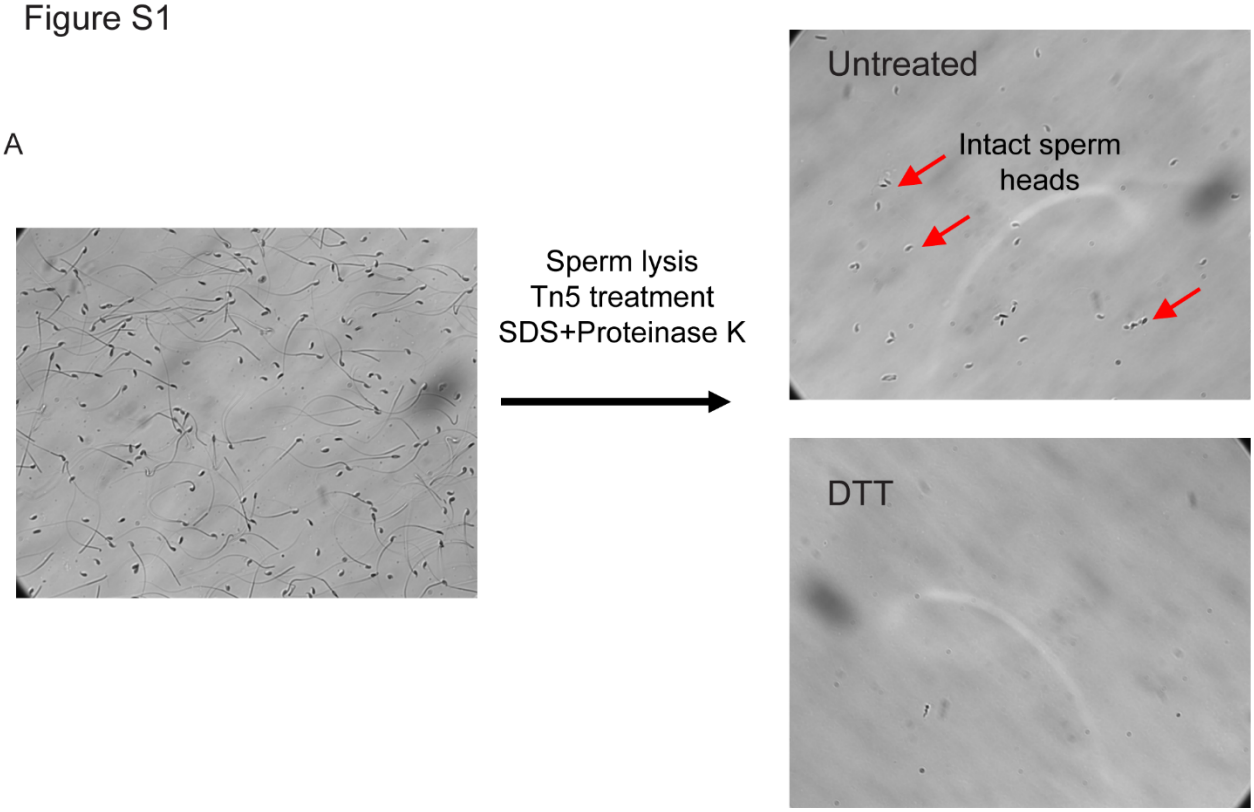
Yin et al, Supplementary Materials:

