Supplementary Materials

Materials and Methods

Mice

All mouse (*Mus musculus)* experiments were carried out with approval from the University of Michigan Institutional Committee on Use and Care of Animals (PRO00006047, PRO00008135, PRO000006047, PRO00008135, PRO00010000, PRO00011691), in accordance with the guidelines established by the National Research Council Guide for the Care and Use of Laboratory Animals. Mice were kept at the University of Michigan animal facility in an environment with controlled light (12 h on/off), temperature (21-23 °C), and humidity (30–70%), with ad libitum access to water and food (Lab Diet no. 5008 for breeding mice, no. 5LOD for non-breeding animals).

$PRM1^{+/V5}$ mice, $PRM2^{+/V5}$ mice were generated on a mixed genetic background (BL6 / SJL hybrid) using CRISPR–Cas9-mediated genome editing. Transgenic mouse models were generated by the University of Michigan Transgenic Animal and Genome Editing Core Facility. All animal procedures were carried out in accordance with guidance from the Institutional Animal Care and Use Committee and approval from the University of Michigan Medical Center. All mice were backcrossed to the C57BL/6J background for at least 6 generations.

Assessment of fertility of mutant mouse models.

To assess fertility, epididymal sperm were collected in 2ml of Donner's media (25 mM $NaHCO_3$, 20 mg/mL Bovine Serum Albumin, 1 mM sodium pyruvate, 0.53% vol/vol sodium DL-lactate in Donner's stock)[1] for 30min at 37°C. Briefly, the collected sperm were used to assess sperm count, motility, and morphology. Sperm count and motility were measured using a Makler chamber. Measurements were taken in triplicate and >100 sperm counted for motility.

Fertility assessment:

Fertility assessment was done as previously described[2]. Briefly, male mice (8 weeks old) were individually housed for 3 days before mating. Each mail was then paired with three females for three independent replicates per genotype. Females were checked daily for copulatory plugs and once detected, were separated into individual cages. Fertility was assessed by recording the number of females successfully impregnated.

Immunofluorescence (IF) and seminiferous tubule staging

Testes were collected and fixed in 4% PFA in 4°C 4°C overnight with rotation, washed in 1X PBS overnight, and stored in 70% ethanol prior to formalin-fixed paraffin embedding (FFPE). Embedded tissues were sectioned into five-micrometer-thick tissue sections and deparaffinized and permeabilized in 0.1% triton X-100 in PBS for 15min. Tissue were then subjected to antigen retrieval by boiling in 10 mM sodium citrate, pH 6.0, for 30 min, followed by 20 min of cooling at room temperature while submerged in the sodium citrate solution. Sections were blocked in 3% BSA in 1X PBS for one hour, followed by incubation in wash buffer (3% BSA, 500 mM glycine in 1X PBS) for two hours and a 10-minute block in Bloxall. Primary antibodies were applied overnight at 4°C. Sections were washed four times in PBST (0.1% BSA, 0.05% Tween-20 in PBS) prior to incubation with secondary antibodies (1:500; Life Technologies/Molecular Probes). PNA-lectin (1:500; GeneTex) and DAPI were added simultaneously with secondary antibodies to label the acrosome and total nuclear DNA, respectively.

For the detection of mouse primary antibodies against Hup1n, Hup2b, and Tnp2, immunofluorescence was performed using the M.O.M. Mouse on Mouse Kit (Vector Laboratories) according to the manufacturer's instructions, with the following modifications: primary antibody incubation was extended to overnight, and sections were washed four times in PBST prior to secondary antibody addition.

Seminiferous tubule stages were assigned to cross-sections based on PNA-lectin staining patterns of the spermatid acrosome and the complement of cell types present within each tubule, categorizing tubules into stages VIII, IX–XII, I–III, and IV–VI as previously described.[3]

Antibodies

For IF staining, antibodies were used at the following dilutions, Hup1n: 1:100, Hup2b:1:100, mouse V5: 1:500, rabbit V5: 1:100, TNP1: 1:100, TNP2: 1:50, histone H3: 1:500, histone H4ac: 1:500.For western blot, PRM1 custom ab was used at 1:500 and histone H3 at 1:1000. For Cut&Tag, antibodies were used in the following dilutions: H4: 1:100, H2B: 1:50, H4ac:100, H3K27me3:1:100.

Catalog no of antibodies used:

Antibodies	Catalog No:
V5 (mouse)	MCA1360
V5 (rabbit)	V8137
Hup1n	MAb-Hup1N-150
Hup2b	MAb-Hup2B-150
Tnp1	17178-1-AP
Tnp2	sc-393843
H3	ab1791
H4ac	06-866
H3k27me3	13-0055
H2B	07-371
H4	16047-1-AP

Synchronization of spermatogenesis in juvenile mice

Spermatogenesis was synchronized in male pups by feeding them a retinoic acid inhibitor WIN 18446, starting at 2 days post-partum for 7 days followed by an injection of retinoic acid dissolved in DMSO, as previously described [4] with slight modifications. Retinoic acid dosage was reduced from 200ug to 100ug to reduce retinoic acid-mediated toxicity. Testes were collected beginning with the second wave of spermiogenesis from day 31 onwards. For each animal, 1.5 testis was processed for experimental analysis while the remaining ½ was reserved for stage verification. Synchronization was confirmed by PNA-lectin staining of acrosome morphology on cross-sections of the reserved testis, ensuring that all samples collected and analyzed corresponded to the intended spermatogenic stage.

Salt fractionation of chromatin

Testes were synchronized as described above and collected from the second wave of spermiogenesis, after retinoic acid injection. Snap-frozen testes were homogenized in a Dounce homogenizer, washed with PBS, and incubated with 0.1% CTAB on ice for 5 minutes to remove flagella. Cells were washed twice in 50 mM Tris-HCl (pH 8.0) and lysed on ice for 15 minutes,

then centrifuged at 800g for 15 minutes. The supernatant was retained as the cytoplasmic fraction. Pelleted nuclei were digested with 1 U MNase at 37°C for 30 minutes, quenched with EGTA, and centrifuged at 400g for 10 minutes; the supernatant was collected as the MNase fraction[5]. Remaining chromatin was subjected to sequential washes with 0.5 M, 1 M, and 2 M NaCl for 30 minutes each[5], with supernatants collected after centrifugation at 400g for 10 minutes following each wash. Proteins in each fraction were precipitated by addition of 20% TCA and overnight incubation at −20°C with rotation, pelleted at 12,000g for 10 minutes at 4°C, and resuspended in water prior to immunoblotting.

Isolating spermatids from specific stages for ATAC-seq:

Single cell dissociation: Testes from adult mice were collected from adult mice as previously described[6,7], and the tunica albuginea was removed. S Seminiferous tubules were transferred to 10 ml of Digest Buffer 1 (Advanced DMEM: F12 supplemented with 2 mg/ml Collagenase IA (Sigma) and 0.2 mg/ml DNase I (Worthington Biochemical Corp)), inverted 5–10 times, and incubated horizontally at 35°C with shaking at 215 rpm for 5 minutes. Tubules were allowed to settle for 1 minute at room temperature and the supernatant was discarded. Tubules were then resuspended in 10 ml of Digest Buffer 2 (0.5 mg/ml Trypsin and 0.4 mg/ml DNase I in Advanced DMEM: F12), inverted 5–10 times, and incubated at 35°C with shaking at 215 rpm for 5 minutes. Tubules were further dissociated by pipetting 3–5 times with a 1 ml pipette tip, and trypsin activity was quenched by addition of 3 ml fetal bovine serum (FBS; Sigma). The cell suspension was filtered through 70 micrometer strainer and centrifuged at 300g for 5min at 4 °C. Cells were then resuspended in 6ml of MACS buffer containing 0.5% BSA (Miltenyi Biotec).

Fluorescent activated cell sorting: Dissociated cells in MACS buffer containing 0.5% BSA were quantified using a hemocytometer. 0.06µL (0.6µg) of Hoechst 33342 dye was added per 1million cells alongside 4µL of Syto 16 dye. Cells were shaken in the incubator at 35 °C/215rpm for 20min. An equal amount of Hoechst 33342 and Syto16 dyes was added again, and cell suspensions were shaken in the incubator at 35 °C/215rpm for another 20min. Cells were then centrifuged at 300g for 5min at 4 °C and resuspended in 6-10ml of MACS buffer. 10ul of Draq7 dye was added, and cells were filtered through 70 micrometer filters and used for FACS. Cells were gated as previously described in Gill et al. 2022[8] to sort for round spermatids and elongating spermatids.

ATAC-seq experiments and sequencing

ATAC-seq was performed as described in Corces et al. 2017[9] with minor modifications. Briefly, FACS-sorted spermatids were centrifuged at 500g for RS and 2500g for ES for 10 minutes at 4 °C. 220,000 cells were resuspended in DnaseI buffer (500ul 1X HBSS, 2.5ul of 1M MgCl2, 0.5ul of 1M CaCl2) with 2.5ul of 20mg/ml DNaseI. Cells were treated at 37 °C for 20min. Cells were washed with 1X PBS twice during the second wash, 10% Drosophila spike-in cells were added prior to centrifugation. Cells were resuspended in ATAC resuspension buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$) supplemented with 0.1% IGEPAL, 0.1% Tween-20, and 0.01% digitonin, and monitored for lysis by trypan blue exclusion. C Lysis was quenched by addition of 900 µl ATAC resuspension buffer containing 0.1% Tween-20, and nuclei were pelleted at 2,500g for 10 minutes at 4°C. Transposition reaction was performed by resuspending nuclei in transposition mix (50 µl 2X TD buffer, 5 µl transposase, 33 µl PBS, 1 µl 1% digitonin, 1 µl 10% Tween-20, 10 µl H_2O) and incubating at 37°C for 30 minutes with 1,000 rpm mixing. Reactions were cleaned using Qiagen MinElute Kit according to the manufacturer's instructions. Libraries were amplified as previously described[9] with 5 additional cycles in the final amplification step. Final libraries were purified with the Qiagen MinElute Kit, and further size-selected using 1.5X AMPure XP beads. Final libraries were sequenced as 150-bp paired-end reads on an Illumina NovaSeq X. Note: later-stage spermatids beyond stage 4 were very difficult to permeabilize and generate reliable samples without significantly perturbing chromatin. We

therefore opted to stop at these stages. We know DTT treatment in sperm affects the DNA binding ability of certain modified protamine molecules.

