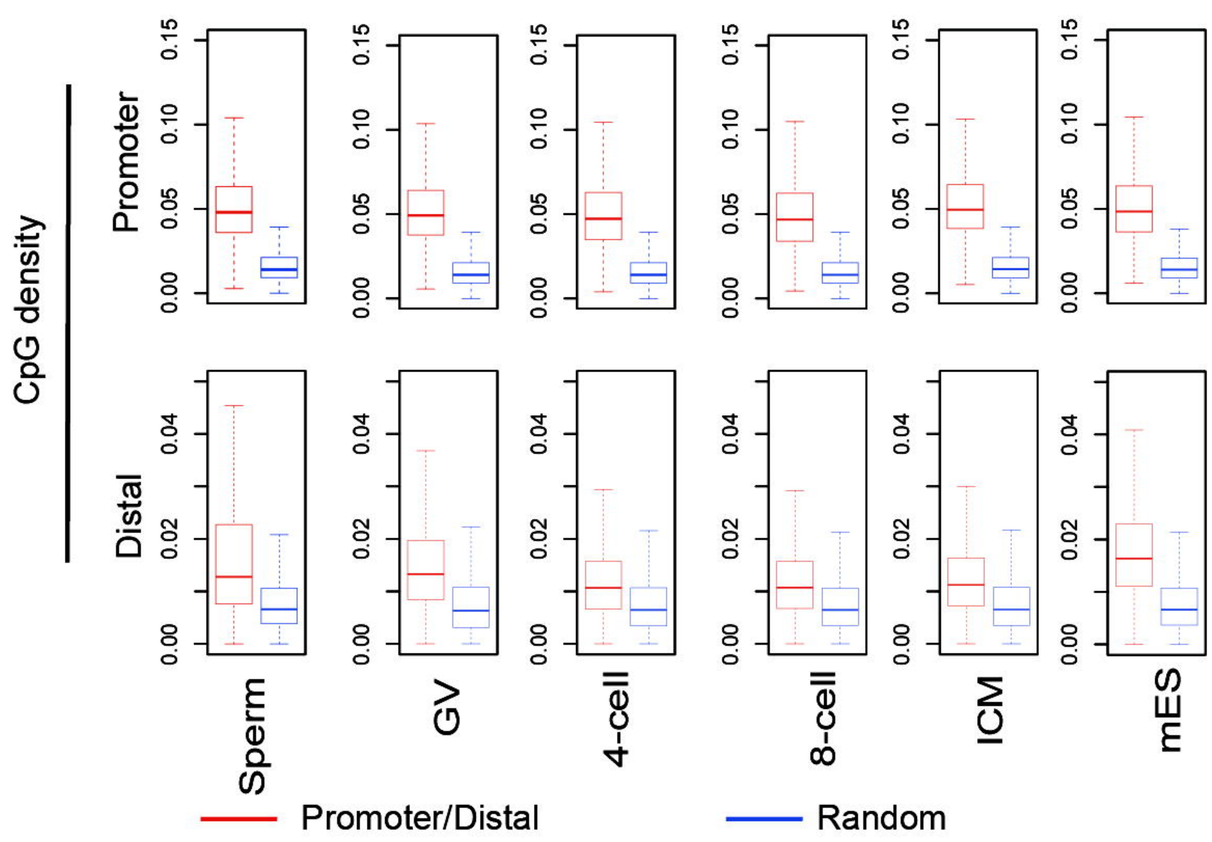
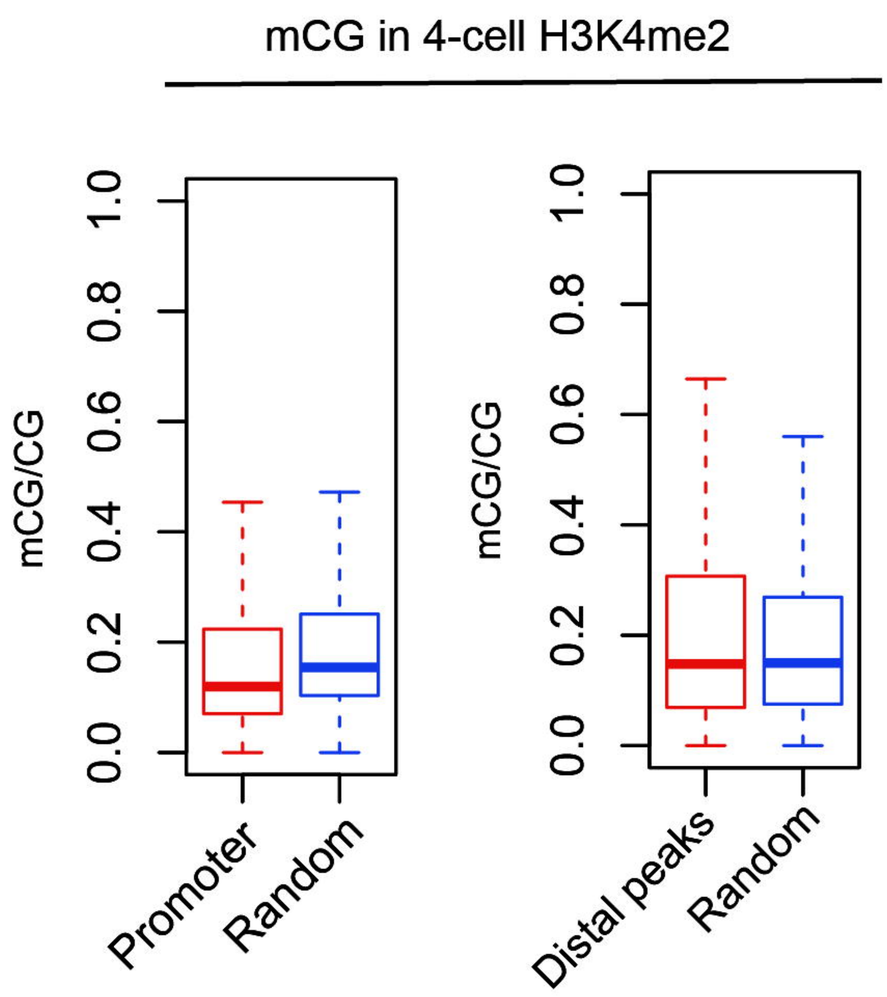
