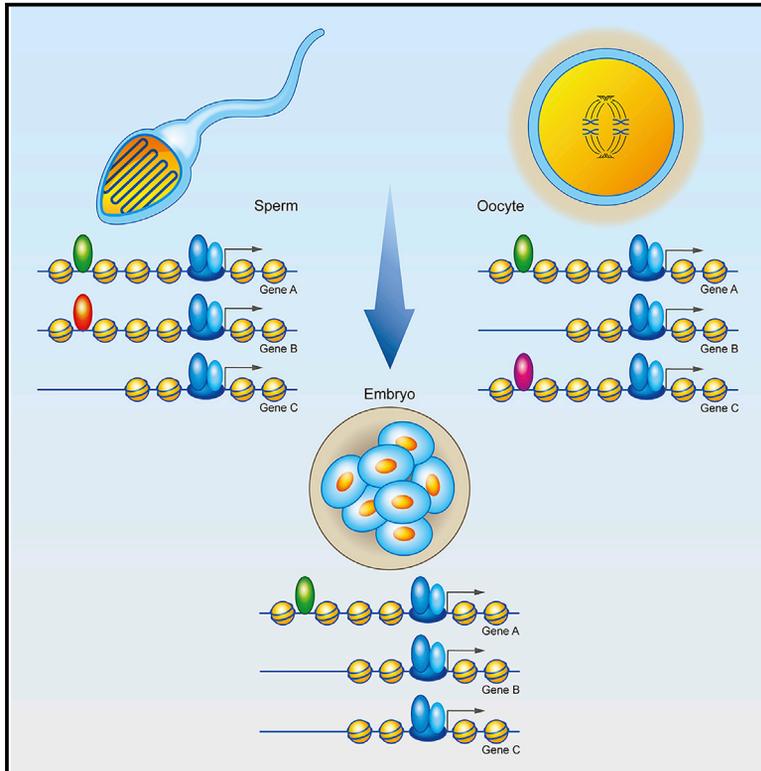# Maintenance of CTCF- and Transcription Factor-Mediated Interactions from the Gametes to the Early Mouse Embryo

Graphical Abstract

Authors

Yoon Hee Jung, Isaac Kremsky, Hannah B. Gold, ..., Brianna J. Bixler, Anthony W.S. Chan, Victor G. Corces

Correspondence

vgcorces@gmail.com

In Brief

Jung et al. show that sperm and oocyte promoters and enhancers contain ATAC-seq accessible sites suggesting the presence of Pol II, Foxa1, ERα, and AR. Accessibility in gametes persists in early embryos. Interactions mediated by CTCF in gametes are inherited from each parent and become the same in 8-cell embryos.

Highlights

- ATAC-seq accessibility at sperm and oocyte promoters is maintained in the embryo

- Sperm enhancers containing transcription factors are conserved in mammals

- Accessible sperm enhancers are also open in oocytes and preimplantation embryos

- Interactions mediated by FoxA1 and CTCF and cohesin persist from gametes to embryos

CellPress

Maintenance of CTCF- and Transcription Factor-Mediated Interactions from the Gametes to the Early Mouse Embryo

Yoon Hee Jung,[1] Isaac Kremsky,[1] Hannah B. Gold,[1] M. Jordan Rowley,[1] Kanchana Punyawai,[2,3] Alyx Buonanotte,[2,3,4] Xiaowen Lyu,[1] Brianna J. Bixler,[1,4] Anthony W.S. Chan,[2,3,4] and Victor G. Corces[1,4,5,*]
[1]Department of Biology, Emory University, 1510 Clifton Rd. NE, Atlanta, GA 30322, USA
[2]Yerkes National Primate Research Center, 954 Gatewood Rd. NE, Atlanta, GA 39329, USA
[3]Department of Human Genetics, Emory University School of Medicine, 615 Michael St., Atlanta, GA 30322, USA
[4]Genetics and Molecular Biology Program, Graduate Division of Biological and Biomedical Sciences, 1462 Clifton Rd., Atlanta, GA 30322, USA
[5]Lead Contact
*Correspondence: vgcorces@gmail.com
https://doi.org/10.1016/j.molcel.2019.04.014

SUMMARY

The epigenetic information present in mammalian gametes and whether it is transmitted to the progeny are relatively unknown. We find that many promoters in mouse sperm are occupied by RNA polymerase II (Pol II) and Mediator. The same promoters are accessible in GV and MII oocytes and preimplantation embryos. Sperm distal ATAC-seq sites containing motifs for various transcription factors are conserved in monkeys and humans. ChIP-seq analyses confirm that Foxa1, ERα, and AR occupy distal enhancers in sperm. Accessible sperm enhancers containing H3.3 and H2A.Z are also accessible in oocytes and preimplantation embryos. Furthermore, their interactions with promoters in the gametes persist during early development. Sperm- or oocyte-specific interactions mediated by CTCF and cohesin are only present in the paternal or maternal chromosomes, respectively, in the zygote and 2-cell stages. These interactions converge in both chromosomes by the 8-cell stage. Thus, mammalian gametes contain complex patterns of 3D interactions that can be transmitted to the zygote after fertilization.

INTRODUCTION

Terminally differentiated gametes are the foundational material from which a complete organism is formed. Emerging evidence suggests that phenotypes induced by environmental cues or exposure to toxicants can be inherited by individuals in subsequent generations (Heard and Martienssen, 2014). Environmental factors may affect the epigenome during gametogenesis, since this process provides an opportune window for epigenetic reprogramming of the germline when chromatin undergoes major alterations of epigenetic information. However,

it remains unclear how acquired phenotypes can be maintained through the reprogramming events taking place in the germline and after fertilization to be transferred from parents to offspring. Studies examining sperm chromatin revealed that sperm retain nucleosomes and post-translationally modified histones in different regions of the genome (Brykczynska et al., 2010; Carone et al., 2014; Hammoud et al., 2009). Analyses of nucleosome retention in mouse sperm using assay for transposase-accessible chromatin sequencing (ATAC-seq) and distribution of covalent histone modifications using native chromatin immunoprecipitation sequencing (ChIP-seq) concluded that approximately 60% of all promoters are in an active epigenetic state, including well-positioned nucleosomes with H3K4me3, H3K27ac, H3K9ac, and H3K36me3 (Jung et al., 2017). Sperm also retain nucleosomes containing H3K4me1 and H3K27ac at specific intergenic or intronic regions, suggesting that the sperm genome may contain epigenetically active enhancers, despite the absence of transcription (Jung et al., 2017). Furthermore, the presence of regions protected from the insertion of Tn5 transposase suggests that some transcription factors may remain bound to sperm DNA. This has been confirmed by mapping the presence of CCCTC-binding factor (CTCF) and Smc1 using ChIP-seq. Results from these analyses suggest that CTCF and cohesin are bound to the sperm genome in many of the same sites where these proteins are present in the genome of embryonic stem cells (ESCs) (Jung et al., 2017). Analysis of Hi-C data in sperm indicates the presence of CTCF and cohesin loops preferentially formed when CTCF sites are present in convergent orientation, and many of these loops are also present in ESCs and somatic cells. Sperm chromatin is folded into A and B compartmental domains corresponding to regions containing active or repressive histone modifications, respectively (Jung et al., 2017). It is possible that this wealth of epigenetic information present in the sperm is a leftover of prior gene expression during sperm maturation without functional significance during embryonic development. However, it is also possible that proteins bound to the sperm genome play an instructive role in the regulation of transcription in the early embryo.
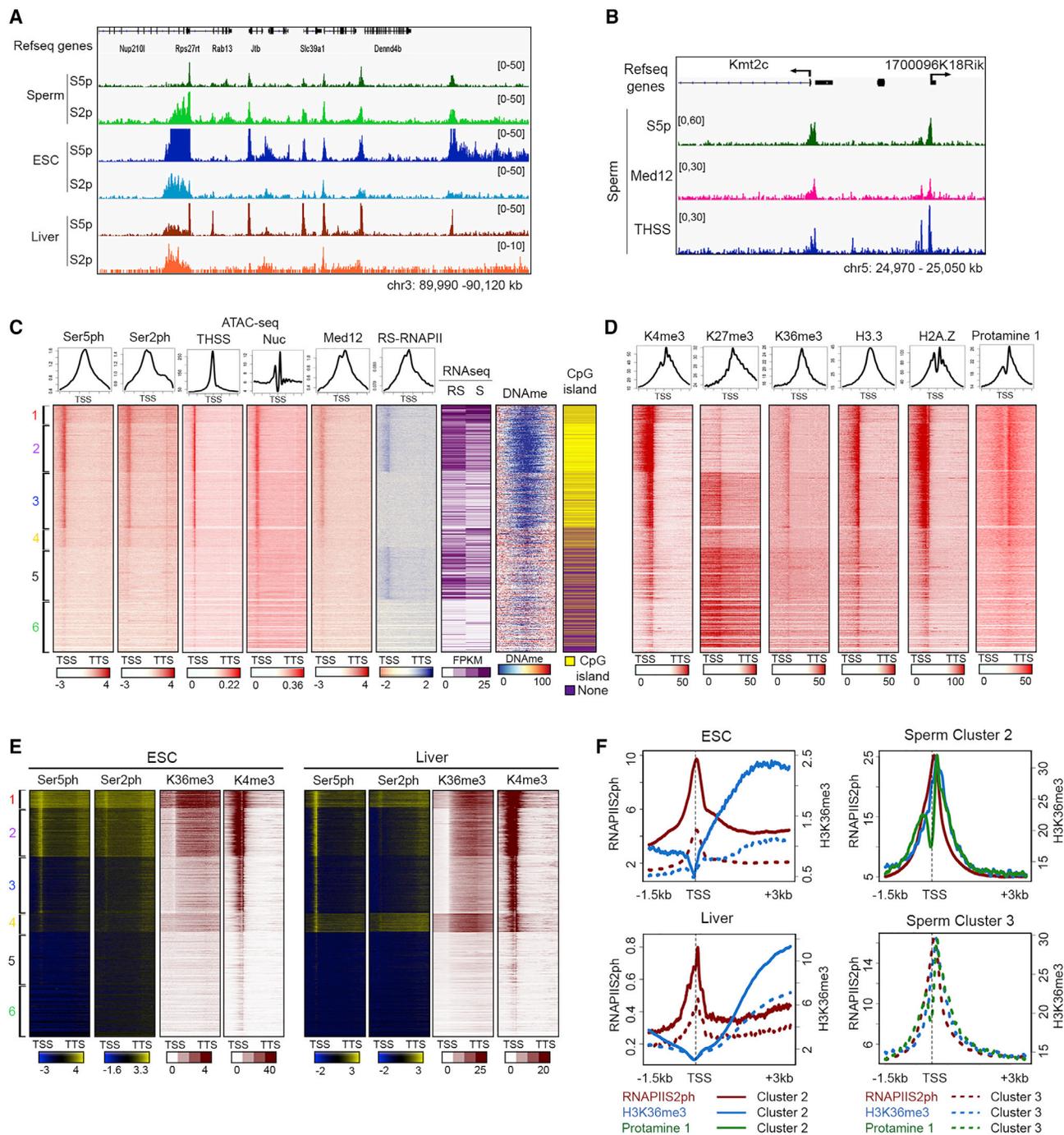
Figure 1. Pol II and Mediator Are Present at Sperm Promoters

(A) Genome browser view showing the distribution of RNAPIISer5ph and RNAPIISer2ph in a specific region of the mouse sperm genome. As a comparison, RNAPIISer5ph and RNAPIISer2ph in ESCs and liver are shown.

(B) Genome browser view showing Med12 distribution with respect to that of RNAPIISer5ph and ATAC-seq THSSs in a specific region of the mouse sperm genome.

(C) Heatmaps and average plots showing RNAPIISer5ph and RNAPIISer2ph enrichment at all mouse Refseq annotated genes. Sites are ordered by k-means clustering of RNAPIISer5ph and RNAPIISer2ph signal between TSSs and TTSs. The enrichment of Pol II in different clusters correlates with the enrichment of the Med12 and ATAC-seq THSS signal. As a comparison, pan-Pol II signal and RNA-seq in round spermatids and mature sperm are shown. RS, round spermatid; S, mature sperm.

(legend continued on next page)

Here, we explore the state of mouse sperm chromatin and find that many sperm promoters contain components of the transcription complex, including RNA polymerase II (Pol II) phosphorylated in Ser2 and Ser5, and the Mediator complex. Footprint analyses using ATAC-seq data suggest the presence of specific transcription factors, and these factors are conserved in syntenic regions of the macaque and human sperm genomes. In the female germline, based on the presence of ATAC-seq sites containing the CTCF motif, CTCF sites appear to be present in the germinal vesicle (GV) oocyte, where they form loops, and remain in mature MII oocytes without contributing to the 3D organization of metaphase chromosomes at this stage. CTCF-dependent organization is passed on to the zygote in a parent-of-origin-dependent manner until the 8-cell stage, when CTCF loops become the same in both homologous chromosomes. Furthermore, enhancer-promoter interactions mediated by transcription factors in the gametes also persist in the early embryo. These findings suggest that the contribution of information contained in sperm and oocyte chromatin to early embryonic development is greater than was previously recognized, suggesting that acquired phenotypes in the parents may be transmitted through the germline in the form of DNA-bound proteins and their interactions.