CUT&Tag experiment and sequencing

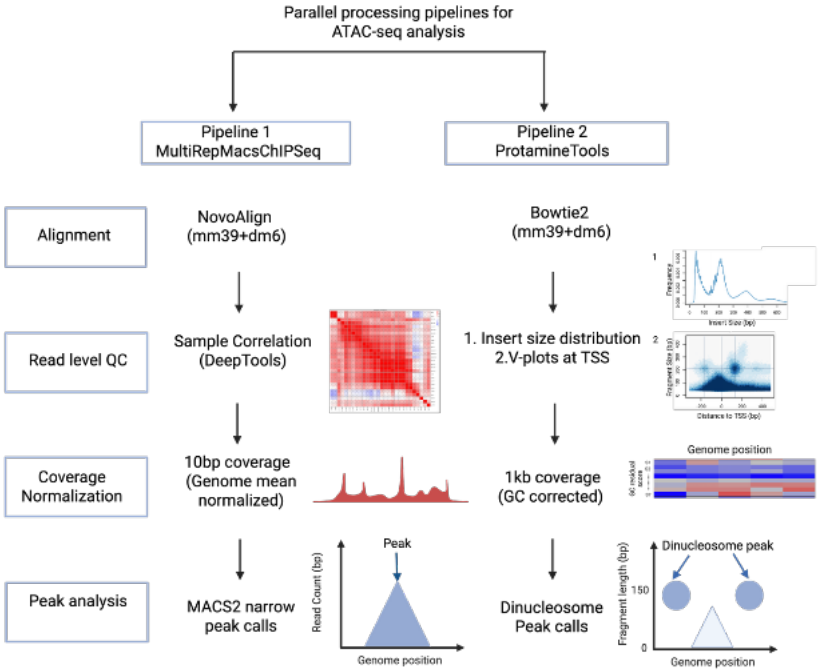
CUT&Tag protocol for spermatids was adopted from EpiCypher® CUTANA™ Direct-to-PCR CUT & Tag Protocol and Kaya Okur et al 2019[10] with modifications. Briefly, FACS-sorted spermatids were centrifuged at 500g for RS and 2500g for ES for 10 minutes at 4 °C. 220,000 cells were resuspended in DnaseI buffer (500ul 1X HBSS, 2.5ul of 1M MgCl2, 0.5ul of 1M CaCl2) with 2.5ul of 20mg/ml DNase I. Cells were treated at 37 °C for 20min. Cells were then washed twice with 1X PBS. Round spermatids were lysed with NEBuffer + 0.1% triton 10min on ice. For elongating spermatids, cells were treated with 0.05% lysolecithin for 10min. Cells were then resuspended in NEBuffer + 0.1% Triton X-100 + 0.1% digitonin and monitored for lysis using trypan blue exclusion. 100ul nuclei were aliquoted into 10ul of activated conA beads in NEBuffer + 0.1% triton and incubated together for 10min at room temperature. Nuclei were incubated with primary antibodies overnight on a nutator. Nuclei were incubated with species-specific secondary antibodies at a 1:100 dilution for 1 hour at room temperature. All subsequent steps were done according to EpiCypher® CUTANA™ Direct-to-PCR CUT&Tag Protocol. All libraries were amplified using 14 PCR cycles. Libraries were cleaned with 1.3X AmpureXP beads and sequenced with 150 bp paired end reads on an Illumina NovaSeq X.

Hi-C experiment

One million FACS-sorted round spermatids were centrifuged at 500g for 10min at 4°C. Cells were resuspended in 1ml PBS and 2% formaldehyde. All subsequent steps were according to the Arima Hi-C protocol (Arima HiC-Kit, User Guide Mammalian Cell Lines, Doc A160134 v01). Libraries were generated across 4 biological replicates and sequenced using 150bp paired end reads on an Illumina NovaSeq X.

DATA ANALYSIS:

ATAC-seq data generated across the seven key stages of spermiogenesis were processed in parallel using two different pipelines. Key differences between the two methods are outlined below and will be referred to as Pipeline 1 and Pipeline 2.



Parallel processing pipelines used for ATAC-seq analysis.

ATAC-seq analysis – Pipeline1: MultiRepMacsChIPSeq

Alignment and Normalization

ATAC-seq reads were aligned using Novoalign (version 4.04.01, http://novocraft.com) to a combined mouse (UCSC version mm39) and *Drosphila* (UCSC version BDGP6) reference genome using default-tuned parameters for NOVASEQ and including built-in adapter trimming with sequence CTGTCTCTTATACACATCT. Alignments were tagged with mate scores using Samtools fixmate (version 1.21, https://www.htslib.org/doc/1.21/samtools-fixmate.html). Alignments were split into separate genome-specific files using Picard ReorderSam (version 2.23.3, https://broadinstitute.github.io/picard/). Alignment files from samples and replicates were put through the MultiRepMacsChIPSeq pipeline (https://huntsmancancerinstitute.github.io/MultiRepMacsChIPSeq/, release 20) for fragment-based peak calling. Properly-paired alignments were filtered based on insertion length (40-300 bp), exclusion of mitochondrial and contig chromosomes, removal of optical duplicate alignments (pixel distance of 250), and partial PCR de-duplication via random subsampling to a final total duplication rate of 5%. Genome-wide fragment coverage was generated (in 10 bp bins), normalized to a depth of 1 million fragments. Coverage was scaled to the median sequencing depth of the given samples, statistical enrichment (q-value) genome-wide tracks generated using MACS2[11] and the empirical genomic mean coverage, and peaks called using MACS2[11] using a minimum peak size of 200 bp, gap of 50 bp, and minimum threshold of 0.01 (1% FDR).

Delta coverage tracks were generated by taking bedgraph file format of the coverage tracks using MACS2[11] (version 2.2.7.1) bdgcmp function to substract the genome mean of the sample, which was calculated using the custom script generate_mean_bedGraph.pl (https://github.com/tjparnell/HCI-Scripts/blob/master/ChIPSeq/generate_global_mean_bedGraph.pl). The resulting bedgraph files were visualized by converting to bigwig files. Correlation heatmap between aligned bam files

were generated using the function multibamsummary of deeptools[12] (version 3.5.4) and python (version 3.11.5) with a bin size of 10kb.

Gene ontology analysis:
MACS 2 peak calls per stage were merged using bedtools2[13] (v2.31.1) 'merge' and intersected with transcriptional start sites of mm39 reference genome using bedtools2[13] (v2.31.1) 'intersect'. The resulting unique gene names were used to perform gene ontology analysis using David Bioinformatics[14] (https://davidbioinformatics.nih.gov/workspace.html).

ATAC-seq analysis – Pipeline2: Protamine-tools
ATAC-seq data from spermatids presents distinct challenges relative to the application of the method to other cell types because chromatin states are altered on a genome-wide scale during spermiogenesis. Regions with increased insert counts can represent both typical positioned nucleosomes with localized hyper-accessible nucleosome-free regions, e.g., those associated with changing transcriptional activity, as well as extended highly accessible regions during protamine deposition. We created new tools for integrating this unique data class with other data types in the 'protamine-tools' pipeline (https://github.com/HammoudWilson/protamine-tools). The workflow is divided between stage 1 data analysis pipelines and stage 2 interactive plotting apps created using R Shiny. Detailed instructions are provided, although much of this purpose-specific codebase will only apply to similar spermatid data sets. The codebase includes normalization approaches that were initially explored but were not used in any presented figures; these are discussed more briefly below.

Read alignment to a mouse-fly composite genome

We first created a composite mouse-fly FASTA file with chromosomes from the mm39 and dm6 reference assemblies using protamine-tools actions `genome download` and `genome composite`. Chromosome names were suffixed with the genome as 'chr1_mm39', etc. Read pairs were prepared for alignment by `atac align` by trimming them to 150 bases per read followed by quality filtering and merging of overlapping read pairs using `fastp` v0.24[15] with settings `--dont_eval_duplication --length_required 35 –correction --merge --overlap_len_require 20 --overlap_diff_limit 10 --overlap_diff_percent_limit 20 --trim_poly_g --adapter_sequence CTGTCTCTTATACACATCT`. Prepared reads were aligned to the composite genome using `bowtie2` v2.5 with setting `-x -U --end-to-end` for merged single reads and `--end-to-end --maxins 650 --no-mixed --no-discordant` for read pairs from longer inserts and alignments merged to a single BAM file. Merged reads and concordant read pairs were deduplicated such that no two DNA inserts shared the same alignment endpoints. Reads were filtered to require a minimum mapping quality (MAPQ) of 30 to ensure high confidence placements.

Binned data analysis

Due to the rapidly changing accessibility landscape of large genomic regions during spermiogenesis, we first assessed ATAC-seq insert yield at low resolution by binning the composite genome into 1kb bins using the pipeline action `genome bin`. The approach sought to assess regional Tn5 accessibility. Throughput, bins in the mm39 and dm6 GENCODE blacklists[16] were masked in files as "not included" in statistics and plots. Various genome and sample level scores were collected for each bin for correlation and plotting as described below.

Raw bin counts and insert size distributions

Filtered and deduplicated ATAC-seq inserts were counted in the bin that contained their leftmost aligned ends by pipeline action `atac collate`. Insert counts were normalized to reads per kb per million reads (RPKM) for inter-sample comparison. RPKM is synonymous with count per million reads (CPM) at 1kb bins, but bins were often aggregated further below. Due to the relative sizes of bins and inserts, especially aggregated bins, most inserts were fully contained in one bin or region. Some inserts crossed into adjacent bins, which was tracked during next steps.

Insert size distributions were assembled per sample library and per genome, so that insert sizes could be compared as a function of both spermatid stage and origin from the mouse vs. Drosophila spike-in DNA.

Mappability and other insert size-dependent scores

An unusual feature of our ATAC-seq data is that the sequenced insert sizes vary considerably across different spermatid stages, leading to potential latent insert size-dependent biases during sample comparison. We assessed the mappability of the composite genome as a function of insert size from 35 to 650 bp, the range used for ATAC-seq analysis, using action `genome mappability` in a manner that considered insert size, since small inserts are inherently less mappable in repetitive sequences. Insert sizes were stratified into twenty groups from beginning at 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300 bp. The `genmap map` utility, v1.3[17], was used to calculate the unique mappability of inserts at the smallest insert size in each group with settings `-K <insert_size>-E <n_errors> -bg`, where n_errors was 1 for insert_size <= 40, 2 for insert_size <= 100, 3 for insert_size <= 200, and 4 for insert_size >= 200, to allow for increased numbers of sequencing errors in longer inserts. Bin mappability was calculated on a per sample basis as an average of the mappability of the different insert size levels weighted by the sample insert size distribution. A bin's raw insert count was divided by the dynamically calculated mappability to yield adjusted insert counts. Bin counts were rescaled to the same summed count after mappability correction to avoid inflating observed counts.

We explored using insert size profiles as a primary bin score. Specifically, we used samples from early RS and intermediate ES stages to establish emission probabilities for nucleosome-bound and transitional and/or protamine-bound states, respectively. We calculated the relative log likelihood for each bin for the early and late-stage emission profiles and normalized it to the bin count. The resulting normalized relative log likelihoods (NRLL) were visualized in R Shiny but not depicted in figures as the information proved redundant and less stable than other scores.

