Micrococcal nuclease sequencing of porcine sperm suggests enriched co-location between retained histones and genomic regions related to semen quality and early embryo development

Marta Gòdia[1,2,*], Yu Lian[1,*], Marina Naval-Sanchez[3], Inma Ponte[4], Joan Enric Rodríguez-Gil[5], Armand Sanchez[1,6] and Alex Clop[1,7]

[1] Centre for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Cerdanyola del Vallés, Catalonia, Spain
[2] Animal Breeding and Genomics, Wageningen University and Research, Wageninger, Netherlands
[3] Agriculture & Food, CSIRO, Brisbane, Queensland, Australia
[4] Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Catalonia, Spain
[5] Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Catalonia, Spain
[6] Animal and food sciences, Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Catalonia, Spain
[7] Consejo Superior de Investigaciones Científicas, Barcelona, Catalonia, Spain
* These authors contributed equally to this work.

Submitted 3 October 2022
Accepted 16 May 2023
Published 21 June 2023

Corresponding authors
Marta Gòdia, marta.godia@wur.nl
Alex Clop, alex.clop@csic.es

Academic editor
Motoki Takaku

Additional Information and Declarations can be found on page 20

DOI 10.7717/peerj.15520

Copyright
2023 Gòdia et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

ABSTRACT

The mammalian spermatozoon has a unique chromatin structure in which the majority of histones are replaced by protamines during spermatogenesis and a small fraction of nucleosomes are retained at specific locations of the genome. The sperm's chromatin structure remains unresolved in most animal species, including the pig. However, mapping the genomic locations of retained nucleosomes in sperm could help understanding the molecular basis of both sperm development and function as well as embryo development. This information could then be useful to identify molecular markers for sperm quality and fertility traits. Here, micrococcal nuclease digestion coupled with high throughput sequencing was performed on pig sperm to map the genomic location of mono- and sub-nucleosomal chromatin fractions in relation to a set of diverse functional elements of the genome, some of which were related to semen quality and early embryogenesis. In particular, the investigated elements were promoters, the different sections of the gene body, coding and non-coding RNAs present in the pig sperm, potential transcription factor binding sites, genomic regions associated to semen quality traits and repeat elements. The analysis yielded 25,293 and 4,239 peaks in the mono- and sub-nucleosomal fractions, covering 0.3% and 0.02% of the porcine genome, respectively. A cross-species comparison revealed positional conservation of the nucleosome retention in sperm between the pig data and a human dataset that found nucleosome enrichment in genomic regions of importance in development. Both gene ontology analysis of the genes mapping nearby the mono-nucleosomal peaks and the identification of putative transcription factor binding motifs within the mono- and

How to cite this article Gòdia M, Lian Y, Naval-Sanchez M, Ponte I, Rodríguez-Gil JE, Sanchez A, Clop A. 2023. Micrococcal nuclease sequencing of porcine sperm suggests enriched co-location between retained histones and genomic regions related to semen quality and early embryo development. **PeerJ 11:e15520 DOI 10.7717/peerj.15520**

the sub- nucleosomal peaks showed enrichment for processes related to sperm function and embryo development. There was significant motif enrichment for Znf263, which in humans was suggested to be a key regulator of genes with paternal preferential expression during early embryogenesis. Moreover, enriched positional intersection was found in the genome between the mono-nucleosomal peaks and both the RNAs present in pig sperm and the RNAs related to sperm quality. There was no co-location between GWAS hits for semen quality in swine and the nucleosomal sites. Finally, the data evidenced depletion of mono-nucleosomes in long interspersed nuclear elements and enrichment of sub-nucleosomes in short interspersed repeat elements.These results suggest that retained nucleosomes in sperm could both mark regulatory elements or genes expressed during spermatogenesis linked to semen quality and fertility and act as transcriptional guides during early embryogenesis. The results of this study support the undertaking of ambitious research using a larger number of samples to robustly assess the positional relationship between histone retention in sperm and the reproductive ability of boars.