RESULTS

Components of the Transcription Complex Are Present in Sperm Promoters

We recently reported that approximately 60% of transcription start sites (TSSs) of annotated genes in sperm contain a nucleosome-free region flanked by well-positioned nucleosomes marked by active histone modifications (Jung et al., 2017). The presence of a hypersensitive region between the −1 and +1 nucleosomes raises the question of whether TSSs in sperm are occupied by the transcription machinery. To explore this issue, we investigated the presence of Pol II in sperm by performing western analyses using antibodies to Pol II phosphorylated in Ser5 (RNAPIISer5ph) and Pol II phosphorylated in Ser2 (RNAPIISer2ph). Sperm was isolated from the cauda epididymis with a purity of <1 somatic cells per 1,000 sperm. Western analyses of histone H3 and protamine 1 in sperm and somatic cells are in agreement with a pure population of sperm cells (Figure S1A). Results from western analyses show that both forms of RNA polymerase are present in sperm (Figure S1B). Furthermore, results from immunofluorescence confocal microscopy also indicate that RNAPIISer5ph and RNAPIISer2ph are present in the sperm nucleus and that the observed staining can be competed with peptides used to prepare the corresponding antibodies (Figure S1C). Staining with a non-specific immunoglobulin G (IgG) or with secondary antibodies failed to detect any signal (Figure S1D). Finally, immunofluorescence confocal microscopy of testis cryosections show the presence of RNAPIISer5ph and RNAPIISer2ph in mature sperm present in

the seminiferous tubules (Figure S1E). We therefore performed ChIP-seq in sperm with antibodies to these 2 phosphorylated forms of Pol II and identified 12,740 and 13,547 peaks of RNAPIISer5ph and RNAPIISer2ph, respectively (Table S1). Analysis of their genome-wide distribution with respect to gene features indicates that a large number of RNAPIISer5ph and RNAPIISer2ph sites are located at TSSs and that their distribution is similar to that in ESCs or somatic cells of the liver (Figure 1A). Both forms of Pol II are also located at distal intergenic regions (Figures S1F and S1G), where they overlap with H3K4me1 and H3K27ac, suggesting that they may represent active enhancers (Figure S1H). To further explore the significance of the distribution of Pol II in the mouse sperm genome, we performed ATAC-seq using the Omni-ATAC protocol, which allows a higher signal-to-noise ratio when generating chromatin accessibility profiles (Corces et al., 2017). We obtained 2 biological replicates (Table S1) and separated ATAC-seq reads into the 50- to 115-bp range, corresponding to sites where bound proteins protect the DNA from the Tn5 transposase (Tn5 hypersensitive sites [THSSs]), and the 180- to 247-bp range corresponding to mono-nucleosomes (Schep et al., 2015). We identified a total of 61,238 THSSs. We then used TSSs and transcription termination sites (TTSs) of all mouse annotated genes as anchors and carried out k-means clustering using the reads of RNAPIISer5ph and RNAPIISer2ph ChIP-seq (Figure 1C). Results show that approximately 14,756 TSSs are enriched with RNAPIISer5ph and RNAPIISer2ph (clusters 1–4), and the levels of Pol II at these TSSs correlate with the accessibility to Tn5 transposase and the nucleosome signal obtained by ATAC-seq (Figure 1C). TSSs occupied by RNAPIISer5ph and RNAPIISer2ph are hypo-methylated on DNA and are enriched in CpG islands. Using the same TSS order as in Figure 1C, we found that flanking nucleosomes carry the active histone modifications H3K4me3 and H3K36me3 (Figure 1D). We also performed ChIP-seq with the histone variants H3.3 and H2A.Z (Table S1) and found that these nucleosomes also contain H3.3 and H2A.Z, while the +1 nucleosome is enriched in protamine 1 (Figure 1D). TSSs in cluster 3 are also marked by H3K27me3, suggesting a bivalent state (Figure 1D).

Mediator-promoter interactions stabilize the formation of transcription complexes, helping to position the Pol II complex at the TSS. To test whether Mediator is also present in sperm, we performed western and immunofluorescence confocal microscopy analyses using antibodies to the Med12 subunit and confirmed the presence of Mediator in sperm nuclei (Figures S1B and S1C). We then performed ChIP-seq with Med12 antibodies (Table S1). Results show that Med12 co-localizes with RNAPIISer5ph and RNAPIISer2ph at specific gene promoters (Figure 1B), and it is present at the same TSSs in the sperm genome (Figure 1C). These results suggest that, in addition to nucleosomes, sperm also retain components of the transcription complex at approximately 60% of all annotated promoters, despite the lack of transcription.

(D) Heatmaps and average plots showing sperm histone modifications H3K4me3, H3K27me3 and H3K36me3, histone variants H3.3 and H2A.Z, and protamine 1. Clusters and gene order are the same as in (C).

(E) Enrichment of RNAPIISer5ph, RNAPIISer2p, H3K36me3, and H3K4me3 in ESCs and liver at gene promoters clustered in the same order as in (C).

(F) Average profiles of RNAPIISer2ph, H3K36me3, and protamine 1 for clusters 2 and 3 shown in (E).

See also Figure S1.

To test whether the presence of Pol II in mature sperm correlates with transcriptional activity at earlier stages during spermatogenesis, we compared mature sperm Pol II ChIP-seq data with published Pol II ChIP-seq and RNA-seq in round spermatids. Results show that the presence of Pol II in sperm does not correlate well with Pol II or RNA levels in round spermatids. For example, TSSs in clusters 1 and 2 also contain Pol II and are transcribed in round spermatids, but TSSs in clusters 3 and 4 lack Pol II in round spermatids, and TSSs in cluster 5 are transcribed and contain Pol II in these cells but not in sperm (Figure 1C). These results suggest that some or all of the accessible promoters enriched with Pol II in sperm may prepare genes for expression at later stages of development after fertilization. Most RNAPIISer2ph and H3K36me3 in sperm appear to be present at TSSs but not extending through gene bodies (Figures 1C and 1D). This is distinct from the canonical patterns of RNAPIISer2ph and H3K36me3 presence throughout coding regions, including the 3′ end of transcribed genes, and correlating with active elongation. To explore differences in the location of Pol II and H3K36me3 between sperm and cell types undergoing active transcription, we compared their distribution in sperm with that in ESCs and liver cells using published data from these 2 cell types (Bunch et al., 2014; Ji et al., 2015; Li et al., 2014; Wamstad et al., 2012). In ESCs, RNAPIISer5ph and RNAPIISer2ph are present at TSSs in clusters 1–2, but extend into the gene bodies in ESCs and liver cells, coincident with the enrichment of H3K36me3 through the genes (Figures 1E and 1F). Therefore, although sperm promoters are poised in an active chromatin state with Pol II, Mediator, and active histone modifications, these promoters are not transcribed. One possible explanation for the lack of transcription in sperm despite the apparent similarity of the chromatin state to that of transcriptionally active cell types is the presence of protamines on sperm DNA, specifically enriched just downstream of the TSS and Pol II (Figures 1D and 1F). This raises the possibility that protamines may be involved in preventing poised Pol II in sperm from elongating into the gene body.

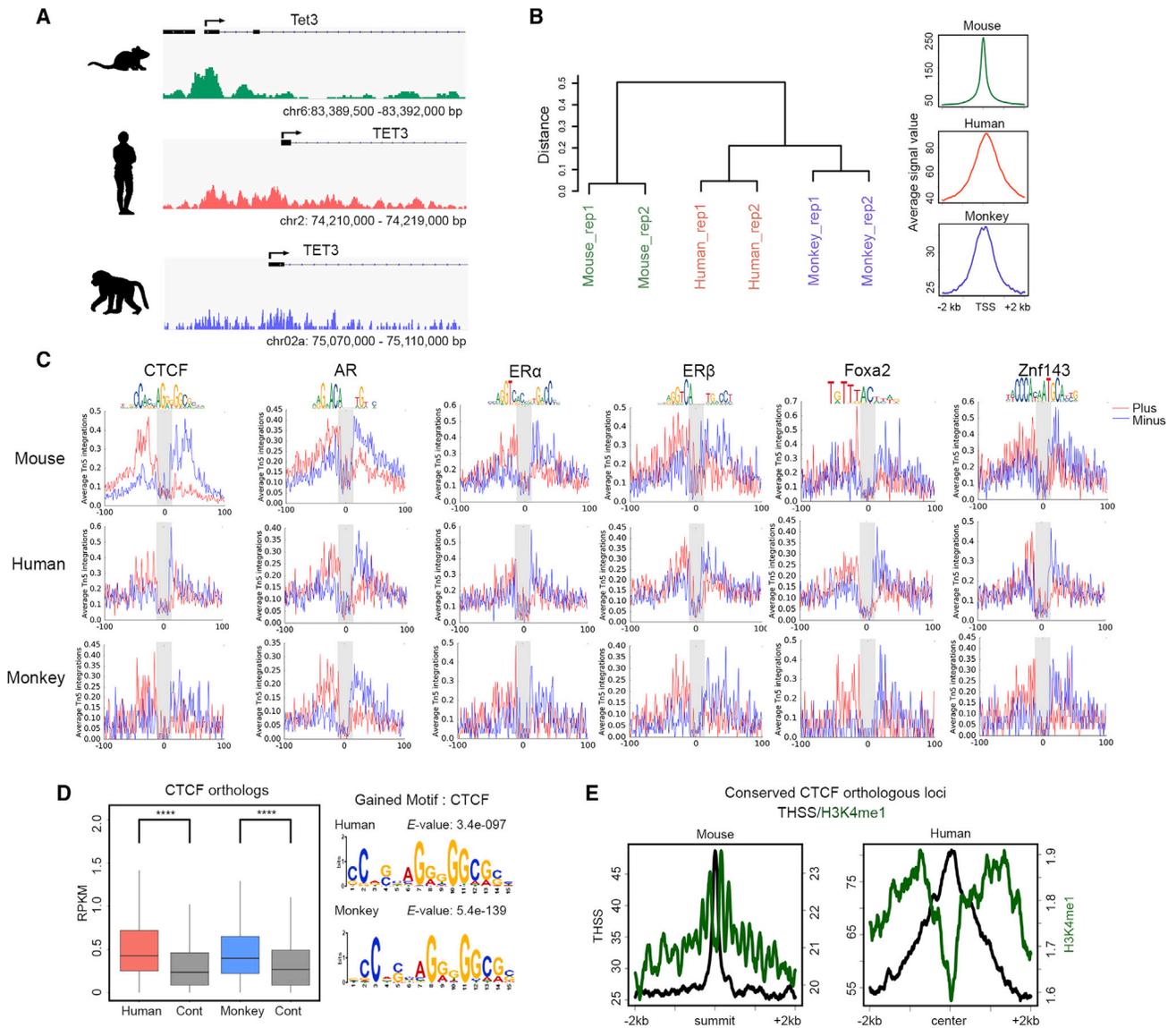## Conservation of the Sperm Chromatin Regulatory Landscape in Monkeys and Humans

Analysis of accessible regions detected by subnucleosome-sized reads obtained by Omni-ATAC-seq indicates the occurrence of thousands of THSSs at non-TSS sequences, suggesting the presence of DNA binding proteins at promoter-distal regions, including introns and intergenic regions. We identified 50,082 distal THSSs corresponding to putative transcription factor binding sites excluding TSSs in sperm, and searched these sites for DNA footprints of known transcription factor binding motifs using Wellington (Piper et al., 2013). We found significant enrichment of DNA footprints for numerous transcription factors (TFs), including members of the Forkhead family such as Foxa1, Zn finger proteins such as CTCF and Znf143, and nuclear hormone receptors including androgen receptor (AR) and estrogen receptor (ER) (Figure S2A). To test whether this accessibility landscape may have functional significance, we explored whether it is conserved in the sperm of other mammalian species. We performed Omni-ATAC-seq with sperm samples from humans and rhesus macaques (*Macaca mulatta*). We obtained

2 independent biological replicates for each species with high correlation coefficients (Table S1). We combined replicates and analyzed reads separately in the 50- to 115-bp and 180- to 247-bp ranges, corresponding to the presence of TFs and mono-nucleosomes, respectively. We identified 89,713 THSSs in humans and 73,384 THSSs in monkeys. We then identified conserved THSSs around TSSs corresponding to the presence of the transcription complex between the +1 and −1 nucleosomes, as described above. We considered only genes having syntenic locations in all 3 species. Tet3 is an example of a conserved gene with a THSS at the promoter in all 3 species analyzed (Figure 2A). Gene Ontology (GO) term analysis of genes whose TSS accessibility is maintained among mice, monkeys, and humans is enriched in terms for embryonic development. The hierarchical clustering of homologous genes by accessibility indicates that humans show a greater similarity to monkeys than to mice (Figure 2B), which is consistent with the evolutionary distances between the 3 species. Furthermore, analysis of footprints identified using ATAC-seq reads corresponding to distal THSSs indicates that human and monkey sperm also contain footprints for numerous TFs, including CTCF and Znf143, members of the Fox family, and androgen and estrogen receptors (Figure 2C).

To gain further insight into the significance of the conservation of TF binding sites across mammalian species, we examined mouse CTCF ChIP-seq peaks and determined whether human and monkey sperm have THSSs at CTCF syntenic sites. We defined syntenic locations of mouse CTCF in human and monkey using LiftOver (mm9 to hg19 and hg19 to rheMac8). Notably, CTCF syntenic locations in both species showed significant chromatin accessibility compared to random control sequences, and consensus CTCF motifs were also found at these conserved sequences (Figure 2D). Furthermore, CTCF syntenic sites are flanked by nucleosomes containing H3K4me1 in mouse and human sperm (Figure 2E). Similar analyses show the conservation of Foxa1 and Znf143 sites (Figure S2B). Finally, distal THSSs present in human sperm overlap with previously defined tissue-specific enhancers (Figure S2C) (Gao et al., 2016). These observations suggest that putative protein binding sites in mammalian sperm chromatin are evolutionarily conserved among mouse, monkey, and human, suggesting that TFs may remain bound to the sperm genome and may have functional relevance during embryonic development.