GC bias normalization and GC-related bin properties

Tn5 libraries have a notable bias in mapped read recovery based on local base content. The extent of this GC bias differed significantly between ATAC-seq samples collected at different spermatid stages, in part due to their different insert size distributions since smaller inserts tend to be more GC rich. For low-resolution analysis this technical effect becomes a substantial determinant of bin coverage, although some component of true biological differences between samples may correlate with GC content.

The GC base content of bins was assessed for DNA inserts that could productively map to each bin for each insert size group and a net effective GC value calculated for each sample-bin as for mappability above. We correlated mappability-adjusted bin insert counts to this bin GC content and performed regression analysis using the negative binomial distribution using R expression `MASS::glm.nb(formula, control = glm.control(maxit = 100))`, where formula was `readsPerAllele ~ fractionGC + fractionGC2` and `fractionGC2` was the square of the bin GC fraction to achieve a quadratic fit of the trend toward higher coverage at higher GC content and the degree of count over-dispersion. We calculated the "GC residual Z-score" (GCRZ) as the deviation from the trend line in Z units by projecting the negative binomial quantiles onto a normal distribution. Thus, samples with a high GCRZ were overrepresented relative to peer bins with a similar insert size-adjusted GC content and runs of bins with high GCRZ scores were judged to be enriched independently of bin GC content. For regression we only considered bins with GC fraction between 0.25 and 0.6 to prevent outlier effects from rare bins with extreme values.

Two additional factors were considered as contributors to the strong GC bias in spermatid ATAC-seq. CpG dinucleotides are more frequent in higher GC content bins. Indeed, the relative

insert yield in bins with high CpG content changed substantially as spermiogenesis progressed and signal shifted away from gene promoters toward the rest of the genome. However, this effect was not a substantial contributor to the changing GC bias across stages since bins with high CpG content are a small fraction of the genome and those bins went against the trend of higher GC preference in later spermatid stages. Tn5 also has a well-described base preference, although not a strict requirement, that leads to non-random base utilization at insert ends[18]. Its preferred motif is GC-rich, an effect which becomes increasingly important as insert size decreases. Pipeline action `atac sites` used observed inserts across all samples to establish a discontinuous 9-based Tn5 preference motif at the bases with the strongest Tn5 preference, 5'-**|*-***-***, where* is a preferred base and | is the cleaved bond. Bins were assigned weights based on their relative enrichment for preferred sites, and the GC bias fit was repeated with weight-adjusted insert counts. GC bias was lower after independently accounting for Tn5 site preferences, but the net results did not change, presumably because the unweighted analysis already accounted for this effect.

Transcription level of bins and annotated transcription start sites

We used mouse round spermatid PRO-seq data[19] , obtained as BigWig files from GEO series GSE228452, as our baseline to understand the transcription levels of genes and bins. Data representing the 3' terminus of PRO-seq reads were lifted over from the mm10 to mm39 reference assemblies, aggregated over both strands, counted per bin, and normalized to yield log10 RPKM by pipeline action `nascent bin`. Bins with PRO-seq signal below a log10 RPKM of -3 were having no evidence of transcription. Separately, we examined the 5' ends of genes in the GENCODE M36 mouse genome annotation, i.e., putative transcription start sites (TSS), for evidence of transcriptional activity and positioned nucleosomes to use as references points for chromatin changes during spermiogenesis. TSSs were scored per-sample as "active" by pipeline action `nascent tss` when the 1kb downstream of the 1st base of the gene had a PRO-seq RPKM of at least 0.5 and there was at least a 10-fold increase in RPKM when compared to the 1kb upstream of that base. Genes whose 1st 1kb had RPKM < 0.05 were considered inactive for TSS assessments. For active TSS, we attempted to find evidence for positioned a positioned nucleosome just downstream of the TSS using pipeline action `atac tss` when there were at least 25 mononucleosome-sized ATAC-seq inserts (150 to 275 bp) covering the region within 250 bp of the gene start. The position of the inferred position nucleosome was taken as the median of the insert midpoints. Metadata plots over all active or inactive TSS were registered based on either the position of the TSS or the inferred nucleosome center.

CUT&Tag score integration

BigWig files from H2B, H4, H3K27Me3 and H4Ac CUT&Tag data were handled in a manner like PRO-seq data by summing insert counts over both strands per bin in pipeline action `cuttag score`. Here, for some histone targets many or most bins had an RPKM of zero.

Bin aggregation to lower resolution plots (horizontal aggregation)

Many plots and analyses compared wider genome regions at considerably lower resolution than 1kb bins. Horizontal bin aggregation was achieved by summing the counts or averaging the normalized scores, such as RPKM, of the 1kb bins that crossed each pixel in the final image, weighted by the fraction of the bin overlapping that pixel. Thus, each pixel faithfully reports the aggregated data it contained at the resolution of the image.

Sample aggregation into spermatid stages and stage types (vertical aggregation)

Working from prior knowledge of the biological properties of the spermatid samples and visual examination of ATAC-seq data, we defined seven time-ordered stages of spermiogenesis: early round spermatids (RS; early_RS), intermediate RS (int_RS), late RS (late_RS), earliest elongating spermatids (ES; earliest_ES), early ES (early_ES), intermediate ES (int_ES), and late ES (late_ES). RS stages showed differences mainly in the location of typical chromatic ATAC-seq peaks, earliest_ES behaved as a transition stage, and the later ES stages showed large

differences from typical ATAC-seq manifest as broadly increased accessibility and smaller insert sizes. Accordingly, we further defined two stage types, RS and ES, that grouped the stage types with matching round/elongating labels. To simplify plots, we performed a vertical aggregation of at least two samples per stage or stage type by summing ATAC-seq insert counts prior to making score calculations by the same methods as for individual samples. We finally defined a stage type delta as RS – ES to yield a single score per bin with high values when RS had more signal than ES.

Stage mean calculation

We devised a further way of assessing bin scores, especially ATAC-seq RPKM, that considered all samples together in a single bin or region to reveal which stages along the timed series were most enriched for higher scores. For this "stage mean" metric, we indexed the spermatid stages from 1 (early_RS) to 7 (late_ES) and took an average of the stage indices weighted by the average score values of each stage's samples. Thus, low and high stage means indicate that early or late-stage samples, respectively, showed higher signal relative to the rest of the genome in that bin or region. Because mappability and GC biases are largely the same across samples in a bin or region (aside from insert size-dependent effects noted above), stage mean values are robust to local genomic effects and provide a reliable score for the timed changes occurring during spermiogenesis. Stage mean scores cannot distinguish U-shaped patterns; higher scores in both early and late stages will give a middle stage mean value. However, most bins showed either decreasing or increasing scores, or peaks centered on a specific stage, where stage mean is a faithful reporter.

Normalization for heat map plotting

Heatmap plots used a color scheme where red colors were "hot" and correlated with increased signal, or greater signal in round vs. elongating spermatids, and blue colors were "cold" and the opposite of red/hot. Some scores were inherently centered and scaled and required no further normalization, including GCRZ and Hi-C compartment scores. PRO-seq log10 RPKM represents a varying intensity of positive signals so only red/hot colors were used to reflect the level of local transcription up to a fully red color at a log10 RPKM of 10. For other scores, heatmap plots were normalized using Z-scores of normally distributed score values, e.g., bin GC fraction, or as quantiles of non-normal score distributions centered around the median. Importantly, most ATAC-seq, CUT&Tag and Hi-C scores are relative values across the genome.

Base-resolution positioned dinucleosome peak calling

Most spermatid stages also revealed changing local accessibility peaks associated with positioned nucleosomes that we sought to track. Many algorithms are challenged in finding these peaks in samples with much broader accessibility, where accumulations of mapped inserts were often called as peaks that did not represent the same biological mechanisms as accessibility peaks at promoters, enhancers, and other functional genomic elements. We developed an efficient base-level algorithm in Rust to locate "dinucleosome peaks" characterized by (i) two nucleosome-size spans frequently crossed by longer ATAC-seq inserts but with few endpoints mapping within them, and (ii) intervening and flanking nucleosome-free regions with a higher density of insert endpoints and coverage by smaller inserts. Such peaks are distinct from other generally hyper-accessible regions in showing a specific insert pattern that is evident on V plots and characteristic of genome elements that coerce nucleosomes to specific locations.

Peaks were called *ab initio* working only from ATAC-seq insert endpoints. The algorithm finds the most likely positions of positioned dinucleosomes in local regions, if there were any, and filters those positions against minimal criteria for dinucleosome peak calling. Working per chromosome, three base-level maps of inserts locations are accumulated over all samples of a specific spermatid stage. The first map counts the number of insert endpoints at every base. The second counts the centers of mononucleosome-size inserts from 147 to 320 bp. The third counts the centers of dinucleosome-size inserts from 320 to 485 bp, which are expected to fall in the

nucleosome-free regions between two positioned nucleosomes. A series of counting masks represent different candidate gap lengths of the intervening nucleosome-free region from 51 to 351 bp with a step size of 4. Nucleosomes occupy 147 bp flanking each gap, and an additional 75 bp are included on each nucleosome flank as regions expected to have high endpoint counts.

For every third base along the chromosome, the insert maps and counting masks are used to generate a positive score where (i) insert endpoints within nucleosome-free bases are counted with a weight of 1, (ii) mononucleosome-size insert centers in nucleosome-bound bases are counted with a weight of 2 (the weight of two endpoints), and (ii) dinucleosome-size insert centers in nucleosome-free bases are counted with a weight of 1 (making them less important than mononucleosomes). In this way, many inserts will not be counted when the gap length is sub-optimal or nucleosomes are not positioned as endpoints and centers will not conform to the mask. The highest scoring combination of genome position and dinucleosome spacing establishes a local maximum that is called as the first peak. No other combination in the vicinity that would override the gap of the first, highest scoring combination is considered further. The portions of the chromosome flanking each peak are then analyzed recursively in the same way until no additional high-scoring peaks remain.

After calling peaks for a spermatid stage, adjacent peaks that overlap in their nucleosomes or flanking nucleosome-free regions are merged into a "nucleosome chain", i.e., a span with three or more positioned nucleosomes, by taking an average of the nucleosome positions and gap lengths weighted by the peak scores. Chains grow iteratively until no more overlaps are found. Chains are then intersected between all samples to find "peak regions" as the widest coordinate span of all overlapping chains from different stages. RPKM and stage mean values are calculated for all samples for each chain called in individual stages as well as for the overlap groups, yielding the signal in each stage relative to all peaks called by one or more samples. A final level of filtering retains only those peak regions with a maximum RPKM in any stage of 2.0 and a difference from the lowest to the highest stage RPKM of at least 0.75 times the maximum stage value.