Supplemental Figures S1-7

Figure S1



Supplemental Figure S1. DTT is required to lyse mature spermatozoa

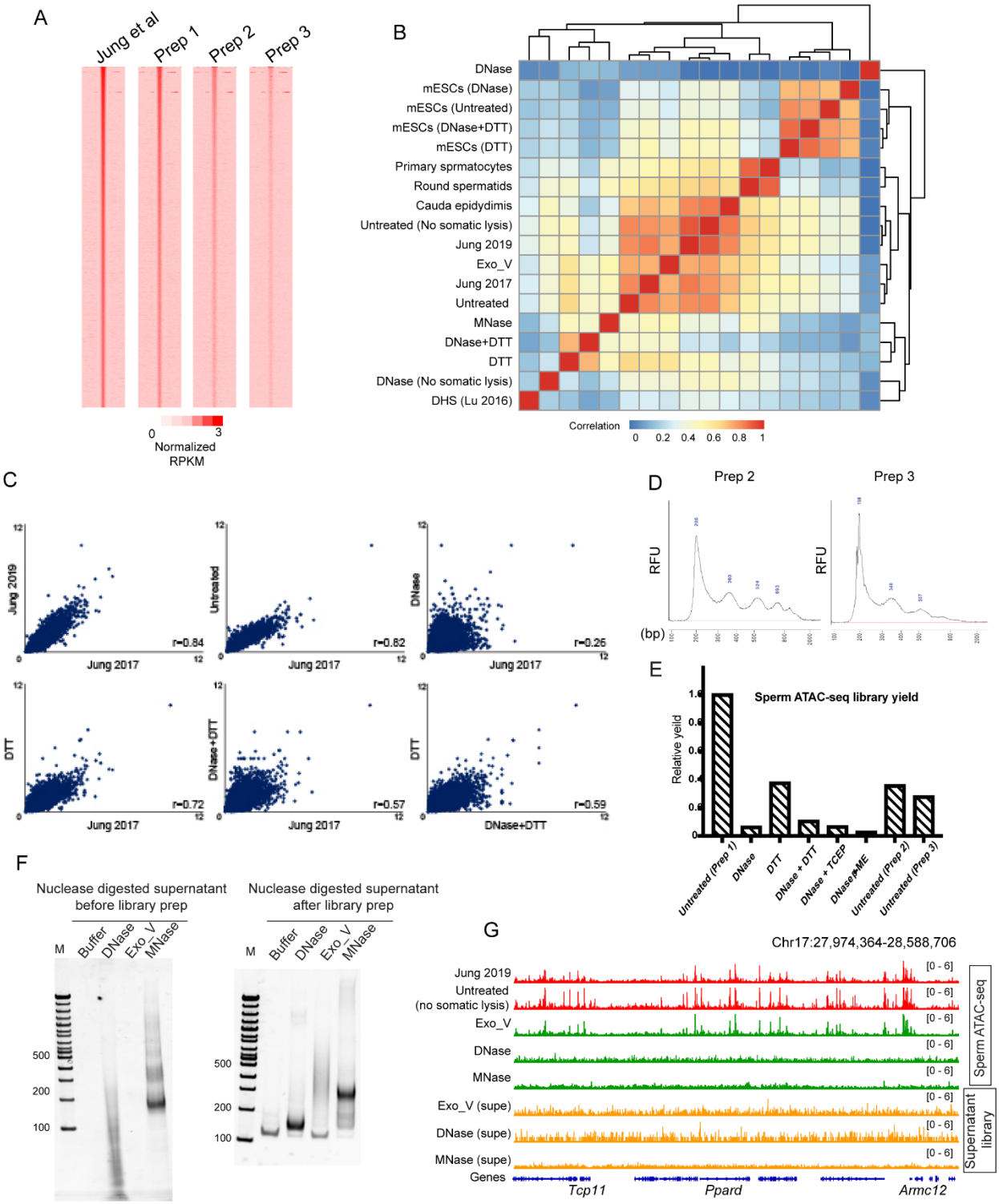
A) DIC Images of typical cauda epididymal sperm samples before and after processing for ATAC-seq. Top panels show sperm processed without the use of DTT, revealing unbroken sperm heads even after overnight incubation in SDS and Proteinase K. Bottom panels show efficient sperm lysis for DTT-treated sperm under the same conditions.

B) DTT is required to recover genomic DNA from sperm. Native PAGE gel shows the gDNA of sperm isolated from different methods. Here and in **Supplemental Fig. S2**, we compared sperm obtained from the cauda epididymis using three different preparation methods. In each case the cauda epididymis was removed via two gentle incisions at the junctions with the corpus and the vas deferens. Preparations differed by how the epididymis was further treated (**Methods**) – briefly, in Prep 1, epididymis was subject to multiple incisions and sperm were squeezed out, in Prep 2 the epididymis was punctured by needle and sperm were squeezed out, and in Prep 3 the epididymis was incised and sperm were allowed to swim out without any squeezing. In all three cases tissue was incubated in Donners medium at 37°C for 1 hr, after which the sperm-containing supernatant was carefully recovered. Sperm were then washed and either left untreated or incubated with 50 mM DTT for 1 hour, then quenched with NEM. Samples were then treated identically with lysis buffer, incubated in SDS and Proteinase K at 55°C for 16 hours, and lysed in PCI (phenol:chloroform:isoamyl alcohol) followed by ethanol precipitation for genomic DNA extraction.

C) Genomic DNA recovery following permeabilization with varying levels of DTT. Maximal DNA recovery is observed at 50 and 80 mM DTT, with somewhat lower gDNA recovery from the 20 mM DTT condition. Consistent with our estimates from cytosine methylation studies (Galan et al. 2021), contaminating cell-free DNA represents ~5-6% of the DNA recovered from a fully (50 or 80 mM) DTT-permeabilized sperm prep.

D) Genomic DNA recovered from DTT-permeabilized sperm, with or without pre-treatment with DNase I. Notably, DNase I pretreatment does not impact the integrity of gDNA recovered from sperm.

Figure S2


A
Jung et al
Prep 1
Prep 2
Prep 3
0
3
Normalized RPKM
B
DNase
mESCs (DNase)
mESCs (Untreated)
mESCs (DNase+DTT)
mESCs (DTT)
Primary sprmatocytes
Round spermatids
Cauda epidydimis
Untreated (No somatic lysis)
Jung 2019
Exo_V
Jung 2017
Untreated
MNase
DNase+DTT
DTT
DNase (No somatic lysis)
DHS (Lu 2016)
Correlation
0 0.2 0.4 0.6 0.8 1
C
Jung 2019
Untreated
DNase
DTT
DNase+DTT
Jung 2017
r=0.84
r=0.82
r=0.26
r=0.72
r=0.57
r=0.59
D
Prep 2
Prep 3
RFU
(bp)
E
Sperm ATAC-seq library yield
Relative yeild
Untreated (Prep 1)
DNase
DTT
DNase + DTT
DNase + TCEP
DNase+ME
Untreated (Prep 2)
Untreated (Prep 3)
F
Nuclease digested supernatant before library prep
Nuclease digested supernatant after library prep
M
Buffer
DNase
Exo_V
MNase
500
200
100
G
Chr17:27,974,364-28,588,706
Jung 2019
Untreated (no somatic lysis)
Exo_V
DNase
MNase
Exo_V (supe)
DNase (supe)
MNase (supe)
Genes
[0 - 6]
Sperm ATAC-seq
Supernatant library
Tcp11
Ppard
Armc12