The Pioneer Factor Foxa1 Co-localizes with Nuclear Receptors at Putative Sperm Enhancers

Analysis of footprints generated by ATAC-seq subnucleosome-sized reads suggests the presence of TFs bound to the sperm genome, including the pioneer factor Foxa1 and the estrogen and androgen receptors, both of which require Foxa1 for recruitment to chromatin (Hurtado et al., 2011). To directly test the possibility that Foxa1 is bound to the sperm genome, we first confirmed its presence in sperm nuclei using western analysis (Figure S3A). Immunofluorescence confocal microscopy also shows the presence of this protein in sperm nuclei, and the observed staining can be competed with peptides used to prepare the corresponding antibodies (Figure S3B). Furthermore, immunofluorescence confocal microscopy of testis cryosections show the presence of Foxa1 in mature sperm present in the

Figure 2. Conservation of the Sperm Chromatin Regulatory Landscape in Mammals

(A) Genome browser view showing ATAC-seq THSS signal between species at a syntenic region of the genome.

(B) Hierarchical clustering of genes in syntenic regions of the genome by ATAC-seq accessibility between mouse, monkey, and human.

(C) Average footprint profile of ATAC-seq THSSs at regions containing the given TF-binding motif.

(D) Boxplots showing the RPKM of ATAC-seq THSSs present at distal accessible regions in mouse sperm containing the CTCF motif in syntenic regions of humans and monkeys. The motifs found at these syntenic regions in monkeys and humans are also shown.

(E) Average profiles of ChIP-seq signal for H3K4me1 mouse and human sperm relative to conserved CTCF syntenic loci.

See also Figure S2.

seminiferous tubules (Figure S3C). We then carried out ChIP-seq with antibodies to Foxa1 (Table S1) and identified 7,116 peaks in the mouse sperm genome (Figure 3A). Some of these sites are present at promoters, but most are located in intergenic regions or introns, likely corresponding to putative enhancers (Figure 3B). Analysis of motifs present at Foxa1 binding sites indicates the presence of binding sequences for estrogen, androgen, and glucocorticoid receptors; grainyhead-like TF1; nuclear factor I X (also known as CCAAT-binding TF); and TF activator protein

(AP)-2 gamma (Figure S3D). Since Foxa1 has been shown to facilitate the binding of nuclear hormone receptors such as ERα and AR, we confirmed that ERα and AR are present in sperm using western analysis (Figure S3A) and immunofluorescence confocal microscopy (Figures S3B and S3C). We then examined the genomic distribution of ERα and AR using ChIP-seq (Figure 3A; Table S1). Like Foxa1, ERα and AR sites are located at promoters, introns, and distal intergenic regions (Figure 3B). The 3 proteins Foxa1, ERα, and AR co-localize at 52% of all
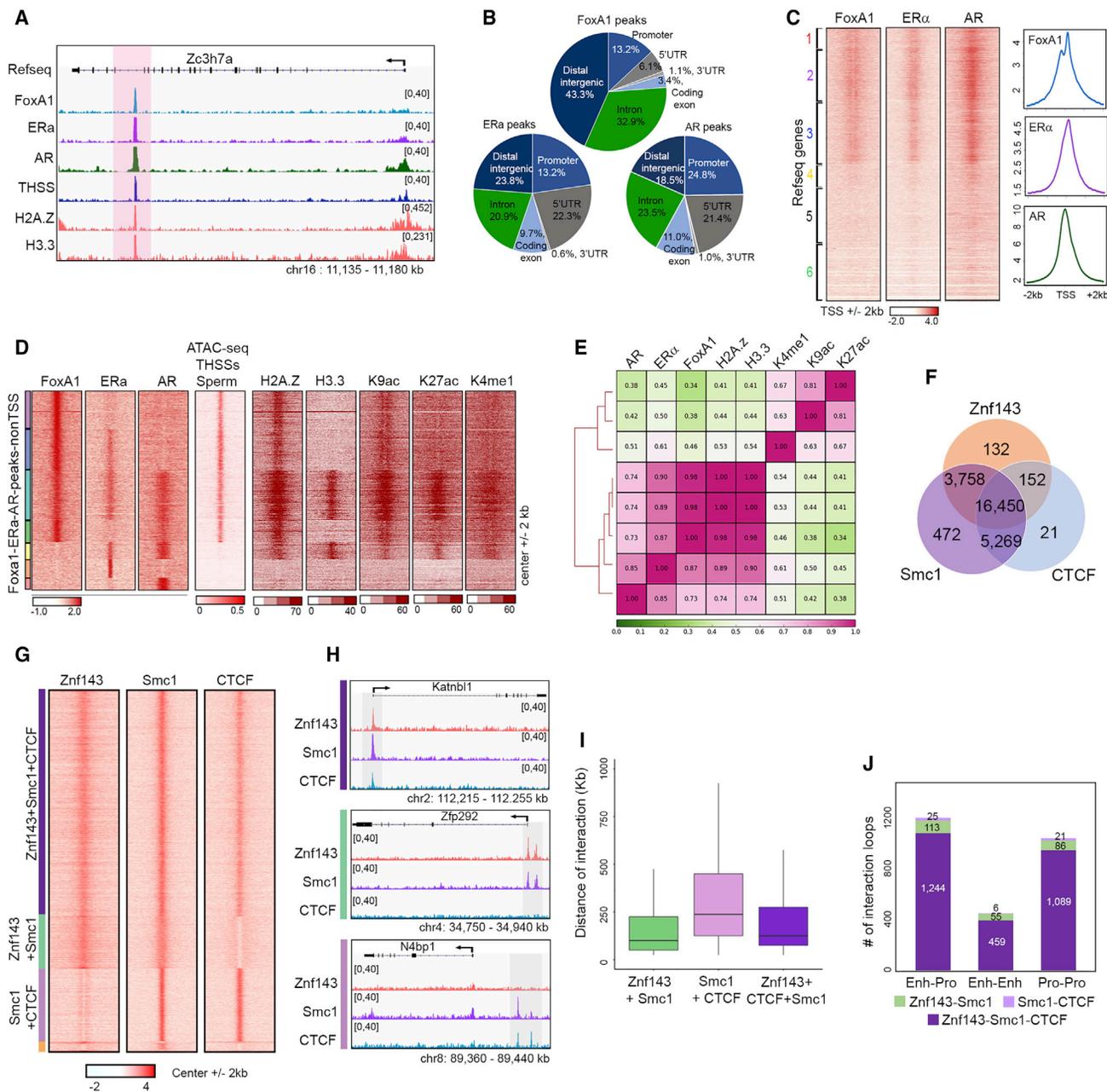
Figure 3. Foxa1, Nuclear Hormone Receptors, and Znf143 Are Present in Sperm Chromatin

(A) Genome browser view showing co-localization of Foxa1, ERα, and AR in a region of the mouse sperm genome.

(B) Genome-wide distribution of Foxa1, ERα, and AR peaks identified by MACS. TSS ± 500 bp are listed as promoters.

(C) Heatmaps and average profiles showing Foxa1, ERα, and AR ChIP-seq signal around TSSs (±2 kb).

(D) Heatmaps showing Foxa1, ERα, and AR ChIP-seq signal at ATAC-seq distal THSS non-TSS peaks. Clusters derived by k-means clustering of ChIP-seq reads. Histone modifications related to potential enhancers and histone variants at these sites are also shown (center ±2 kb).

(E) Pearson's correlation of Foxa1, ERα, and AR ChIP-seq read density genome-wide with histone modifications related to potential enhancers and histone variants. The clustering method is hierarchical complete linkage.

(F) Venn diagram showing overlap of non-TSS Znf143 peaks with Smc1 and CTCF.

(G) Heatmaps showing Znf143, Smc1, and CTCF at distal non-TSS peaks from ChIP-seq signal for each protein. Clusters were derived by k-means clustering of ChIP-seq reads.

(H) Genome browser views of specific examples corresponding to the 3 types of sites shown in (G).

(legend continued on next page)

annotated TSSs in the mouse genome (Figure 3C). These proteins can also be found together at a large fraction of TSSs in somatic cells (Figure S3E).

In addition to TSSs, Foxa1, ERα, and AR are present at a large number of distal sites in the genome. To analyze in detail the relative distribution of these 3 proteins, we combined the summits of all distal peaks for the 3 proteins and used them as anchors to perform k-means clustering. Results show that Foxa1 is present at many distal sites containing ERα and AR (Figure 3D). A subset of these sites contains all 3 proteins, but other sites contain only ERα or AR, others contain Foxa1 alone, and still others contain ERα and/or AR but lack Foxa1 (Figure 3D). In somatic cells, Foxa1 preferentially binds to lineage-specific enhancers that are defined by H3K27ac, H3K4me1, and histone variant H2A.Z (Jozwik et al., 2016). In sperm, all distal Foxa1 sites are flanked by nucleosomes containing H2A.Z, and those including ERα and AR also contain H3.3, H3K9ac, H3K27ac, and H3K4me1 (Figure 3D). To further analyze the relation among these proteins and histone modifications, we used ChIP-seq data to perform unsupervised hierarchical cluster analysis, and the results are shown in Figure 3E. These results suggest that Foxa1 and associated proteins may remain bound to sperm chromatin to mark enhancer elements involved in the regulation of gene expression in different tissues of the developing embryo. In agreement with this hypothesis, Foxa1 sites in sperm overlap with enhancers defined in ESCs; neural precursor cells (NPCs); and heart, lung, kidney, and other cell types of the developing embryo (Figure S3F) (Gao et al., 2016).

Znf143 Co-localizes with CTCF and Cohesin and Is Involved in Loop Formation

Analysis of footprints using ATAC-seq subnucleosome-sized reads suggests the presence of multiple Zn finger proteins bound to the mouse sperm genome (Figure S2A). One of these proteins, Znf143, is of special interest because it has been shown to co-localize with CTCF and cohesin and to mediate interactions between promoters and distal regulatory sequences in somatic cells (Bailey et al., 2015). Western analysis of sperm protein extracts supports the presence of this protein in mouse sperm (Figure S3A). To examine the distribution of Znf143 in mouse sperm, we performed ChIP-seq with antibodies to this protein (Table S1). We observed a total of 20,492 peaks for Znf143, most of which overlap with those of CTCF, Smc1, or both (Figure 3F). Many of these peaks are present at promoters, but also in introns and distal intergenic regions and are enriched in motifs for Znf143 and CTCF. To examine the relative distribution of these 3 proteins in more detail, we combined all of the peaks and used them as anchors to perform k-means clustering of Znf143, Smc1, and CTCF ChIP-seq signal. Most Znf143 peaks overlap with Smc1 and CTCF, and small fractions contain Znf143 and Smc1 or CTCF and Smc1 (Figure 3G). Therefore, most CTCF and cohesin sites in the sperm genome also contain Znf143. Examples of each class are shown in Figure 3H.

To analyze the possible role of Znf143 in CTCF- and cohesin-mediated interactions with high resolution, we performed *in situ* Hi-C with the restriction enzyme DpnII and obtained 2 biological replicates with a total of 520 million intra-chromosomal contacts (Table S2). We then used FitHi-C (Ay et al., 2014) to define significant interactions in sperm and examined the presence of Znf143, CTCF, and Smc1 at anchors mediating these interactions. Genomic sites containing the observed combinations of these proteins mediate interactions in the sperm genome, including sites containing Znf143 and Smc1, CTCF, and Smc1 and Znf143, CTCF, and Smc1 (Figure 3I). Interactions mediated by Znf143 and Smc1 take place over shorter distance ranges than those mediated by CTCF and Smc1 or the combination of the 3 proteins (Figure 3I). Furthermore, most contacts in sperm contain all 3 proteins at interacting anchors, and interacting sites containing Znf143 are present at both enhancers and promoters (Figure 3J). Examples of the different types of interactions are shown in Figure S3G.

GV and MII Stage Oocytes Contain Accessible Sites Corresponding to Specific TFs

To examine whether female gametes also contain accessible promoters and putative enhancers, we performed Omni-ATAC optimized for a small number of nuclei in mouse oocytes at the GV and MII stages of oogenesis. During fetal development, the mammalian oocyte enters meiosis and is arrested at the diplotene stage of prophase I, termed the GV stage, from birth to puberty. After a hormone surge, the oocyte undergoes meiotic maturation and is arrested again in metaphase of meiosis II (MII), which is the mature stage fertilized by sperm (Von Stetina and Orr-Weaver, 2011). We generated biological replicates of Omni-ATAC-seq libraries using 300–400 oocytes for each sample (Table S1). For subsequent analyses, we used a previously published list of oocyte TSSs that reflect the usage of alternative promoters in this cell type (Veselovska et al., 2015). Using model-based analysis of ChIP-seq (MACS) and THSS reads, we identified 28,439 and 31,597 THSS peaks in GV and MII oocytes, respectively. These sites are distributed across various gene features such as promoters, introns, and distal intergenic regions (Figure 4A). Therefore, condensed MII oocyte chromosomes contain a rich map of accessible sites similar to those of GV oocytes. Using nucleosome-size reads to map the location of nucleosomes, we found that oocytes at the GV and MII stages contain accessible TSSs flanked by well-positioned +1 and −1 nucleosomes (Figure 4B). Furthermore, using previously published H3K4me3 ChIP-seq data (Dahl et al., 2016; Zhang et al., 2016), we found good correlation between promoter accessibility and the presence of H3K4me3 (Figure 4C). Specific examples of genes with different degrees of correlation between these 3 features are shown in Figure S4A. These results suggest that a large number of promoters are accessible in GV oocytes, and these promoters remain accessible in MII metaphase chromosomes.

(I) Boxplot showing the distance of interactions mediated by Znf143, Smc1, and CTCF combinations present at anchors. These significant interactions are defined by FitHi-C using sperm Hi-C data.

(J) Number of interaction loops defined by HiCCUPS mediated by combinations of Znf143, Smc1, and/or CTCF. Enhancer-promoter interactions are defined based on the presence of H3K27ac and TSS ± 500 bp.