Peak cluster/trajectory analysis

To assess whether peak regions formed distinct clusters or conformed better to a continuous trajectory of types, we used the umap2 function from the R uwot package to perform various types of UMAP analysis of peak region values for each stage. One UMAP type used stage RPKM values and Euclidean distances and thus took ATAC-seq magnitude into account. For other analyses, region RPKM scores were scaled as the log2 fold change from the mean RPKM and analyzed using both Euclidean and correlation distances. All outputs were consistent with a continuous trajectory of peak region types that correlated well with stage mean.

Peak lag analysis

We performed lag analysis to understand how peak regions correlated to each other when they co-localized in the genome. We calculated scaled log2 fold RPKM values as above and calculated the degree of variance between peak regions as a function of the binned log10 distance between them. Lower variance at lower inter-peak distances reveals the degree of auto-correlation of nearby peaks and the distance over which it manifests.

Re-analyzing bins without dinucleosome peaks

A possible confounder of the 1 kb bin analyses above is that bins overlapping hyper-accessible peaks might behave differently than the rest of the genome. To assess the magnitude of this effect we used pipeline action `atac recollate` to repeat the bin-level analyses after masking peak-containing bins as "not included". Doing so had little impact on binned results because there are many more non-peak than peak-containing bins in the genome.

Code availability

The protamine-tools pipeline is available on GitHub at https://github.com/HammoudWilson/protamine-tools. Within the repository, (i) the 'templates' folder contains example job files to support application of the pipeline code to other similar data sets, (ii) the 'resources' folder contains a small files needed to run the pipeline, including the compiled Rust executable for ab_initio peak calling, and (iii) the 'logs' folder contains files with complete job file configuration and logged output from jobs as we ran them. No additional installed programs are required to run the protamine-tools pipeline as it contains commands to build a conda environment with appropriate program versions.

Comparison between peak calls from pipeline 1 and pipeline 2

The number of peaks called from methods discussed above were plotted as bar plots using ggplot2 geom_bar function in R. Overlapping peaks between the two methods were identified using the findoverlaps function from the package GenomicRanges and visualized as Venn diagrams plotted using the R package called VennDigram. Peaks were assigned to different genomic regions using the annotatePeak function from the package ChIPseeker in R.

ATAC MACS2 and Di-nucleosome Heatmap

To visualize temporal chromatin accessibility dynamics, a union peakset was generated by merging MACS2 narrowPeak files across seven developmental stages (early RS to late ES) and reducing overlapping regions using the GenomicRanges[20] (v1.58.0) R package. GC-residual z-score signal intensities were extracted by overlapping these union peaks with previously mentioned 1kb genomic bins and calculating the mean signal per peak. Following signal centering, the dataset was filtered to include only dynamic regions, defined as peaks with a signal range (max – min) greater than 1.0 across all stages. For visualization, a representative subset of 50,000 peaks was ordered by their stage mean. The final signal matrix was row-standardized using z-score normalization and plotted using the pheatmap (v1.0.13)[21] R package, with the color scale truncated at ±3 to emphasize relative accessibility shifts across the developmental trajectory.

To visualize chromatin accessibility dynamics using dinucleosome peak calls, adjacent peaks that overlap in their nucleosomes or flanking nucleosome-free regions are merged into a "nucleosome chain", and gap lengths weighted by the peak scores are referred to as stage mean. The combined matrix was ordered by stage mean (calculation described in earlier section) and plotted using pheatmap (v1.0.13)[21] R package

Processing CUT&Tag data:

Cut&Tag reads were aligned using Novoalign (version 4.04.01) to a combined mouse (UCSC version mm39) and *Drosphila* (UCSC version BDGP6) reference genome using default-tuned parameters for NOVASEQ and including built-in adapter trimming with sequence CTGTCTCTTATACACATCT. Alignments were tagged with mate scores using Samtools fixmate (version 1.21). Alignments were split into separate genome-specific files using Picard ReorderSam (version 2.23.3).

Alignment files from samples and replicates were analyzed using the MultiRepMacsChIPSeq pipeline (https://huntsmancancerinstitute.github.io/MultiRepMacsChIPSeq/, release 20). This included fragment filtering based on insertion length (40-300 bp), proper pairs, exclusion of mitochondrial and contig chromosomes, removal of optical duplicate alignments (pixel distance of 250), and partial PCR de-duplication via random subsampling to a final total duplication rate of 5%.

The fragment coverage bigWig tracks were generated by the MultiRepChIPSeq pipeline after filtering alignments for proper pairs, mapping quality (> 10), insert size (40-300 bp), exclusion from unwanted chromosomes (chrM and unmapped contigs) and exclusion intervals (extreme coverage over repetitive sequence elements), and fragment de-duplication (removal of all optical

duplicates and random subsampling of remaining biological/PCR duplicates to 5 or 10% of non-duplicate levels. The tracks were scaled to Reads (Fragments) Per Million (RPM).

The pipeline was run with the --independent flag to generate independent peak calls of each sample replicate. Replicate peak calls were then merged by the pipeline, retaining only those intervals called by 2 or more replicates as the final sample peak call set.

Hi-C Data Processing and Analysis

Data Processing and Quality Control

Raw sequencing data, compressed in ORA format, were decompressed using Orad (v2.7.0-c9253; Python 3.11) referencing the NCBI RefSeq assembly GCF_000001635.27. Subsequent processing was performed using a custom implementation of the Juicer pipeline[22](v1.6) adapted for a cluster environment. Reads were aligned to the mm39 reference genome (ENCODE no-alt analysis set) using BWA-MEM (v0.7.17). Ligation contacts were validated against restriction sites for Arima-HiC (motifs GATC and GANTC). PCR duplicates and reads with a MAPQ score < 30 were removed to retain only valid contact pairs.
Hi-C contact matrices were generated using Juicer Tools (v1.7.6; OpenJDK 1.8.0_452) and balanced using Knight-Ruiz (KR) normalization. Replicate files were merged using the Juicer mega.sh script to increase read depth for downstream analysis. The HiCRep reproducibility measure was run using 3DChromatin_ReplicateQC[23,24] v0.0.1 with parameters `h=5, maxdist=5000000` at 50 kb Hi-C resolution.

Data Analysis

Compartments (A/B) were identified using the Juicer Tools eigenvector command at 250kb resolution. To ensure consistency across samples, the sign of the eigenvector was oriented such that positive values corresponded to the A compartment in the ES-E14 mESC cell line[25]. This orientation and subsequent comparative plotting were performed using custom Python scripts (Python 3.11, pandas 2.3.1, NumPy 2.3.1, Matplotlib 3.10.3, SciPy 1.16.0). Genomic coordinates were converted from mm10 to mm39 using the UCSC liftOver tool with the mm10ToMm39.over.chain.gz chain file. 2D pairwise contacts of Hi-C data were visualized with Juicebox (v2.15)[26] at a 500 kb resolution with a coverage normalization, unless specified otherwise.

Assessing ATAC-seq Peaks across Stages

For each stage, we sought to assess the significance of the number of MACS2 peaks per chromosome. We constructed a null distribution by repeatedly sampling with replacement (1,000 permutations) peak files at random (numpy v2.3.1) for each chromosome and summing their counts, ensuring equal representation across stages. Observed counts were then compared to this null distribution to calculate z-scores and p-values (scipy v1.16.0), which were adjusted for multiple testing using the Benjamini-Hochberg[7] method with FDR of 0.05 (statsmodels v0.14.5). To account for correlation between stages, pairwise Jaccard indices[8,9] (pybedtools v0.12.0) between peak sets were used as proxies for correlation and we applied Brown's method[10], a correlation-corrected extension of Fisher's combined probability test[11], to evaluate the global deviation of observed data from the null. The covariance terms required for Brown's method were estimated using the Kost & McDermott polynomial approximation[12]. Results were visualized as boxplots (matplotlib v3.10.3, seaborn v0.13.2) of the null distributions with observed values and statistical significance annotated. The global p-values indicate deviance from an expected random background if there were not differences between stages. The analysis above was conducted in a python (v3.11.13) notebook using pandas v2.3.1.

Linear Mixed Model Regression Analyses Data Processing

We integrated transcription and chromatin accessibility data for the linear mixed model regression analysis. Specifically, scRNA-seq transcripts were annotated using the *Mus musculus* genome assembly (GRCm38.81). To remove gene duplicates, we kept the sources in order ensembl_havana, havana, then ensembl. If there were still duplicates, we kept the lower gene_version. We utilized PyRanges[13] (v0.1.4) join to intersect transcription data with ATAC-seq GC-corrected residuals. For each gene, chromatin accessibility was calculated as the mean signal at the bin containing the promoter the bins upstream and downstream of the promoter. Data was formatted into a longitudinal structure where each gene-stage observation constituted a single data point. The analysis above was conducted in python (v3.11.13).

Quantifying Transcription-Accessibility Coupling Strength (Fig 5a)

Data Preprocessing

Genes encoding protamines (*Prm1, Prm2*) and transition proteins (*Tnp1, Tnp2*) are hyper-upregulated during sperm packaging in the early ES stage from driving global regression trends and were resultingly removed from subsequent analysis. Additionally, inactive genes with mean expression across stages below the 25th percentile were excluded.

Statistical Modeling

We fitted a Linear Mixed-Effects Model (LMM) using the lme4 package[14] (v1.1.38) in R (v4.5.2). To assess the relationship between transcription and chromatin accessibility across stages, the model was specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 S_{2ij} + \beta_3 S_{3ij} + \beta_4 S_{4ij} + \beta_5 (S_{2ij} * X_{ij}) + \beta_6 (S_{3ij} * X_{ij}) + \beta_7 (S_{4ij} * X_{ij}) + b_{0i} + \varepsilon_{ij}$$

where

- Y_{ij} is the ATAC GC residual corrected Z-score for gene i at stage j.
- X_{ij} is the log(Transcription + 1) for gene i at stage j.
- $S_{2ij}, S_{3ij}, S_{4ij}$: Binary indicator variables taking the value 1 if observation *j* for gene *i* corresponds to stages int RS, late RS, or early ES, respectively, and 0 otherwise. (Reference stage: early RS).
- b_{0i} is the random intercept for gene *i*, accounting for baseline heterogeneity.
- β_1: The slope of the transcription-accessibility relationship for the reference stage.
- β_{5-7} : The interaction coefficients representing the difference in the transcription slope for stages int RS, late RS, and early ES relative to the reference stage.
- ϵ_{ij}: The residual error for gene i at stage j, assumed $N(0, {\sigma_\epsilon}^2)$

Significance Testing and Diagnostics

P-values for fixed effects were obtained using Satterthwaite's degrees of freedom method via the lmerTest package[15]. Model fit was assessed using residual plots and QQ-plots. Visual inspection indicated minor heteroscedasticity; however, we decided to continue with the model because parameter estimates in LMMs remain unbiased even under heteroscedasticity, though precision may be affected[16].