Supplemental Figure S2. Effects of sperm purification methods on chromatin accessibility landscapes

A) Heatmaps showing ATAC-seq enrichment for the three sperm preparations (see **Supplemental Fig. S1B**), aligned over peaks from Jung *et al*. (along with Jung *et al* data in the leftmost panel). Importantly, enrichment for open chromatin at peaks reported by Jung *et al*, and a nucleosomal insert landscape (see panel (D)), were observed for all three methods, but these peaks were stronger for the more disruptive tissue dissections. This suggests that at least some of the cell free chromatin contamination explored here may arise from the process of tissue disruption, but that this is unavoidable even under the gentlest dissection methods.

B) Correlation matrix between the indicated genome-wide datasets. RPKM was calculated for 2 kb genome-wide bins and pairwise spearman correlations were assessed. Two large blocks of well-correlated datasets are apparent. The first shows strong correlation between ES cell ATAC-seq datasets obtained either from untreated cells or from cells following DNase, DTT, or DNase+DTT treatments (see also **Fig. 3**). The other group of well-correlated datasets include public data (Jung et al. 2017; Jung et al. 2019), data from untreated sperm (this study), and, importantly, ATAC-seq data from the cauda epididymal epithelium (see **Fig. 4C**). Moreover, we find that skipping the somatic cell lysis steps – as conducted in Jung *et al* – resulted in still better agreement between our "untreated" ATAC-seq profile and published ATAC-seq datasets. This suggests that contaminating chromatin is not exclusively produced as a result of somatic cell lysis, and that detergent washing does help to remove contaminating material, albeit inefficiently.

Data for immature sperm populations are distinct from ESCs and untreated sperm, while DTT-treated sperm (whether DNase-pretreated or not) cluster separately from immature sperm, untreated sperm, or ESCs. (Exo_V: Exonuclease V; DHS: DNase-seq)

C) Scatterplots showing local ATAC-seq enrichment at 1 kb surrounding all TSSs, for the indicated pairs of libraries. All three datasets from untreated sperm (this study and Jung 2017 and 2019) are highly-correlated, while data from DNase+DTT-treated sperm are distinct from any untreated samples. DTT-only sperm exhibits intermediate correlations with both untreated and DNase+DTT treated sperm samples.

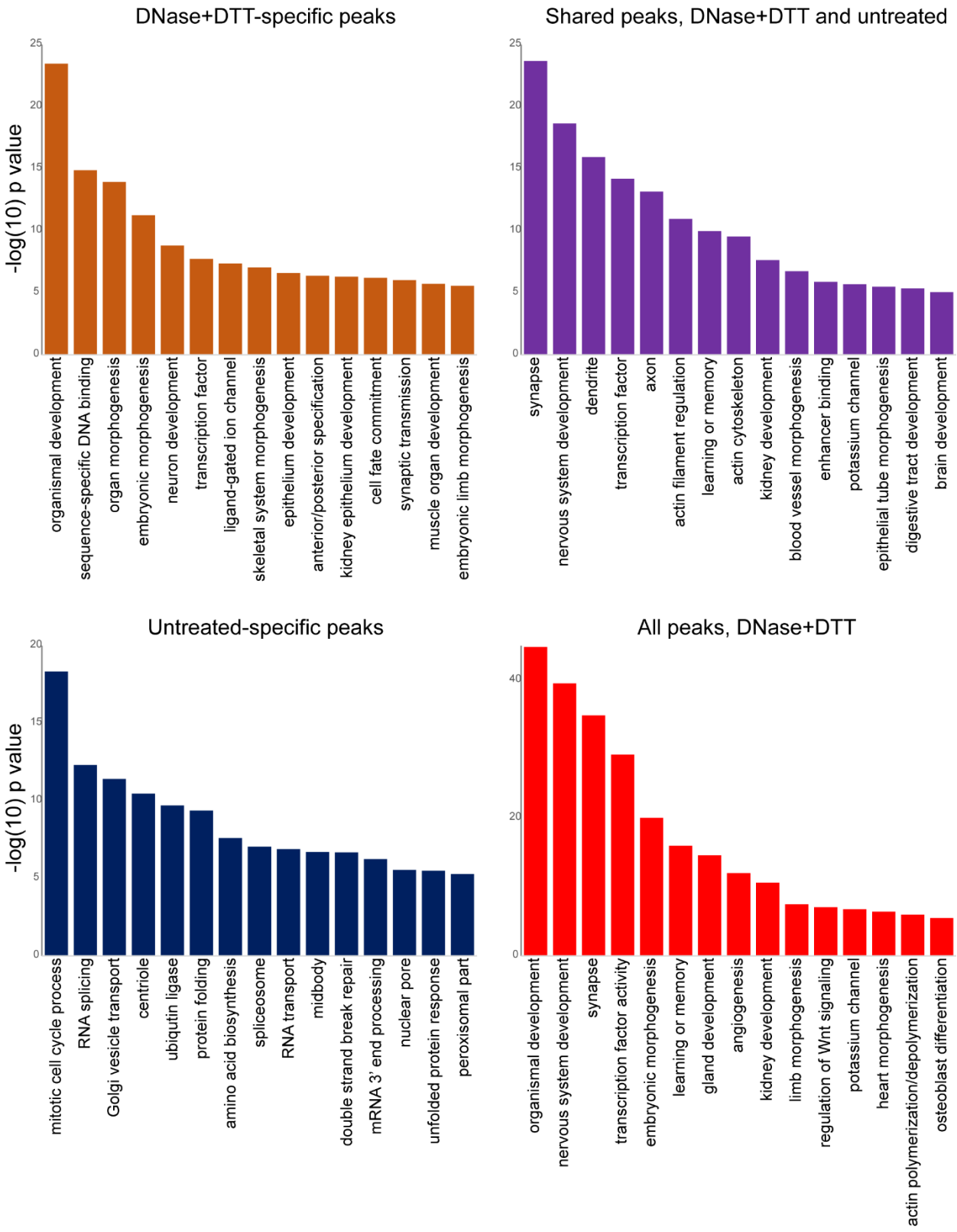
D) Insert length distributions for the sperm preparations 2 and 3 (corresponding insert lengths for Prep 1 are shown in **Fig. 1B**).