See also Figure S3.
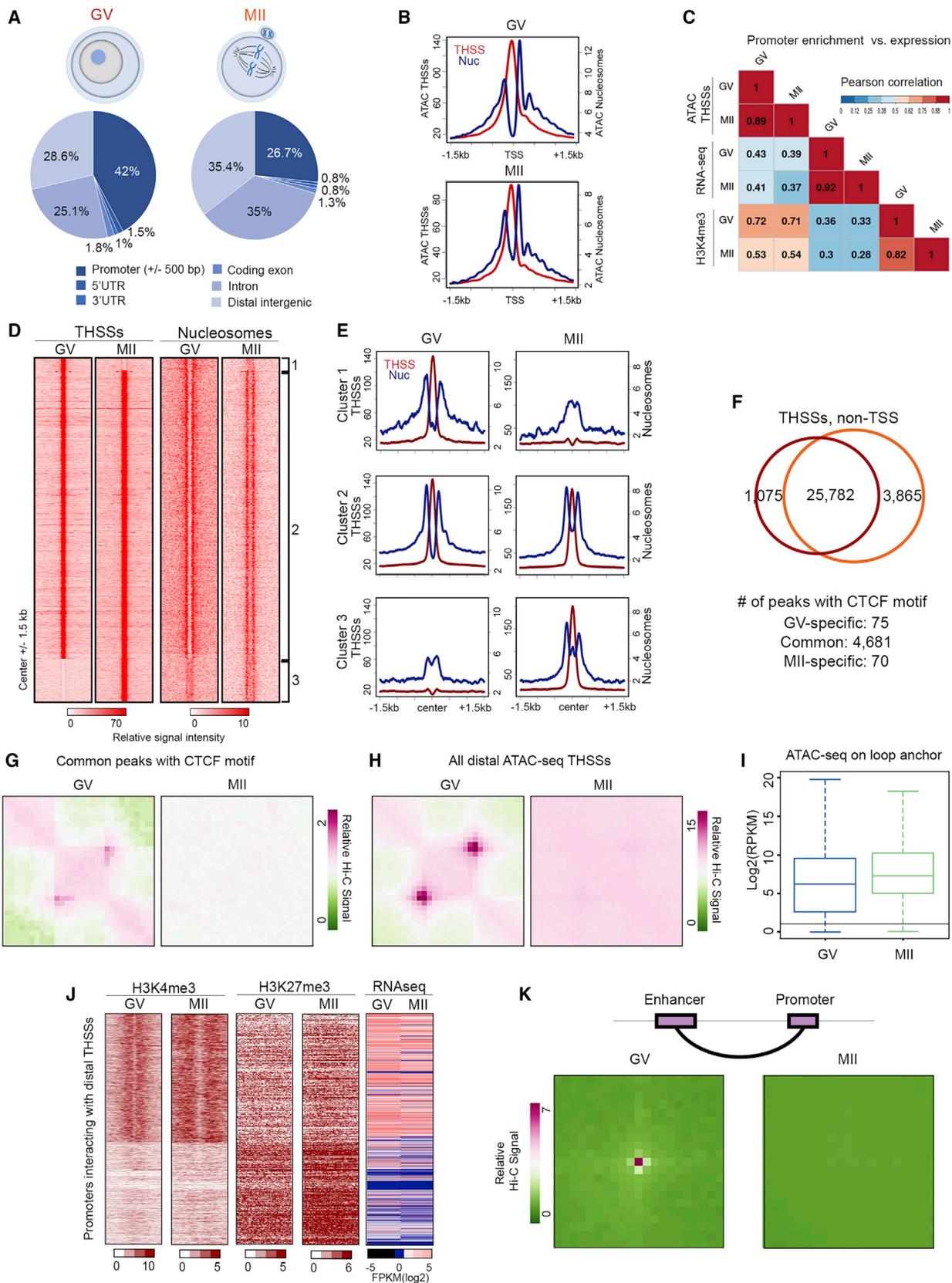
(legend on next page)

Although we have no direct evidence to suggest that accessible promoters in oocytes contain components of the transcription complex, the good correlation observed in sperm between ATAC-seq THSSs and ChIP-seq signal for Pol II and Med12 suggests that these or other components of the transcription complex may also be present at oocyte promoters.

In addition to THSSs present at the promoter, GV and MII oocytes contain a large number of THSSs in distal intergenic regions and introns (Figure 4A). We identified 25,782 distal THSSs shared by the 2 stages, 1,075 THSSs only present in the GV stage, and 3,865 THSSs newly gained in the MII stage (Figure 4D). These sites are flanked by well-positioned nucleosomes in both stages (Figures 4D and 4E). These nucleosomes lack H3K4me3 but contain H3K27ac in MII oocytes, suggesting that they may correspond to enhancers in an active state (Figure S4B). In agreement with this hypothesis, these sites have been identified as enhancers in various tissues of the developing embryo (Figure S4C). To gain insights into the nature of the TFs bound at these sites, we performed motif discovery analyses using common, GV-specific, and MII-specific distal THSSs using Multiple EM for Motif Elicitation (MEME) (Figure S4D). The most prevalent motifs at GV-specific peaks correspond to the TFs Nr5a2, Esrra, Sp1, and Gata3. MII-specific peaks—those appearing *de novo* as the oocyte goes into metaphase—correspond to the AP-1 factors Fos and Junb, Runx1, and Zfp691, whereas common peaks contain Junb, CTCF, and Nr5a2 (Figure S4D). CTCF, Nr5a2, and Esrra are known to be essential proteins for early development (Wu et al., 2016). Although we lack direct evidence for the presence of these TFs at oocyte THSSs, the results raise the intriguing possibility that TFs that are important for early embryo development may remain bound to the maternal chromatin through meiosis.

Of the 25,782 THSSs found in both GV and MII oocytes at distal sites, 4,681 contain the CTCF motif (Figure 4F). This is surprising, since CTCF has been shown to be evicted from metaphase chromosomes (Oomen et al., 2019), concomitant to the dramatic reorganization of chromosome structure that accompanies entry into mitosis (Gibcus et al., 2018). Analysis of 3D chromatin organization using Hi-C suggests that GV oocytes show the typical higher-order structures, such as topologically associating domains (TADs) and compartments, whereas MII oocytes lack these structures and instead show a similar organization to that observed in mitotic somatic cells (Du et al., 2017). Though it is formally possible that a different protein, rather than CTCF, is bound to THSSs containing the CTCF motif, we asked whether these sites present in GV and MII oocytes are associated with the formation of loops by analyzing published Hi-C data from these stages of oogenesis using Hi-C Computational Unbiased Peak Search (HiCCUPS) (Durand et al., 2016b). Although chromatin accessibility at many putative CTCF binding sites in GV oocytes is preserved in the MII stage, chromatin interactions mediated by these common sites are lost in MII oocytes (Figure 4G). We also examined Hi-C-derived interaction loops anchored by all distal ATAC-seq THSSs. Again, interaction loops present in the GV stage were not detectable in MII oocytes (Figure 4H), even though accessibility at THSSs present at loop anchors defined in GV show no change between the 2 stages (Figure 4I). To further analyze changes in interactions between the GV and MII oocytes, we defined active promoters as TSSs containing H3K4me3 and lacking H3K27me3, and inactive promoters as TSSs containing opposite histone modifications (Figure 4J). Promoters in an active epigenetic state in GV oocytes maintain this state in MII, and their activity correlates with levels of RNA in these 2 stages (Figure 4J). Specific examples are shown in Figure S4E. We also defined putative enhancers as the THSSs present in cluster 2 in Figure 4D. We then identified significant interactions in the Hi-C data using FitHi-C and determined enhancer-promoter interactions. We found that putative enhancers contact active promoters in GV oocytes, but these interactions disappear in the MII stage (Figure 4K), although TFs and the transcription complex remain bound to these sequences in MII chromosomes. These results suggest that the compaction of chromosomes in MII oocytes does not affect the binding of most proteins present in the GV stage, but interactions mediated by these proteins, including CTCF or a putative different protein that binds to the same motif and other distal sites presumed to be enhancers, are reprogrammed during chromosome condensation.

Chromatin Accessibility at Promoters in the Gametes Persists in Preimplantation Embryos

It is possible that promoters in an active state in the gametes play an instructional role in establishing transcription early in

---

Figure 4. Differences in the Distribution of TFs and Their Interactions in GV and MII Oocytes

(A) Genome-wide distribution of ATAC-seq THSSs from GV and MII oocytes.

(B) Average profiles of ATAC-seq THSSs and nucleosome signals from GV and MII oocytes at TSSs.

(C) Pearson's correlation of ATAC-seq THSSs with RNA-seq and H3K4me3 ChIP-seq at promoters in GV and MII oocytes.

(D) Heatmaps showing ATAC-seq THSSs and nucleosome signal at distal non-TSS peaks. Clusters derived by k-means clustering of ATAC-seq THSS reads. Cluster 1, GV specific; cluster 2, common between GV and MII; cluster 3, MII specific.

(E) Average profiles of ATAC-seq THSS and nucleosome signals from GV and MII oocytes at non-TSS peaks corresponding to clusters shown in (D).

(F) Venn diagram showing overlap of distal ATAC-seq THSS non TSS peaks between GV and MII oocytes. The lower part shows the number of peaks containing the CTCF motif.

(G) Metaplots of GV and MII oocyte Hi-C significant interactions between anchors occupied by ATAC-seq THSSs (cluster 2 in D) containing the CTCF motif.

(H) Metaplots of GV and MII oocyte Hi-C significant interactions between anchors occupied by all distal ATAC-seq THSSs present in GV oocytes.

(I) Boxplot comparing ATAC-seq THSS reads on the loop anchors shown in (H) between GV and MII oocytes. RPKM, reads per kilobase million.

(J) Heatmaps showing RNA-seq, H3K4me3, and H3K27me3 ChIP-seq from GV and MII oocytes at gene promoters interacting with common THSSs (cluster 2 in D) analyzed in (K).

(K) Metaplots of GV and MII oocyte Hi-C significant interactions between enhancers defined by the presence of THSSs and H3K27ac, and active promoters containing H3K4me3.

See also Figure S4.

embryogenesis. To test this hypothesis, we examined the correlation between ATAC-seq accessibility of sperm promoters and the levels of RNAPIISer5ph and found a high degree of correlation, suggesting that the strength of THSS peaks at promoters may be used as a direct measure of transcription complex occupancy (Figure 5A). We then classified sperm TSSs into 4 groups based on THSS signal and RNAPIISer5ph levels (Figure 5A). We compared ATAC-seq accessibility in these 4 clusters of sperm promoters with that of GV and MII oocytes. The results suggest that accessible promoters in sperm are also accessible in oocytes (Figure 5B). To examine whether these promoters remain accessible during embryogenesis, we made use of DNase-seq obtained at various stages of embryonic development (Lu et al., 2016). We found that promoters with high accessibility in sperm and oocytes are also accessible in all examined embryonic stages from zygote to morula (Figure 5B). Promoters in cluster 1 consistently have the strongest signal, which becomes progressively weaker in clusters 2–4, with little to no signal in cluster 4, and ATAC-seq accessibility correlates well with density of CpG islands (Figure 5B). A similar result was obtained using ATAC-seq data from preimplantation embryos (Wu et al., 2016) (Figure S5A). We confirmed these qualitative observations on chromatin accessibility in promoter regions in a quantitative manner by calculating reads per kilobase million (RPKM) for DNase-seq and fragments per kilobase million (FPKM) for ATAC-seq data in each stage from the gametes to the inner cell mass (ICM). The accessibility of promoters in each cluster remains approximately the same between the gametes, the PN3 and PN5 stages of the zygote, 2-cell and 4-cell stages, and increases during the 8-cell and morula stages (Figures 5C and S5B). A detailed quantitative measurement of the degree of correlation of normalized levels of DNase-seq and ATAC-seq at promoters in sperm, oocytes, and preimplantation embryos is shown in Figure S5C. An example showing the maintenance of promoter accessibility between the gametes and the early embryonic stages in the *Dnpep* gene is shown in Figure 5D.

Widespread transcription in the mouse embryo does not occur until the 2-cell stage, suggesting that the existence of accessible promoters containing components of the transcription complex is not sufficient for gene expression. To understand the relation between epigenetic modifications and accessibility at promoters, we re-analyzed published H3K4me3 and H3K27me3 data obtained in preimplantation embryos and gametes (Zhang et al., 2016; Zheng et al., 2016). For allele-specific comparisons, we distinguished between the paternal and maternal allele reads using SNP information and processed only normalized SNP-trackable reads in each allele (see STAR Methods). Results show that the majority of the H3K4me3 signal present in sperm-accessible promoters (clusters 1–3) is lost in the zygote and begins reestablishment in the late 2-cell stage, perhaps as a consequence of the start of major zygotic genome activation (ZGA) (Figure 5F). An example of the distribution of H3K4me3 during embryonic development at the promoter of the *Dnpep* gene is shown in Figure 5E. H3K4me3 is present at low levels in the promoters of MII oocytes, the maternal chromosomes of the PN5 zygote, and, as is the case for the paternal chromosome, increases in the late 2-cell stage (Figure 5F). In contrast,

H3K27me3 shows a negative correlation with accessibility at gene promoters in sperm, with those in cluster 4 showing the highest levels of this modification (Figure S5D). Accessible sperm promoters also lose this histone modification in the zygote, and appreciable levels are restored only in the ICM (Figure S5D). However, H3K27me3 is retained in the metaphase chromosomes of the MII oocyte and persists during all stages of early embryonic development in the maternal chromosome (Figure S5D). To further explore the relation between promoter accessibility in the gametes and preimplantation embryos and epigenetic modifications, we examined previously published genome-wide bisulfite sequencing (GWBS) data (Jung et al., 2017; Wang et al., 2014). We found that most accessible promoters in sperm have low DNA methylation levels (clusters 1–3) and remain unmethylated until embryonic day (E) 6.5, at which time methylation levels increase slightly. Most methylation changes in the early embryo take place in non-accessible promoters (cluster 4), which become demethylated up to the blastocyst stage and then are rapidly remethylated in both maternal and paternal chromosomes (Figure 5G).