Data Visualization

To dissect the interaction effects, we used the emmeans package (v2.0.1, https://rvlenth.github.io/emmeans/). Specifically, we used emtrends to estimate the slope of the transcription-accessibility relationship and its asymptotic 95% confidence intervals for each stage. Subsequently, we performed pairwise contrasts of these slopes to explicitly test for significant differences in coupling strength (Δβ) between all possible stage pairs, corrected for

multiple comparisons using the Tukey method[18]. We also used emmip function to calculate marginal predicted ATAC values across the range of transcription levels.

Two types of visualizations were generated using ggplot2 (v4.0.1):

1. Marginal Prediction Plots: Fitted regression lines with confidence ribbons were plotted to visualize the global trend of chromatin coupling across stages.
2. Slope Comparison Plots: Point-range plots were generated to directly compare the estimated slopes (β) and their precision across stages.

Low vs High Expression Regression (Fig 5b)

All statistical analyses were conducted in R (v4.5.2). To analyze accessibility trajectories, genes were stratified based on their transcriptional activity in the early RS stage. "High" transcription genes were defined as the top 5th percentile, while "Low" genes were defined as those in the 25th–50th percentile range using dplyr (v1.1.4).

To control for initial differences in chromatin openness, we calculated the change in accessibility (Δ) from baseline for every gene at every timepoint ($Y_{ij} - Y_{i0}$). We employed a LMM to characterize the differential rate of remodeling over time, accounting for the developmental inflection point at the transition from Round Spermatids to Elongating Spermatids. Developmental stages were mapped to a continuous time variable t (Early RS = 0, Int RS = 1, Late RS = 2, Early ES = 3). A knot was placed at t=2 to allow for distinct rates of change before and after the onset of elongation.

The model was specified as follows:

$$\Delta Y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_j + \beta_3 (t_j - k)_+ + \beta_4 (G_i * t_j) + \beta_5 (G_i * (t_j - k)_+) + b_{0i} + \varepsilon_{ij}$$

where

- ΔY_{ij} is the change in ATAC Z-score relative to the Early RS baseline for gene i at stage j.
- G_i is the transcription group indicator (1 for High, 0 for Low).
- t_j is the time point.
- $(t_j - k)_+$ represents the linear spline function (0 before the knot k=2, and t−2 after the knot).
- $\beta_4$and β_5 capture difference in remodeling rate (velocity) between transcription groups before and after the knot, respectively.
- b_{0i} is the random intercept for gene i, accounting for gene-specific deviations from the group trajectory.

The model was fitted using the lme4 package[14] (v1.1.38). We tested the null hypothesis H_0: $\beta_4 = \beta_5 = 0$, interpreted as there is no difference in the rate of remodeling between low and highly transcribed genes using the car package's[19] (v3.1.3) linearHypothesis to calculate the Wald test statistic.

Data Visualization

Model-predicted means were generated for each group across all time points. 95% Confidence Intervals were calculated using the standard errors derived from the fixed-effects variance-covariance matrix. Visualization was performed using ggplot2 (v4.0.1).

Histone Modifications by Compartments **Supp Figure 5c/d**

A python (v3.9.23) script was used to annotate peak regions with compartment (A/B) and extract signal values from bedGraph files using pyBedGraph[20] (v0.5.43). For each chromosome, the mean ATAC signal across each peak was computed in parallel for all peaks, and each region was assigned to a compartment based on the sign of the compartment score. The resulting annotated dataframes were then used for downstream statistical analysis and visualization of signal distributions by compartment. For each stage, we performed a two-sided Kolmogorov-Smirnov test[21] (scipy v1.13.1) to assess differences between compartments, displaying the results on boxplots and Empirical Cumulative Distribution Function plots with annotated sample sizes and p-values using seaborn (v0.13.2) and matplotlib (v3.9.4).

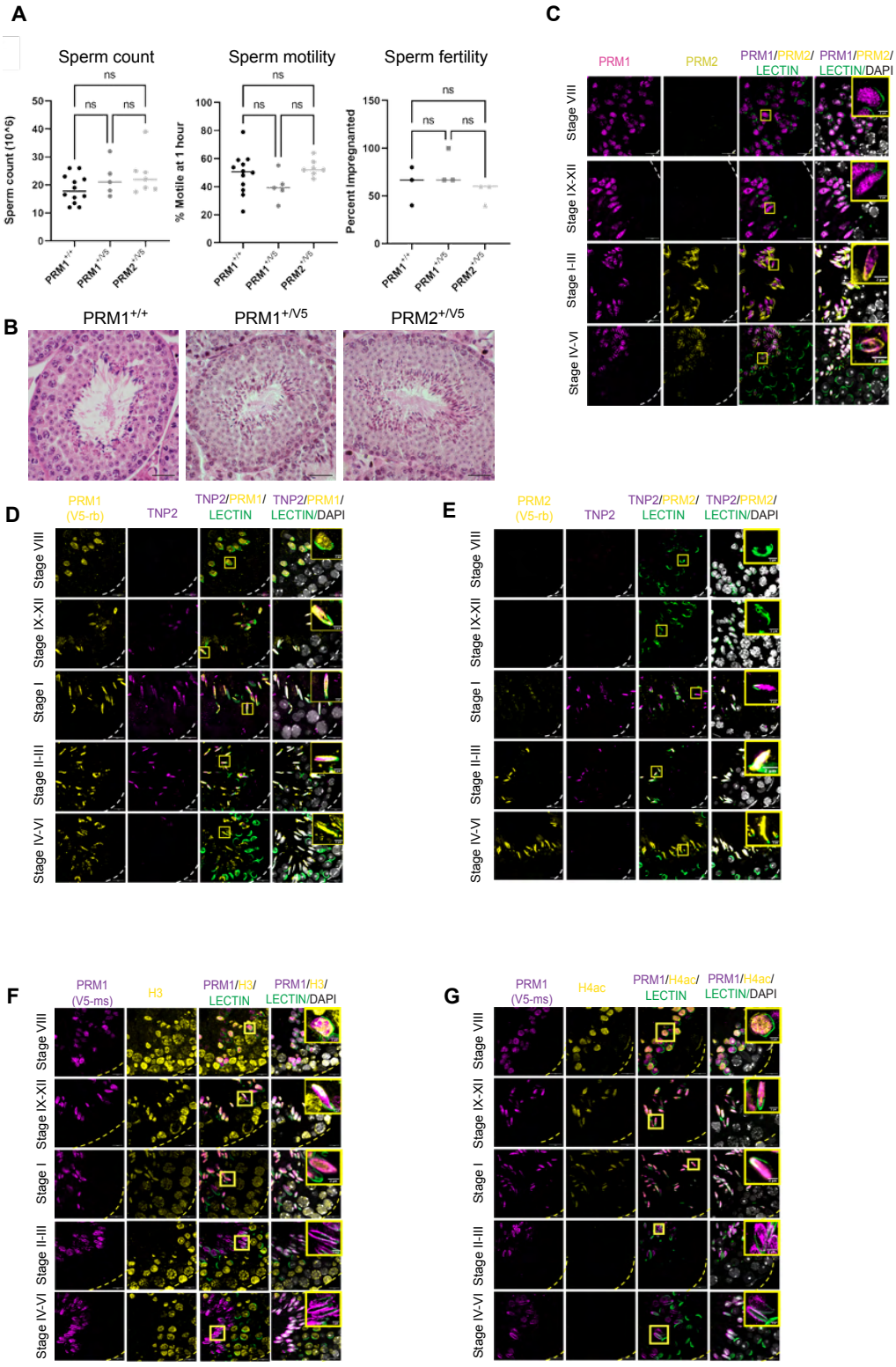
Median ATAC-seq signal trace across stages

In a python notebook, we converted the ATAC .rds file to a pandas (v2.3.1) dataframe using the rpy2 (v3.6.3) python package. Each ATAC bin was matched to its corresponding compartment using genomic coordinates. For each stage and compartment, the median GC residual z-score was calculated.

To establish a baseline for comparison, a theoretical null distribution was generated by sampling 1,000 points from a standard normal distribution $N(0,1)$, using a fixed random seed (42) from NumPy (v2.3.1) for reproducibility. The choice of $N(0,1)$ as the background distribution is predicated on the properties of the GC residual z-scores; by definition, z-scores are standardized transformations where the global population mean is 0 and the standard deviation is 1.

Under the null hypothesis - assuming no systematic enrichment or depletion of ATAC signal within specific genomic compartments - the GC residuals should remain centered around the expected mean of 0. By comparing the observed compartment-specific medians against this theoretical expectation, we can identify biological deviations from the global genomic background. To ensure a fair comparison and account for sampling noise in the null model, we applied the same bootstrapping procedure (1,000 resamples, n=1,000) to both the synthetic and observed data. This involved subsampling 1,000 points from the observed data to match the size of the expected dataset, providing a robust estimate of the median and its associated 95% confidence intervals. The resulting confidence intervals and medians were visualized using matplotlib (v3.10.3).

Supplementary Figures S1-S7



Supplementary Figure 1: Tagging the endogenous PRM1 and PRM2 loci has no effect on fertility or nuclear localization.

(A) *Prm1*$^{+/V5}$ and *Prm2*$^{+/V5}$ males have normal sperm counts, motility, and fertility relative to wild-type *Prm1*$^{+/+}$ controls. Each data point represents an individual animal. Comparisons were made by one-way ANOVA with correction for multiple testing; ns, not significant.