Subjects Agricultural Science, Developmental Biology, Genomics, Molecular Biology, Zoology
Keywords Swine, Sperm, MNase-seq, Chromatin, Reproduction

INTRODUCTION

Current research is showing evidences that, in addition to the paternal genome, sperm also carries other molecular information to the zygote. This includes a wide repertoire of RNAs, proteins and epigenetic marks in the form of DNA methylation, retained nucleosomes and histone modifications that can play a role in spermatogenesis, fertility, early embryo development and even the transmission of inter-generational information (*Krawetz, 2005*; *Gòdia, Swanson & Krawetz, 2018*; *Spadafora, 2017*; *Hammoud et al., 2009*; *Oikawa et al., 2020*). During spermatogenesis, male germ cells undergo a series of profound morphological and functional changes that conclude in mature spermatozoa. In the last stages of spermatogenesis, when the cell is transcriptionally silent, genomic DNA (gDNA) is highly condensed to fit into the sperm head (*Ward & Coffey, 1991*) and ensure genomic integrity, early embryo development and ultimately, fertility (*Ward, 2010*). In mammals, the condensation of the spermatozoon chromatin occurs through the progressive replacement of histones by protamines. However, a small fraction of the sperm DNA remains organized in histones as shown in humans (5 to 15% of histones are retained in the sperm chromatin) (*Gatewood et al., 1987*; *Hammoud et al., 2009*; *Arpanahi et al., 2009*; *Brykczynska et al., 2010*; *Castillo et al., 2014*; *Samans et al., 2014*), cattle (13.4%) (*Samans et al., 2014*) and mice (1 to 2%) (*Balhorn, Gledhill & Wyrobek, 1977*; *Erkek et al., 2013*; *Carone et al., 2014*; *Johnson et al., 2016*). Several studies indicate that these nucleosomes are distributed along the genome following a programmatic pattern with preferential retention at gene promoters and developmental related loci (*Hammoud et al., 2009*; *Arpanahi et al., 2009*; *Brykczynska et al., 2010*; *Castillo et al., 2014*; *Jung et al., 2017*; *Yoshida et al., 2018*). However, these results are inconclusive because at least three other

publications reported that sperm nucleosomes are enriched at gene-poor regions or repeat elements (*Samans et al., 2014*; *Carone et al., 2014*; *Yamaguchi et al., 2018*) while they may even be depleted at promoters (*Carone et al., 2014*; *Samans et al., 2014*) and exons (*Samans et al., 2014*). The reasons underlying these discrepancies, which are not mutually exclusive, are unclear and could be attributable to technical differences in the protocols used in each study. Carone and colleagues found that nucleosomes differ in their resistance to micrococcal nuclease (MNase) digestion and that the most fragile nucleosomes tended to map near repeat elements while nucleosomes at promoters were more stable (*Carone et al., 2014*). Thus, protocols using large amounts of MNase would artifactually result in nucleosomal enrichment at promoters (*Carone et al., 2014*). However, other studies, which used alternative protocols not based on MNase and that claimed to have unbiasedly characterized all nucleosomes present in the sperm chromatin, found enriched nucleosomes at promoters or gene bodies (*Johnson et al., 2016*; *Jung et al., 2017*; *Yoshida et al., 2018*). Another study based on nucleoplasmin instead of MNase to solubilize the sperm's chromatin concluded enrichment at both, promoters and gene bodies as well as gene deserts (*Yamaguchi et al., 2018*). They hypothesized that MNase based protocols do not retain all nucleosomes (*Yamaguchi et al., 2018*).

Nucleosome positioning in the genome and chromatin accessibility is critical in the regulation of gene expression and the alteration of this epigenomic architecture has been linked to multiple phenotypes in a range of tissues and cell types (*Lai & Pugh, 2017*). Nucleosomes in sperm may either be leftovers of gene expression during spermatogenesis or provide transcriptional instructions upon fertilization for gamete recognition and early embryo development. *Jung et al. (2017)* found that a proportion of histone modifications in gene promoters of the mouse sperm chromatin, recapitulates the chromatin structure and transcriptional activity of late spermatids. Also, a wealth of studies reported that at least a proportion of histone modifications in sperm is maintained (*van der Heijden et al., 2006*; *Lismer et al., 2020*) and even that nucleosome positioning (*Ihara et al., 2014*) and the location of histone modifications (*Oikawa et al., 2020*) in sperm are related to gene expression in early embryos. Thus, this information could be of great value to shed light into the catalog of genomic regions and genes related to sperm biology and early embryo development and thus help identifying elusive molecular markers for traits related to sperm quality and fertility.

In livestock, the architecture of the sperm chromatin at the genomic level has only been investigated in cattle (*Samans et al., 2014*). In pigs (*Sus scrofa*), research has been carried to study the sperm RNA (*Gòdia et al., 2019*; *Gòdia et al., 2020a*; *Ablondi et al., 2021*) and protein (*Mendonça et al., 2017*) populations as well as the DNA methylation patterns (*Khezri et al., 2019*; *Pértille et al., 2021*), but the sperm's chromatin structure remains unexplored.