Comparison of Chromatin Accessibility at Regulatory Regions between Sperm and Oocytes

A large fraction of THSS ATAC-seq peaks in sperm and oocytes display enrichment in distal regions or introns, which are presumed to be regulatory elements (Figures 2C and 4A). We therefore asked whether these distal sites, which are bound by various TFs in sperm, are different between male and female gametes. Visual inspection of ATAC-seq data suggests the existence of common and gamete-specific distal sites (Figure 6A). Therefore, we combined the distal ATAC-seq sites in sperm, GV, and MII oocytes and used them as anchors to perform k-means clustering. We found 23,168 sperm-specific, 10,058 oocyte-specific, and 32,465 common distal THSSs (Figure 6B). All of the sites are flanked by well-positioned nucleosomes containing H3K27ac in the corresponding cell type. MII oocyte-specific sites are flanked by better-positioned nucleosomes containing higher levels of H3K27ac (Figures 6B and 6C). To examine whether there are other chromatin features that distinguish sperm-specific versus common sites, we examined the distribution of histone variants around these sites. Results show that sperm distal sites that are also present in oocytes contain H3.3 and their flanking nucleosomes contain H2A.Z and high levels of H3K27ac in sperm, whereas those that are sperm- or oocyte-specific lack histone H3.3 (Figure 6C).

We next used ATAC-seq data obtained in preimplantation embryos (Wu et al., 2016) to ask whether distal sites accessible to Tn5 in sperm and oocytes are also accessible during early embryonic development. We found that sperm- and oocyte-specific distal sites are not maintained in preimplantation embryos. However, common sites present in both sperm and oocytes show robust and similar accessibility at all embryonic stages beginning 7.5 h after fertilization (Figure 6B). Using nucleosome-sized reads from ATAC-seq data, we found that these sites are flanked by well-positioned nucleosomes in all preimplantation stages (Figures 6C and 6SA). Gamete-specific distal sites that lack ATAC-seq THSSs in the embryo, suggesting the absence of bound TFs at these sites, still contain nucleosome-sized reads
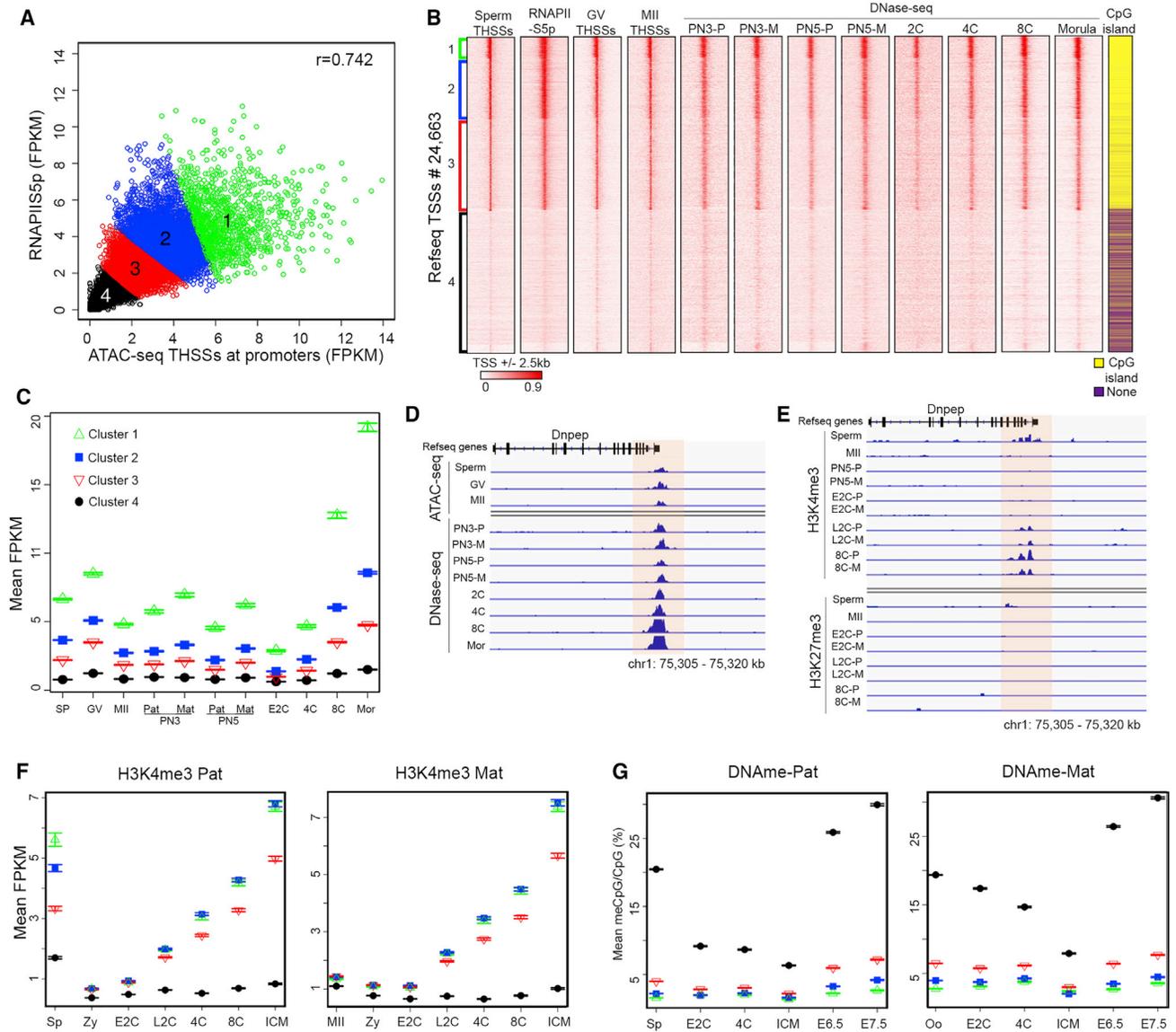
Figure 5. Promoter Accessibility Is Maintained between Germ Cells and Early Embryos

(A) Scatterplot showing the Pearson's correlation between sperm ATAC-seq THSSs and RNAPIISer5ph ChIP-seq at promoters (TSS ± 500 bp). Four clusters were defined by k-means clustering.

(B) Heatmaps showing sperm and oocyte ATAC-seq THSSs at promoters clustered as shown in (A). Also shown is DNase-seq signal enrichment at the same promoters in different stages of preimplantation embryos. GV, GV oocyte; MII, MII oocyte; PN3, zygote pronuclear stage 3; PN5, zygote pronuclear stage 5; 2C, 2-cell stage embryos; 4C, 4-cell stage embryos; 8C, 8-cell stage embryos; P, paternal; M, maternal.

(C) Plot showing the average ATAC-seq or DNase-seq signal in the different clusters and cell types shown in (B). Error bars indicate SEMs. SP, sperm; Mor, morula; Pat, paternal; Mat, maternal. Other symbols as in (B).

(D) Track view of ATAC-seq and DNase-seq signal at an example gene for the different cell types shown in (C). E2C, early 2-cell stage embryo; L2C, late 2-cell stage embryo. Other symbols as in (B).

(E) H3K4me3 and H3K27me3 ChIP-seq signal at an example gene for the different cell types shown in (C). Symbols as in (B).

(F) Plot showing average allele-specific signal of H3K4me3 ChIP-seq at the TSS clusters defined in (A). The paternal signal is shown on the left side of each panel, and the maternal signal is shown on the right side. SP, sperm; Zy, zygote; E2C, early 2-cell stage embryo; L2C, late 2-cell stage embryo; 4C, 4-cell stage embryos; 8C, 8-cell stage embryos; ICM, inner cell mass. The number of TSSs present in each cluster are as follows: cluster 1, 1,127; cluster 2, 3,196; cluster 3, 4,809; cluster 4, 7,563.

(G) Plot showing average allele-specific signal of BS-seq data at the TSS clusters defined in (A). The paternal signal is shown on the left side of each panel, and the maternal signal is shown on the right side. Oo, oocytes. Other symbols as in (F). The number of TSSs present in each cluster are as follows: cluster 1, 513; cluster 2, 1,498; cluster 3, 2,290; cluster 4, 3,325. Other symbols as in (F).

See also Figure S5.
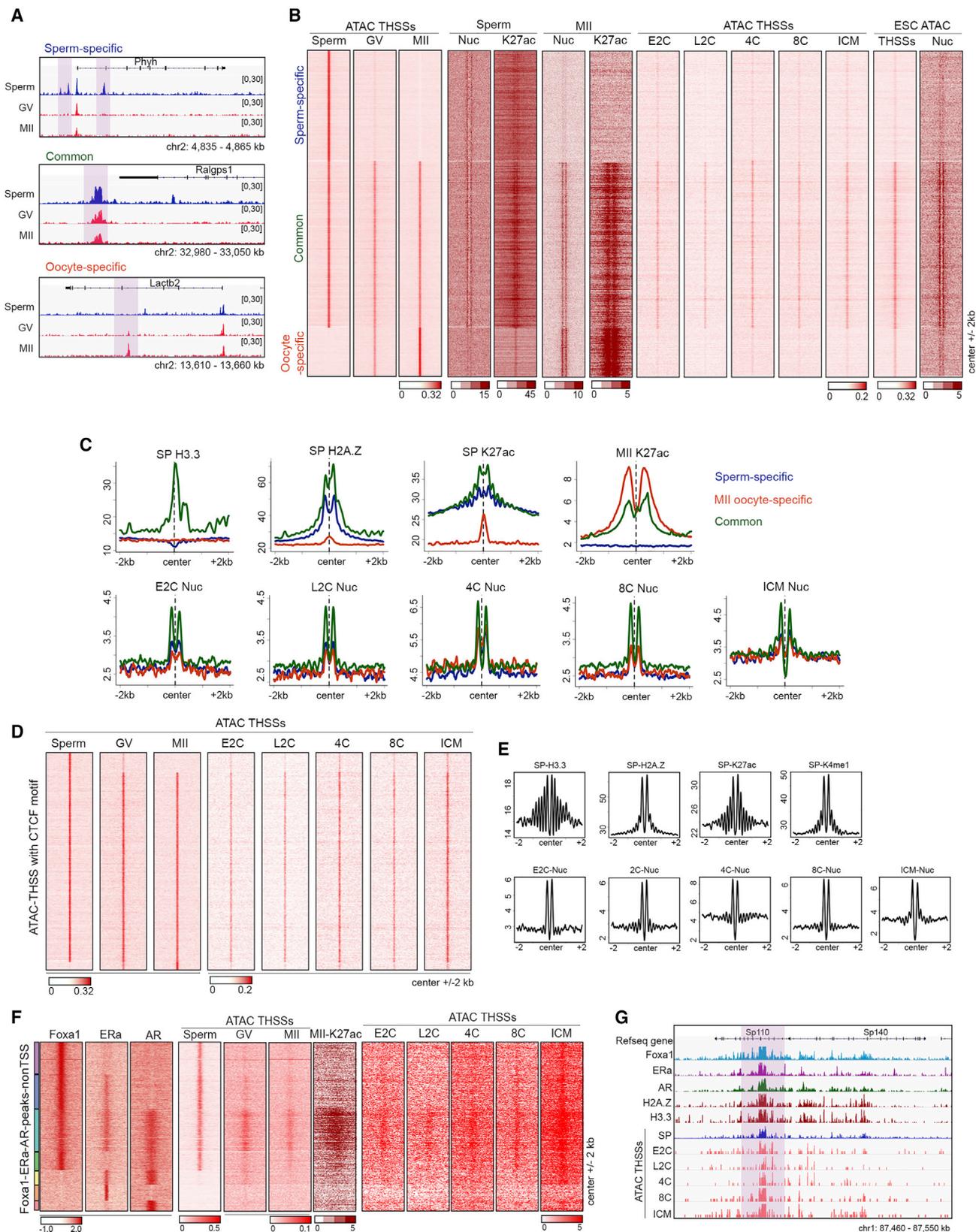
(legend on next page)

suggestive of positioned nucleosomes, although the intensity of this signal is lower than at common sites (Figures 6C and S6A). These findings suggest that distal THSSs present in both gametes, which presumably represent regulatory sequences, may be transmitted from germ cells to preimplantation embryos, where they maintain an open chromatin state throughout early embryonic development. In sperm, where this information is available, sites that persist in the embryo are marked by H3.3 and flanking nucleosomes containing H2A.Z (Figure 6C). To explore the possibility of additional epigenetic differences between distal THSSs in the gametes that persist in the embryo and those that do not, we examined the DNA methylation state at these sites. Sperm- and oocyte-specific sites are methylated at intermediate levels in the gametes, become demethylated up to the blastocyst stage, and are remethylated again to levels similar to those present in the gametes (Figure S6B). However, accessible regions common to the sperm and oocyte, which remain accessible in the embryo, have methylation levels <20% at all stages from gamete through epiblast (Figure S6B).

To determine which TFs are present at THSSs that persist between the gametes and the early embryo, we performed a motif analysis with MEME-ChIP on these sites. We found a number of TF motifs known to be essential for early development, including CTCF, ZIC4, NFYA, NFYB, AR, and ESR2 (Figure S6C). To further analyze the possible persistence of these proteins, we examined the maintenance of binding sites for CTCF. We combined all ATAC-seq THSSs present in sperm or oocytes containing a CTCF motif and used them as anchors to perform k-means clustering. Only a small fraction of CTCF sites are gamete specific, and those present in both sperm and oocytes are maintained in preimplantation embryos (Figure 6D). CTCF sites maintained in the early embryo are flanked by H3.3 and H2A.Z, contain H3K27ac and H3K4me1, and are flanked by well-positioned nucleosomes in all stages of early embryonic development (Figure 6E). In addition to CTCF, distal THSSs common to sperm and oocytes and maintained in the early embryo contain motifs for AR and ERs. Many of these sites also contain Foxa1 in sperm. Analysis of ATAC-seq subnucleosome-sized reads suggests that ChIP-seq sites containing Foxa1, AR, and ERα in sperm are also present in GV and MII oocytes (Figure 6F). Furthermore, these sites co-occupied by Foxa1, AR, and ERα in both sperm and oocytes are maintained in preimplantation embryos based on the presence of THSSs observed in ATAC-seq signal (Figure 6F). An example of a specific region of the

genome showing this maintenance of accessibility is shown in Figure 6G.