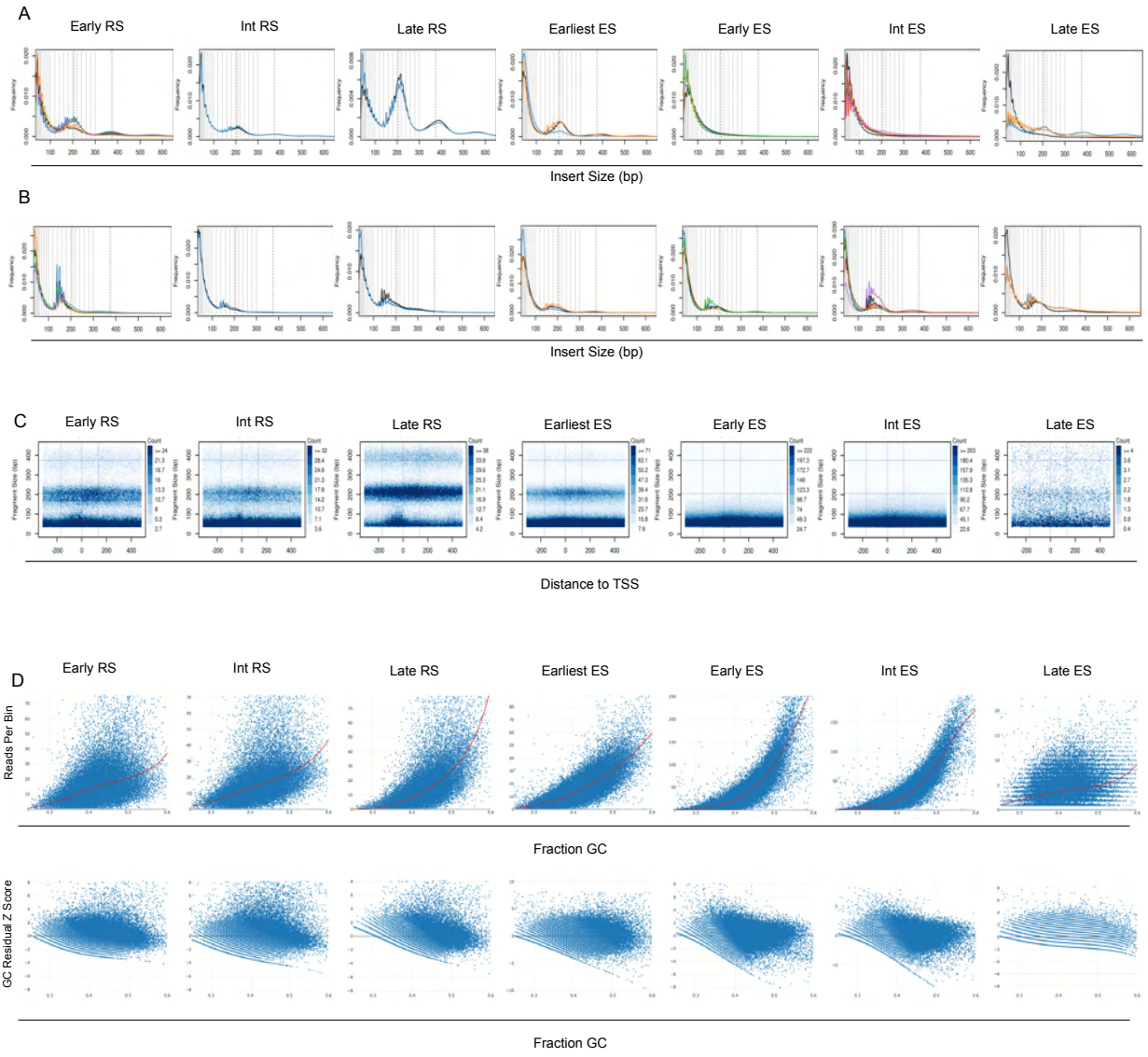
(B) Hematoxylin and eosin staining of representative testis cross-sections from adult wild-type, *Prm1*$^{+/V5}$ and *Prm2*$^{+/V5}$ males. Scale bar, 20 μm.

(C) Testis cross-sections from *Prm2*$^{+/V5}$ mice were co-stained with anti-V5 (PRM2) and Hup1N (PRM1). Magenta, PRM1 (Hup1N); yellow, PRM2 (V5); green, lectin; gray, DAPI. Scale bar, 10 μm; inset scale bar, 2 μm.

(D) *Prm1*$^{+/V5}$ testis cross-sections were co-stained with anti-V5 (PRM1) and anti-TNP2. Magenta, PRM1 (V5); yellow, TNP2; green, lectin; gray, DAPI. Scale bar, 10 μm; inset scale bar, 2 μm.

(E) *Prm2*$^{+/V5}$ testis cross-sections were co-stained with anti-V5 (PRM2) and anti-TNP2. Magenta, PRM2 (V5); yellow, TNP2; green, lectin; gray, DAPI. Scale bar, 10 μm; inset scale bar, 2 μm.

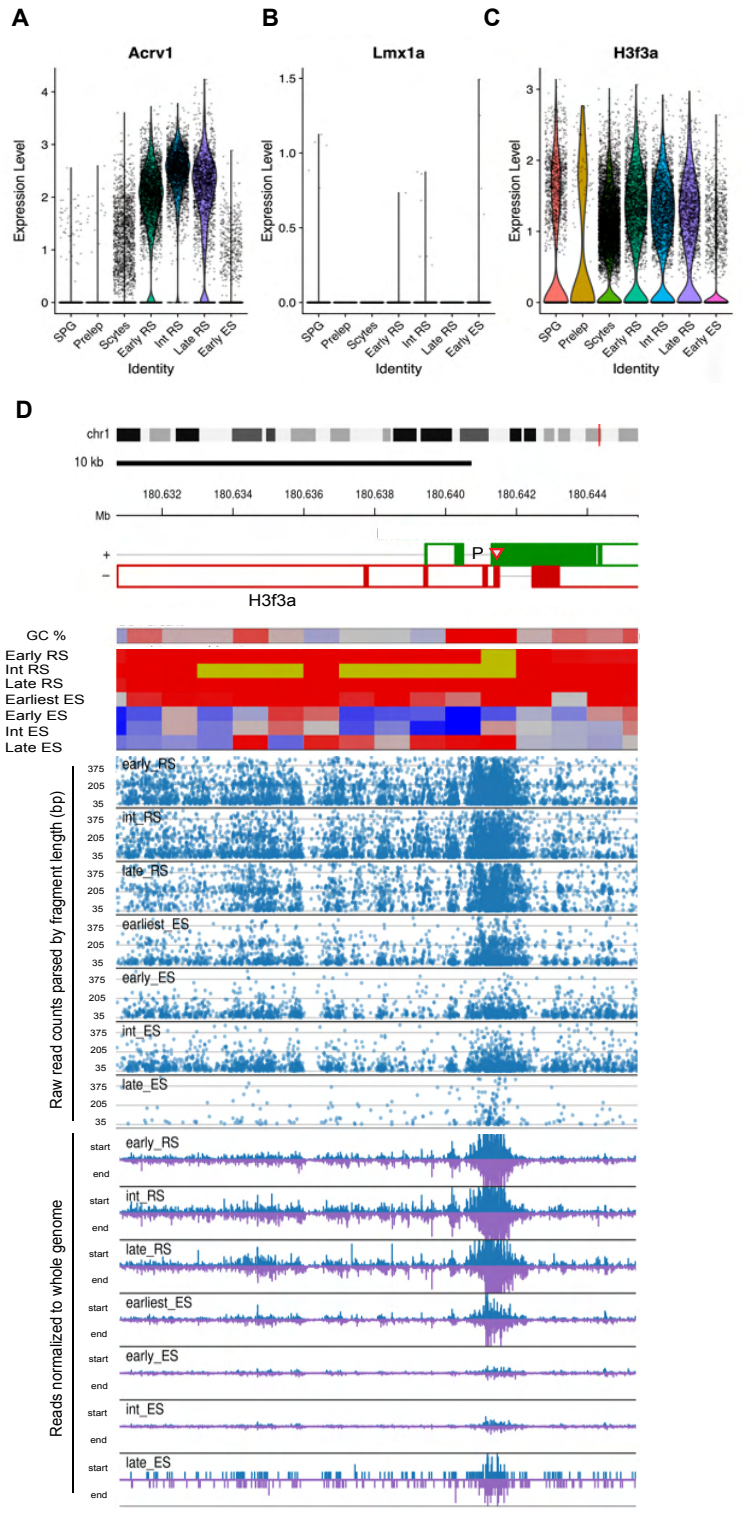
(F,G) *Prm1*$^{+/V5}$ testis cross-sections were co-stained with anti-V5 (PRM1) and anti-H3 or anti-H4ac. Magenta, PRM1 (V5); yellow, H4ac; green, lectin; gray, DAPI. Scale bar, 10 μm; inset scale bar, 2 μm.



Supplementary Figure 2: ATAC-seq replicates display concordant fragment length distributions.

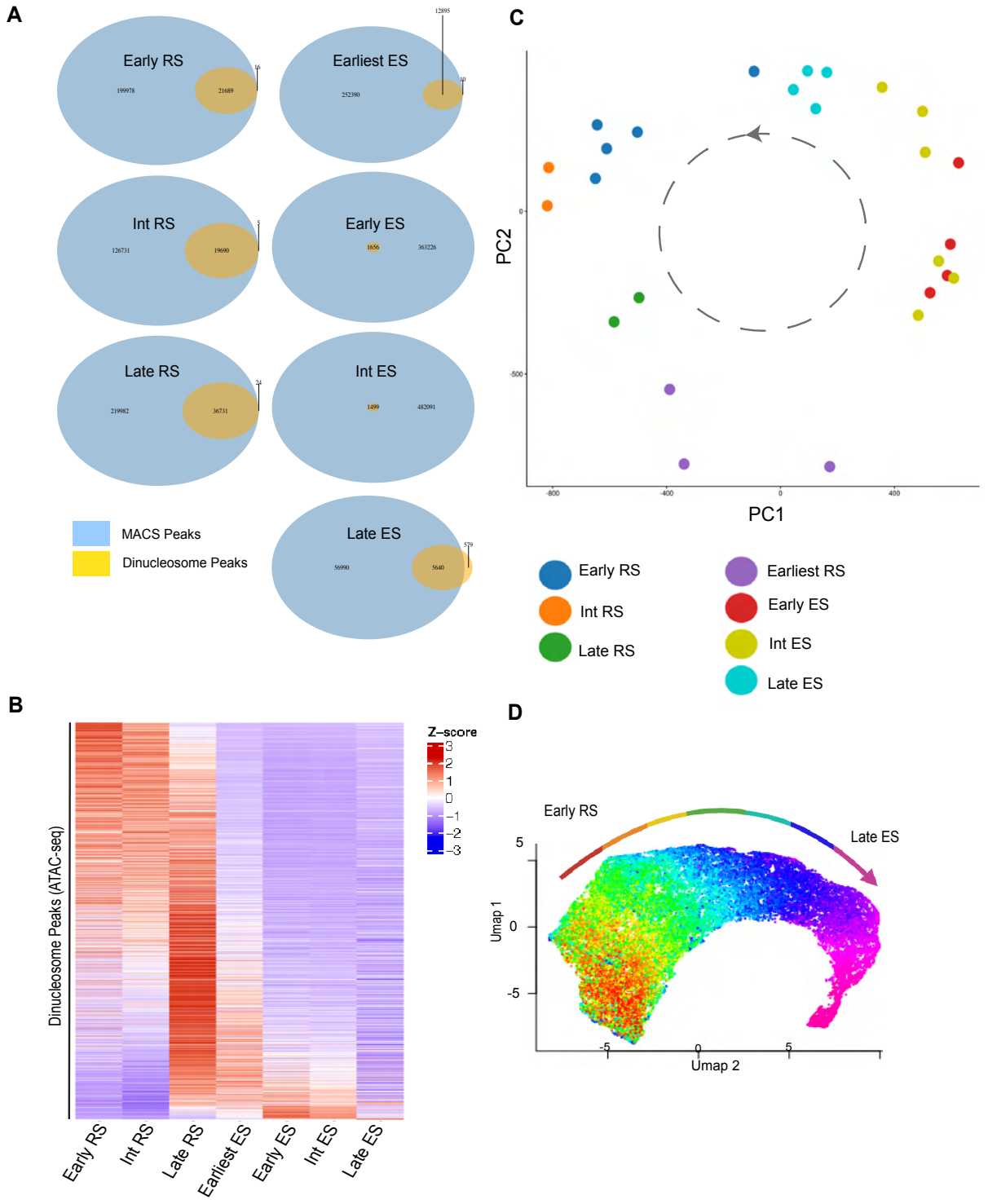
(**A**) ATAC-seq fragment length distributions for all biological replicates at each stage of spermiogenesis. Each line represents an individual replicate. Replicate numbers: Early RS, $n = 5$; Int RS, $n = 2$; Late RS, $n = 2$; Earliest ES, $n = 3$; Early ES, $n = 5$; Int ES, $n = 6$; Late ES, $n = 3$. (**B**) Fragment length distributions of *Drosophila* spike-in reads corresponding to the samples in (A), used for normalization across stages. (**C**) V-plots of ATAC-seq fragment lengths across all seven stages as a function of distance from the annotated TSS of inactive genes based on PRO-seq (Kaye et al., 2024). Compared to active