E) ATAC-seq library yields for sperm treated with the indicated conditions prior to Tn5 treatment. All yields (mean of DNA library yields from two replicates) from other conditions were normalized to untreated group (Prep 1).

F) Untreated sperm were treated with the indicated nucleases, then sperm were pelleted by centrifugation and genomic DNA was recovered from the supernatant and visualized by gel electrophoresis (left panel). Note the nucleosomal bands in the MNase-digested material, consistent with cell-free contamination by chromatin rather than naked DNA. Right panel visualizes sequencing libraries prepared from supernatant following the indicated treatments.

G) Deep sequencing of the sperm preps from panel (F). Bottom three panels show the supernatant material from (F), revealing that contaminating material arises from the entire genome, rather than specific loci. ATAC panels for the nuclease-treated samples were prepared using sperm pelleted following the indicated nuclease treatments, but not permeabilized with DTT. Note that no enrichment is seen for sperm treated with the endonucleases DNase I or MNase, while exonuclease treatment leaves cell-free DNA available for ATAC-seq. Importantly, the continued presence of ATAC peaks in this material confirms that cell-free chromatin cosediments with sperm through a gentle centrifugation step.

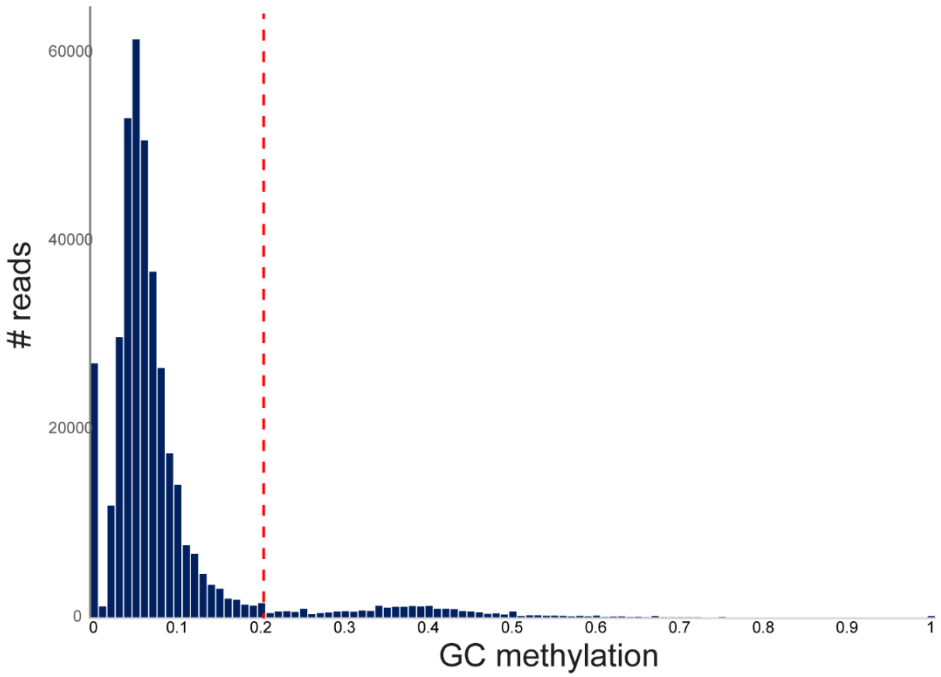
Figure S3


DNase+DTT-specific peaks
-log(10) p value
organismal development
sequence-specific DNA binding
organ morphogenesis
embryonic morphogenesis
neuron development
transcription factor
ligand-gated ion channel
skeletal system morphogenesis
epithelium development
anterior/posterior specification
kidney epithelium development
cell fate commitment
