Revisiting chromatin packaging in mouse sperm

Qiangzong Yin, Chih-Hsiang Yang, Olga S. Strelkova, et al.

Genome Res. published online December 21, 2023
Access the most recent version at doi:10.1101/gr.277845.123

Revisiting chromatin packaging in mouse sperm

Qiangzong Yin,[1] Chih-Hsiang Yang,[1] Olga S. Strelkova,[1] Jingyi Wu,[2,3,4] Yu Sun,[5] Sneha Gopalan,[6] Liyan Yang,[7] Job Dekker,[7,8] Thomas G. Fazzio,[6] Xin Zhiguo Li,[5] Johan Gibcus,[7] and Oliver J. Rando[1]

[1]*Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, Massachusetts 01605, USA;* [2]*Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA;* [3]*Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA;* [4]*Departments of Cell Biology and Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA;* [5]*Center for RNA Biology: From Genome to Therapeutics, Department of Biochemistry and Biophysics, Department of Urology, University of Rochester Medical Center, Rochester, New York 14642, USA;* [6]*Department of Molecular Cellular and Cancer Biology,* [7]*Department of Systems Biology, University of Massachusetts Chan Medical School, Worcester, Massachusetts 01605, USA;* [8]*Howard Hughes Medical Institute, University of Massachusetts Chan Medical School, Worcester, Massachusetts 01605, USA*

Mammalian sperm show an unusual and heavily compacted genomic packaging state. In addition to its role in organizing the compact and hydrodynamic sperm head, it has been proposed that sperm chromatin architecture helps to program gene expression in the early embryo. Scores of genome-wide surveys in sperm have reported patterns of chromatin accessibility, nucleosome localization, histone modification, and chromosome folding. Here, we revisit these studies in light of recent reports that sperm obtained from the mouse epididymis are contaminated with low levels of cell-free chromatin. In the absence of proper sperm lysis, we readily recapitulate multiple prominent genome-wide surveys of sperm chromatin, suggesting that these profiles primarily reflect contaminating cell-free chromatin. Removal of cell-free DNA, and appropriate lysis conditions, are together required to reveal a sperm chromatin state distinct from most previous reports. Using ATAC-seq to explore relatively accessible genomic loci, we identify a landscape of open loci associated with early development and transcriptional control. Histone modification and chromosome folding profiles also strongly support the hypothesis that prior studies suffer from contamination, but technical challenges associated with reliably preserving the architecture of the compacted sperm head prevent us from confidently assaying true localization patterns for these epigenetic marks. Together, our studies show that our knowledge of chromosome packaging in mammalian sperm remains largely incomplete, and motivate future efforts to more accurately characterize genome organization in mature sperm.

[Supplemental material is available for this article.]

The packaging of an organism's genome into the limiting confines of a typical nucleus requires multiple levels of organization, allowing relatively free access to gene regulatory elements in the face of extensive physical compaction. In eukaryotes, the key principles of chromatin organization have been elucidated over the past half-century: A repeating nucleoprotein complex known as the nucleosome serves to compact ~150 bp of DNA, resulting in the famous "beads on a string" polymer that is then further organized via loop extrusion and phase separation into larger-scale structures such as topologically associating domains (TADs) and genome "compartments" (Dekker and Misteli 2015; Friedman and Rando 2015; Klemm et al. 2019).

Although we have a relatively mature understanding of chromatin organization in typical somatic cell types, genomic organization in the highly compact sperm nucleus remains less well understood. At the most basic level, it has been known for decades that the vast majority of histones are removed from the genome during the process of spermatogenesis in mammals to be replaced with small highly basic peptides known as protamines (Bellvé et al. 1975; Calvin 1976; Balhorn et al. 1977; Balhorn 1982). Intriguingly, a small subset of nucleosomes is retained in mature sperm, between ~2% and ~15%, depending on the mammal (Gaucher et al. 2010; Gaspa-Toneu and Peters 2023). A number of studies have mapped the locations of retained nucleosomes, with the majority of reports in humans and mice finding nucleosomes enriched at CpG island regulatory elements associated with key developmental genes (Arpanahi et al. 2009; Hammoud et al. 2009; Brykczynska et al. 2010; Erkek et al. 2013; Johnson et al. 2016; Jung et al. 2017, 2019), although a handful of studies instead find nucleosomes retained primarily in long stretches over "gene deserts" of low gene density (Carone et al. 2014; Yamaguchi et al. 2018). It appears at least one explanation for this discrepancy may be the extent of nuclease digestion used in mapping nucleosomes in sperm, although many other technical details, including fixation and "preswelling" steps before chromatin preparation, also differ between studies. Other features of sperm genome compaction that have been investigated in detail include covalent modifications of histones (Hammoud et al. 2009; Brykczynska et al. 2010; Erkek et al. 2013; Jung et al. 2017, 2019; Lismer et al. 2020; Yoshida et al. 2020; Lismer et al. 2021; Bedi et al. 2022), localization of various non-histone proteins

Corresponding author: Oliver.Rando@umassmed.edu
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.277845.123.

(Jung et al. 2017, 2019), and three-dimensional folding of the sperm genome (Battulin et al. 2015; Jung et al. 2017, 2019; Ke et al. 2017; Alavattam et al. 2019; Vara et al. 2019; Wang et al. 2019b; Gou et al. 2020). In addition to the inherent interest in understanding how the genome folds under conditions of extreme compaction, it has been suggested that the organization of the sperm genome—the locations and modifications of retained histones, most notably—may also play a role in programming early gene expression in the preimplantation embryo and may thereby influence later phenotypes in offspring (Siklenka et al. 2015; Lesch et al. 2019).

In the mouse system, sperm are typically collected from the distal, or cauda, region of the epididymis, an epithelial tube where sperm mature for ~10 days following testicular spermatogenesis. Several groups have recently reported the presence of abundant cell-free chromatin in the proximal, or caput, epididymis (Galan et al. 2021; Chen et al. 2022), where it results in ~20%–30% contamination of sperm cytosine methylation profiles. Although both of these studies focused on the abundant contamination found in the caput epididymis, there is also a small amount of contaminating chromatin present in cauda sperm preparations. We typically observe ~2%–5% contamination of these sperm preps; the extent of contamination differs somewhat from investigator to investigator. Importantly, ~2% contamination of cytosine methylation data is comparable to the precision of a typical cytosine methylation survey and is therefore essentially insignificant for most purposes. However, we reasoned that this contamination could be a much greater problem for measurements of sperm chromatin, as 2% genome equivalents of a fully nucleosomal genome would compete with and potentially overwhelm the signal from the bona fide nucleosomal complement of mouse sperm.

Here, we explored the hypothesis that many or most published studies of mouse sperm chromatin organization reflect contaminating cell-free chromatin. Overall, our data strongly imply that current views of the mouse sperm chromatin landscape are flawed, and motivate renewed focus on this important area of chromosome biology and epigenetics.

Results

Probing sperm chromatin architecture by ATAC-seq

In the course of studies focused on dietary effects on sperm chromatin, we performed assay for transposase-accessible chromatin by sequencing (ATAC-seq) (Buenrostro et al. 2015) to identify relatively accessible genomic loci across the sperm genome. Here and through the rest of this paper, we isolate sperm from the cauda epididymis via tissue dissection and sperm swim out, and we then wash sperm with a somatic cell lysis buffer. After washing, purified cauda sperm show no detectable (<0.1%) contamination by intact somatic cells (Supplemental Fig. S1A, left panel). Although our original goal was to exactly reproduce the conditions used by Jung et al. (2017), we noticed that the investigators in this study did not add any reducing agent to properly lyse sperm. This was of concern as sperm are subject to extensive disulfide cross-linking during maturation, and failure to reverse these cross-links via treatment with a reducing agent (DTT, β-ME, or TCEP) results in significant difficulties in recovering genomic DNA from sperm. This is illustrated in Supplemental Figure S1, B and C; even after overnight SDS and Proteinase K treatment, scant genomic DNA (~5% of the levels recovered from DTT-treated sperm) is recovered in the absence of reducing agent. Moreover, microscopic examination of untreated sperm revealed abundant intact sperm heads even after extensive incubation in detergent and Proteinase K, whereas DTT-treated samples were entirely lysed under the same conditions (Supplemental Fig. S1A, right panels). Taken together, these findings suggest the concerning possibility that many assays of sperm chromatin architecture, including but not limited to those using inadequate levels (Supplemental Fig. S1C) of reducing agent to allow access to the sperm genome, could be contaminated by cell-free chromatin.

To test the hypothesis that published assays of sperm chromatin architecture might be contaminated by cell-free chromatin, we set out to characterize the impact of several treatments on sperm ATAC-seq profiles. Four conditions were considered. First, purified and washed cauda sperm were subjected to the exact same ATAC-seq protocol reported by Jung et al. (2017) and used in several additional studies (Jung et al. 2017, 2019; Gou et al. 2020): We refer to this as the "untreated" case; given that no reducing agent was used to help permeabilize sperm, we predict that the resulting ATAC-seq profiles should primarily—indeed, can only—reflect the organization of cell-free chromatin. Second, we pretreat sperm with DNase I to eliminate any cell-free chromatin before probing with Tn5; we predict that little to no genomic DNA should be available for transposition after removing cell-free chromatin but without properly opening sperm. Third, we add appropriate levels (50 mM) of DTT before Tn5 treatment to allow access to the sperm genome; under these conditions, we anticipate a mixed profile resulting from both cell-free chromatin and the bona fide sperm genome. Finally, by pretreating sperm with DNase I and then adding DTT before Tn5 probing, we hope to reveal the landscape of accessible chromatin in mature spermatozoa without any confounding contamination. Importantly, under these conditions we find no evidence that residual DNase activity subsequently affects sperm genomic DNA (Supplemental Fig. S1D).

Sperm ATAC-seq profiles are sensitive to cell-free chromatin contamination

Our initial efforts to probe sperm chromatin architecture were performed using the same "untreated" ATAC-seq conditions reported by Jung et al. (2017). Under these conditions, we robustly reproduce the ATAC-seq landscapes detailed in several prior reports (Figs. 1A, 2; Supplemental Fig. S2A–C), as expected. Analysis of library insert lengths revealed the typical pattern observed in somatic cell ATAC-seq libraries, with a ~75-bp peak (after accounting for sequencing adaptor lengths) corresponding to paired Tn5 insertions into open regulatory elements, followed by a series of peaks every ~150 bp corresponding to insertions flanking mono/di/trinucleosomes (Fig. 1B; Supplemental Fig. S2D).

Turning next to sperm samples pretreated with DNase I to remove cell-free chromatin before Tn5 probing, we generally failed to recover enough transposed DNA to build a sequencing library. When we did obtain sequencing libraries, we recovered very low yields, with <10% of the levels of amplified DNA that could be recovered in the untreated condition (Supplemental Fig. S2E). Analysis of insert lengths revealed a complete loss of the nucleosome signature (Fig. 1B), and sequencing profiles from these libraries were almost entirely flat; almost all the ATAC-seq peaks recovered in untreated sperm were lost after DNase treatment (Fig. 2A,B; Supplemental Fig. S2F,G). These findings are consistent with the hypothesis that published ATAC-seq profiles generated from untreated sperm are most likely contaminated by cell-free chromatin.