Conservation of CTCF- and TF-Mediated DNA Loops from Gametes through Preimplantation Embryos

It is possible that interactions mediated by CTCF and other TFs in sperm are maintained in preimplantation embryos and contribute to the establishment of specific transcription patterns during early embryogenesis. To test this hypothesis, we examined publicly available Hi-C data obtained with oocytes, preimplantation embryos (Du et al., 2017; Flyamer et al., 2017), sperm (the present study), and mouse brain cortex (Du et al., 2017). Hi-C data obtained in mouse embryos were separated into interactions present in the paternal or maternal chromosomes using SNP information. We determined significant interactions in the different Hi-C datasets using FitHi-C (Ay et al., 2014) and then analyzed CTCF-mediated interactions using CTCF motif-containing THSSs common to sperm, GV, and MII oocytes. These sites mediate interactions in both sperm and GV oocytes (Figure S7A). Furthermore, some or all of the sperm or oocyte interactions are also present in the paternal or maternal chromosomes of preimplantation embryos, respectively (Figure S7A). Based on ChIP-seq experiments, we have identified 3 types of cohesin-containing sites: those containing CTCF, Znf143, or CTCF plus Znf143. All of these 3 types of cohesin-containing sites mediate specific interactions in sperm and are maintained in the paternal chromosomes from the 1-cell stage to the ICM of the blastocyst (Figure S7B). To examine the persistence of interactions between the gametes and the early embryo in a different manner, we called paternal chromosome-specific loops at the PN5 zygote stage using a derivative of the HiCCUPS method (Cubeñas-Potts et al., 2017) and examined the presence of these paternal-specific loops in sperm and embryos. We found that paternal-specific loops present in the zygote are also present in the sperm and in the paternal chromosomes up to the blastocyst stage, as well as in the brain cortex (Figure 7A). These loops are not present in the GV oocyte or in the maternal chromosomes of 1- and 2-cell embryos. However, by the 8-cell stage, the maternal chromosomes acquire the same loops as those present in the paternal alleles (Figure 7A). Similarly, CTCF loops present in the PN5 maternal chromosome are present in the GV oocyte and are conserved in the maternal chromosomes throughout early embryonic development. These loops are not present in the paternal chromosomes until the 8-cell

Figure 6. Chromatin Accessibility at Regulatory Regions in Gametes and Preimplantation Embryos
(A) Track view of distal ATAC-seq THSSs in sperm, GV, and MII oocytes at allele-specific or common accessible distal regulatory regions (highlighted).
(B) Heatmaps comparing distal ATAC-seq THSSs and nucleosomes between sperm, oocytes, and preimplantation embryos. Also shown is the distribution of H3K27ac ChIP-seq signal in sperm and MII oocytes and ATAC-seq THSSs in preimplantation embryos and ESCs. Clusters were derived by k-means clustering of sperm and oocytes ATAC-seq THSS reads.
(C) Average profiles of ChIP-seq signal for histone variants H3.3 and H2A.Z in sperm and H3K27ac in sperm and MII oocytes at the same distal regulatory regions, as in (B). Lower panels show average plots of ATAC-seq nucleosome signals from preimplantation embryos.
(D) Heatmap showing ATAC-seq THSSs from sperm, GV, and MII oocytes and preimplantation embryos at all distal regulatory regions containing the CTCF motif.
(E) Average profiles of ChIP-seq signal for histone variants H3.3 and H2A.Z in sperm and H3K27ac in sperm and MII oocytes at the same distal regulatory regions as in (D). Average plots of ATAC-seq nucleosome signals from preimplantation embryos are also shown.
(F) K-means clustering of Foxa1, ERα, and AR ChIP-seq signal at distal THSSs overlapping a peak for at least 1 of these 3 proteins. ATAC-seq THSS signal from sperm, GV, and MII oocytes, and preimplantation embryos is shown at the same clusters.
(G) Track view of a specific genomic region showing the maintenance of accessibility at Foxa1 sites in sperm during embryo development.
See also Figure S6.
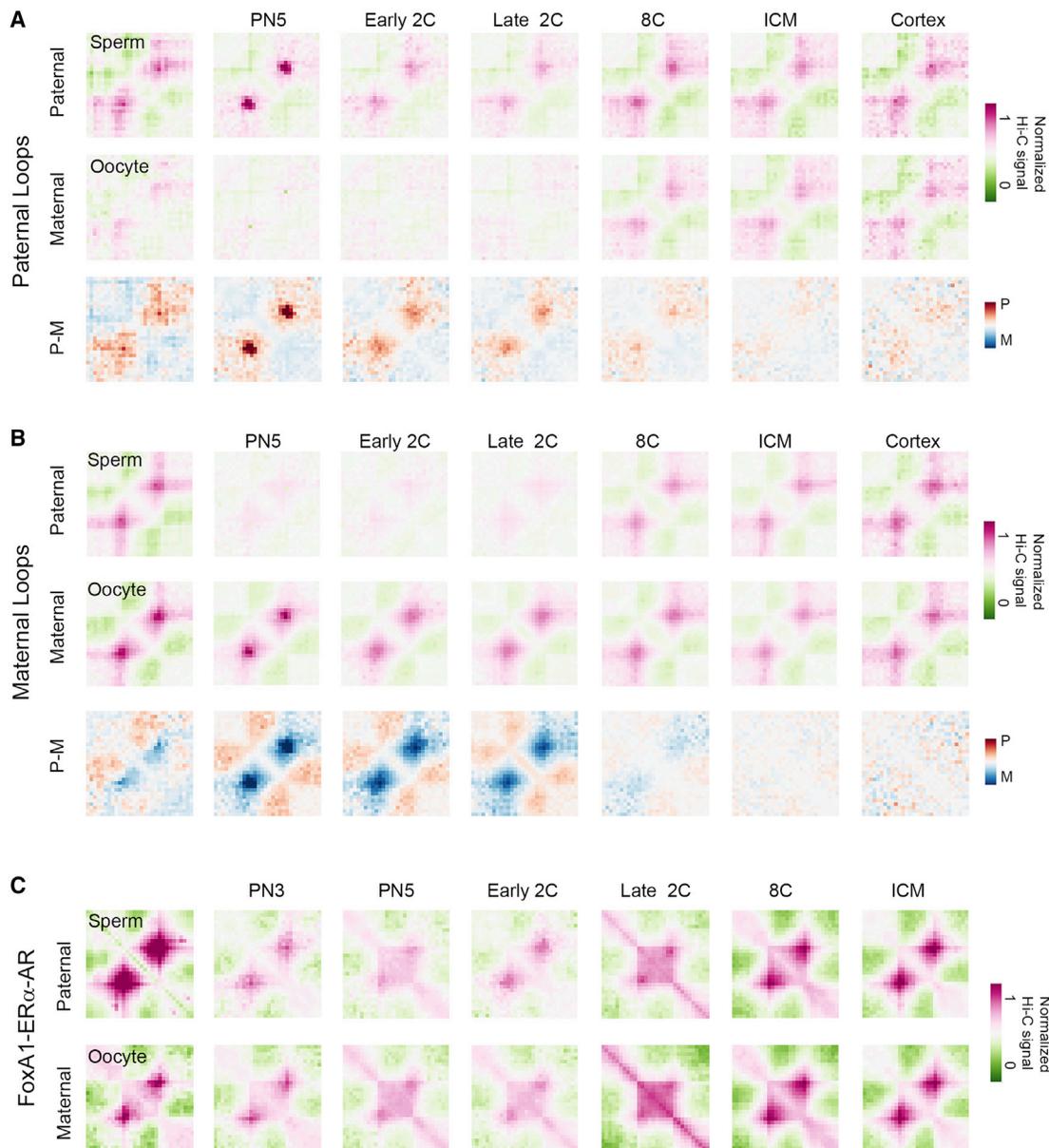
Figure 7. Conservation of CTCF- and TF-Mediated DNA Loops from Gametes through Preimplantation Embryos

(A) Metaplot showing median, distance-normalized Hi-C of sperm, GV oocytes, and allele-specific preimplantation embryos at loops defined in paternal PN5 chromosomes using a derivative of the HiCCUPS method. The bottom row shows the difference between paternal and maternal median, distance-normalized Hi-C interactions.

(B) Metaplot showing median, distance-normalized Hi-C of sperm, GV oocytes, and allele-specific preimplantation embryos at loops defined in maternal PN5 chromosomes using a derivative of the HiCCUPS method. The bottom row shows the difference between paternal and maternal median, distance-normalized Hi-C interactions.

(C) Metaplot showing median, distance-normalized Hi-C of sperm, GV oocytes, and allele-specific preimplantation embryos at Foxa1-mediated enhancer-promoter significant interactions determined by FitHi-C.

See also Figure S7.

stage, when both chromosomes appear to acquire the same 3D organization.

In addition to CTCF, a subset of distal Foxa1 sites containing the AR and ERα receptors are present in the gametes and persist in different stages of preimplantation embryos (Figure 6F). Using

significant interactions determined by FitHi-C, we examined contacts between these putative enhancers and promoters. These Foxa1 sites mediate interactions with promoters in the chromatin of sperm and GV oocytes (Figure 7C). Furthermore, these contacts appear to be maintained in the paternal and

maternal chromosomes during embryonic development between the zygote and the ICM of the blastocyst (Figure 7C). These results suggest that chromosomal interactions mediated by architectural proteins and TFs in the germline persist in the chromosomes of the embryo and could play a significant role in the expression of the genome during early embryogenesis and in the inheritance of transcriptional states between the gametes and the early embryo.

DISCUSSION

Here, we present evidence for the persistence of TF occupancy between the gametes and the early mouse embryo. The molecular mechanisms responsible for the transmission of acquired epigenetic information between males and their offspring via sperm remain uncertain, despite increasing reports of environment-altering exposures producing heritable phenotypes in the next generation. Current models to explain transgenerational inheritance emphasize DNA methylation, non-coding RNAs, and histone modifications, which undergo extensive reprogramming during early development. The results described here expand the repertoire of information present in the gametes to include TFs, architectural proteins, and the long-range contacts they mediate, and show that this information persists in preimplantation embryos, making DNA-bound proteins ideal candidates to explain the inheritance of acquired epiphenotypes.

The chromatin state of the mammalian gametes has been difficult to explore in detail due to the high compaction of their chromatin. However, using Omni-ATAC-seq we have been able to map thousands of sites in the mouse sperm genome, where the occurrence of Tn5 transposase hypersensitive sites suggests the presence of bound proteins. A subset of these sites is located at TSSs of approximately 60% of all mapped genes. Despite the transcriptionally inactive state of sperm, these promoters contain Mediator and Pol II phosphorylated in Ser5 and Ser2 (i.e., at the initiation and elongation stages of transcription), respectively. Oocytes at the prophase GV stage also contain accessible promoters, which remain accessible in the MII metaphase stage. Furthermore, the accessibility state is quantitatively maintained from the gametes to various stages of the preimplantation embryo; in other words, promoters with the highest subnucleosome-sized ATAC-seq reads in embryos correspond to those with the highest levels of ATAC-seq and Pol II in sperm and oocytes. It is possible that the apparent maintenance of the transcription complex at these promoters is related to a requirement for their expression in preimplantation embryos. Alternatively, the persistence of promoter accessibility between gametes and the early embryo may be related to the reprogramming of DNA methylation of the paternal and maternal genomes after fertilization. This hypothesis is supported by the low methylation levels of these promoters in the gametes, and the maintenance of this state in preimplantation embryos, which is in contrast to the high methylation and dramatic reprogramming of methylation in non-accessible promoters.

In addition to promoters, Omni-ATAC-seq allows the identification of thousands of sites located in distal regions of the sperm and oocyte genomes. These sites may correspond to

regulatory sequences such as enhancers, and their conservation in syntenic regions of the rhesus macaque and human genomes suggests a functional role for these sequences. In agreement with this, analysis of DNA binding motifs at these regions suggests the presence of TFs that remain bound to the sperm and oocyte genomes, including MII oocytes. A subset of these sites persists in the genome of the embryo at different stages up to the time of implantation. The sites that are maintained correspond to those present in both sperm and oocytes. In sperm, where this information is known, sites that persist in the embryo are flanked by nucleosomes containing H3.3, H2A.Z, and high levels of H3K27ac, whereas those that are not maintained lack H3.3 and have lower levels of H2A.Z and H3K27ac. The results suggest a regulated and determined assembly of sperm chromatin with the possible objective of controlling information passed on to the embryo. These observations agree with results obtained in zebrafish, in which H2A.Z-containing placeholder nucleosomes correlate with the persistence of epigenetic information between sperm and the embryo (Murphy et al., 2018). Sperm-specific sites that fail to persist still contain nucleosomes around these sites in the embryo, but their positioning becomes less marked as development proceeds. These results raise the interesting question of whether the presence of the TF determines the presence of the H2A.Z nucleosomes or conversely.

As proof of concept to show that TFs are actually present in sperm chromatin at ATAC-seq-accessible regions, we used ChIP-seq to analyze the distribution of Foxa1 and nuclear hormone receptors. We found that these proteins are present at thousands of TSSs and distal sites in the sperm genome flanked by nucleosomes containing H2A.Z and active histone modifications. A subset of these sites is also present in GV and MII oocytes, and these common sites persist in preimplantation embryos. Persistent sites contain H3.3 and higher levels of H2A.Z and active histone modifications, including H3K4me1, in sperm. These inheritable distal sites also contain AR and ERα, supporting the idea that they constitute regulatory elements that may play a role in controlling gene expression during early embryo development. In agreement with this, Foxa1-mediated interactions between these sites and gene promoters can be observed in Hi-C data obtained in preimplantation embryo stages. These interactions are already present in the gametes, suggesting the transmission of 3D chromatin organization between the gametes and the embryo.