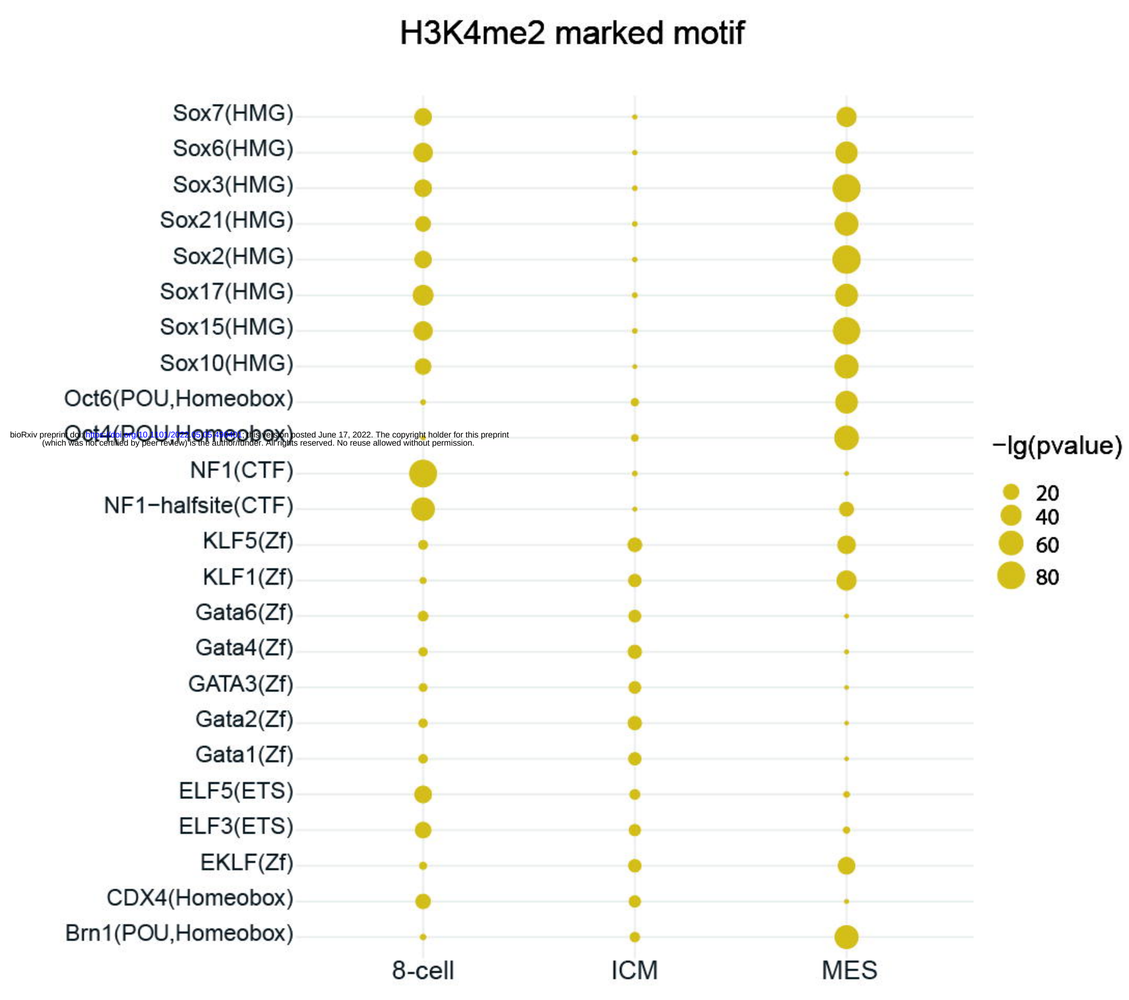


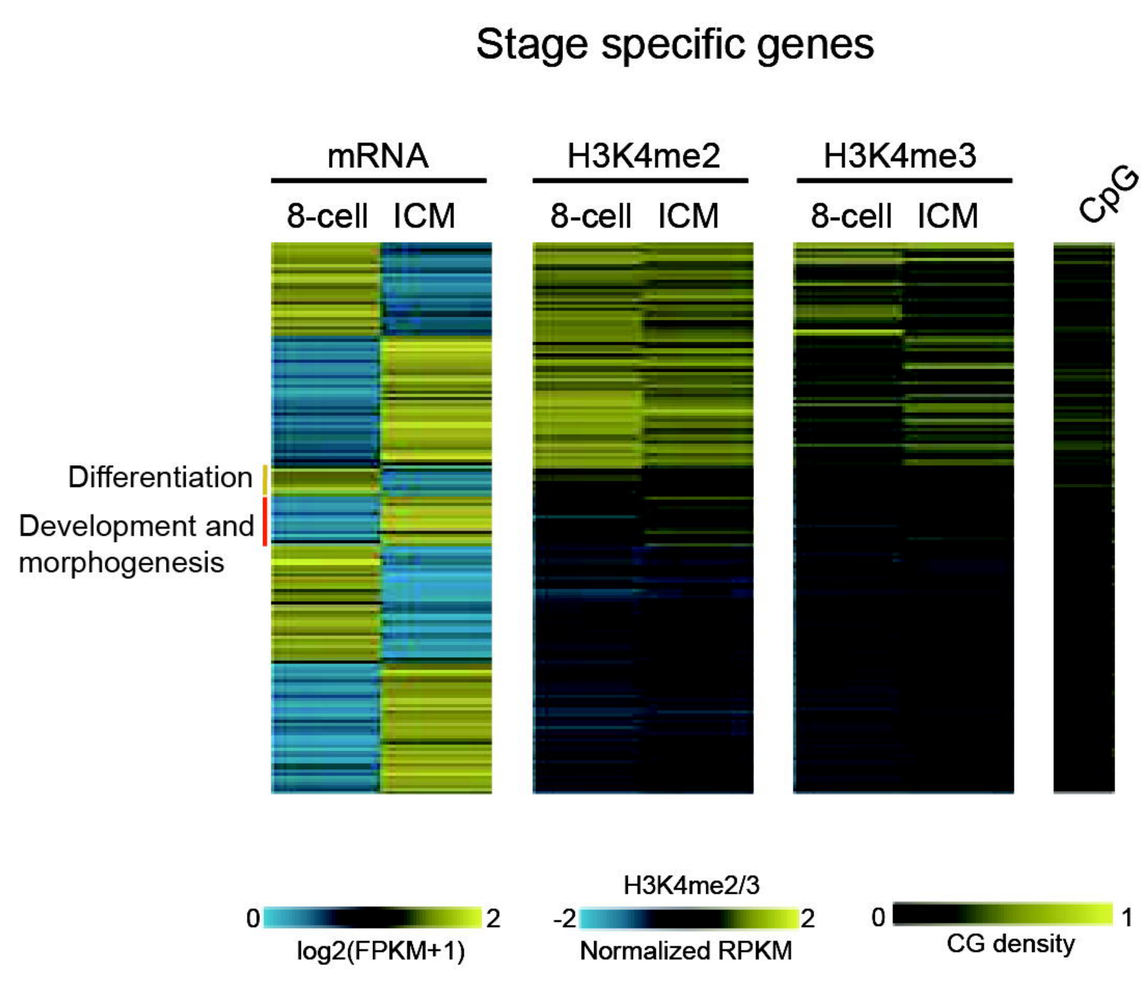
Fig. S4. H3K4me2 in mouse 4/8-cell embryos