synaptic transmission
muscle organ development
embryonic limb morphogenesis
Shared peaks, DNase+DTT and untreated
synapse
nervous system development
dendrite
transcription factor
axon
actin filament regulation
learning or memory
actin cytoskeleton
kidney development
blood vessel morphogenesis
enhancer binding
potassium channel
epithelial tube morphogenesis
digestive tract development
brain development
Untreated-specific peaks
-log(10) p value
mitotic cell cycle process
RNA splicing
Golgi vesicle transport
centriole
ubiquitin ligase
protein folding
amino acid biosynthesis
spliceosome
RNA transport
midbody
double strand break repair
mRNA 3' end processing
nuclear pore
unfolded protein response
peroxisomal part
All peaks, DNase+DTT
organismal development
nervous system development
synapse
transcription factor activity
embryonic morphogenesis
learning or memory
gland development
angiogenesis
kidney development
limb morphogenesis
regulation of Wnt signaling
potassium channel
heart morphogenesis
actin polymerization/depolymerization
osteoblast differentiation

Supplemental Figure S3. Functional enrichment in sperm open chromatin

Bar graphs show p values (expressed as -log10 of the p value) for various Gene Ontology categories enriched in genes found near the indicated sperm ATAC-seq peaks.

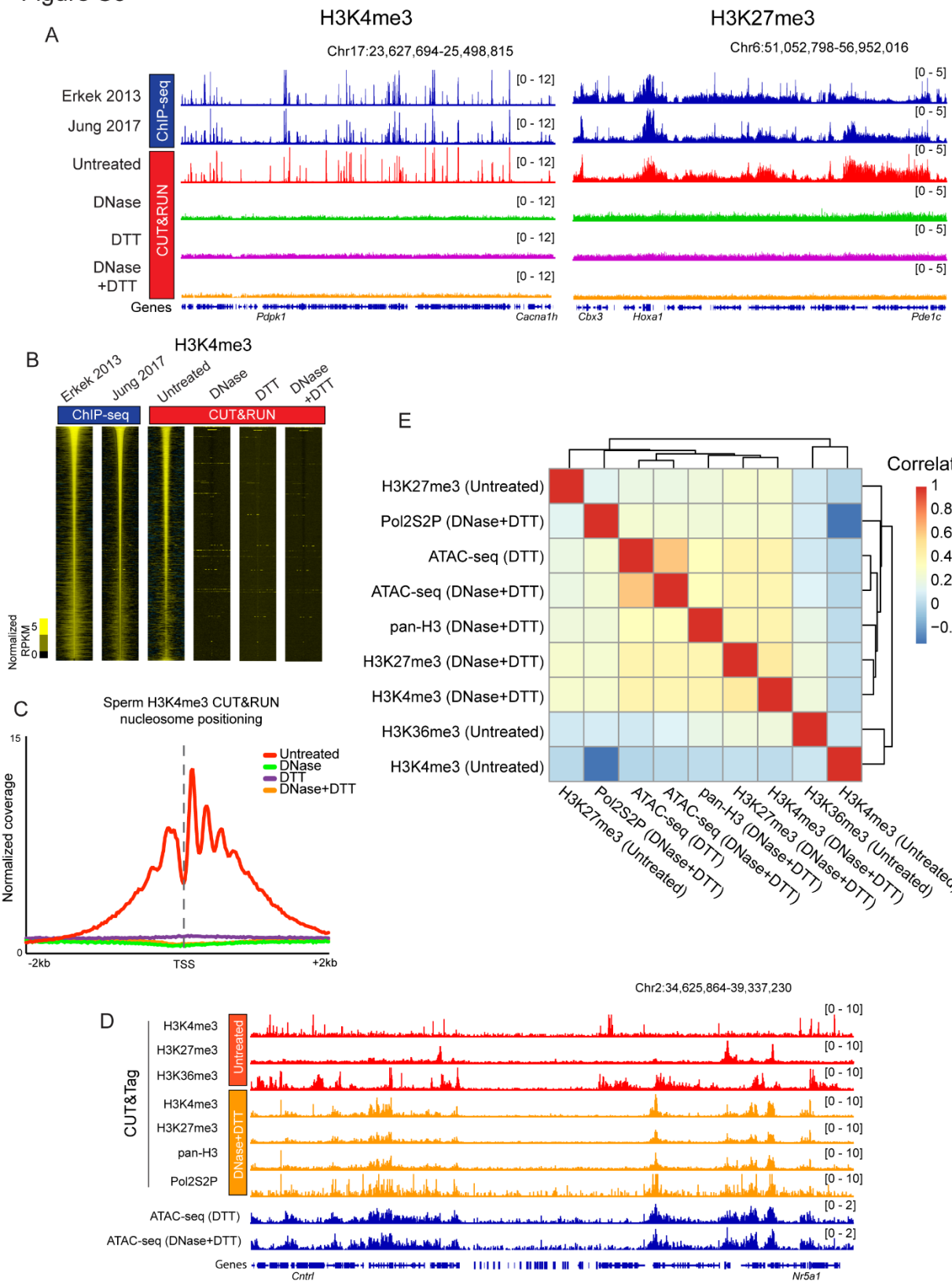
Figure S4



Supplemental Figure S4. Distinguishing methylase-accessible and inaccessible DNA by Nano-NOMe-seq

Histogram showing number of Nanopore reads (y axis) exhibiting a given level (x axis) of GpC methylation (for reads with a minimum of 20 GpC dinucleotides). Dotted line indicates divider separating inferred sperm DNA reads (protected from GC methylation) from inferred cell-free DNA (accessible to the methylase).