This study is sustained by the hypothesis that nucleosome retention in the pig sperm is not stochastic and owes to reminiscences of the spermatogenesis program and future information for the embryogenic program. Our objective was to map the retained nucleosomes in a pool of two pig sperm samples and evaluate their relative location in relation to different genomic features, some of which were associated to sperm biology,

semen quality and fertility. To fulfil this purpose, pig spermatozoa were digested with MNase and the resulting nucleosome-associated DNAs were subjected to high throughput sequencing. The mono-nucleosomal (MN) and sub-nucleosomal (SN) chromatin fractions were characterized and their correlation with sperm RNA levels and co-location with genomic regions associated to semen traits was assessed.

MATERIALS AND METHODS

Sample collection

Two ejaculates, each from a different healthy Pietrain boar, MN_1 (16 months of age) and MN_2 (9 months old), from two artificial insemination centres were obtained by specialists during their routine sample collection using the gloved hand method (*King & Macpherson, 1973*) and were immediately diluted (1:2) in commercial extender. Both ejaculates showed good semen quality as evidenced by the evaluation of several sperm parameters. The percentages of sperm cell viability, structurally altered acrosomes and morphological abnormalities were measured by staining the samples with the eosin-nigrosin technique after 5 min incubation at 37°C. These analyses showed 94.3% and 92.2% of viable cells for sample MN_1 and MN_2 respectively; 96.7% and 96.3% of cells with normal acrosomes, respectively; 2.2% and 1.1%, 3.1% and 1.1% and 2.2% and 4.6%, of head, neck and tail abnormalities, respectively. Semen samples were purified with BoviPure^{TM} (Nidacon, Moelndal, Sweden) to remove potentially present somatic cells and processed as described in *Gòdia et al. (2018)*. Briefly, a variable volume of sperm according to its concentration, with a maximum of 1 billion cells and not exceeding 11 mL, were layered over 3 mL of BoviPure^{TM} diluted to a final ratio of 60% (v/v) with BoviDilute^{TM} (Nidacon, Moelndal, Sweden) and centrifuged at 300 × g for 20 min at 20 °C with slow acceleration and deceleration rates. The resulting cell pellet was washed with PBS and resuspended in 1 mL of RNase-free PBS and further pelleted by centrifugation. The pellets were stored at −80 °C until further use.

Micrococcal nuclease digestion, library construction and sequencing

Chromatin digestion was performed as previously described (*Zalenskaya, Bradbury & Zalensky, 2000*) with minor modifications. Forty million purified spermatozoa cells were pelleted by centrifugation (3,500 rpm for 10 min) and washed twice with 1 mL of 1x phosphate buffered saline (PBS), and 1 mM phenylmethylsulfonyl fluoride (PMSF). Pelleted cells were resuspended with 1 mL of 0.1% Lysolecithin (dissolved in 1x PBS and 1 mM PMSF), incubated on ice for 10 min and centrifuged (3,500 rpm for 10 min). The sperm head was decondensed in 1 mL of 1x PBS, 1 mM PMSF with 10mM Dithiothreitol (DTT) and incubated on ice for 12 min. After centrifugation (2,000 rpm for 10 min), nuclei were washed with 1 mL of 1x PBS and 2mM DTT and centrifuged again (3,500 rpm for 10 min). The pellet was resuspended in 100 µl of 1x PBS, 1 mM DTT and $CaCl_2$ was added to a final concentration of 0.6 mM. At this time, 5 U of MNase (Sigma-Aldrich) were added and the samples were incubated for 7 min at 37 °C. Digestion was stopped by the addition of ethylene glycol-bis (β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (EGTA) to a concentration of 5 mM. The resulting digested fraction corresponding to