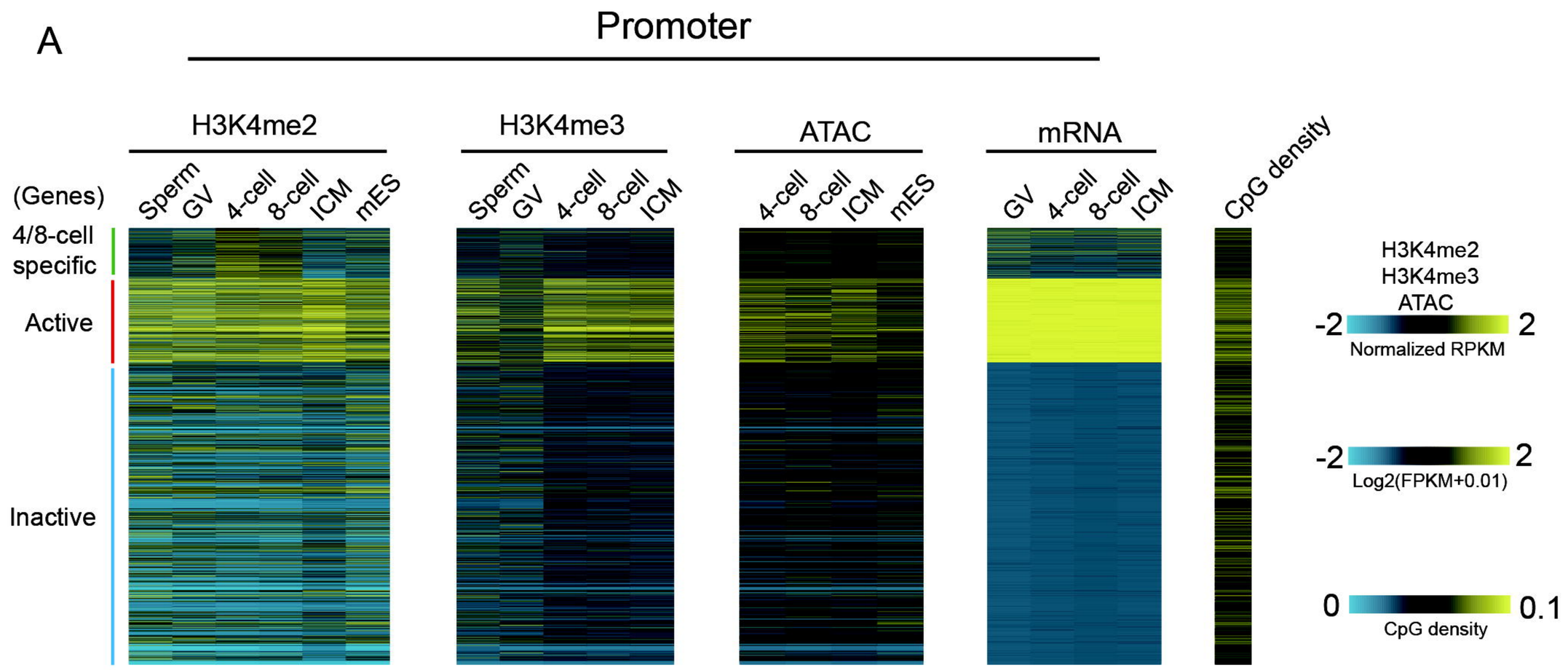


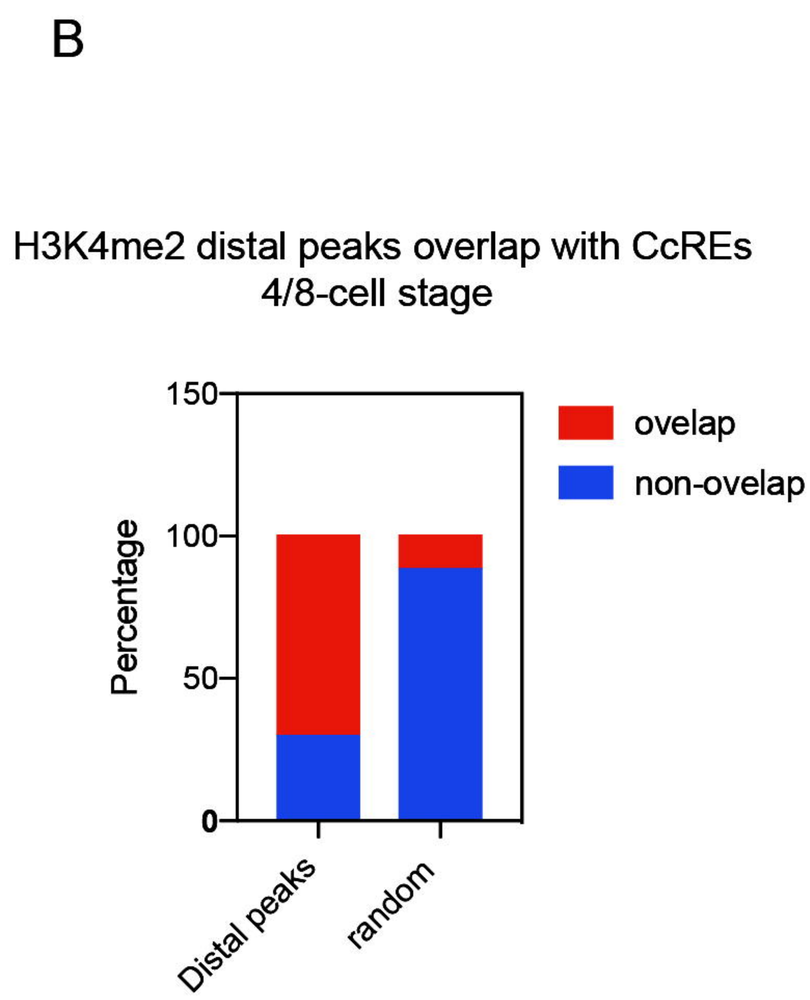