Figure S5


A
H3K4me3
H3K27me3
Chr17:23,627,694-25,498,815
Chr6:51,052,798-56,952,016
Erkek 2013
Jung 2017
Untreated
DNase
DTT
DNase +DTT
ChIP-seq
CUT&RUN
Genes
Pdpk1
Cacna1h
Cbx3
Hoxa1
Pde1c
B
H3K4me3
Normalized RPKM
C
Sperm H3K4me3 CUT&RUN nucleosome positioning
Normalized coverage
-2kb
TSS
+2kb
D
CUT&Tag
H3K4me3
H3K27me3
H3K36me3
pan-H3
Pol2S2P
ATAC-seq (DTT)
ATAC-seq (DNase+DTT)
Chr2:34,625,864-39,337,230
Cntrl
Nr5a1
E
Correlation
H3K27me3 (Untreated)
Pol2S2P (DNase+DTT)
ATAC-seq (DTT)
ATAC-seq (DNase+DTT)
pan-H3 (DNase+DTT)
H3K27me3 (DNase+DTT)
H3K4me3 (DNase+DTT)
H3K36me3 (Untreated)
H3K4me3 (Untreated)

Supplemental Figure S5. Lack of antibody specificity in sperm CUT&RUN and CUT&Tag profiles

A) CUT&RUN profiles for the indicated histone modifications, for sperm either untreated or treated with DNase I and DTT. Published ChIP-seq profiles from the indicated studies (Erkek et al. 2013; Jung et al. 2017) are shown above the CUT&RUN libraries for comparison.

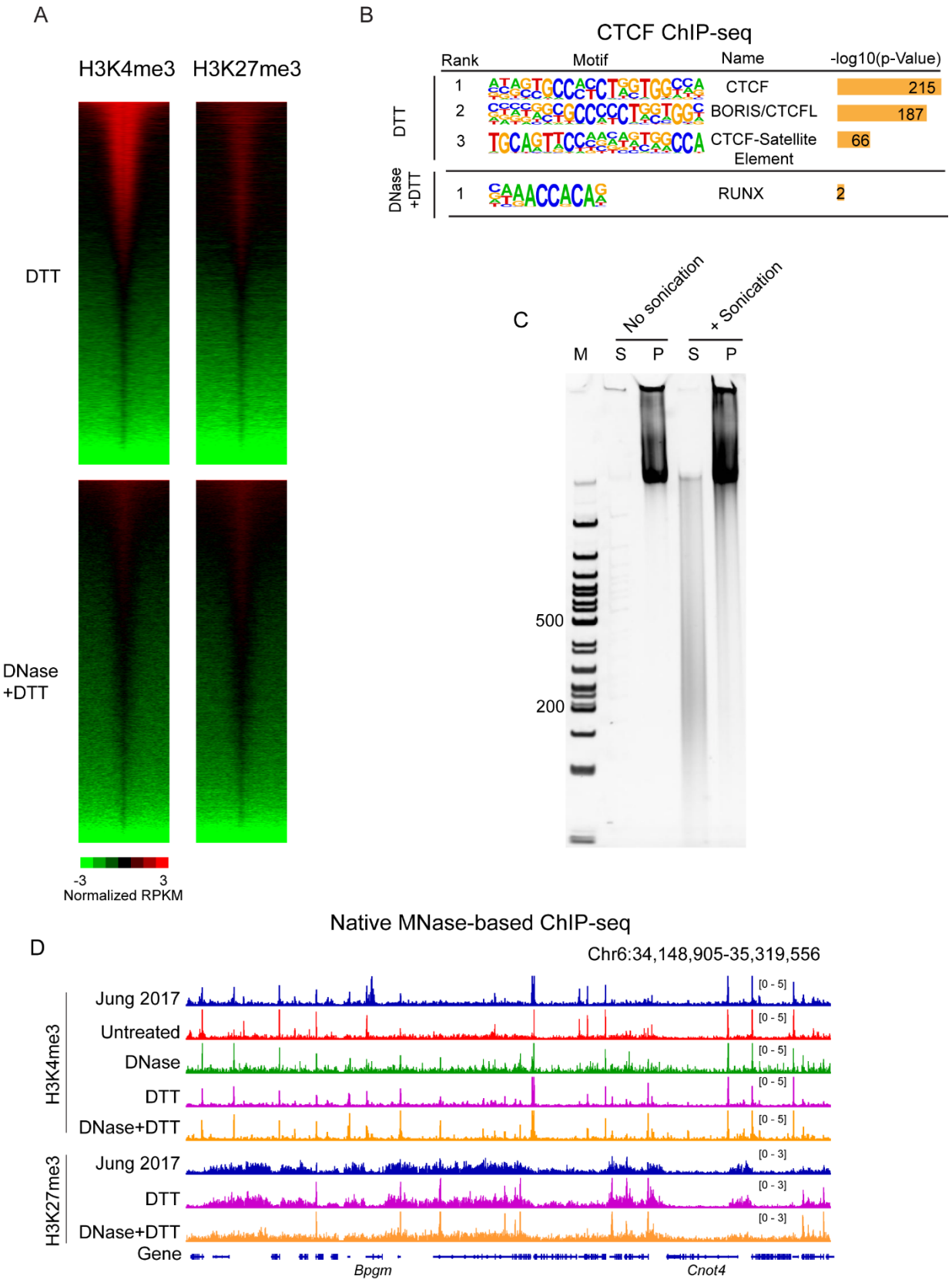
B) Heatmaps showing H3K4me3 signal for the indicated libraries, aligned over peaks called from Erkek *et al* (Erkek et al. 2013)