nucleosome-bound chromatin (soluble) was separated from the undigested protamine-bound fraction (pellet) by centrifugation at 20,000 rpm for 10 min. An aliquot of the digested DNA fraction (which included both the MN and SN fractions) was treated with sodium dodecyl sulfate (SDS) and proteinase K and evaluated in a 1.5% agarose 1x tris-acetate-EDTA (TAE) gel electrophoresis. Another aliquot of the nuclease digested DNA was used to extract DNA with a phenol-chloroform based protocol and was directly used for library prep after purification with the Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The MN and SN fractions were separated after sequencing and read mapping, using bioinformatics tools as described below. Purified DNA was subjected to quality control including quantification with the Qubit™ DNA HS Assay kit (Invitrogen, Waltham, MA, USA) and Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) as well as size and concentration assessment with a 2100 Bioanalyzer and the Agilent DNA 1000 Kit (Agilent Technologies, Santa Clara, CA, USA). gDNA from the two sperm samples was also extracted as done by Hammoud and colleagues (*Hammoud et al., 2009*). Briefly, sperm cells were thawed and pelleted at 10,000 × g for 1 min.
The supernatant was removed and the pellet was resuspended in 1 mL Lysis Reagent (1% Triton X-100, 5 mM MgCl2, 320 mM Sucrose and 10 mM Tris pH 7.5). After centrifugation (20,000 × g for 5 min), cells were resuspended again with 0.5mL Lysis Reagent two times, centrifuged and the supernatant was discarded. The sperm pellet was dispersed with 305.5µl of Enzyme Master Mix (1.25 mM $MgCl_2$, 1.25 mM Deferoxamine Mesylate, 12.5 mM Tris pH 8.0). Then, 1.6 µl RNase A-20 mg/mL -, 1.6 µl of 1 M DTT and 6.5 µl of 5M NaCl was added to each sample. This cell suspension was kept at 37 °C for 30 min. The cells were then dispersed with gentle vertexing and subsequently, 40 µl of 10% SDS and 20 µl of Proteinase K (20 mg/µl) were added. Samples were then incubated at 50 °C for 60 min. Following this step, 1 mL of 100% Isopropanol was added and the tubes were inverted several times until a gel-like DNA precipitate appeared. Then, 0.6 mL of NaI solution was added (40 mM Tris pH 8.5, 20 mM EDTA and 7.6 M NaI), and the tubes were inverted until DNA formed a white precipitate. The DNA pellet was washed three times with 70% EtOH with a final wash with 100% EtOH. The DNA pellet was solubilized with 300 µl of H_2O. The two resulting gDNA samples were pooled to be used as input. Pooled gDNA were sheared to obtain 100 bp long fragments using a Covaris S2 instrument (Covaris Inc, Woburn, MA, USA), according to the manufacturer's instructions. Sequencing libraries of the two MNase treated samples and the input gDNAs were prepared with the PrepX™ DNA Library Kit (Takara, Kusatsu, Japan). The libraries were used as a template to generate 50 bp paired end (PE) reads in an Illumina HiSeq2500 system.

MNase-Seq data preprocessing and quality evaluation

Raw sequencing data was filtered to remove low quality bases, adaptor sequences and short reads (<25 bp length) with Trimmomatic v.0.36 (*Bolger, Lohse & Usadel, 2014*). Trimmed reads were aligned to the porcine genome (Sscrofa11.1) with Bowtie2 v.2.4.1 (*Langmead & Salzberg, 2012*) fitting default parameters except "—very sensitive". Pearson correlation of the read coverages for genomic regions along the genome was calculated using the

multiBamSummary file based on the two MNase samples with the tool deepTools v.3.3.2 (*Ramírez et al., 2016*) with options: "plotCorrelation—removeOutliers–skipZeros". Since the correlation between the samples was very high, the reads from the two replicates (MN_1: replicate A; MN_2: replicate B) were merged and processed as a pool with the aim to increase sequencing depth and peak detection power and accuracy. Then, the MN and SN fractions were bioinformatically separated based on the genomic distance between the two paired reads. The reads from the SN fraction, with a length slightly below 100 bp, were extracted using the parameter "–maxins 110", which specifies that the mapped paired-end reads should not exceed 110 bp. To obtain the reads from the MN fraction, with a length around 150 bp, the parameters "—minins 111" and "—maxins 1000" were applied. Subsequently, duplicate reads of these fractions and the input sample were removed with Picard-Tools MarkDuplicates v.1.56 (http://picard.sourceforge.net). Finally, for both fractions, peak calling was performed with MACS2 v.2.1.0 (*Zhang et al., 2008*) with "-q 0.05 -B -g hs" and compared against the input sample. Visualization of the genomic heatmaps from the MNase-Seq signals using transcription start sites (TSS) as reference points was carried with the deepTools (*Ramírez et al., 2016*) computeMatrix tool with parameters: "reference-point -b 500 -a 500 –skipZeros" and plotHeatmap tool using "—refPointLabel 'TSS'".