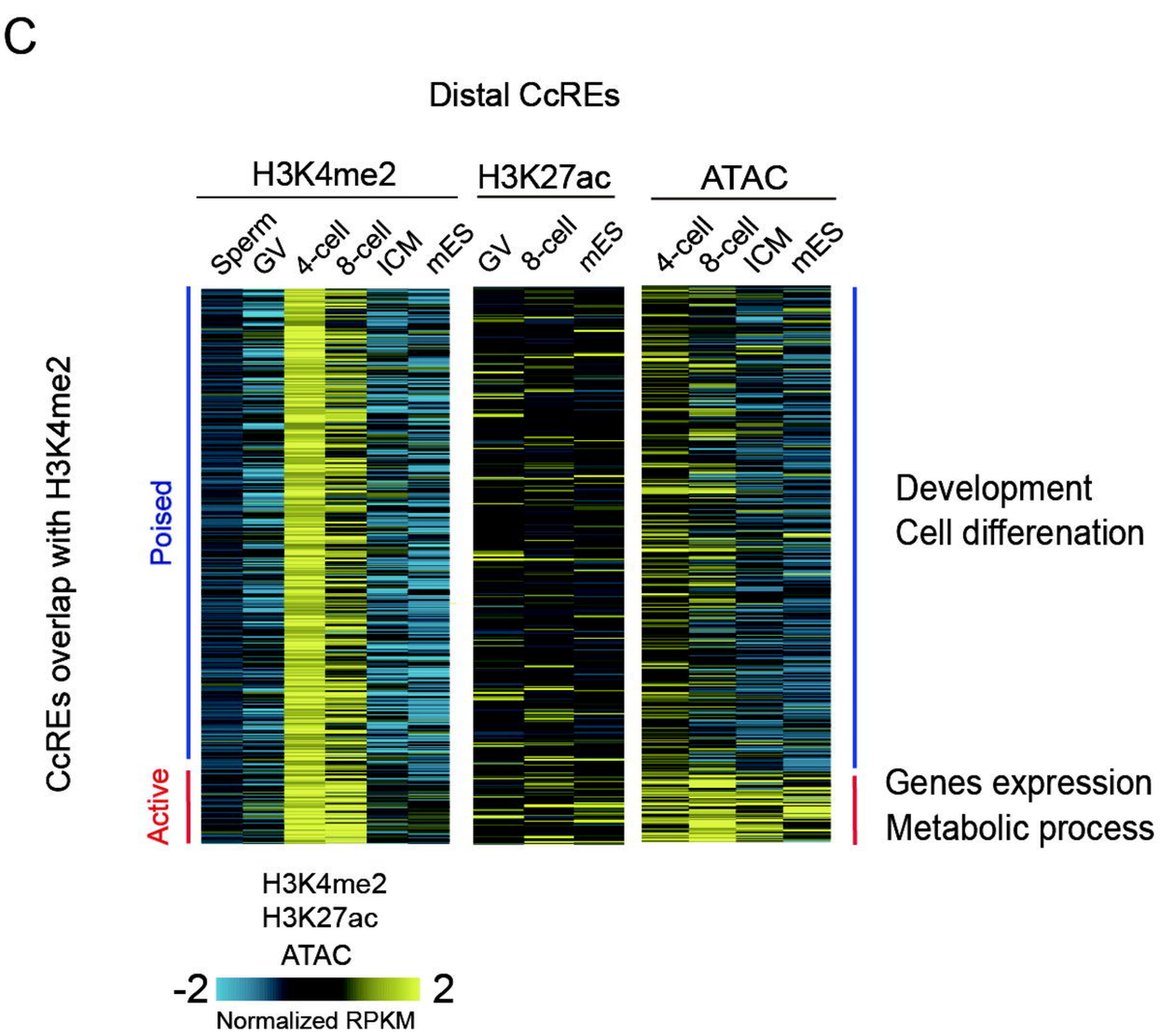


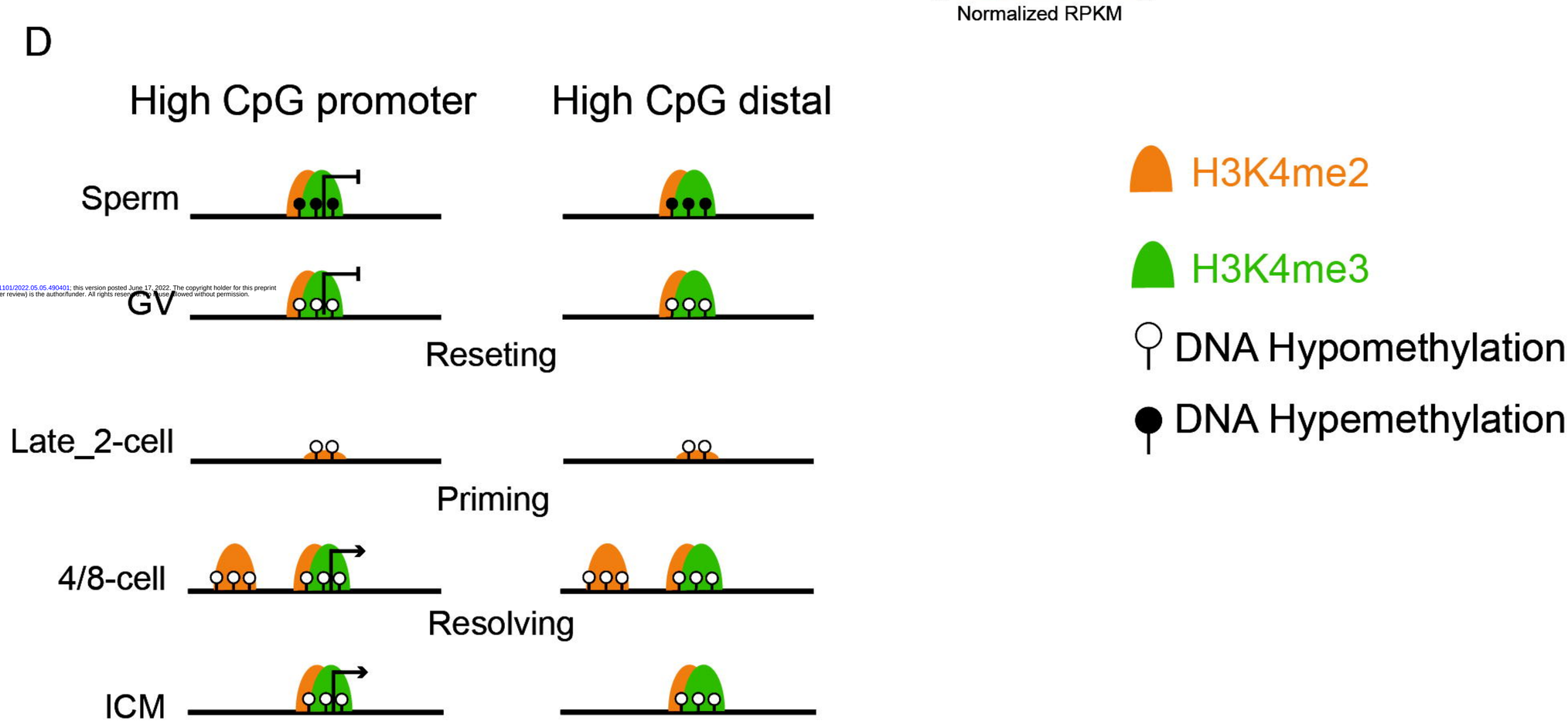
bioRxiv preprint doi: https://doi.org/10.1101/2022.06.16.496456 posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 5. Chromatin states at specific stages and H3K4me2 mode chart during maternal-to-zygotic transition.