The organization of the chromatin fiber in the nuclear space arises as a consequence of interactions between compartmental domains in the same transcriptional state but point-to-point contacts mediated by CTCF and cohesin are also important contributors to the establishment of this organization. Other proteins with less clear roles, such as YY1 and Znf143, are also present at CTCF sites and may regulate the specificity or frequency of these interactions. In agreement with the finding of the Znf143 binding motif at accessible sites, ChIP-seq experiments demonstrate the presence of this protein at thousands of sites in the genome where it co-localizes with CTCF and/or cohesin. Some Znf143 sites lack CTCF but contain cohesin, and these sites appear to mediate shorter-range interactions among enhancers and promoters than classical CTCF and cohesin loops. Most of

the CTCF sites present in sperm are accessible to Tn5 transposase in both GV and MII oocytes, and these sites shared by the gametes persist during early embryogenesis up to the blastocyst stage. In spite of their persistence in MII oocytes, putative CTCF sites present at this stage do not mediate interactions in MII metaphase chromosomes. However, CTCF loops present in the GV oocyte have been reestablished in the maternal chromosome of the zygote by the PN5 stage, persist during early embryonic development, and can be observed in adult brain cortex. These interactions are absent from the paternal chromosome in the zygote and 2-cell stage, but are established by the 8-cell stage. Similarly, paternal chromosome-specific interactions present in sperm and the early embryo are absent in the maternal chromosomes, which become the same as the sperm by the 8-cell stage. The mechanisms by which parent-of-origin specific interactions are established in the sperm or the zygote and converge in the 2 parental chromosomes by the 8-cell stage are unknown and an important issue for future discovery.

In summary, we have shown that sperm and oocytes contain far more information with the potential to encode epigenetic memory than was previously recognized. Specific sites on gamete chromatin are poised with TFs, despite the lack of transcriptional inactivity, and chromatin accessibility at these sites and at distal regulatory elements is maintained in the embryo until at least the ICM stage. Since global remethylation occurs after the ICM stage, persistent sites with bound TFs inherited from gametes and retained through early embryogenesis may inhibit remethylation at their binding motifs. These sites may then remain accessible to the transcription machinery later in development as differentiation ensues. These observations open the possibility that TFs, whose distribution in the genome may be altered by environmental effects, are the basis for the transmission of epiphenotypes between generations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Isolation of mouse sperm and oocytes
  - Preparation of monkey semen
  - Preparation of human semen
  - Assay for transposase-accessible chromatin using sequencing (ATAC-seq)
  - ChIP-seq in sperm cells
  - Western analysis
  - *In situ* Hi-C
  - Data Processing
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.04.014.

AUTHOR CONTRIBUTIONS

Y.H.J. planned and performed experiments, analyzed the data, and wrote the manuscript. I.K. analyzed the data and contributed to writing the manuscript. H.B.G. performed experiments and contributed to writing the manuscript. M.J.R. and B.J.B. analyzed the data. K.P., A.B., and X.L. performed experiments. A.W.S.C. planned experiments. VGC planned the experiments and contributed to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res. *24*, 999–1011.

Bae, W.K., Kang, K., Yu, J.H., Yoo, K.H., Factor, V.M., Kaji, K., Matter, M., Thorgeirsson, S., and Hennighausen, L. (2015). The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration. FASEB J. *29*, 1653–1662.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369-73.

Bailey, S.D., Zhang, X., Desai, K., Aid, M., Corradin, O., Cowper-Sal Lari, R., Akhtar-Zaidi, B., Scacheri, P.C., Haibe-Kains, B., and Lupien, M. (2015). ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat. Commun. *2*, 6186.

Bliss-Moreau, E., Theil, J.H., and Moadab, G. (2013). Efficient cooperative restraint training with rhesus macaques. J. Appl. Anim. Welf Sci. *16*, 98–117.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schübeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat. Struct. Mol. Biol. *17*, 679–687.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bunch, H., Zheng, X., Burkholder, A., Dillon, S.T., Motola, S., Birrane, G., Ebmeier, C.C., Levine, S., Fargo, D., Hu, G., et al. (2014). TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nat. Struct. Mol. Biol. *21*, 876–883.

Carone, B.R., Hung, J.H., Hainer, S.J., Chou, M.T., Carone, D.M., Weng, Z., Fazzio, T.G., and Rando, O.J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev. Cell *30*, 11–22.

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res. *23*, 341–351.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., and Sinnott-Armstrong, N.A. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., and Corces, V.G. (2017). Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res. *45*, 1714–1730.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., He, J., Xiang, Y., Wang, Q., Li, Y., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature *547*, 232–235.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110–114.

Gao, T., He, B., Liu, S., Zhu, H., Tan, K., and Qian, J. (2016). EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. Bioinformatics *32*, 3543–3551.

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., et al. (2018). A pathway for mitotic chromosome formation. Science *359*, eaao6135.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

Hammoud, S.S., Low, D.H., Yi, C., Carrell, D.T., Guccione, E., and Cairns, B.R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell Stem Cell *15*, 239–253.

Hasegawa, K., Sin, H.S., Maezawa, S., Broering, T.J., Kartashov, A.V., Alavattam, K.G., Ichijima, Y., Zhang, F., Bacon, W.C., Greis, K.D., et al. (2015). SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. Dev. Cell *32*, 574–588.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109.

Hewitt, S.C., Li, L., Grimm, S.A., Chen, Y., Liu, L., Li, Y., Bushel, P.R., Fargo, D., and Korach, K.S. (2012). Research resource: whole-genome estrogen receptor α binding in mouse uterine tissue revealed by ChIP-seq. Mol. Endocrinol. *26*, 887–898.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. *34*, D590–D598.

Hisano, M., Erkek, S., Dessus-Babus, S., Ramos, L., Stadler, M.B., and Peters, A.H. (2013). Genome-wide chromatin analysis in mature mouse and human spermatozoa. Nat. Protoc. *8*, 2449–2470.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. *43*, 27–33.

Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E., and Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins asso-ciated with histone-marked genomic regions. Proc. Natl. Acad. Sci. USA *112*, 3841–3846.

Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. Cell Rep. *17*, 2715–2723.

Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., and Corces, V.G. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. Cell Rep. *18*, 1366–1382.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, R., Mav, D., Grimm, S.A., Jothi, R., Shah, R., and Wade, P.A. (2014). Fine-tuning of epigenetic regulation with respect to promoter CpG content in a cell type-specific manner. Epigenetics *9*, 747–759.

Liu, T. (2014). Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol. Biol. *1150*, 81–95.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell *165*, 1375–1388.

Maza, I., Caspi, I., Zviran, A., Chomsky, E., Rais, Y., Viukov, S., Geula, S., Buenrostro, J.D., Weinberger, L., Krupalnik, V., et al. (2015). Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. Nat. Biotechnol. *33*, 769–774.

Moran, S.P., Chi, T., Prucha, M.S., Agca, Y., and Chan, A.W. (2016). Cryotolerance of Sperm from Transgenic Rhesus Macaques (Macaca mulatta). J. Am. Assoc. Lab. Anim. Sci. *55*, 520–524.

Murphy, P.J., Wu, S.F., James, C.R., Wike, C.L., and Cairns, B.R. (2018). Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. Cell *172*, 993–1006.e13.

Oomen, M.E., Hansen, A.S., Liu, Y., Darzacq, X., and Dekker, J. (2019). CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. Genome Res. *29*, 236–249.

Pihlajamaa, P., Sahu, B., Lyly, L., Aittomäki, V., Hautaniemi, S., and Jänne, O.A. (2014). Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. EMBO J. *33*, 312–326.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201.

Putkhao, K., Chan, A.W., Agca, Y., and Parnpai, R. (2013). Cryopreservation of transgenic Huntington's disease rhesus macaque sperm-A Case Report. Cloning Transgenes. *2*, 1000116.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1–11.12.34.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Sahu, B., Pihlajamaa, P., Dubois, V., Kerkhofs, S., Claessens, F., and Jänne, O.A. (2014). Androgen receptor uses relaxed response element stringency for selective chromatin binding and transcriptional regulation in vivo. Nucleic Acids Res. *42*, 4230–4240.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. *25*, 1757–1770.

Shao, Z., Zhang, Y., Yuan, G.C., Orkin, S.H., and Waxman, D.J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. *13*, R16.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol. *16*, 209.

Von Stetina, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. Cold Spring Harb. Perspect. Biol. *3*, a005553.

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell *151*, 206–220.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979–991.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol. Cell *63*, 1066–1079.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Znf143	Proteintech	Cat#16618-1-AP; RRID:AB_2218324
Anti-RNAPIISer5ph	Abcam	Cat#ab5131; RRID:AB_449369
Anti-RNAPIISer2ph	Abcam	Cat#ab5095; RRID:AB_304749
Anti-Med12	Bethyl laboratories	Cat#A300-774A; RRID:AB_699756
Anti-Foxa1	Abcam	Cat#ab5089; RRID:AB_304744
Anti-ERα	Santa Cruz	Cat#sc-543; RRID:AB_631471
Anti-AR	Santa Cruz	Cat#sc-816; RRID:AB_1563391
Anti-H3.3	Abcam	Cat#ab176840; RRID:AB_2715502
Anti-H2A.Z	Abcam	Cat#ab4174; RRID:AB_304345
Deposited Data		
Raw and processed data of Hi-C libraries	This study	GEO: GSE116856
Raw and processed data of ChIP-seq libraries	This study	GEO: GSE116855
Raw and processed data of ATAC-seq libraries	This study	GEO: GSE116854
DNase-seq in mouse embryos	Lu et al., 2016	GEO: GSE76642
ATAC-seq in mouse embryos	Wu et al., 2016	GEO: GSE66390
Hi-C in mouse embryos	Du et al., 2017	GEO: GSE82185
Hi-C in oocytes	Flyamer et al., 2017	GEO: GSE80006
H3K4me3 ChIP-seq in mouse embryos	Zhang et al., 2016	GEO: GSE71434
H3K27me3 ChIP-seq in mouse embryos	Zheng et al., 2016	GEO: GSE76687
WGBS in mouse embryos	Wang et al., 2014	GEO: GSE56697
RNA-seq in GV and MII oocytes	Zhang et al., 2016	GEO: GSE71434
H3K4me3 ChIP-seq in GV oocytes	Zhang et al., 2016	GEO: GSE71434
H3K27me3 ChIP-seq in GV and MII oocytes	Zheng et al., 2016	GEO: GSE76687
H3K4me3 and H3K27ac ChIP-seq in MII oocytes	Dahl et al., 2016	GEO: GSE72784
RNAPIISer5ph & RNAPIISer2ph ChiP-seq in liver	Li et al., 2014	GEO: GSE41472
H3K36me3 ChIP-seq in liver	Li et al., 2014	GEO: GSE41472
H3K4me3 ChIP-seq in liver	Bae et al., 2015	GEO: GSE53627
RNAPIISer5ph ChiP-seq in ESC	Wamstad et al., 2012	GEO: GSE47949
RNAPIIS2ph ChiP-seq in ESC	Bunch et al., 2014	GEO: GSE48253
H3K36me3 & H3K4me3 ChIP-seq in ESC	Ji et al., 2015	GEO: GSE62380
RNAPII2 ChIP-seq in spermatids	Hammoud et al., 2014	GEO: GSE49621
H3K4me3, H3K27me3, H3K36me3, H3K27ac, H3K9ac, H3K4me1, and Protamine1 ChIP-seq in sperm	Jung et al., 2017	GEO: GSE79230
RNA-seq in spermatids	Hasegawa et al., 2015	GEO: GSE55060
RNA-seq in sperm	Hammoud et al., 2014	GEO: GSE49622
BS-seq in sperm	Jung et al., 2017	GEO: GSE79230
H3K4me1 ChIP-seq in human sperm	Hammoud et al., 2014	GEO: GSE57095
Smc1 and CTCF ChIP-seq in sperm	Jung et al., 2017	GEO: GSE79230
ATAC-seq in ESC	Maza et al., 2015	GEO: GSE67298
Foxa1 ChIP-seq in mouse caput epididymis	Pihlajamaa et al., 2014	GEO: GSE47192
ERα ChIP-seq in mouse uterus	Hewitt et al., 2012	GEO: GSE56501
AR ChIP-seq in mouse prostate	Sahu et al., 2014	GEO: GSE51106

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Juicer, HiCCUPS	Durand et al., 2016b	https://github.com/aidenlab/juicer/wiki
Juicebox	Durand et al., 2016a	http://aidenlab.org/juicebox/
MACS2.0	Liu, 2014	https://github.com/taoliu/MACS
Bowtie	Langmead et al., 2009	https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.2.1.1
Samtools	Li et al., 2009	https://sourceforge.net/projects/samtools/files/
Picard Tools	N/A	http://picard.sourceforge.net; https://broadinstitute.github.io/picard/
TOPHAT2	Kim et al., 2013	https://github.com/infphilo/tophat
Cufflinks	Trapnell et al., 2010	http://cole-trapnell-lab.github.io/cufflinks
DANPOS	Chen et al., 2013	https://sites.google.com/site/danposdoc/install
bedtools	Quinlan, 2014	https://bedtools.readthedocs.io/en/latest/
ngsplot	Shen et al., 2014	https://bioconductor.org/packages/release/bioc/html/edgeR.html
MAnorm	Shao et al., 2012	http://bcb.dfci.harvard.edu/~gcyuan/MAnorm/MAnorm.htm
Wellington-bootstrap	Piper et al., 2013	https://pythonhosted.org/pyDNase/
Java treeview	N/A	https://sourceforge.net/projects/jtreeview/files/
Cluster3	N/A	http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm
FIMO	Grant et al., 2011	http://meme-suite.org/
LiftOver	Hinrichs et al., 2006	http://genome.ucsc.edu/cgi-bin/hgLiftOver
MEME	Bailey et al., 2006	http://meme-suite.org/
FitHi-C	Ay et al., 2014	https://noble.gs.washington.edu/proj/fit-hi-c/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact Victor Corces, email: vgcorces@gmail.com, Phone: 404-727-4250, Fax: 404-727-2880.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experiments presented in this study make use of sperm isolated from mice, monkeys, and humans. All experiments were conducted according to the animal research guidelines from NIH and all protocols for animal usage were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) or the Institutional Review Board (IRB) of Emory University. Human sperm was obtained from anonymous volunteers and no informed consent was required in the protocol approved by the IRB. Mice were maintained and handled in accordance with the Institutional Animal Care and Use policies at Emory University. Mice were housed in standard cages on a 12: 12 h light:dark cycle and given ad lib access to food and water. Healthy 8-week old CD1 mice (Charles River Labs) not involved in previous procedures were used for sperm isolation. No genotyping was performed.