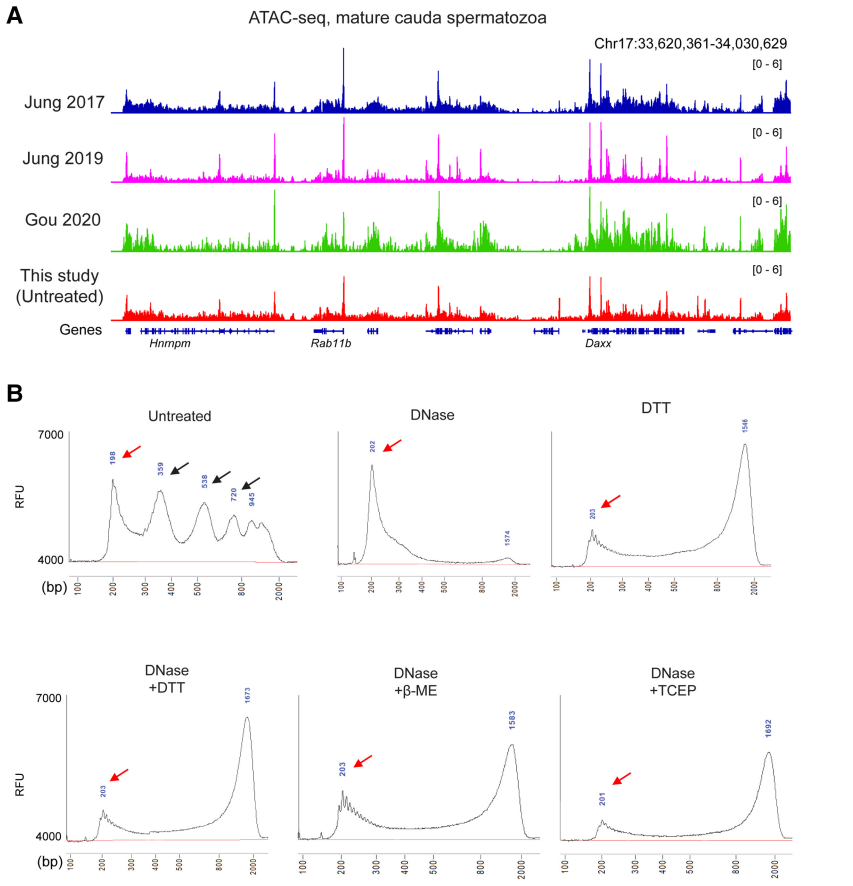

Figure 1. ATAC-seq profiles from sperm following DNase and DTT treatment. (*A*) Genome browser tracks for ATAC-seq data from untreated sperm (this study) and from three published ATAC-seq data sets for sperm (Jung et al. 2017, 2019; Gou et al. 2020). (*B*) Bioanalyzer traces of ATAC-seq sequencing libraries prepared from sperm samples pretreated with the indicated conditions before Tn5 transposition. Red arrows show ~75-bp inserts (after accounting for sequencing adaptors) corresponding to open chromatin regions, and black arrows in "untreated" libraries show insert lengths corresponding to mono-, di-, and trinucleosome length inserts. Notable in all the DTT-treated libraries is a peak at ~1.3 kb. This peak reflects a landscape of random insertions across the sperm genome, with the size of the insert being an artifact of the PCR conditions used in library preparation. We confirmed this by building a second set of ATAC-seq libraries in which the PCR extension time was extended to 2 min compared with 1 min in the libraries shown. Under these conditions, we find a ~2.5-kb peak, confirming that PCR extension time is responsible for this peak location. (RFU) Relative fluorescence units.

Eliminating cell-free chromatin reveals bona fide open chromatin profiles from mature sperm

We next turn to ATAC-seq libraries prepared in the presence of the high levels of reducing agent required to enable access to Tn5 and other protein probes of chromatin architecture. We focus first on sperm preparations treated with both DNase and DTT. This pretreatment should recover bona fide features of sperm chromatin, because it depletes cell-free chromatin before permeabilizing the sperm head. Analysis of library insert sizes revealed a dominant peak of ~75-bp inserts (Fig. 1B), corresponding to short accessible loci, but no notable signature of nucleosome-sized footprints. This reflects a relatively flat landscape of Tn5 insertions across the primarily protamine-packaged genome, with any ~150-bp nucleosomal footprints expected to result from low level of nucleosome retention (~2%) in murine sperm being overwhelmed by this global background (albeit with potential implications for nucleosome positioning; see Discussion). Importantly, we find similar ATAC-seq insert lengths in DNase-treated sperm that were permeabilized using other reducing agents such as β-ME or TCEP (Fig. 1B), further supporting our hypothesis that prior ATAC-seq studies in murine sperm suffered from a failure to properly reverse disulfide cross-links.

Comparing the DNase/DTT-treated ATAC-seq landscape with the untreated landscape, we found that the majority of "untreated" ATAC peaks were lost, and only a small fraction—associated with very high CpG–density regulatory elements—were retained in DNase/DTT-treated sperm (Fig. 2A,B). Systematic peak calling identified 10,741 accessible loci in DNase/DTT-treated sperm samples compared with 15,680 accessible loci in untreated samples; 2148 peaks were shared between the two data sets at regulatory regions of unusually high CpG density. As in other cell types, ATAC peaks in sperm were overrepresented at promoters (Fig. 2C). However, the promoters that were uniquely identified in DNase-treated sperm were associated with reduced CpG density compared with the much higher CpG density found at the ATAC peaks from cell-free chromatin (Fig. 2B). Interestingly, we found no evidence for well-positioned 150-bp ATAC-seq fragments (often considered diagnostic of nucleosome localization) surrounding accessible loci (Fig. 2C). This argues against a standard organization with accessible regulatory elements flanked by a pair of positioned nucleosomes; instead, we observe a modest and somewhat fuzzy/delocalized enrichment of 150-bp footprints across promoters (see Discussion). In terms of biological functions, accessible peaks in the DNase + DTT data set were enriched near genes involved in transcriptional control and early development, in contrast to various cell cycle and RNA splicing annotations specific to untreated sperm peaks (Supplemental Fig. S3).

Finally, we consider ATAC-seq libraries produced from DTT-only-treated sperm. Consistent with these samples comprising a mixture of cell-free chromatin and bona fide sperm chromatin, we found ATAC signal over both the peaks unique to untreated sperm and the DNase + DTT peaks (Fig. 2). This is true globally, with DTT-treated ATAC-seq landscapes showing a mix of both cell-free accessibility peaks and bona fide sperm accessibility peaks (Fig. 2B). Moreover, metagene plots of open chromatin peaks and nucleosome signatures over transcription start sites revealed that DTT-only samples showed hybrid profiles between the untreated and DNase + DTT profiles (Fig. 2C), again supporting the idea that these libraries represent a mixture of cell-free chromatin and bona fide sperm.

DNase and DTT do not affect open chromatin in mESCs

Our findings in mature sperm support the hypothesis that prior ATAC-seq studies of the sperm genome primarily reported on the status of cell-free chromatin, because this contamination is the only chromatin available in the absence of adequate reducing agent. However, we also considered the hypothesis that the DNase and DTT treatments might somehow drive artifactual changes to global chromatin architecture. To test this hypothesis, we performed ATAC-seq in a commonly studied cell type—mESCs—treated with the same conditions used for sperm preparations. Our ATAC-seq profiles from untreated mESCs broadly recapitulated known features of the ESC chromatin landscape (Fig. 3A,B), with widespread peaks of accessible chromatin, flanked by well-positioned nucleosomes, associated with promoters and enhancers (Fig. 3C). This landscape of open chromatin was almost entirely unaffected by DNase, DTT, or DNase + DTT treatments (Fig. 3B,C), confirming that these treatments do not grossly alter Tn5 transposition or chromatin organization in cell lines.

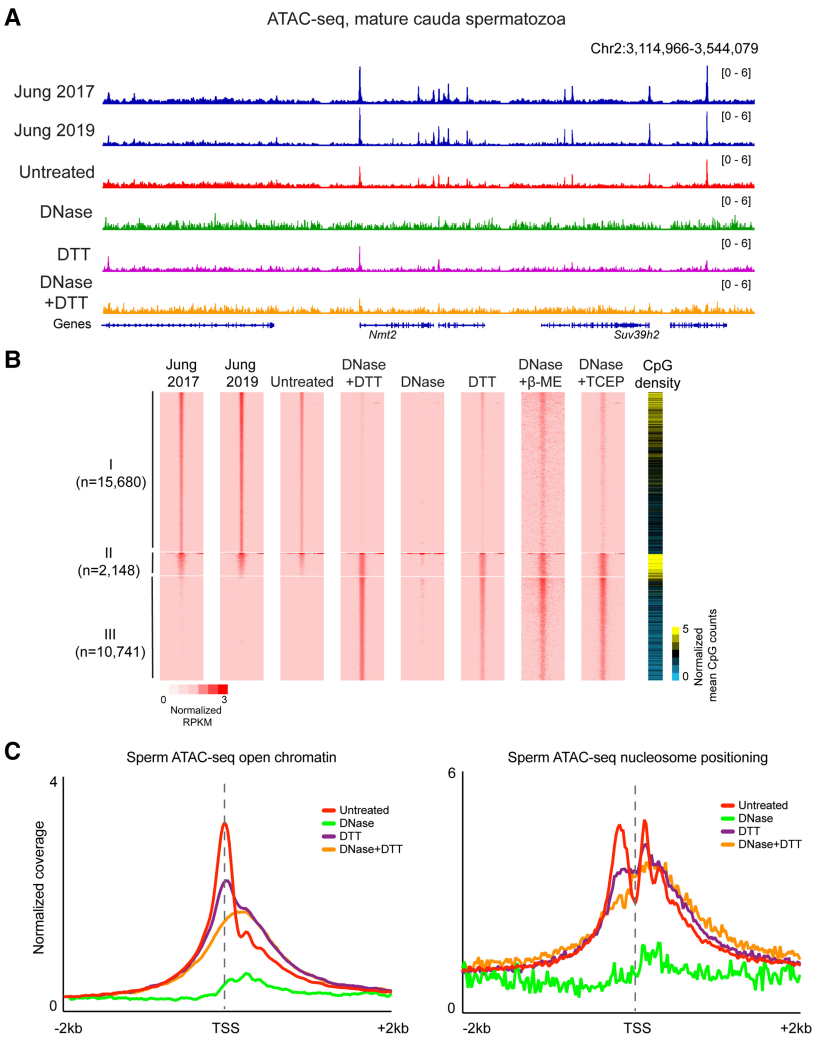

Figure 2. Published ATAC-seq profiles in sperm are dominated by cell-free chromatin contamination. (*A*) ATAC-seq browser tracks for sperm subject to the indicated conditions. (*B*) Heatmaps for the four indicated libraries, sorted into three categories: peaks specific to the untreated data set (*I*; n = 13,605), shared peaks (*II*; n = 2148), and peaks specific to DNase/DTT data sets (*III*; n = 8627). (*C*) Metagenes for <120-bp inserts, diagnostic of open chromatin, or for >150-bp inserts, diagnostic of nucleosome footprints, aligned over all annotated TSSs.

Contaminating chromatin likely derives from the epididymal epithelium

What is the origin of the cell-free chromatin that contaminates cauda epididymal sperm? Based on our previous efforts (Galan et al. 2021) focused on *caput* epididymal sperm, in which the contaminating DNA skewed imprinting control regions (ICRs) from a germ cell profile (e.g., 0% or 100% methylated at ICRs) toward a somatic methylation profile (50% methylation), we favor the hypothesis that contaminating DNA arises from somatic cells rather than cells of the germline lineage.

To test this hypothesis in cauda (rather than caput) epididymal sperm, we turned to NOMe-seq (Kelly et al. 2012), a protocol that leverages the bacterial M.CviPI methyltransferase to methylate accessible cytosines in GpC dinucleotides (rather than the endogenous CpG dinucleotide context). Importantly, the 5mC readout here enables simultaneous analysis of genomic accessibility at GpC dinucleotides, as well as endogenous cytosine methylation at CpG dinucleotides. We assayed cytosine methylation directly by Oxford Nanopore sequencing to avoid bisulfite-related fragmentation of the genome, which would prevent long-distance linkage between methylase accessibility and endogenous methylation on the same DNA molecule (Wang et al. 2019a; Shipony et al. 2020). This approach, known as Nano-NOMe-seq (Lee et al. 2020; Battaglia et al. 2022), provides simultaneous readouts of induced GC methylation and endogenous CG methylation on single DNA molecules with read lengths of tens of kilobases. This allows us to assess the developmental origin of cell-free DNA—soma versus germline—based on the endogenous CpG methylation program at ICRs.

We subjected untreated (e.g., no DNase or DTT) cauda epididymal sperm to M.CviPI methylation, then isolated genomic DNA using DTT treatment to allow recovery of both bona fide sperm DNA along with any cell-free contamination, and assayed cytosine methylation genome-wide by Nanopore sequencing. This initial untargeted analysis covered the genome at about 2–3× depth, revealing that ~5%–10% of sequenced GpC dinucleotides were methylated, consistent with our prediction that only cell-free DNA would be available for M.CviPI methylation. However, despite the good length of the reads (median, ~18–25 kb) in this initial data set, relatively few reads included the small number of ICRs for which we could easily distinguish somatic from germline-derived genomes.