C) Metagene showing averaged H3K4me3 enrichment for a 4 kb window surrounding all transcription start sites.

D) As in (A), for CUT&Tag libraries. Note that four distinct epitopes, including one which should be absent from mature sperm (Pol2S2P), all exhibit highly similar localization landscapes, which also resemble DTT-treated ATAC-seq profiles.

E) Correlation matrix between CUT&Tag and ATAC libraries. Datasets for untreated sperm are all distinct, whereas ATAC-seq and CUT&Tag datasets (regardless of antibody used) for permeabilized sperm are broadly concordant, consistent with CUT&Tag data simply reflecting untargeted Tn5 activity in the CUT&Tag protocol.

Figure S6


A
H3K4me3
H3K27me3
DTT
DNase
+DTT
-3
3
Normalized RPKM
B
CTCF ChIP-seq
Rank
Motif
Name
-log10(p-Value)
DTT
1
CTCF
215
2
BORIS/CTCFL
187
3
CTCF-Satellite Element
66
DNase +DTT
1
RUNX
2
C
No sonication
+ Sonication
M
S
P
S
P
500
200
D
Native MNase-based ChIP-seq
Chr6:34,148,905-35,319,556
H3K4me3
Jung 2017
Untreated
DNase
DTT
DNase+DTT
H3K27me3
Jung 2017
DTT
DNase+DTT
Gene
[0 - 5]
[0 - 3]
Bpgm
Cnot4

Supplemental Figure S6. Features of sperm ChIP-seq datasets

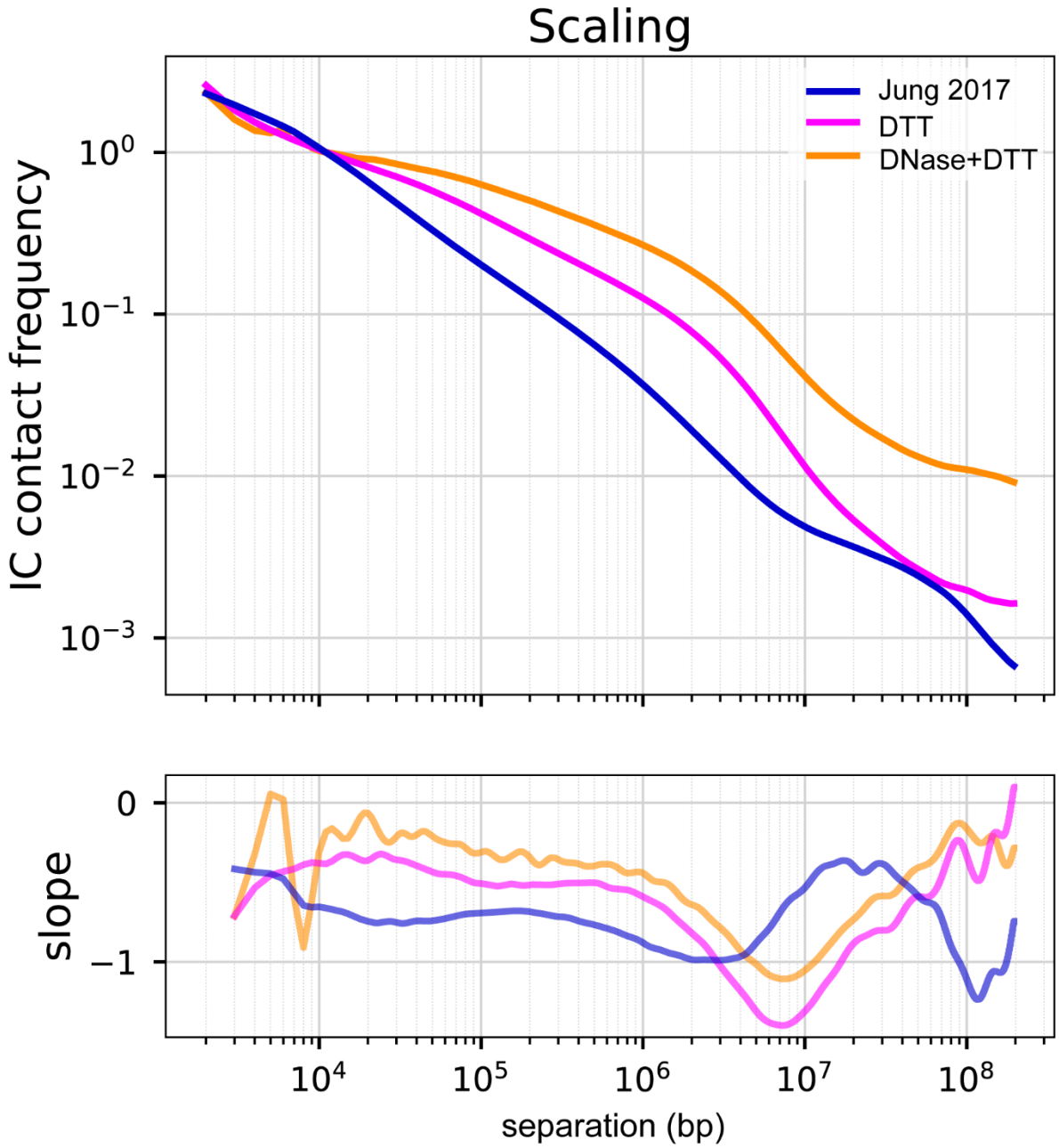
A) Heatmaps of H3K4me3 and H3K27me3 ChIP-seq data for the DTT only and DNase+DTT libraries, aligned over all promoter regions (n = 57,102;TSS +/- 2.5 kb), organized from high to low enrichment of H3K4me3 signals.