METHOD DETAILS

Isolation of mouse sperm and oocytes

Euthanasia was performed by CO_2 asphyxiation and the epididymes were removed. Mature sperm were collected from the dissected cauda epididymis of 8-10-week-old CD1 mice (Charles River Labs). After dissection to eliminate blood vessels and fat, the cauda epididymis was rinsed with PBS, deposited in Donners medium in a cell culture plate, and punctured with a needle. Sperm were then transferred to a tube and allowed to swim up for 1 hr (Hisano et al., 2013). Purity of sperm was determined by examination under a microscope after DAPI staining. After counting 1000 sperm, purity was determined to be at least 99.9% if no contaminating cells were observed.

For the preparation of GV-stage oocytes, the ovaries were harvested from 16-day-old CD1 female pups. GV oocytes were harvested from the ovaries. The zona pellucida was removed with hyaluronidase to avoid any residual cumulus cells. For MII oocyte

collection, CD1 females at 6-8 weeks of age were hormonally stimulated by injection of 7.5 IU of pregnant mare serum gonadotropin (PMSG) followed by 5 IU of human chorionic gonadotropin (hCG) at 48 h post-PMSG injection. MII oocytes were harvested at 13-14 h post-hCG injection. Zona pellucida and polar bodies were removed with hyaluronidase.

Preparation of monkey semen

Ejaculates were collected from two male wild-type rhesus monkeys (*Macaca mulatta*) via electro-ejaculation. Each male was chair trained (Primate Product Inc.) using the "pole and collar" technique (Bliss-Moreau et al., 2013; Moran et al., 2016). The monkey was then lightly sedated with Ketamine (~0.4 mg/kg body weight) administered via intramuscular injection. One pre-sized defibrillator gel electrode was wrapped around the base of the penis and connected to the negative electrode lead. The second gel electrode was positioned immediately behind the glans and connected to the positive lead. The electro-ejaculator device was slowly raised up to 32 V or until ejaculation. Ejaculates were kept at room temperature for 25 min to liquefy. The liquid sample was transferred into a fresh 15 mL conical tube and washed once with TALP-HEPES medium supplemented with 4 mg/ml BSA (Moran et al., 2016; Putkhao et al., 2013), then centrifuged at 400 x g at room temperature for 5 min. Supernatant was removed, and sample was resuspended into 1 mL total volume with TALP-HEPES + BSA. Density gradient centrifugation was carried out using PureSperm® (Nidacon). Briefly, samples were layered on a double density gradient of 2 mL of PureSperm®40 (top) and 2 mL PureSperm®80 (bottom) and centrifuged at 300 x g for 20 min at room temperature. Supernatant was removed and the pellets were washed with 1 mL PureSperm® Wash and centrifuged at 500 x g for 10 min. Samples were then resuspended in 1 mL TALP-HEPES + BSA and an aliquot was examined for sperm concentration and purity.

Preparation of human semen

Human semen samples were obtained from 20-25 year-old heathy donors. After seminal liquefaction, sperm was transferred to sterile 10 mL centrifuge tubes and washed twice with Human Tubal Fluid medium (Irvine Scientific) supplemented with human serum albumin. After washing, isolation and purification of human spermatozoa was carried out as described (Hisano et al., 2013).

Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq was carried out using the Omni-ATAC protocol (Corces et al., 2017). After sperm cells were counted, the nuclei from 100,000 sperm were isolated with Lysis Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin. The purified sperm nuclei pellet was then resuspended in the transposase reaction mix containing 0.05% digitonin and incubated for 30 min at 37°C. Following incubation, sperm were treated with Proteinase K at 55°C for 2 hr, and gDNA was isolated by phenol:chloroform:isoamyl alcohol and EtOH precipitation. Library amplification was done with 2x KAPA HiFi mix (Kapa Biosystems) and 1.25 μM indexed primers using the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and 10-11 cycles at 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min.

ChIP-seq in sperm cells

Chromatin immunoprecipitation to detect the localization of DNA-binding proteins in sperm was performed using the standard ChIP-seq protocol with some modifications. Briefly, 10-20 million sperm cells were crosslinked with 1% formaldehyde in 1x PBS for 10 min at RT, and the reaction was quenched with 125 mM glycine for 10 min at RT. After washing with PBS, crosslinked sperm were lysed with 5 mM PIPES, 85 mM KCl, 0.5% NP40 and 1x protease inhibitors (P8107S, NEB) on ice for 15 min. After centrifugation, sperm cells were resuspended in RIPA buffer (1x PBS; 1% NP40; 0.5% sodium deoxycholate; 0.1% SDS; 1x protease inhibitors) and incubated on ice for 20 min. The purified sperm chromatin was sonicated to 300-1000 bp using a Diagenode Bioruptor. After 25 cycles (30 s on and 60 s off) the supernatant containing sheared chromatin was collected. Immunoprecipitation was performed overnight at 4°C with antibodies against Med12 (A300-774A, Bethyl laboratories), RNAPIISer5ph (ab5131, Abcam), RNAPIISer2ph (ab5095, Abcam), Foxa1(ab5089, Abcam), AR (sc-816, Santa Cruz), Znf143 (16618-1-AP, Proteintech) and ERα (sc-543, Santa Cruz). Libraries for Illumina sequencing were constructed using the following standard protocol. Fragment ends were repaired using the NEBNext End Repair Module and adenosine was added at the 3′ ends using Klenow fragment (3′ to 5′ exo minus, New England Biolabs). Precipitated DNA and input DNA were incubated with adaptors at room temperature for 1 hr with T4 DNA ligase (New England Biolabs) and amplified with Illumina primers. Chromatin immunoprecipitation to detect histone variants was carried out as described (Hisano et al., 2013) using antibodies to H2A.Z (ab4174, Abcam) and H3.3 (ab176840, Abcam).

Western analysis

For Western analysis of chromatin proteins, an equal number of sperm and J1 mESCs were resuspended in 1x Laemmli sample buffer (5% 2-mercaptoethanol, 0.002% bromophenol blue, 10% Glycerol, 2% SDS, 62.5 mM Tris-HCl pH 6.8). Samples were boiled for 5 min and supernatant was loaded onto SDS-PAGE 4%–15% gels at a ratio of 12:1 sperm:mESCs. Because sperm are haploid, sperm lanes contain 6-fold more genome equivalents than lanes with mESCs. Membranes were blocked in TBST (20 mM Tris, pH7.4, 150 mM NaCl, 0.05% Tween 20) with 5% nonfat milk powder and incubated overnight with the antibodies described above. Membranes were washed 3 times with TBST and incubated with secondary antibodies-conjugated to HRP (1:5000, Jackson ImmunoResearch Laboratories) for 1 hr. After three more washes, the presence of different proteins was detected using SuperSignal West Pico/Dura Chemiluminescent substrate (Thermo Scientific).

In situ Hi-C

in situ Hi-C libraries were prepared using DpnII restriction enzyme as previously described (Rao et al., 2014). Briefly, 10 million sperm were crosslinked with 1% formaldehyde, quenched with glycine, washed with PBS, and permeabilized to obtain intact nuclei. Nuclear DNA was then digested with DpnII, the 5′ overhangs were filled with biotinylated dCTPs and dA/dT/dGTPs to make blunt-end fragments, which were then ligated, reverse-crosslinked, and purified by standard DNA ethanol precipitation. Purified DNA was sonicated to 200-500 bp small fragments and captured with streptavidin beads. Standard Illumina TruSeq library preparation steps, including end-repairing, A-tailing, and ligation with universal adaptors were performed on beads, washing twice in Tween Washing Buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20) between each step. DNA on the beads was PCR amplified with barcoded primers using KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) for 5∼12 PCR cycles to obtain enough DNA for sequencing. Generated libraries were paired-end sequenced on Illumina HiSeq2500 v4 or NovaSeq 6000 instruments. Two biological replicates were generated, and replicates were combined for all analyses after ensuring high correlation.

Data Processing

Analysis of ATAC-seq data

All libraries were sequenced using an Illumina Hiseq2500 v4 sequencer and 50 bp paired-end format. Paired reads were aligned to the mouse reference genome mm9, human reference genome hg19 and monkey reference genomes rheMac8 and MacaM using Bowtie2. ATAC-seq reads were aligned using default parameters except -X 2000 -m 1. PCR duplicates were removed using Picard Tools. To adjust for fragment size, we aligned all reads as + strands offset by +4 bp and – strands offset by −5 bp (Buenrostro et al., 2013). For all ATAC-seq datasets except the pre-implantation embryo data, the THSSs (Tn5 hypersensitive sites) and mono-nucleosome fractions were separated by choosing fragments 50-115 bp and 180-247 bp in length, respectively. For the pre-implantation embryo ATAC-seq data, THSS and mono-nucleosome fractions were separated by choosing fragments 50-125 bp and 171-256 bp in length, respectively. Mono-nucleosome reads were analyzed using DANPOS2 (Chen et al., 2013). MACS2 was used for peak calling for THSSs (Liu, 2014).

ChIP-Seq data processing

All reads were mapped to unique genomic regions using Bowtie2 (Langmead and Salzberg, 2012) and the mm9 genome. PCR duplicates were removed using Picard Tools. MACS2 was used to call peaks using default parameters with IgG ChIP-seq data as a control.

Transcription factor footprint analysis

To analyze the footprints of TFs in ATAC-seq data, motifs on a set of peaks were used as anchors for running dnase_average_profile.py scripts of the Wellington program in ATAC-seq mode. The footprint p values of all motifs on a set of peaks were derived using the wellington_footprints.py scripts of the Wellington program in ATAC-seq mode on read-normalized ATAC-seq THSSs (< 115 bp) fragments.

Analysis of publicly available SNP-trackable ATAC-seq, DNase-seq, and ChIP-seq of gametes and embryos

All data were aligned to the mm9 mouse reference genome using Bowtie2 using the–no-mixed –no-discordant flags. Additionally, the -X 2000 flag was used for the ATAC-seq and H3K4me3 ChIP-seq data. All data were aligned in paired-end mode except for DNase-seq, which were aligned in single-end mode. Uniquely mapped reads were extracted and converted to bam format for all downstream analyses. Uniquely mapped reads were inputted into the SNPsplit software in order to determine the parent-of-origin for the data. SNP tables for the crosses for each dataset were downloaded from the Sanger Institute mouse genome project. DNase-seq data were not split according to SNPs.

Analysis of publicly available DNA methylation data

Publicly available bed files containing counts of CpG methylation versus unmethlyated CpGs were separated by parent-of-origin according to SNPs based on methylC-seq data (Wang et al., 2014). Average methylation values were calculated as a weighted average, with the weight at each CpG being equal to the number of reads covering that CpG.

Normalization of SNP-trackable sequencing data

After separating reads by strain according to SNPs, regions were normalized with a modified form of either RPKM or FPKM. The typical definitions for RPKM/FPKM were used, except instead of dividing by the total number of reads for each sample, we divided by the total number of SNP-containing reads. For example, if N_P were the number of SNP-containing reads in the paternal allele, and N_M were the corresponding number of reads on the maternal allele, the number of reads (or fragments) per kilobase in a region of interest were divided by N_P+N_M in millions of reads. For visualization of tracks in IGV, reads were normalized by RPM using the same N_P+N_M factor. This normalization was used for all strain-specific analyses using SNP reads coming from the same sample.

Hi-C data processing

Paired-end reads from Hi-C experiments were aligned to the mouse mm9 reference genome using Juicer (Rao et al., 2014). After PCR duplicates and low-quality reads were removed, high-quality reads were assigned to DpnII restriction fragments, and Hi-C interaction contacts where mapped in a binned matrix to create a hic file. A derivative of the HiCCUPS method (Cubeñas-Potts et al., 2017; Rao et al., 2014) was used to call significant peaks in the Hi-C interaction matrix. All statistically significant peaks were post-filtered for observed values greater than 12 Hi-C contacts, observed over expected values greater than 2 Hi-C contacts, and interactions that

are less than 5 Mb apart. The WashU Epigenome browser was used to obtain arc views of significant loops. FitHi-C (Ay et al., 2014) was used to call significant interaction at 25 kb resolution from the second pass with a q-value threshold of q > 0.001.

DATA AND SOFTWARE AVAILABILITY

ChIP-seq, ATAC-seq, and Hi-C data are available from NCBI's Gene Expression Omnibus (GEO). The accession number for all the datasets reported in this paper is GEO: GSE116857. Reviewers can access these data using token whqtysowzriddkp. Custom scripts were used to separate ATAC-seq reads into subnucleosomal and nucleosome-size ranges; to obtain the line plots shown in Figures 5 and S5; and to obtain metaplots of Hi-C data shown in Figures 7 and S7. These scripts are available without restrictions upon request.