We therefore used CRISPR-Cas9 to target our Nanopore sequencing to more deeply sequence several ~25- to 30-kb regions

surrounding well-characterized ICRs (Xie et al. 2012; Gilpatrick et al. 2020). The resulting data set was far deeper, with several hundred reads spanning each of the targeted ICRs (Fig. 4A,B). To assess the lineage of origin for cell-free DNA, we first separated reads according to the extent of methylation at GC dinucleotides to separately analyze methylase-accessible reads (with average GpC methylation ≥20%; reporting on cell-free DNA with a total of 26,729 reads) and methylase-protected reads (with average GpC methylation < 20%; reporting on sperm with a total of 341,117 reads). The estimated cell-free DNA contamination was thus 7.8% here (Supplemental Fig. S4). As expected of methylation in the germline, ICRs in sperm, defined based on their protection from M.CviPI methylation, were either fully CpG-methylated (paternally imprinted) or fully CpG-unmethylated (maternally imprinted) (Fig. 4A,B). In contrast, accessible (GC-methylated) genomic reads showed mixed CpG methylation profiles (~50%) at ICRs (Fig. 4A,B), confirming our prediction that cell-free DNA is derived from somatic cells rather than sperm. Together, these data provide independent validation of our hypothesis that protein probes (Tn5, M.CviPI, antibodies, MNase, etc.) cannot access the genome of sperm that has not been permeabilized by DTT treatment, and support our prediction that accessible cell-free DNA/chromatin is somatic in origin.

Beyond this, our methylation-based study cannot more precisely define the cell of origin for contaminating DNA in cauda epididymal sperm preparations, in part because to our knowledge there have been no methylation surveys of relevant reproductive support cell types (e.g., Sertoli cells in the testis, principal cells in the epididymis, etc.). That said, comparing ATAC-seq profiles from sperm isolated following relatively gentle versus relatively disruptive dissection protocols (single slice followed by swim out vs. mincing and squeezing) revealed more extensive contamination resulting from more disruptive dissection (Supplemental Fig. S2A,D). We therefore generated ATAC-seq libraries from cauda epididymis epithelium. These profiles broadly agreed with both prior ATAC-seq data sets (Jung et al. 2017, 2019; Gou et al. 2020), as well as our own data set, from untreated sperm (Fig. 4C; Supplemental Fig. S2B). Conversely, contaminated ATAC-seq data—our untreated sperm data or published data sets (Jung et al. 2017, 2019; Gou et al. 2020)—showed poor correlation with data generated from sperm precursors (spermatocytes and round spermatids) (Fig. 4C; Supplemental Fig. S2B; Maezawa et al. 2018). Taken together, these data definitively show that contamination of cauda epididymal sperm arises from somatic tissues, most likely epithelial cells of the epididymis.

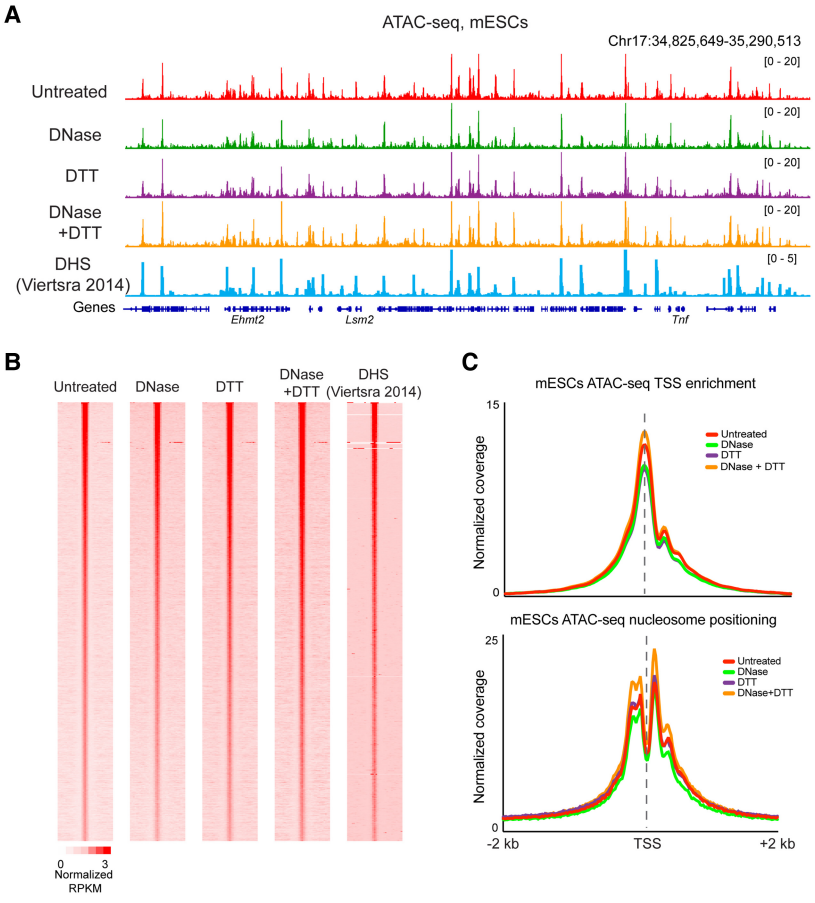

Figure 3. DNase and DTT treatment do not affect ATAC-seq profiles in mESCs. (*A*) ATAC-seq profiles for mESCs treated with the indicated conditions, as well as published ESC DHS data from Vierstra et al. (2014). (*B*) Heatmaps show ATAC read counts aligned over peaks (n = 58,130) called from untreated cells. (*C*) Metagenes for <120-bp inserts, and >150-bp inserts, as in Figure 2C.

Contamination of histone modification profiles by cell-free chromatin

Our ATAC-seq data provide strong evidence that epigenomic studies of untreated sperm report primarily on contaminating cell-free chromatin. This finding naturally raises questions about other published aspects of sperm chromatin architecture. Given that multiple studies have implicated sperm histone modifications in programming early gene expression in the preimplantation embryo (Hammoud et al. 2009; Brykczynska et al. 2010; Erkek et al. 2013; Jung et al. 2017, 2019; Lismer et al. 2020, 2021; Yoshida et al. 2020) and given that perturbing histone modifications during spermatogenesis has phenotypic effects in the next generation (Siklenka et al. 2015; Lesch et al. 2019), we set out to explore whether the histone modification landscape of sperm might also be affected by cell-free chromatin in the epididymis.

We initially performed CUT&RUN and CUT&Tag (Skene and Henikoff 2017; Kaya-Okur et al. 2019) to identify genomic loci associated with H3K4me3 or H3K27me3, histone modifications related to the Trithorax and Polycomb epigenetic memory systems (Schuettengruber et al. 2007) that have been mapped in sperm in multiple prior studies (Hammoud et al. 2009; Brykczynska et al. 2010; Erkek et al. 2013; Jung et al. 2017, 2019; Lismer et al. 2020, 2021). In both cases, we found that libraries

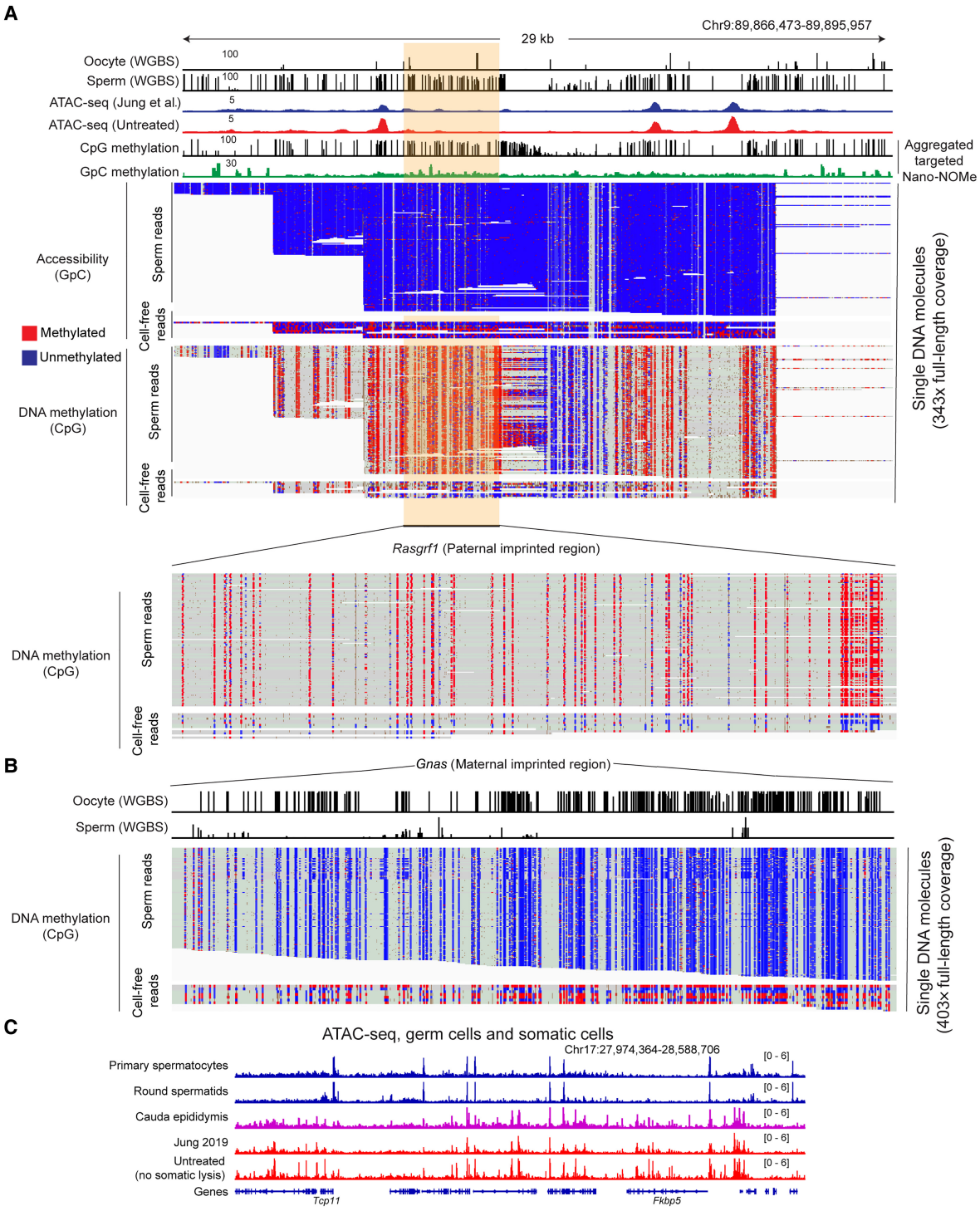

Figure 4. Contaminating cell-free DNA is of somatic origin. (*A*,*B*) Nano-NOMe-seq data for two imprinting control regions in untreated sperm. Untreated sperm were subject to M.CviPI-driven GpC methylation; then genomic DNA was extracted after DTT treatment to obtain DNA from both contaminating material as well as sperm; and resulting DNA was subject to long-read sequencing by Oxford Nanopore. Resulting methylation calls (red and blue represent methylated and unmethylated cytosine, respectively) are shown separately for CpG methylation and GpC methylation (analysis was restricted to HCG and GCH, where H represents A/C/T, to avoid ambiguous GCG methylation), as indicated. In addition, reads are separated based on the extent of GpC methylation. The majority of reads show low (5.5 ± 3.4%) GpC methylation, whereas a small fraction of reads show >20% GpC methylation and represent accessible DNA molecules in untreated sperm preparations. Importantly, DNA molecules protected from M.CviPI, arising from the bona fide sperm genome, show the 0% or 100% methylation expected at imprinting control regions in germline samples. In contrast, the small fraction of GC-methylated reads, reflecting accessible cell-free chromatin, includes a mix of methylated and unmethylated ICRs consistent with a somatic origin for these DNA molecules. (*C*) ATAC-seq data for the indicated samples, including ATAC-seq for cauda epididymal epithelium. Key here is the strong agreement between ATAC profiles for untreated sperm (representing contaminating chromatin) and cauda epididymal epithelium (see also Supplemental Fig. S2B).

prepared from untreated sperm recapitulated modification landscapes previously obtained using ChIP-seq (Supplemental Fig. S5; Erkek et al. 2013; Jung et al. 2017), but these profiles were completely lost following DNase treatment. This again argues that prior histone modification maps were likely contaminated by cell-free chromatin. Curiously, we were unable to elicit antibody-specific CUT&RUN or CUT&Tag profiles from DTT-treated sperm (with or without DNase pretreatment) (Supplemental Fig. S5); we have thus far been unable to define conditions that yield antibody-specific profiles for these protocols.