TSS in Fig. 2H, nucleosomes are not positioned but also disappear in Early ES and Int ES. (**D)** GC bias regression for ATAC-seq data across spermiogenesis stages. **(Top)** Reads per 1kb bin as a function of GC content. Red lines are the regression fit. **(Bottom)** Z-score relative to the GC regression fit as a function of GC content showing flattening of the data. Representative plots are shown for each stage.



Supplementary Figure 3: Chromatin accessibility dynamics at loci with distinct functional properties during spermiogenesis are not affected by genome-wide normalization.
(A–C) Single-cell RNA-seq expression profiles of *Acrv1* (A), *Lmx1a* (B), and *H3f3a* (C) across

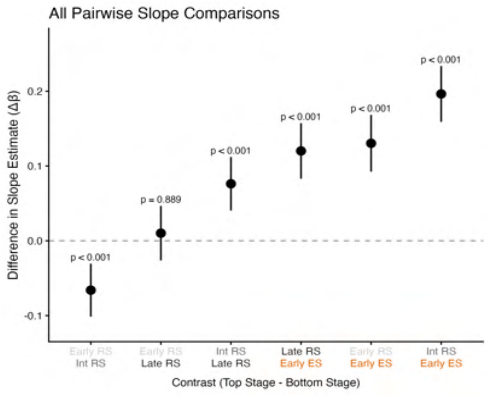
spermatogenic cell types from mouse testis (Green et al., 2018). Cell centroids for each stage are indicated on the x-axis; expression level (normalized counts) on the y-axis.
(D) Chromatin accessibility at the *H3f3a* locus, representing a region that remains accessible throughout spermiogenesis with accessibility retained at the promoter in late elongating spermatids (late ES). **(Top)** Heatmap of GC-residual Z scores across the indicated spermatogenic stages. **(Middle)** V-plots across all seven stages. Each dot is a single mapped insert. Horizontal lines are at 35bp (minimum allowed fragment size), 205 bp (mono-nucleosome peak center), and 375 bp (di-nucleosome peak center). **(Bottom)** Strand-specific read counts normalized to the whole genome, enabling quantitative comparison of accessibility across stages independent of sequencing depth and providing strand-resolution of transcription factor footprints and precise nucleosome positioning at this locus.



Supplementary Figure 4: Comparison of peak-calling strategies across spermiogenic stages. **(A)** Venn diagrams illustrating the overlap between peaks called by MACS2 (blue) and the dinucleosome peak-calling algorithm (gold) at each stage of spermiogenesis. Numbers indicate peak counts within each region. **(B)** Heatmap of dinucleosome peak accessibility scores across spermiogenic stages. Each row is

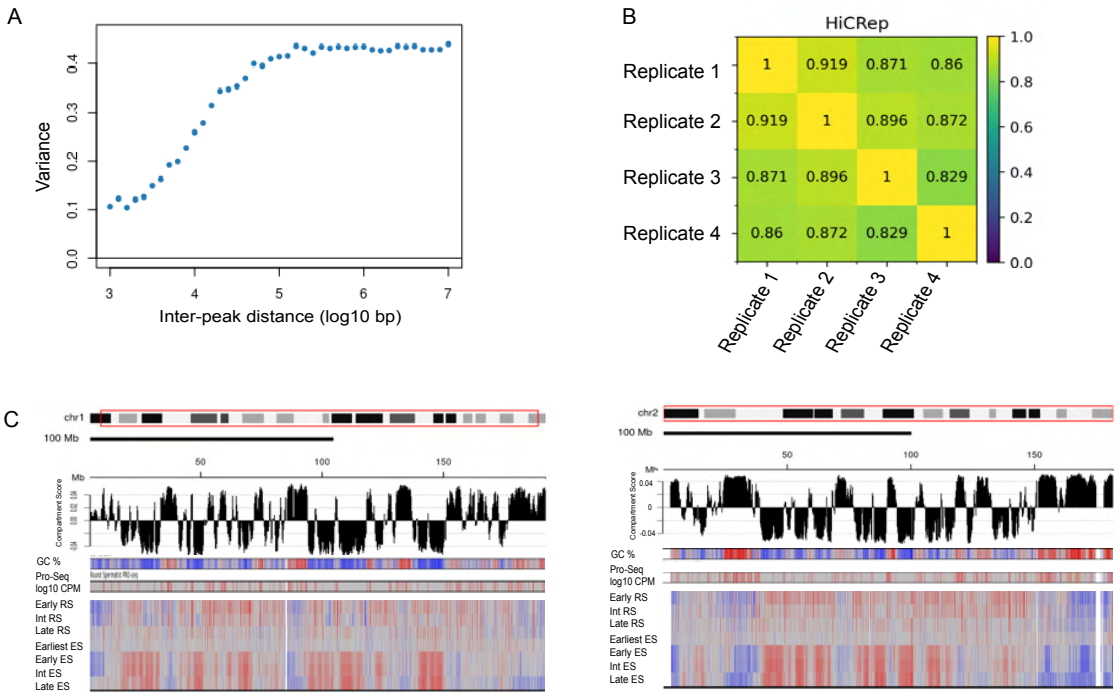
a peak overlap region ordered by the centroid of the stage-specific signals.
(C) Principal component analysis (PCA) of 26 ATAC-seq datasets using GC-residual scores computed over MACS2 peak regions. Each point represents an individual dataset, colored by spermatogenic stage as indicated. The samples cluster into a circular trajectory with Late ES returning to be most like Early RS.
(D) UMAP of di-nucleosome peaks across spermiogenesis, rainbow-colored by stage. Early RS (red), int RS (orange), late RS (yellow), earliest ES (green), early ES (cyan), int ES (blue) and Late ES (magenta). With (B), no discrete clusters are observed within a continuous trajectory of relative accessibility timing.

A

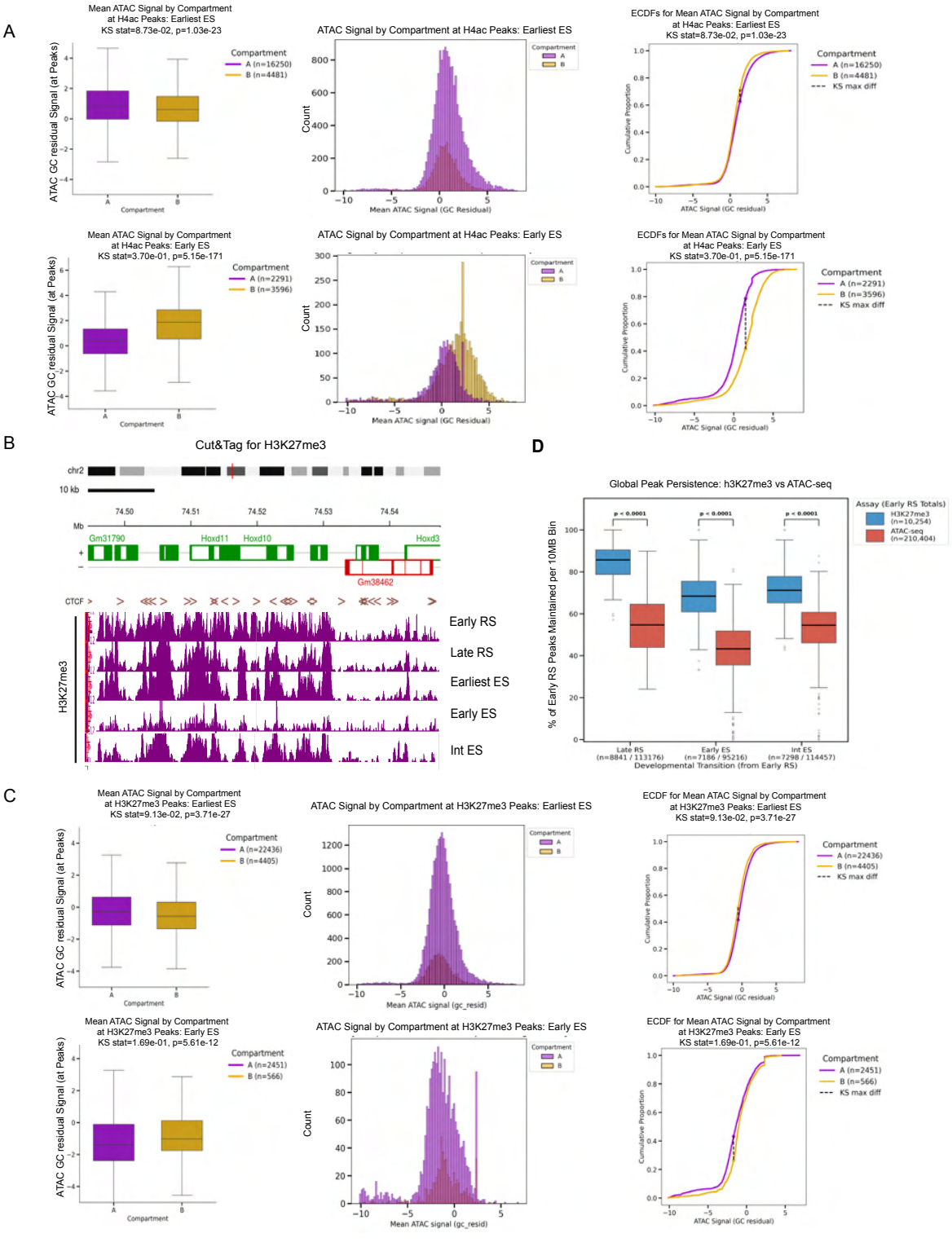


Supplementary Figure 5: Comparing chromatin-transcription relationships during spermiogenesis.