B) Sequence logos for enriched motifs at CTCF ChIP-seq peaks called from DTT-only and DNase+DTT libraries, as indicated.

C) Limited solubilization of the sperm genome following sonication. Formaldehyde-fixed sperm were sonicated to fragment the genome for ChIP-seq. Following sonication, resulting material was centrifuged (10,000×g, 10 min) and DNA was extracted from pellet and supernatant fractions and characterized by gel electrophoresis.

D) Native MNase-ChIP-seq for H3K4 and K27 methylation. Data for untreated sperm closely match previously-reported histone modification profiles; curiously, for this assay – unlike for other histone modification mapping protocols – we find essentially no change following DNase or DTT treatments. This includes DNase treatment without DTT permeabilization, which is concerning given that we should be unable to access the sperm genome under these conditions. We speculate that DNase treatment of cell-free chromatin leaves some nucleosomes intact, which are available for immunoprecipitation but which do not have adjacent linker DNA required for Tn5 insertion in CUT&Tag, or which add to a genome-wide (**Supplemental Fig. S2G**) nonspecific background for CUT&RUN where MNase-released nucleosomes are used for readout.

Figure S7



Supplemental Figure S7. Hi-C scaling plots

Top panel shows interaction frequency between two loci at increasing genomic distances (x axis) for the indicated datasets. Bottom panel shows the derivative of these curves. The relatively flat profile seen in the DNase+DTT dataset is typical for poorly-crosslinked Hi-C libraries, potentially owing to the paucity of lysines in sperm DNA-associated proteins like the protamines.

We note that while the majority of published Hi-C studies of mouse sperm exhibit typical somatic features including A/B compartments and TADs (Battulin et al. 2015; Jung et al. 2017; Ke et al. 2017; Alavattam et al. 2019; Jung et al. 2019; Wang et al. 2019), TADs were notably absent in (Vara et al. 2019). Although this would be explained parsimoniously had DTT been used in that study – generating maps dominated by bona fide sperm rather than contaminating chromatin, as seen in our **Fig. 6** data – this does not appear to have been the case based on the Vara *et al* Methods section.

Instead, two technical choices may provide a potential explanation. First, Vara *et al* state that testis, epididymis, and epididymal sperm were co-incubated during testicular dissociation, conditions which include DNase I treatment (DNase treatment is a typical part of testicular dissociation protocols). Of course this would eliminate the cell-free chromatin contamination present in the cauda epididymis, as demonstrated throughout this manuscript, but raises the question of what material was captured in the resulting Hi-C libraries. This highlights the second technical choice, where Vara *et al* snap freeze FACS-sorted sperm prior to all fixation and permeabilization steps. In our experience, MNase-seq data are dramatically altered when the assay is performed on frozen and thawed sperm vs. freshly isolated sperm used immediately (unpublished). We speculate that this reflects ice damage to sperm frozen without cryoprotectant, potentially allowing later enzyme access to fractured sperm heads. In this scenario Vara *et al* would be able to generate Hi-C maps, free of cell free chromatin, from sperm broken by freezing.

REFERENCES

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. *Nature structural & molecular biology* **26**: 175-184.

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 biology* **16**: 77.

Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schubeler D, van der Vlag J, Stadler MB, Peters AH. 2013. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**: 868-875.

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 genetics* **17**: e1009416.

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. *Molecular cell* **75**: 154-171 e155.

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 reports* **18**: 1366-1382.

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 e320.

Vara C, Paytuvi-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salva-Castro J, Gomez HL, Julia E, Moutinho C, Aiese Cigliano R et al. 2019. Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis. *Cell reports* **28**: 352-367 e359.

Wang Y, Wang H, Zhang Y, Du Z, Si W, Fan S, Qin D, Wang M, Duan Y, Li L et al. 2019. Reprogramming of Meiotic Chromatin Architecture during Spermatogenesis. *Molecular cell* **73**: 547-561 e546.