We obtained similar results using ChIP-seq (Johnson et al. 2007) on formaldehyde-fixed sperm chromatin sheared via sonication (Methods) (Fig. 5; Supplemental Fig. S6). We focused on a sequence-specific DNA-binding protein, CTCF, and the histone modifications H3K4me3 and H3K27me3 (K4/27), as these epitopes were expected to show distinctive localization profiles. For all three epitopes, ChIP-seq profiles from DTT-treated sperm samples, in which we expected to see a mixture of contaminating chromatin and bona fide sperm, yielded profiles consistent with prior mapping studies (Fig. 5A; Erkek et al. 2013; Jung et al. 2017). Peaks from the CTCF data set were strongly enriched for the CTCF binding motif, whereas K4/27 profiles were associated with promoters and repressed developmental regulators, respectively (Fig. 5B,C; Supplemental Fig. S6A).

However, all three ChIP-seq profiles were substantially altered in maps generated from DNase + DTT-treated sperm (Fig. 5A). In the case of CTCF, we obtained a flat landscape, and those peaks that were called from this data set were not enriched for the CTCF motif (Fig. 5B; Supplemental Fig. S6B). This flat landscape suggests either that CTCF is not in fact associated with the mature sperm genome or that our ChIP conditions, whether incomplete fragmentation and solubilization of the sperm genome (Supplemental Fig. S6C) or fixation conditions that impact CTCF epitopes needed for immunoprecipitation, precluded accurate identification of CTCF localization. Similarly, DNase treatment also altered the localization landscape of K4/27. The localization of these histone marks in DNase + DTT-treated sperm was distinct from the CTCF landscape; ChIP-seq in bona fide sperm chromatin therefore showed somewhat antibody-specific profiles, unlike CUT&RUN or CUT&Tag (Supplemental Fig. S5). Nonetheless, all three ChIP data sets revealed relatively flat landscapes, quite distinct from landscapes previously reported and showing little to no enrichment for expected features such as the CTCF DNA binding motif (Supplemental Fig. S6B). As a fourth approach to exploring the histone modification landscape in sperm, we performed native MNase ChIP-seq (Liu et al. 2005; Erkek et al. 2013; Jung et al. 2017) for K4/27 methylation (Supplemental Fig. S6D). Here, we again recover published K4/27 maps when using untreated sperm; however, for this assay, we observe no substantial changes following DNase pretreatment, with or without DTT-dependent permeabilization (see Discussion).

Altogether, we successfully recover prior K4/27 landscapes from untreated sperm using four distinct assays: CUT&RUN, CUT&Tag, sonicated/fixed ChIP-seq, and native MNase ChIP-seq. However, each of these assays shows distinct behavior following DNase or DTT treatment; together, these findings raise substantial concerns regarding affinity-based chromatin analysis in bona fide sperm and motivate continued optimization to allow more efficient recovery of proteins of interest.

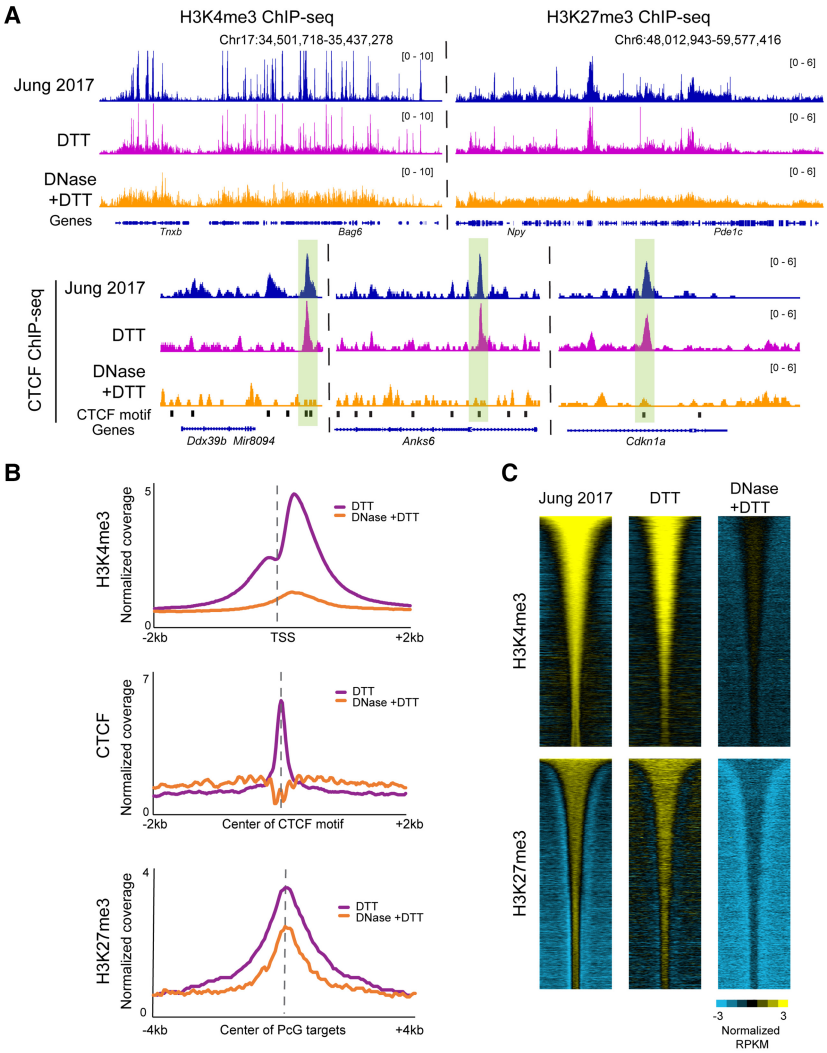

Figure 5. Histone modification profiling in DNase-treated sperm reveals a loss of specific signal. (*A*) Browser tracks for H3K4me3, H3K27me3, and CTCF ChIP-seq in sperm. In all three cases, the *top* panels show data from published sperm data sets (Jung et al. 2017), along with our data from DTT-treated and DNase + DTT-treated sperm underneath. (*B*) Metagenes aligned over the relevant peak locations (TSS, CTCF motif, and Polycomb-group [PcG] targets), showing H3K4me3, H3K27me3, and CTCF enrichment in our DTT-only and DTT + DNase data sets. (*C*) Heatmaps show H3K4me3 and K27me3 enrichment over all peaks from Jung et al. (2017), with data shown for Jung et al. alongside our DTT and DTT + DNase data sets.

Contamination affects published chromosome folding maps from mature sperm

We finally turn to the question of how the sperm genome is folded in three-dimensional space, using the gold-standard molecular approach to chromosome folding, Hi-C (Lafontaine et al. 2021). A number of studies have generated Hi-C maps for mature murine spermatozoa (Battulin et al. 2015; Jung et al. 2017; Ke et al. 2017; Jung et al. 2019; Vara et al. 2019; Wang et al. 2019b), with the majority of these studies (but see Vara et al. 2019; discussed in the Supplemental Fig. S7 legend) reporting maps with typical features of somatic cells: A and B compartments corresponding to euchromatin and heterochromatin, and TADs with flares and loop anchors characteristic of loop extrusion processes. These maps are quite surprising given the compaction required to fit the genome into the small sperm nucleus, as well as the unusual physicochemical properties of the major packaging proteins, the protamines, that organize the sperm genome (Balhorn 1982, 2007). Notably, a review of methods sections of published sperm Hi-C papers revealed that the majority of these efforts neglected to add DTT or other reducing agents required to access the sperm genome (Supplemental Fig. S1B,C) and are thus likely to assay only cell-free or other contaminating somatic cell chromatin architecture.

To test this hypothesis, we generated Hi-C maps for spermatozoa using DNase and/or DTT as in the various experiments detailed above. Figure 6 shows Hi-C contact maps for data from Jung et al. (2017) alongside our DTT-only and DNase + DTT libraries. Data from Jung et al. reveal the typical "checkerboard" pattern of A/B compartment organization at megabase scales (top panels), along with boxes on the diagonal with dots at the corners diagnostic of loop extrusion-mediated organization of gene-scale domains into TADs (middle panels). In our DNase + DTT samples, all of these features are lost, with the exception of a faint signal of compartment organization (Fig. 6, bottom row). Importantly, the absence of TADs seen here has previously been reported in Hi-C studies of human sperm (Chen et al. 2019), as well as in Hi-C studies in both human and *Xenopus* sperm in the accompanying paper (Jessberger et al. 2023). We note that although this absence of specific organization could plausibly arise from certain types of chromosome packaging behavior, a more likely possibility for the limited structure seen in sperm Hi-C is that protamines are inefficiently cross-linked by formaldehyde, and standard Hi-C cross-linking conditions therefore fail to capture those chromosomal contacts that do occur in the sperm nucleus (see Discussion).

Finally, as seen with our ATAC-seq data sets, the DTT-only samples show behavior consistent with a mixture between the featureless diagonal of the DNase + DTT sample with a small contribution from the contaminating signal seen in untreated sperm libraries. This again fits our predictions and further supports our contention that even using DTT to break sperm nonetheless allows for contamination by cell-free chromatin if it has not been eliminated before assay.

Discussion

Here, we revisit the unusual genomic packaging state in mature mouse spermatozoa. Most importantly, we find that a wide variety of published analyses of sperm chromatin compaction are likely to have been contaminated by cell-free chromatin. We show that DNase I treatment of sperm before lysis removes this contaminating material, and further show that inadequate use of reducing agents in many studies precludes analysis of bona fide sperm chromatin by commonly used methods, including ATAC-seq, CUT&RUN, CUT&Tag, ChIP-seq, and Hi-C. Together, our findings force a reappraisal of the packaging of the haploid genome in the compact sperm nucleus and raise new questions about the potential role for sperm chromatin as a carrier of epigenetic information from fathers to offspring.

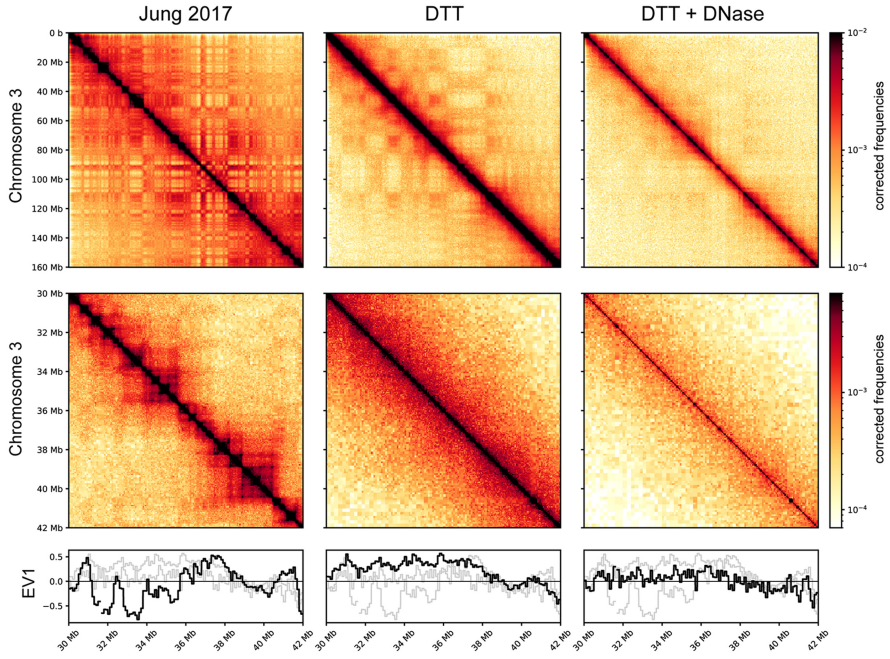

Figure 6. Absence of compartment signals and TADs in bona fide sperm Hi-C maps. Hi-C contact maps for (from *left* to *right*) untreated sperm (data from Jung et al. 2017), DTT-treated sperm, and DNase- and DTT-treated sperm. *Top* panels show a zoomed-out genome view covering Chromosome 3, *middle* panels show a typical ~10-MB zoom-in (Chr 3:30–42 MB). *Bottom* panels show *A*/*B* compartment calls (EV1: eigenvector 1) for the zoomed-in region.