Forest plot displaying all pairwise contrasts of the transcription-accessibility slope estimates between stages. The y-axis represents the difference in slope for each comparison (Top Stage - Bottom Stage). Error bars denote 95% confidence intervals corrected for multiple comparisons (Tukey method). Comparisons involving Early ES consistently show a large difference, confirming that the coupling strength in this stage is significantly lower than in all RS stages. P-values indicate the statistical significance of the difference in slope.



Supplementary Figure 6: Chromatin remodeling during spermiogenesis follows a reproducible, genome-wide, 3D-encoded program.
(A) Variogram, i.e., lag analysis, depicting variance in GC-residual scores in ATAC-seq signal as a function of inter-peak distance (log10 bp).
(B) HiCRep stratum-adjusted correlation coefficients between all pairwise combinations of four Hi-C replicates.
(C) Genome-wide compartment scores and chromatin accessibility dynamics across chromosomes 1 (left) and 2 (right). From top to bottom: chromosome ideogram, compartment score track, GC content, Pro-Seq log10 CPM, and ATAC-seq GC-residual heatmaps across spermatogenic stages (Early RS through Late ES). B compartment (negative compartment scores) regions consistently correlate with higher relative accessibility in ES.


A
Mean ATAC Signal by Compartment at H4ac Peaks: Earliest ES KS stat=8.73e-02, p=1.03e-23
ATAC Signal by Compartment at H4ac Peaks: Earliest ES
ECDFs for Mean ATAC Signal by Compartment at H4ac Peaks: Earliest ES KS stat=8.73e-02, p=1.03e-23
Mean ATAC Signal by Compartment at H4ac Peaks: Early ES KS stat=3.70e-01, p=5.15e-171
ATAC Signal by Compartment at H4ac Peaks: Early ES
ECDFs for Mean ATAC Signal by Compartment at H4ac Peaks: Early ES KS stat=3.70e-01, p=5.15e-171
B
Cut&Tag for H3K27me3
chr2
10 kb
Gm31790
Hoxd11
Hoxd10
Hoxd3
Gm38462
CTCF
H3K27me3
Early RS
Late RS
Earliest ES
Early ES
Int ES
D
Global Peak Persistence: h3K27me3 vs ATAC-seq
% of Early RS Peaks Maintained per 10MB Bin
Developmental Transition (from Early RS)
C
Mean ATAC Signal by Compartment at H3K27me3 Peaks: Earliest ES KS stat=9.13e-02, p=3.71e-27
ATAC Signal by Compartment at H3K27me3 Peaks: Earliest ES
ECDF for Mean ATAC Signal by Compartment at H3K27me3 Peaks: Earliest ES KS stat=9.13e-02, p=3.71e-27
Mean ATAC Signal by Compartment at H3K27me3 Peaks: Early ES KS stat=1.69e-01, p=5.61e-12
ATAC Signal by Compartment at H3K27me3 Peaks: Early ES
ECDF for Mean ATAC Signal by Compartment at H3K27me3 Peaks: Early ES KS stat=1.69e-01, p=5.61e-12

Supplementary Figure 7: Compartment identity overrides histone acetylation in directing compaction order, while H3K27me3 acts as a stabilizing chromatin bookmark throughout spermiogenesis.
(A) Boxplots (left), histograms (middle), and ECDFs (right) of mean ATAC-seq signal (GC residual) within H4ac peaks, stratified by A (purple) and B (gold) compartment identity, at earliest ES (top row) and early ES (bottom row).
(B) Cut&Tag signal for H3K27me3 at the *Hoxd* gene cluster on chromosome 2 across Early RS, Late RS, Earliest ES, Early ES, and Int ES stages. CTCF binding sites are indicated above the gene track.
(C) Boxplots (left), histograms (middle), and ECDFs (right) of mean ATAC-seq signal (GC residual) within H3K27me3 peaks, stratified by A (purple) and B (gold) compartment identity, at earliest ES (top row) and early ES (bottom row). KS statistics and p-values are indicated above each panel.
(D) Boxplots comparing the percentage of early RS peaks maintained per 10 Mb bin across three developmental transitions for H3K27me3 (blue) and ATAC-seq (red).

Table S1

Supplementary Table 1: Classification of ATAC-seq data along stages of spermiogenesis

	Sample ID	Stage VIII	Stage IX-XII	Stage I-III	Stage IV-VI	Stage VII	Staging	Stage	Drosophila Input (%)	Drosophila Recovered (%)
1	26104X4-Day32 RS B		53	47			RS1	early_RS	10.0	45.0
2	26104X8-Day 33 RS		35	65			RS1-2	early_RS	10.0	10.0
3	24290X6-day35-wt-rs-rep2			100			RS1-2	early_RS	10.0	7.3
4	26104X10- Day 35 RS B			96	4		RS3-4	early_RS	10.0	60.0
5	24290X4-day35-wt-rs-rep1			96	4		RS3-4+	early_RS	10.0	11.8
6	26104X12 - Day 36 RS A			30	70		RS4-5	int_RS	10.0	21.0
7	26104X14 - Day 36 RS B			12	79	9	RS4-5	int_RS	10.0	14.0
8	24290X8-day38-wt-rs-rep1				100		RS6	late_RS	10.0	0.9
9	24290X10-day38-wt-rs-rep2				100		RS6-7	late_RS	10.0	1.0
10	23734X2-Day32_ES	23				77	ES8	earliest_ES	10.0	4.5
11	26104X3-Day 31 ES	83	17				ES8	earliest_ES	10.0	3.0
12	26104X1-Day 30 ES	88	13				ES8-9	earliest_ES	10.0	28.0
13	23734X3-Day34_ES	40	60				ES8-9	early_ES	10.0	0.8
14	24290X3-day32-wt-es		95	5			ES10-12	early_ES	10.0	1.1
15	23734X4-Day35_ES	6	35	58			ES12-13	early_ES	10.0	2.8
16	26104X6-Day32-ES B		53	47			ES12-13	early_ES	10.0	17.0
17	26104X9 - Day 33 ES		35	65			ES13-14	int_ES	10.0	3.0
18	26104X5-Day32-ES A		2	98			ES13-14	int_ES	10.0	5.0
19	24290X2-day34-wt-es		18	80	3		ES14-15	int_ES	10.0	0.3
20	24290X7-day35-wt-es-rep2			100			ES14-15	int_ES	10.0	7.5
21	24290X5-day35-wt-es-rep1			96	4		ES14-15	int_ES	10.0	5.6
22	26104X11-Day35 ES B			96	4		ES14-15	int_ES	10.0	29.0
23	26104X13 - Day 36 ES A			30	70		ES15	late_ES	10.0	75.0
24	26104X15 - Day 36 ES B			12	79	9	ES15	late_ES	10.0	96.0
25	24290X9-day38-et-es-rep1				100		ES15	late_ES	10.0	72.2
26	24290X11-day38-wt-es-rep2				100		ES15	late_ES	10.0	57.9

*Table S2 and S3 are large spreadsheets and uploaded separately.

References

1 Pepin, A. S., Jazwiec, P. A., Dumeaux, V., Sloboda, D. M. & Kimmins, S. Determining the effects of paternal obesity on sperm chromatin at histone H3 lysine 4 tri-methylation in relation to the placental transcriptome and cellular composition. *Elife* **13** (2024). https://doi.org/10.7554/eLife.83288

2 Moritz, L. *et al.* 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 (2023). https://doi.org/10.1038/s41594-023-01033-4

3 Nakata, H., Wakayama, T., Takai, Y. & Iseki, S. Quantitative analysis of the cellular composition in seminiferous tubules in normal and genetically modified infertile mice. *J Histochem Cytochem* **63**, 99-113 (2015). https://doi.org/10.1369/0022155414562045

4 Hogarth, C. A. *et al.* Turning a spermatogenic wave into a tsunami: synchronizing murine spermatogenesis using WIN 18,446. *Biol Reprod* **88**, 40 (2013). https://doi.org/10.1095/biolreprod.112.105346

5 Herrmann, C., Avgousti, D. C. & Weitzman, M. D. Differential Salt Fractionation of Nuclei to Analyze Chromatin-associated Proteins from Cultured Mammalian Cells. *Bio Protoc* **7** (2017). https://doi.org/10.21769/BioProtoc.2175

6 Green, C. D. *et al.* A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell* **46**, 651-667 e610 (2018). https://doi.org/10.1016/j.devcel.2018.07.025

7 Shen, Y. C. *et al.* TCF21(+) mesenchymal cells contribute to testis somatic cell development, homeostasis, and regeneration in mice. *Nat Commun* **12**, 3876 (2021). https://doi.org/10.1038/s41467-021-24130-8

8 Gill, M. E., Kohler, H. & Peters, A. Dual DNA staining enables isolation of multiple sub-types of post-replicative mouse male germ cells. *Cytometry A* **101**, 529-536 (2022). https://doi.org/10.1002/cyto.a.24539

9 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017). https://doi.org/10.1038/nmeth.4396

10 Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019). https://doi.org/10.1038/s41467-019-09982-5

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

12 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014). https://doi.org/10.1093/nar/gku365

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

14 Huang, D. W. *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**, W169-175 (2007). https://doi.org/10.1093/nar/gkm415

15 Chen, S. fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *Imeta* **4**, e70078 (2025). https://doi.org/10.1002/imt2.70078

16 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019). https://doi.org/10.1038/s41598-019-45839-z

17 Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687-3692 (2020). https://doi.org/10.1093/bioinformatics/btaa222

18 Green, B., Bouchier, C., Fairhead, C., Craig, N. L. & Cormack, B. P. Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob DNA* **3**, 3 (2012). https://doi.org/10.1186/1759-8753-3-3

19 Kaye, E. G. *et al.* RNA polymerase II pausing is essential during spermatogenesis for appropriate gene expression and completion of meiosis. *Nat Commun* **15**, 848 (2024). https://doi.org/10.1038/s41467-024-45177-3

20 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013). https://doi.org/10.1371/journal.pcbi.1003118

21 Kolde, R. *pheatmap: Pretty Heatmaps*, <https://github.com/raivokolde/pheatmap> (2025).

22 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016). https://doi.org/10.1016/j.cels.2016.07.002

23 Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**, 1939-1949 (2017). https://doi.org/10.1101/gr.220640.117

24 Yardimci, G. G. *et al.* Measuring the reproducibility and quality of Hi-C data. *Genome Biol* **20**, 57 (2019). https://doi.org/10.1186/s13059-019-1658-7

25 Wang, X. *et al.* A generalizable Hi-C foundation model for chromatin architecture, single-cell and multi-omics analysis across species. *bioRxiv* (2024). https://doi.org/10.1101/2024.12.16.628821

26 Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016). https://doi.org/10.1016/j.cels.2015.07.012