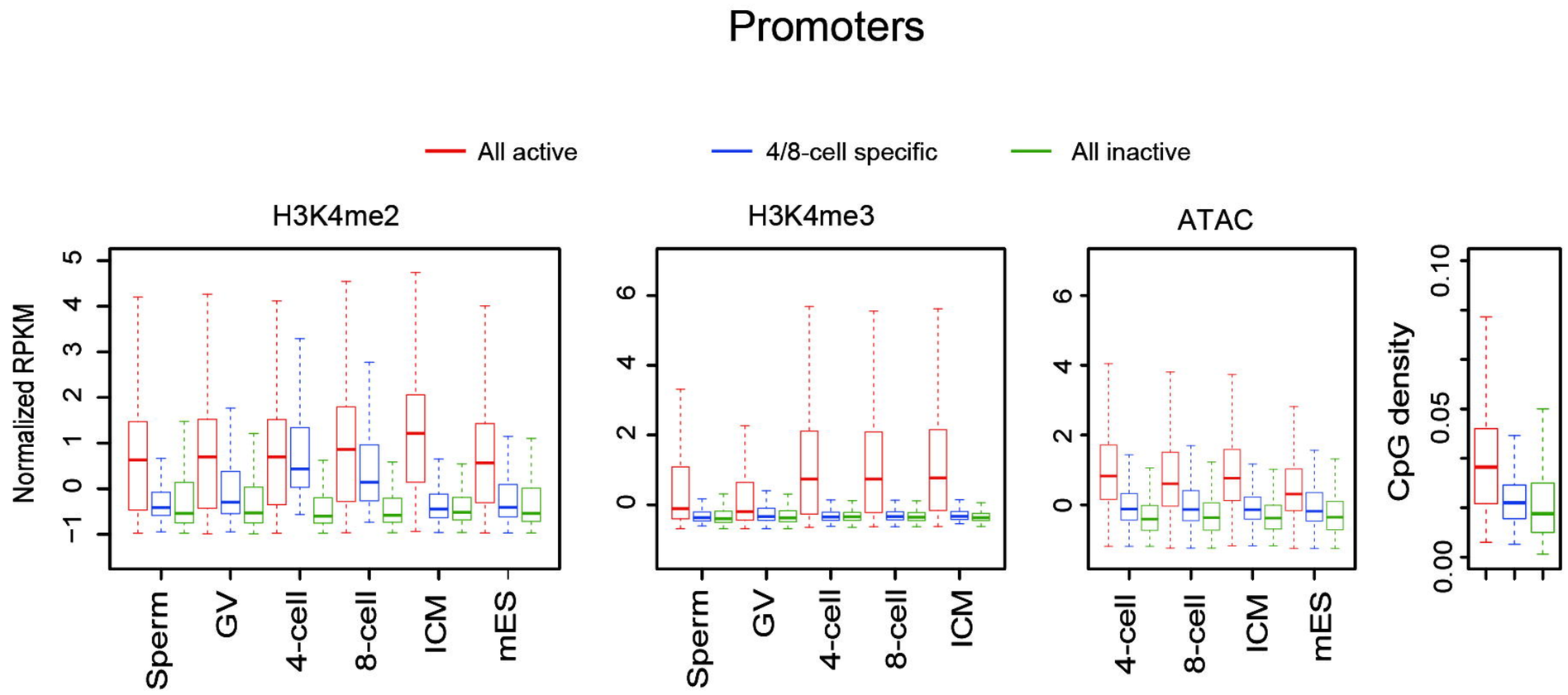




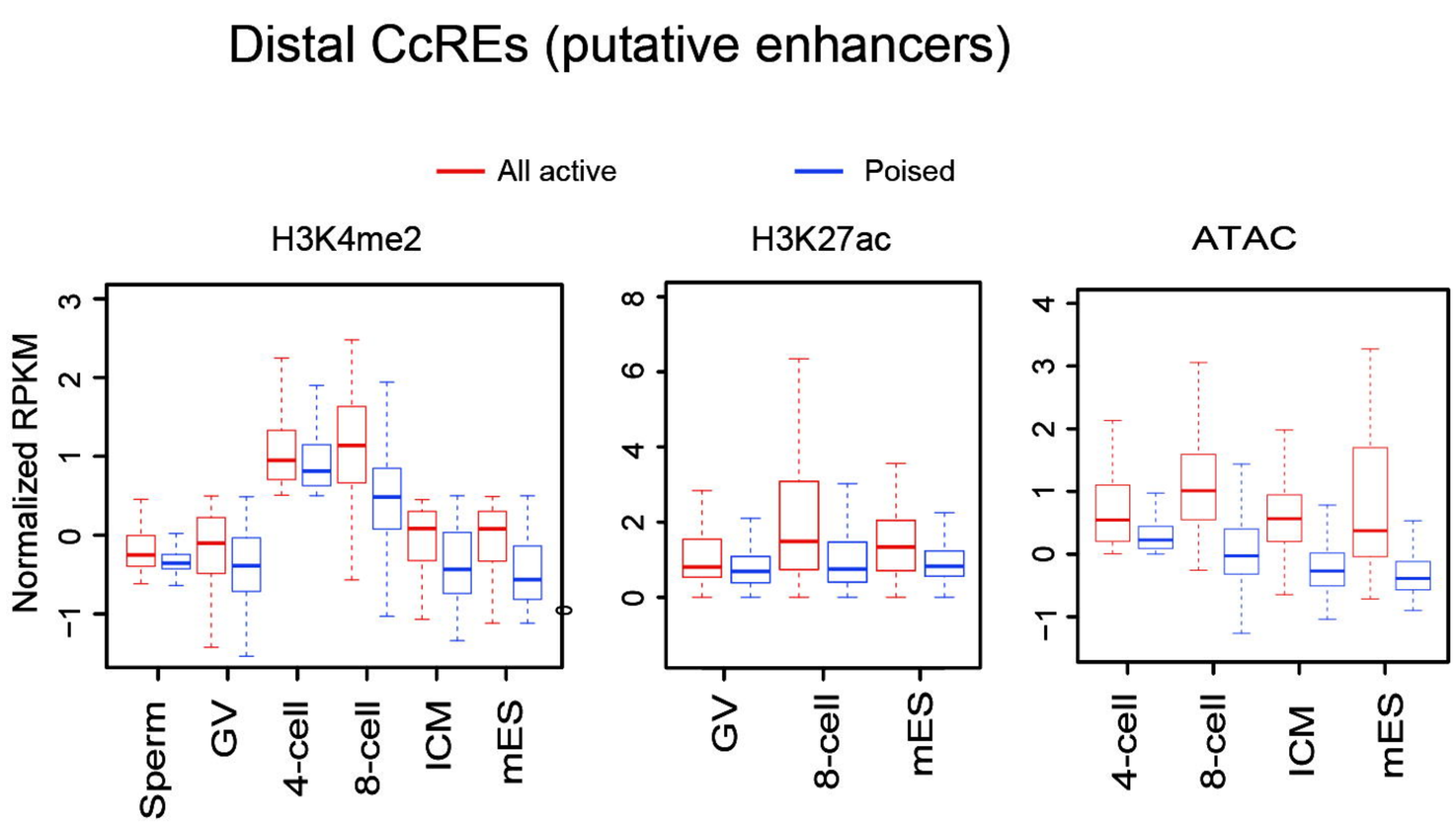
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. Reuse allowed without permission.

Fig. S5. Chromatin states at specific stages and distal CcREs.

A



B



bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490401; this version posted June 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.