Sperm chromatin assays are contaminated by cell-free chromatin

We previously identified cell-free chromatin as a major (~20%–30%) contaminant of immature sperm from the caput epididymis (Galan et al. 2021; Chen et al. 2022) based on cytosine methylation data. Definitive demonstration that unexpected cytosine methylation patterns resulted from cell-free DNA was obtained by showing that pretreating caput sperm with DNase was sufficient to restore cytosine methylation to the levels seen in other germ cell populations. Although we did not highlight potential contamination of cauda sperm, even in this population, we found that DNase treatment resulted in a ~2%–5% reduction in methylation at sites that are hypermethylated in cell-free chromatin compared with sperm (Galan et al. 2021). Further evidence for cell-free

DNA contamination in cauda sperm can be seen in attempts to isolate sperm genomic DNA with and without DTT (Supplemental Fig. S1B,C); in the absence of DTT, a small amount of DNA can be obtained, representing ~5% of the DNA obtained when sperm are properly broken using reducing agents.

The origin of this cell-free chromatin is unclear; in the *caput* epididymis, it is associated with citrullinated histones (Galan et al. 2021) and thus may arise in situ from NETosis (Sollberger et al. 2018) or a similar programmed cell death process. It is worth noting that cell-free chromatin in the caput epididymis copurifies with sperm isolated by FACS (Chen et al. 2022); it is therefore clear that in the absence of enzymatic removal, somatic DNA is likely to contaminate murine sperm preparations no matter how extensive the purification. As seen for caput epididymal sperm, contamination of *cauda* epididymal sperm is also clearly derived from somatic cells (Fig. 4A,B), rather than a small population of easily lysed or otherwise defective sperm populations. Importantly, contamination in the cauda epididymis varies depending on the extent of tissue disruption during dissection (Methods) (Supplemental Fig. S2A,D) and thus could partly or entirely arise during this process. Consistent with this, ATAC-seq profiles from untreated sperm are best correlated with ATAC data generated from the epididymal epithelium (Fig. 4C; Supplemental Fig. S2B). Nonetheless, we still observed artifactual ATAC-seq signal even in sperm obtained using the gentlest dissections (Supplemental Fig. S2A,D), indicating that contaminating material, whether generated during the process of dissection or already present in the epididymal lumen, is present no matter how gentle the dissection.

This subtle contamination is of little importance to studies of cytosine methylation (assuming sperm are obtained from a well-defined cauda epididymis dissection without substantial vas deferens), as extremely deep sequencing is required to obtain a precision >2% in whole-genome bisulfite sequencing. In contrast, studies of cauda sperm *chromatin* are particularly susceptible to this contamination owing to two idiosyncrasies of sperm biology. The first is the aforementioned requirement for a reducing agent for effective sperm lysis, driven by the extensive disulfide cross-linking that occurs during sperm maturation (Calvin and Bedford 1971; Saowaros and Panyim 1979). Thus, any studies of sperm chromatin performed in the absence of DTT or other reducing agents, including the majority of sperm Hi-C studies, as well as a substantial number of one-dimensional chromatin studies, can only possibly survey the landscape of contaminating chromatin. The second challenge, affecting even studies performed using DTT, is that the vast majority of nucleosomes are evicted during spermatogenesis, leaving a small population (~2% in mouse) of nucleosomes retained in mature spermatozoa. Thus, contamination by ~2%–5% cell-free chromatin, presumably completely nucleosome-associated or nearly so (Supplemental Fig. S2F), is sufficient to compete with or even overwhelm any true signal from sperm chromatin. This is of course even more problematic in scenarios in which the bona fide sperm genome cannot be assayed for other technical reasons, as may occur for Hi-C (see below).

Our ATAC-seq data precisely confirm our predictions based on the considerations above. In the absence of DNase or DTT, we robustly recapitulate prior ATAC-seq studies performed under these same conditions (Fig. 1). Our contention is that these accessibility peaks reflect cell-free chromatin, a hypothesis that is supported by the nearly complete loss of signal obtained when using sperm samples pretreated with DNase I or MNase. Perhaps most compelling are the results from samples treated with DTT only, for which we confirm the prediction that ATAC-seq libraries should reveal a superposition of the cell-free chromatin peaks and the bona fide sperm peaks identified in DNase + DTT samples. Altogether our data are completely and parsimoniously explained by our hypothesis that sperm chromatin assays generated from untreated sperm samples can only capture aspects of cell-free chromatin contamination.

An updated view of genomic accessibility in murine sperm

Considering our ATAC-seq data set that was generated from properly broken sperm, free of cell-free chromatin, we found thousands of peaks of genomic accessibility that were only observed in DTT-treated sperm but not in cell-free chromatin (Fig. 2B). These peaks were enriched at regulatory elements associated with genes involved in transcriptional control, early development, and neurogenesis (Supplemental Fig. S3). Whether the chromatin accessibility here is related to prior reports of H3K27me3 marking genes "poised" for early embryonic transcription (Hammoud et al. 2009; Brykczynska et al. 2010) bears further exploration; the difficulty of recovering specific histone modification profiles from uncontaminated sperm (Fig. 5; Supplemental Figs. S5, S6) prevents us from drawing any clear conclusions. In any case, this landscape of chromatin accessibility is of course completely distinct from the ATAC-seq profiles of contaminating cell-free chromatin (Figs. 1, 2; Supplemental Fig. S3).

Our data also provide some indirect evidence regarding nucleosome organization in mature sperm. Specifically, ATAC-seq libraries not only capture relatively accessible genomic loci as relatively short ~75-bp DNA fragments but also provide information on nucleosomes flanking accessible regions (Buenrostro et al. 2013). These ~140- to 150-bp fragments result from a pair of Tn5 transposition events, with one event occurring in a relatively open genomic locus along with a second insertion occurring at the far end of an adjacent nucleosome. Indeed, ~140- to 150-bp ATAC peaks have been used in several studies of mammalian sperm to infer the existence of nucleosomes flanking regulatory elements (CpG islands, etc.) (Jung et al. 2017, 2019). However, in our DNase + DTT ATAC-seq libraries, we find no discernable ~150-bp peak in the insert length distribution (Fig. 1B), and analysis of mapped 150-bp footprints reveals diffuse enrichment rather than any evidence for positioned peaks surrounding accessible genomic loci (Fig. 2C). These data are not evidence against the existence of nucleosomes in sperm—nucleosomes not adjacent to open chromatin require two Tn5 insertions into short linkers and are heavily underrepresented in ATAC-seq libraries of somatic cells—but instead suggest that any sperm-retained nucleosomes may not be highly enriched surrounding accessible regulatory elements. That said, the locations of MNase footprints in sperm are heavily dependent on technical variables from fixation to MNase digestion extent, and it is likely that the same will hold true for ATAC-based accessibility metrics, raising the difficult question of how to independently substantiate the packaging of the sperm genome inferred from assays using large molecular probes.

Together, these considerations raise the question of the molecular nature of accessible genomic loci in mature sperm. Are these regions relatively depleted of protamines? Are they occupied by other proteins, or broadly protein-depleted? Given our concerns with affinity-based chromatin assays in sperm (below), answering these questions in molecular detail will be challenging and will require optimization of affinity-based chromatin assays suited to mature sperm.

Absence of compartments and TADs in mature murine sperm

We next consider the three-dimensional folding of the sperm genome, for which multiple previous studies have reported the a priori unlikely finding that the sperm genome is organized nearly identically to the genomes of fibroblasts and ES cells. As noted in the Introduction, all of the studies making this claim were performed in the absence of sufficient reducing agent and thus can *only* represent cell-free chromatin. Consistent with this model, we note that multiple studies (Alavattam et al. 2019; Patel et al. 2019; Wang et al. 2019b) have explored developmental changes in chromosome folding during spermatogenesis, reporting that typical architectural features (e.g., compartments, TADs, loops) are present in spermatogonia and are progressively lost during sperm development, only to be inexplicably regained in mature spermatozoa and then immediately lost in the paternal pronucleus of the zygote. These data are most consistent with extensive alterations to chromosome folding occurring during spermatogenesis, combined with a failure to appropriately measure the chromosome folding state in mature spermatozoa owing to contaminating extracellular chromatin.

Indeed, as with all the other assays of sperm chromatin, we find that the organization of the sperm genome into compartments, TADs, and loops cannot be reproduced when sperm are properly broken (Fig. 6). We instead find an essentially featureless diagonal in the DNase + DTT condition. Importantly, here again—as with ATAC-seq—we show that DTT-only data sets show a mixture of the bona fide sperm profile with the cell-free chromatin profile, which provides compelling evidence for our overarching hypothesis.

The featureless diagonal seen in our DNase + DTT samples has two potential explanations. The first, which we do not favor, is that the sperm genome is compacted without any consistent architectural features, perhaps by protamine-mediated aggregation/liquid–liquid phase separation as observed in vitro (Moritz et al. 2023). Alternatively, we speculate that standard Hi-C methodology fails to effectively assay interactions between genomic loci in physical proximity in sperm. We suspect that this results from the requirement for formaldehyde cross-linking to capture interactions between genomic loci in most 3C-derived protocols. Formaldehyde reacts efficiently with primary amines including the ε-amine moiety of lysine and primary amine groups on the nucleotide bases. However, formaldehyde is expected to react much less efficiently with the electron pair on arginine thanks to the higher pKa (~12.5 vs 10.5) of the guanidino group making this residue much less nucleophilic. Thus, although formaldehyde very effectively cross-links DNA bases to histone lysines, as well as histone lysines to one another, in the case of protamines, which are highly enriched for arginine and in many species are completely devoid of lysines, formaldehyde may not capture protamine–protamine interactions. Indeed, the relatively flat interaction decay curves for our Hi-C data are consistent with poor cross-linking (Supplemental Fig. S7).

We therefore consider our data extremely strong evidence arguing that prior sperm Hi-C studies in mammals have been contaminated by cell-free chromatin. Beyond this, at present we feel our data do not provide any insight into the three-dimensional organization of the mammalian sperm genome. For example, we find no evidence for the 50- to 80-kb solenoidal structure seen in in vitro studies of protamine–DNA complexes, which would be readily visible in Hi-C as a second diagonal 50–80 kb away from the primary diagonal. Whether such a structure occurs in vivo will be of considerable interest for future studies. Productive avenues for future chromosome folding studies in sperm can be envisioned using specific arginine-reactive or DNA-reactive cross-linkers (Jones et al. 2019; You et al. 2021), cross-linking-independent chromosome folding assays such as Hi-TrAC (Liu et al. 2022) or GAM (Beagrie et al. 2017), or superresolution microscopy "walks" to probe the organization of megabase-scale domains (Boettiger et al. 2016).

Implications for sperm chromatin architecture studies in humans

We note one intriguing finding that may provide important insights into differences between studies of human and mouse sperm chromatin. Briefly, Chen et al. (2019) generated Hi-C maps from human sperm, finding a largely featureless diagonal like the one seen in our mouse DNase + DTT data set. The investigators performed a valuable and insightful mixing experiment in which they performed Hi-C in a mixture of ejaculated human sperm and cauda epididymal mouse sperm, recovering the compartments and TADs reported in mouse sperm along with the flat diagonal they reported for humans. Moreover, in the accompanying paper (Jessberger et al. 2023), Hi-C studies of human and *Xenopus* sperm both failed to detect TADs, again highlighting the question of why human and mouse sperm differ so extensively in their apparent chromosome folding behavior.

We consider two potential explanations for the discrepancy between human and mouse sperm Hi-C. The first is that it has been reported that human sperm are more readily permeabilized in the absence of DTT (Hisano et al. 2013), suggesting that under the reported conditions the mouse sperm genome was inaccessible, and so, mouse contacts could only be generated from cell-free chromatin, whereas, in contrast, the human sample was perhaps somewhat permeabilized (albeit capturing few bona fide chromosomal contacts thanks to poor fixation of protamines with formaldehyde), thereby yielding maps similar to our DTT-only data in Figure 6. Alternatively, we believe it may be the case that ejaculated sperm are less contaminated by cell-free chromatin, as murine seminal fluid carries high levels of DNASE2B (Smyth et al. 2022), which could play a role in clearing any cell-free chromatin carried forward from the epididymis. In addition, at least some contaminating chromatin is likely to be produced during the process of epididymal dissection in the mouse (Fig. 4C; Supplemental Fig. S2A,D), again suggesting that ejaculated sperm samples obtained without tissue disruption might show less cell-free chromatin contamination. Regardless of the explanation for the discrepancy between human and mouse Hi-C experiments, the fact that TADs are only observed in mouse sperm but not in human, along with the elimination of TADs in mouse sperm by removal of cell-free chromatin, shows that the mouse sperm genome is unlikely to be organized in TADs as previously reported.

The state of histone modification landscapes in mature sperm

We finally discuss the implications of our study for the histone modification landscape in sperm, where we consider our findings less definitive. On the one hand, here again our data suggest that the prevailing view of sperm histone modifications (Erkek et al. 2013; Zheng et al. 2016; Jung et al. 2017) might also reflect contamination by cell-free chromatin. Using four common affinity-based methods for histone modification mapping—ChIP-seq (fixed/sonicated, and native/MNase), CUT&RUN, and CUT&Tag—in untreated sperm, we readily reproduce the findings of multiple prior histone modification surveys (Fig. 5; Supplemental Figs.

S5, S6). For three of these assays, as in the case of ATAC-seq, these patterns are substantially altered when sperm are pretreated with DNase I to eliminate cell-free chromatin, again consistent with extant histone modification profiles in sperm reflecting cell-free chromatin contamination.

The only condition in which we see no effect of DNase I treatment is for native MNase-ChIP-seq, in which unfixed sperm are digested with MNase before ChIP (Supplemental Fig. S6D). At first glance, this finding might suggest that prior reports were in fact able to accurately measure sperm chromatin uncontaminated by cell-free chromatin contamination. However, DNase treatment has dramatic effects on the other three histone modification assays here, and more broadly, our data show that protein probes are generally unable to access sperm in the absence of DTT treatment, raising the question of what is being mapped in the DNase-only MNase ChIP-seq data set. We speculate that some nucleosomes in the cell-free DNA may remain intact following DNase treatment. The absence of adjacent linker DNA in such material would prevent Tn5 insertion in ATAC or CUT&Tag methods, but resulting nucleosomes would remain available for immunoprecipitation. Whatever the explanation, the extraordinary differences in behavior for the four mapping approaches used here—CUT&RUN, CUT&Tag, ChIP-seq from fixed and sonicated chromatin, and ChIP-seq from native MNase-digested chromatin—raise significant concerns regarding extant histone modification mapping efforts in sperm.

That said, three issues prevent us from confidently rejecting the prevailing model of "bivalent" nucleosomes present at developmental promoters. First, K4/27 bivalent domains at CpG islands have been reported not only in mouse but also in human sperm studies (Hammoud et al. 2009). As noted above, Hi-C studies diverge extensively between these two species; if the explanation for this difference is that ejaculated sperm are not contaminated by cell-free DNA, this would suggest that K4/27-marked CpG islands might be a bona fide feature of uncontaminated mature spermatozoa in humans, and the similar findings in the mouse would thus support this view of conserved mammalian sperm chromatin. Second, analysis of histone modifications during fetal germ cell development and spermatogenesis in the testis revealed that a subset (47 of 92 promoters) of K4/27-marked poised genes overlap with those reported in mature sperm, suggesting a progressive shaping of histone modifications patterns toward the landscape reported for mature sperm (Lesch et al. 2013). As DNase treatment is commonly performed during testicular dissociation into single cells (thus presumably preventing cfDNA contamination), this again might support the prevailing CGI/bivalency model. That said, none of the testicular populations analyzed in that study will have completed the histone to protamine transition, so these data could be compatible with either model; after all, K4/27 domains occur at CpG islands in other cell types and thus could occur both in spermatocytes/spermatids before the histone to protamine transition, as well as in the cells giving rise to contaminating cfDNA, but not in mature sperm. Finally, unlike the alternative set of peaks identified for ATAC-seq in DNase-treated sperm, we cannot be confident of a putative bona fide sperm histone modification landscape as we have yet to identify conditions that allow for reliable affinity-based chromatin localization studies in DNase-treated sperm, having obtained relatively flat or nonspecific landscapes for ChIP-seq, CUT&RUN, and CUT&Tag using several antibodies.

Taken together, our results should motivate some skepticism regarding the prevailing model for the sperm chromatin landscape, especially given the extensive changes to modification localization following DNase treatment, but this model may yet prove correct. Further optimization will therefore be required to properly assess protein localization across the sperm genome using antibody-based approaches.

Methods

Ethics statement

Animal husbandry and experimentation were reviewed, approved, and monitored under the University of Massachusetts Medical School Institutional Animal Care and Use Committee (Protocol ID: A-1833-18).

Mouse husbandry and tissue collection

All samples were obtained from male mice of the C57BL/6J strain background, consumed control diet Ain-93g, and were euthanized at 12–20 wk of age.

Cauda sperm isolation and purification

For most assays, we used prep 1 (see Supplemental Fig. S2): Cauda epididymis was briefly dissected in PBS to remove fat tissues. Cauda tissues were transferred to Donner's medium and sperm were released by several cuts and gentle squeeze of the cauda tissues. The cauda was discarded, and medium was transferred to a 1.5-mL tube. Sperm swim-up was conducted for 1 h at 37°C. Next, the upper ~1.3 mL medium was recovered and filtered through a 40-µm strainer. Sperm were pelleted, washed with PBS, and resuspended/incubated in somatic lysis buffer (0.01% SDS, 0.005% Triton X-100) for 10 min on ice. Purified sperm were washed with PBS and subjected to downstream treatment. For prep 2, based on the method of Chen et al. (2021), cauda epididymis was punctured by a needle, and the sperm were squeezed out using two forceps. The sperm clot was collected in 250 µL of Donner's medium in a 1.5-mL tube, followed by gently adding 1 mL of Donner's medium on top. Sperm swim-up was conducted for 1 h at 37°C, and the upper ~1 mL medium was recovered for experiments. For prep 3, based on the method of Hisano et al. (2013), cauda epididymis was cut four to five times and directly put in 1 mL of Donner's medium in a round-bottom Falcon tube. Then, 4 mL Donner medium was added along the wall of tube, and sperm were allowed to swim up for 1 h at 37°C. Next, the upper ~4 mL of medium was recovered for sperm experiments.

Cauda epididymis cell suspension

Cauda epididymis tissue was placed in PBS and cut into small pieces. Sperm were squeezed out as much as possible followed by 10 min of incubation at 37°C to further allow sperm swim-out. Only big chunks of epithelium tissues were recovered through a 200-µm cell strainer and washed with in IMDM + DNase media (Thermo Fisher Scientific 31980030; Sigma-Aldrich 1109711 3001). The chopped tissues were further digested with complete media (10% FBS in IMDM), ECM (extracellular matrix digestion media), and IMDM + DNase for 20 min followed by 0.25% trypsin, 50 µg/mL DNase for another 15 min. Reaction was neutralized by FBS, and single-cell suspension was washed with PBS and quality-checked under a microscope.

Nuclease and DTT treatment of cauda sperm

Purified sperm pellet was resuspended in PBS. For DNase treatment, RDD buffer and DNase I (Qiagen 79254) was added at 1:10

and 1:40 (~0.1 unit per μL) dilution in 1× PBS, respectively. DNase treatment was performed at room temperature for 10 min. For MNase treatment, sperm were incubated in 1× PBS containing 85 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 2 mM $CaCl_2$, 0.1 U/μL of MNase (Worthington LS004798) for 5 min at 37°C. For Exo_V treatment, sperm were incubated in 1× PBS containing 1× NEBuffer 4, 1 mM ATP, 0.1 U/μL of RecBCD (NEB M0345) for 20 min at 37°C. After washing with PBS, sperm were incubated in 50 mM DTT (50 mM TCEP or 1% β-ME) in 1× PBS for 1 h at RT and quenched by 100 mM *N*-ethylmaleimide (NEM). Pretreated sperm were pelleted and washed with PBS and subjected to ATAC-seq.

ATAC-seq

ATAC-seq was performed using the Omni-ATAC-seq protocol (Corces et al. 2017) with several modifications. About 100,000 sperm were on-ice-lysed with hypotonic lysis buffer (10 mM Tris-HCl at pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, and 0.01% digitonin) for 10 min. Sperm were then resuspended in tagmentation buffer (10 mM TAPS-NaOH at pH 8.5, 5 mM $MgCl_2$, 10% DMF, 0.05% digitonin), and tagmentation was conducted by adding Tn5 transposase (Diagenode C01070012) at 1:20 and incubated for 30 min at 37°C. Reaction was stopped by SDS and Proteinase K and incubated overnight at 55°C. SDS was quenched by Tween-20, and library amplification was performed with 1× NEBNext high-fidelity PCR master mix containing 0.5 μM indexed primers using the following PCR conditions: 5 min at 72°C; 3 min at 98°C; and 14 cycles of 15 sec at 98°C, 30 sec at 63°C, and 1 min at 72°C. For mESCs and cauda epididymis, DNase/DTT treatment (if applied) was performed without prior somatic lysis buffer (0.01% SDS, 0.005% Triton X-100) treatment. Cell lysis was performed in tagmentation buffer containing 0.02% digitonin. Following lysis, ATAC was performed as described for sperm. In all cases, libraries were paired-end sequenced on an Illumina NextSeq 500 platform.

ATAC-seq data analysis

Paired-end ATAC-seq reads were adapter-trimmed and aligned to the mm10 genome using Bowtie 2 (version 2.3.2) (Langmead and Salzberg 2012) with the parameters: -t -q -N 1 -L 25 -X 2000 no-mixed no-discordant. All unmapped reads, nonuniquely mapped reads, and PCR duplicates were removed. MACS2 (Zhang et al. 2008) was used for peak calling with parameters: --nolambda --nomodel. Reads with insert size >150 bp were separated and used for nucleosome positioning analysis. TSS enrichment analysis was performed using HOMER annotatePeaks (Heinz et al. 2010).

Nano-NOMe-seq

Nano-NOMe libraries were prepared as previously described (Battaglia et al. 2022). Briefly, about 2 million sperm were on-ice-lysed with hypotonic lysis buffer (10 mM Tris-HCl at pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, and 0.01% digitonin) for 10 min. Sperm were pelleted down and resuspended in 500 μL of 1× GpC buffer containing 150 μL of 1 M sucrose, 1.5 μL of 32 mM S-adenosylmethionine (SAM; NEB B9003), and 50 μL of M.CviPI (NEB M0227L). The suspension was carefully mixed and incubated for 7.5 min at 37°C, followed by a boost with an additional 25 μL of M.CviPI and 1.5 μL of SAM for 7.5 min. The reaction was stopped by the addition of 25 μL of 10% SDS, 5 μL of 0.5 M EDTA, 6 μL of 20 mg/mL Proteinase K, and 28 μL of 1 M DTT followed by overnight incubation at 55°C. The total sperm DNA was recovered by PCI and ethanol precipitation. For whole-genome nanopore sequencing, 1 μg of sperm DNA was used for library preparation using a Oxford Nanopore Ligation–based library prep kit (ONT SQK-LSK110). For target-enriched sequencing, single-stranded CRISPR–Cas9 DNA oligonucleotides (20 nt) were designed using the ChopChop v.3 designer tool (https://bitbucket.org/valenlab/chopchop/src/master/) and purchased from IDT. sgRNA were prepared using EnGen sgRNA synthesis kit (NEB E3322S). RNP complexes were assembled individually by combining 8 μL of nuclease-free water, 1.5 μL of NEBuffer r3.1 (NEB), 3 μL of 300 nM sgRNA, 1 μL of Cas9 Nuclease (NEB M0386), followed by a 20-min incubation at 25°C. The different RNP complexes were then pooled in a single tube. Per experiment, 3–4 μg of sperm gDNA was dephosphorylated with 10 μL of rSAP (NEB M0371S) and 16 μL of 10× rCutSmart buffer (NEB B6004S) for 30 min at 37°C and then heat inactivated for 5 min at 65°C. DNA was then purified through PCI and ethanol precipitation. The dephosphorylated DNA sample was then combined with the pool of RNP complexes in a 1:9 ratio. The reaction was incubated for 30 min at 37°C to enable Cas9 cleavage. After digestion, Cas9 was inactivated by the addition of a 1/25 volume of 20 mg/mL Proteinase K followed by a 15-min incubation at 55°C. Cleaved and purified DNA was purified by PCI and ethanol precipitation and was subjected to dA-tailing using Klenow fragment (NEB E6053). Reactions were incubated for 30 min at 37°C and immediately subjected to adapter ligation as per the manual of the Oxford Nanopore Ligation–based library prep kit (ONT SQK-LSK110). All nanopore libraries were sequenced using MinION flow cell (ONT R9.4.1).

Nanopore data processing

Base calling was performed on the raw FAST5 files with Guppy (v.3.0.3 or v.5.0.11, ONT), using a configuration file for high-accuracy DNA base calling on an R9.4.1 pore at 450 bases/sec. The resulting reads were then mapped to the mm10 mouse reference genome without alternate contigs using minimap2 v.2.11 (Li 2018) with default settings for alignment of nanopore reads (-x map-ont). Reads that mapped with a quality score of less than 50 were then filtered out using SAMtools v.1.7 (Li et al. 2009). CpG and GpC methylations were simultaneously called on the remaining reads using nanopolish v.0.11.1. Average CpG methylation was calculated for each group of CpGs that did not contain any GCGs. To visualize the methylation patterns of the reads with the Integrative Genomics Viewer (IGV) v.2.12.2 (Thorvaldsdottir et al. 2013), we modified the individual reads in the alignment (BAM) files. All cytosines called as unmethylated were converted to thymine to simulate bisulfite conversion. This was achieved using code adapted from the Timp Laboratory's nanopore-methylation-utilities (https://github.com/timplab/nanopore-methylation-utilities). Once these converted files were loaded in IGV, we were then able to visualize methylation using IGV's bisulfite mode. Tiled data file tracks showing aggregated CpG methylation levels were generated with igvtools v.2.4.16 (Thorvaldsdottir et al. 2013).

CUT&RUN

CUT&RUN libraries were prepared following the same protocol as previously described (Chen et al. 2019) with minor modifications. Specifically, 0.1% NP-40, 0.1% Tween-20, and 0.05% digitonin were used in the antibody incubation buffer and wash buffer, and SPRI beads were used for DNA extraction after Proteinase K treatment. For H3K27me3 CUT&RUN, a rabbit anti-H3K27me3 antibody (1:100; Diagenode C15410195) was used. For H3K4me3 CUT&RUN, a rabbit anti-H3K4me3 antibody (1:100;

Diagenode C15410003) was used. All CUT&RUN libraries were sequenced on an Illumina NextSeq 500 platform.

CUT&Tag

CUT&Tag libraries were prepared following the same protocol as previously described (Kaya-Okur et al. 2019). Antibodies used were against H3K27me3 (1:50; Thermo Fisher Scientific MA5-11198), H3K4me3 (1:50; abcam ab8580), histone H3 (1:50; abcam ab1791), and serine 2 Phospho-Rpb1 CTD (1:50; Cell Signaling Technology 13499S). All CUT&Tag libraries were sequenced on an Illumina NextSeq 500 platform.

ChIP-seq

ChIP-seq experiments were conducted following the protocol previously described (Hisano et al. 2013; Chen et al. 2021) with some modifications. For the fixed-sonication-based method, sperm were first fixed with 1% formaldehyde in PBS for 10 min at room temperature, followed by quenching with glycine. Next, sperm were pretreated with 50 mM DTT and 1 mg/mL heparin for 2 h at 37°C. Sperm were then quenched with *N*-ethylmaleimide (NEM) and washed with PBS twice. After lysing sperm in complete buffer containing detergent, an equal volume of 2× RIPA buffer was supplied to make a 1× RIPA condition. Chromatin shearing was performed using a Covaris S220 platform with the following parameters: peak power 140 W, duty factor 5%, and 200 cycles/burst for a total of 30 min. For the native MNase–based method, sperm were directly lysed in complete buffer containing detergent and digested with 15 unit of MNase (Worthington LS004798) for 5 min at 37°C. The reaction was stopped by EDTA and supplied with an equal volume of 2× RIPA buffer. Sonicated or MNase-digested sperm chromatin was recovered from the supernatant by spinning at 10,000*g* for 10 min at 4°C and was directly subjected to immunoprecipitation with corresponding antibodies (H3K4me3: Diagenode C15410003; H3K27me3: Diagenode C15410195; CTCF: Millipore 07-792). Library preparation was performed using NEBNext ultra II DNA library prep kit (NEB E7645). All ChIP-seq libraries were sequenced on an Illumina NextSeq 500 platform.

Histone modification data analysis.

For both the CUT&RUN and ChIP-seq data sets, paired-end reads were trimmed using Trim Galore! (version 0.4.5; https://github.com/FelixKrueger/TrimGalore), followed by read alignment to the mm10 genome using Bowtie 2 (version 2.3.2) (Langmead and Salzberg 2012) with the following parameters: -t -q -N 1 -L 25 -X 700 no-mixed no-discordant. PCR duplicates were removed using "MarkDuplicates" from Picard Tools (version 2.8.0; https://broadinstitute.github.io/picard/). Only non-PCR duplicates and uniquely aligned reads (alignment records without "XS" tag) were used for downstream analyses. The RPKM values for each 100-bp bin were calculated following the formula "read counts/[(bin_length/1000) × (total_reads/10^6)]. Peak calling for the H3K4me3 and H3K27me3 data sets was performed by MACS2 with following parameters: --nolambda --nomodel, and --broad-cutoff 0.1 ---nolambda --nomodel, respectively.

Hi-C library construction

The sperm library collection was made as described previously (Lafontaine et al. 2021) with some modifications. Briefly, cells were fixed in 1% formaldehyde and, after washing in PBS, left for 1 h in 50 mM DTT. After quenching with 100 mM NEM and washing in PBS, cells were incubated with DpnII overnight at 37°C. After restriction digestion, cells were incubated for 4 h at 23°C with the Klenow DNA polymerase I to fill in the 5′ overhang and incorporate biotin-14-dATP followed by chromatin ligation for 4 h at 16°C with T4 DNA ligase (Invitrogen). Then cells were left overnight in 400 ng/mL Proteinase K and 0.5% SDS overnight at 60°C to reverse cross-linking. After DNA purification and precipitation, DNA was sonicated to 200- to 300-bp fragments, and an additional size selection was performed with Ampure beads. To repair DNA ends after sonication, T4 DNA polymerase (NEB) and Klenow DNA polymerase (NEB) were used together with T4 polynucleotide kinase (NEB). Biotinylated ligation products were pulled down using streptavidin beads (Invitrogen) followed by A-tailing and adaptor ligation for library construction. All Hi-C libraries were sequenced on an Illumina NextSeq 500 platform.

Hi-C data analysis

Hi-C data was analyzed using the Open Chromosome Collective suite (https://github.com/open2c). Briefly, 50-bp paired-end FASTQ files were mapped to the mm10 mouse reference genome using the *distiller-nf* (https://github.com/mirnylab/distiller-nf), invoking BWA-MEM to map FASTQ pairs in a single-side regime (-SP). Aligned reads were allowed a 1-bp flexibility for the removal of optical and PCR duplicates using *pairtools* (https://github.com/mirnylab/pairtools). After removal of identical positions and strand orientations, the remaining valid pairs were binned into contact matrices of various resolutions using *cooler* (Abdennur and Mirny 2020). An iterative balancing procedure (Imakaev et al. 2012) was applied to all matrices, ignoring the first two diagonals to avoid short-range ligation artifacts at a given resolution, and bins with low coverage were removed using MADmax filter with default parameters. The resultant ".cool" contact matrices were used in downstream analyses using *cooltools* (https://github.com/mirnylab/cooltools) and uploaded to a HiGlass server (Kerpedjiev et al. 2018) via the Reservoir Genome Browser (https://resgen.io/).

Contact probability ($P(s)$) plots & derivatives

We produced $P(s)$ plots per chromosome arm as outlined in the "contacts vs distance" section of cooltools (https://cooltools.readthedocs.io/en/latest/notebooks/contacts_vs_distance.html). Briefly, mm10 chromosome sizes and arm locations were downloaded as ViewFrames from the UCSC database using Bioframe (https://bioframe.readthedocs.io/en/latest/index.html). For each arm, interaction counts (diagsum) were generated from a 1-kb cooler file for each distance from the self–self diagonal. Counts for all chromosome arms were aggregated and logbin-smoothed ($\sigma = 0.05$), and $P(s)$ curves were normalized for the total number of valid interactions in each data set. Corresponding derivative plots were generated using a numpy gradient to estimate the slope from each $P(s)$ plot.

Compartment analyses

When present, compartmentalization is the strongest signal and will be represented by the first eigenvector (EV1). We performed eigenvector decomposition on observed-over-expected contact maps at 100-kb resolution separated for each chromosomal arm using the cooltools package derived scripts (https://cooltools.readthedocs.io/en/latest/notebooks/compartments_and_saddles.html).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO;

https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE226684 (CUT&RUN), GSE226685 (ATAC-seq), GSE226686 (ChIP-seq), GSE240247 (CUT&Tag), GSE240248 (Hi-C), and GSE240249 (Nanopore).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank N. Krietenstein for invaluable technical advice on sperm lysis conditions and Hi-C, B. Bernstein and W. Flavahan for helpful advice on Nano-NOMe-seq experiment design, and S. Hammoud, B. Lesch, and A. Peters for discussions and insightful comments on sperm chromatin architecture. This work was supported by National Institutes of Health R01AG073238 (O.J.R.), R01HD072122 (T.G.F.), R01HG003143 (J.D.), and R35GM128782 (X.Z.L.). J.D. is an investigator of the Howard Hughes Medical Institute.

Author contributions: Q.Y. and O.J.R. designed the majority of experiments in this manuscript, and Q.Y. performed them, with the following exceptions. C.-H.Y. assisted with sperm collection and treatment and with ES cell experiments. O.S.S., Y.S., L.Y., and J.G. performed Hi-C experiments, which were analyzed by J.D., X.Z.L., and J.G. J.W. and Q.Y. performed and analyzed Nanopore experiments. S.G. and T.G.F. performed CUT&Tag experiments.

References

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36:** 311–316. doi:10.1093/bioinformatics/btz540

Alavattam KG, Maezawa S, Sakashita A, Khoury H, Barski A, Kaplan N, Namekawa SH. 2019. Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat Struct Mol Biol* **26:** 175–184. doi:10.1038/s41594-019-0189-y

Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D. 2009. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* **19:** 1338–1349. doi:10.1101/gr.094953.109

Balhorn R. 1982. A model for the structure of chromatin in mammalian sperm. *J Cell Biol* **93:** 298–305. doi:10.1083/jcb.93.2.298

Balhorn R. 2007. The protamine family of sperm nuclear proteins. *Genome Biol* **8:** 227. doi:10.1186/gb-2007-8-9-227

Balhorn R, Gledhill BL, Wyrobek AJ. 1977. Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16:** 4074–4080. doi:10.1021/bi00637a021

Battaglia S, Dong K, Wu J, Chen Z, Najm FJ, Zhang Y, Moore MM, Hecht V, Shoresh N, Bernstein BE. 2022. Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nat Genet* **54:** 1504–1513. doi:10.1038/s41588-022-01188-8

Battulin N, Fishman VS, Mazur AM, Pomaznoy M, Khabarova AA, Afonnikov DA, Prokhortchouk EB, Serov OL. 2015. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol* **16:** 77. doi:10.1186/s13059-015-0642-0

Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, et al. 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543:** 519–524. doi:10.1038/nature21411

Bedi YS, Wang H, Thomas KN, Basel A, Prunier J, Robert C, Golding MC. 2022. Alcohol induced increases in sperm histone H3 lysine 4 trimethylation correlate with increased placental CTCF occupancy and altered developmental programming. *Sci Rep* **12:** 8839. doi:10.1038/s41598-022-12188-3

Bellvé AR, Anderson E, Hanley Bowdoin L. 1975. Synthesis and amino acid composition of basic proteins in mammalian sperm nuclei. *Dev Biol* **47:** 349–365. doi:10.1016/0012-1606(75)90289-4

Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu CT, Zhuang X. 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529:** 418–422. doi:10.1038/nature16496

Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schübeler D, Stadler MB, Peters AH. 2010. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* **17:** 679–687. doi:10.1038/nsmb.1821

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 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. doi:10.1038/nmeth.2688

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109:** 21–29. doi:10.1002/0471142727.mb2129s109

Calvin HI. 1976. Comparative analysis of the nuclear basic proteins in rat, human, guinea pig, mouse and rabbit spermatozoa. *Biochim Biophys Acta* **434:** 377–389. doi:10.1016/0005-2795(76)90229-4

Calvin HI, Bedford JM. 1971. Formation of disulphide bonds in the nucleus and accessory structures of mammalian spermatozoa during maturation in the epididymis. *J Reprod Fertil Suppl* **13: Suppl 13:** 65–75.

Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, Fazzio TG, Rando OJ. 2014. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* **30:** 11–22. doi:10.1016/j.devcel.2014.05.024

Chen X, Ke Y, Wu K, Zhao H, Sun Y, Gao L, Liu Z, Zhang J, Tao W, Hou Z, et al. 2019. Key role for CTCF in establishing chromatin structure in human embryos. *Nature* **576:** 306–310. doi:10.1038/s41586-019-1812-0

Chen Z, Djekidel MN, Zhang Y. 2021. Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat Genet* **53:** 551–563. doi:10.1038/s41588-021-00821-2

Chen H, Scott-Boyer MP, Droit A, Robert C, Belleannée C. 2022. Sperm heterogeneity accounts for sperm DNA methylation variations observed in the caput epididymis, independently from DNMT/TET activities. *Front Cell Dev Biol* **10:** 834519. doi:10.3389/fcell.2022.834519

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14:** 959–962. doi:10.1038/nmeth.4396

Dekker J, Misteli T. 2015. Long-range chromatin interactions. *Cold Spring Harb Perspect Biol* **7:** a019356. doi:10.1101/cshperspect.a019356

Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schübeler D, van der Vlag J, Stadler MB, Peters AH. 2013. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* **20:** 868–875. doi:10.1038/nsmb.2599

Friedman N, Rando OJ. 2015. Epigenomics and the structure of the living genome. *Genome Res* **25:** 1482–1490. doi:10.1101/gr.190165.115

Galan C, Serra RW, Sun F, Rinaldi VD, Conine CC, Rando OJ. 2021. Stability of the cytosine methylome during post-testicular sperm maturation in mouse. *PLoS Genet* **17:** e1009416. doi:10.1371/journal.pgen.1009416

Gaspa-Toneu L, Peters AH. 2023. Nucleosomes in mammalian sperm: conveying paternal epigenetic inheritance or subject to reprogramming between generations? *Curr Opin Genet Dev* **79:** 102034. doi:10.1016/j.gde.2023.102034

Gaucher J, Reynoird N, Montellier E, Boussouar F, Rousseaux S, Khochbin S. 2010. From meiosis to postmeiotic events: the secrets of histone disappearance. *FEBS J* **277:** 599–604. doi:10.1111/j.1742-4658.2009.07504.x

Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W. 2020. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol* **38:** 433–438. doi:10.1038/s41587-020-0407-5

Gou LT, Lim DH, Ma W, Aubol BE, Hao Y, Wang X, Zhao J, Liang Z, Shao C, Zhang X, et al. 2020. Initiation of parental genome reprogramming in fertilized oocyte by splicing kinase SRPK1-catalyzed protamine phosphorylation. *Cell* **180:** 1212–1227.e14. doi:10.1016/j.cell.2020.02.020

Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. 2009. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460:** 473–478. doi:10.1038/nature08162

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Hisano M, Erkek S, Dessus-Babus S, Ramos L, Stadler MB, Peters AH. 2013. Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nat Protoc* **8:** 2449–2470. doi:10.1038/nprot.2013.145

Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9:** 999–1003. doi:10.1038/nmeth.2148

Jessberger G, Varnai C, Stocsits RR, Tang W, Stary G, Peters J-M. 2023. Cohesin and CTCF do not assemble TADs in *Xenopus* sperm and male pronuclei. *Genome Res* (this issue) doi:10.1101/gr.277865.123

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316:** 1497–1502. doi:10.1126/science.1141319

Johnson GD, Jodar M, Pique-Regi R, Krawetz SA. 2016. Nuclease footprints in sperm project past and future chromatin regulatory events. *Sci Rep* **6:** 25864. doi:10.1038/srep25864

Jones AX, Cao Y, Tang YL, Wang JH, Ding YH, Tan H, Chen ZL, Fang RQ, Yin J, Chen RC, et al. 2019. Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nat Commun* **10:** 3911. doi:10.1038/s41467-019-11917-z

Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, Corces VG. 2017. Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. *Cell Rep* **18:** 1366–1382. doi:10.1016/j.celrep.2017.01.034

Jung YH, Kremsky I, Gold HB, Rowley MJ, Punyawai K, Buonanotte A, Lyu X, Bixler BJ, Chan AWS, Corces VG. 2019. Maintenance of CTCF- and transcription factor-mediated interactions from the gametes to the early mouse embryo. *Mol Cell* **75:** 154–171.e5. doi:10.1016/j.molcel.2019.04.014

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10:** 1930. doi:10.1038/s41467-019-09982-5

Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, et al. 2017. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170:** 367–381.e20. doi:10.1016/j.cell.2017.06.029

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22:** 2497–2506. doi:10.1101/gr.143008.112

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* **19:** 125. doi:10.1186/s13059-018-1486-1

Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20:** 207–220. doi:10.1038/s41576-018-0089-8

Lafontaine DL, Yang L, Dekker J, Gibcus JH. 2021. Hi-C 3.0: improved protocol for genome-wide chromosome conformation capture. *Curr Protoc* **1:** e198. doi:10.1002/cpz1.198

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N, Sedlazeck FJ, Hansen KD, Simpson JT, Timp W. 2020. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat Methods* **17:** 1191–1199. doi:10.1038/s41592-020-01000-7

Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. 2013. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc Natl Acad Sci* **110:** 16061–16066. doi:10.1073/pnas.1315204110

Lesch BJ, Tothova Z, Morgan EA, Liao Z, Bronson RT, Ebert BL, Page DC. 2019. Intergenerational epigenetic inheritance of cancer susceptibility in mammals. *eLife* **8:** e39380. doi:10.7554/eLife.39380

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34:** 3094–3100. doi:10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Lismer A, Siklenka K, Lafleur C, Dumeaux V, Kimmins S. 2020. Sperm histone H3 lysine 4 trimethylation is altered in a genetic mouse model of transgenerational epigenetic inheritance. *Nucleic Acids Res* **48:** 11380–11393. doi:10.1093/nar/gkaa712

Lismer A, Dumeaux V, Lafleur C, Lambrot R, Brind'Amour J, Lorincz MC, Kimmins S. 2021. Histone H3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring. *Dev Cell* **56:** 671–686.e6. doi:10.1016/j.devcel.2021.01.014

Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. 2005. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* **3:** e328. doi:10.1371/journal.pbio.0030328

Liu S, Cao Y, Cui K, Tang Q, Zhao K. 2022. Hi-TrAC reveals division of labor of transcription factors in organizing chromatin loops. *Nat Commun* **13:** 6679. doi:10.1038/s41467-022-34276-8

Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH. 2018. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* **46:** 593–608. doi:10.1093/nar/gkx1052

Moritz L, Schon SB, Rabbani M, Sheng Y, Agrawal R, Glass-Klaiber J, Sultan C, Camarillo JM, Clements J, Baldwin MR, et al. 2023. Sperm chromatin structure and reproductive fitness are altered by substitution of a single amino acid in mouse protamine 1. *Nat Struct Mol Biol* **30:** 1077–1091. doi:10.1038/s41594-023-01033-4

Patel L, Kang R, Rosenberg SC, Qiu Y, Raviram R, Chee S, Hu R, Ren B, Cole F, Corbett KD. 2019. Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat Struct Mol Biol* **26:** 164–174. doi:10.1038/s41594-019-0187-0

Saowaros W, Panyim S. 1979. The formation of disulfide bonds in human protamines during sperm maturation. *Experientia* **35:** 191–192. doi:10.1007/BF01920608

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128:** 735–745. doi:10.1016/j.cell.2007.02.009

Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, Greenleaf WJ. 2020. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods* **17:** 319–327. doi:10.1038/s41592-019-0730-2

Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M, et al. 2015. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350:** aab2006. doi:10.1126/science.aab2006

Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6:** e21856. doi:10.7554/eLife.21856

Smyth SP, Nixon B, Anderson AL, Murray HC, Martin JH, MacDougall LA, Robertson SA, Skerrett-Byrne DA, Schjenken JE. 2022. Elucidation of the protein composition of mouse seminal vesicle fluid. *Proteomics* **22:** e2100227. doi:10.1002/pmic.202100227

Sollberger G, Tilley DO, Zychlinsky A. 2018. Neutrophil extracellular traps: the biology of chromatin externalization. *Dev Cell* **44:** 542–553. doi:10.1016/j.devcel.2018.01.019

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192. doi:10.1093/bib/bbs017

Vara C, Paytuví-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salvà-Castro J, Gómez HL, Julià E, Moutinho C, Aiese Cigliano R, et al. 2019. Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep* **28:** 352–367.e9. doi:10.1016/j.celrep.2019.06.037

Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, et al. 2014. Mouse regulatory DNA landscapes reveal global principles of *cis*-regulatory evolution. *Science* **346:** 1007–1012. doi:10.1126/science.1246426

Wang Y, Wang A, Liu Z, Thurman AL, Powers LS, Zou M, Zhao Y, Hefel A, Li Y, Zabner J, et al. 2019a. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res* **29:** 1329–1342. doi:10.1101/gr.251116.119

Wang Y, Wang H, Zhang Y, Du Z, Si W, Fan S, Qin D, Wang M, Duan Y, Li L, et al. 2019b. Reprogramming of meiotic chromatin architecture during spermatogenesis. *Mol Cell* **73:** 547–561.e6. doi:10.1016/j.molcel.2018.11.019

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B. 2012. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148:** 816–831. doi:10.1016/j.cell.2011.12.035

Yamaguchi K, Hada M, Fukuda Y, Inoue E, Makino Y, Katou Y, Shirahige K, Okada Y. 2018. Re-evaluating the localization of sperm-retained histones revealed the modification-dependent accumulation in specific genome regions. *Cell Rep* **23:** 3920–3932. doi:10.1016/j.celrep.2018.05.094

Yoshida K, Maekawa T, Ly NH, Fujita SI, Muratani M, Ando M, Katou Y, Araki H, Miura F, Shirahige K, et al. 2020. ATF7-dependent epigenetic changes are required for the intergenerational effect of a paternal low-protein diet. *Mol Cell* **78:** 445–458.e6. doi:10.1016/j.molcel.2020.02.028

You Q, Cheng AY, Gu X, Harada BT, Yu M, Wu T, Ren B, Ouyang Z, He C. 2021. Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nat Biotechnol* **39:** 225–235. doi:10.1038/s41587-020-0643-8

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

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. doi:10.1016/j.molcel.2016.08.032

Received March 3, 2023; accepted in revised form October 31, 2023.