Peak location relative to gene annotation

Peaks were categorized by their position in relation to gene features as annotated in Ensembl (v96) using BEDtools v.2.29.2 (*Quinlan & Hall, 2010*) intersect. The peaks were classified as mapping to TSSs, promoters, overlapping 5′ untranslated region (UTR), 3′ UTR, exonic, intronic and intergenic as carried in *Halstead et al. (2020)*.

In order to determine any potential positional preference or avoidance of the peaks within the gene features, 1,000 iterative permutations of the peak genomic locations with the same number of peaks following the same peak width distribution but with randomized genomic locations, were carried using BEDtools (*Quinlan & Hall, 2010*) shuffle. Subsequently, genomic features were assigned to the randomized peaks based on their genomic position within these features. These results were subsequently compared against our results on real data using a permutation test. This test calculates the proportion of permutations in which the number of peaks mapping to a specific feature deviates from the mean of the distribution more than the number of overlaps observed in the real data. In other words, it determines whether the peaks in the permutations are located more towards the tails of the distribution compared to the real data. The fold change (FC) was also measured. FC was calculated by dividing the number of overlaps between the real MN or SN peaks and each of the genomic features by the equivalent mean of the distribution of the number of overlaps obtained with the shuffled data. Comparisons resulting in *P-value* < 0.001 and FC ≥ 1.5 or FC ≤ 0.66 were classified as enriched or depleted, respectively.

Gene ontology analysis of the genes near MNase peaks

The genes mapping within or less than 5 kbp apart from the identified peaks were extracted from Ensembl v96 with BEDtools (*Quinlan & Hall, 2010*) closestBed and were

used for Gene Ontology (GO) enrichment analysis. GO was carried out with Cytoscape v.3.8.2 plugin ClueGO v.2.5.7 (*Bindea et al., 2009*) using the Cytoscape's porcine dataset and default settings. Only the significant Bonferroni corrected *P*-values were considered.

Motif enrichment analysis at the MN and SN underlying sequences

Motif enrichment analysis was carried out with the software HOMER v.4.10.0 (*Heinz et al., 2010*) findMotifsGenome.pl function with default parameters.These default parameters include: (i) random selection of the genome's background regions to compare against the input sequence data and using the background frequency of nucleotides in the human genome, hg19; (ii) the default motif length for the search is 8-12 nucleotides long; (iii) motif search on both the forward and reverse strands of the input sequences.

Positional conservation of chromatin-associated DNA with human and cattle sperm

A comparable human sperm MNase-Seq dataset that sequenced the mono-nucleosomal fraction (GSE15690) from *Hammoud et al. (2009)*, obtained from a pool of four donors was downloaded from the Gene Expression Omnibus database. The genomic coordinates for the human sperm MN peaks were liftover to Sscrofa11.1 using the UCSC liftover tool (*Kuhn, Haussler & Kent, 2013*) with default parameters, except for the parameter "-minMatch = 0.1". To evaluate enrichment, the genomic locations of the porcine MNase peaks were randomized 1,000 times using BEDtools (*Quinlan & Hall, 2010*) shuffle and overlapped to the MNase peaks of human or cattle using BEDtools v.2.29.2 (*Quinlan & Hall, 2010*) intersect. Thereafter, the results were contrasted to the overlap of the real data with the permutation test that calculates the proportion of permutations in which the number of overlaped peaks lies further toward the tails of the distribution compared to the observed value in the real data. FC was also calculated by dividing the number of overlaps between the real MN or SN peaks and human MNase peaks by the mean of the distribution of the number of overlaps obtained from the shuffled data. To determine co-location enrichment or depletion, a threshold of P-value < 0.001 and FC ≥ 1.5 or FC ≤ 0.66, was set to determine enrichment and depletion, respectively.

The same approach was used for two cattle (GSE47843) comparable MNase-Seq datasets that sequenced the mono-nucleosomal fractions from two bulls (samples s_1_1_bovine and s_5_1_bovine) from the Holstein breed (*Samans et al., 2014*). Genome coordinates from bull (busTau7) genomes were first liftover to hg19, and then from hg19 to Sscr11.

Integration with other -omics data

Although mature sperm is assumed to be transcriptionally inactive, RNAs in sperm may both reflect preceding events in transcriptionally active spermatogenic precursor cells during male germ cell development and have a role in early development after fertilization (*Gòdia, Swanson & Krawetz, 2018*; *Ostermeier et al., 2004*). Thus, the existence of a relationship between the location of retained nucleosomes in sperm and the sperm transcriptome profile was hypothesized. To test this hypothesis, sperm transcriptome data

from 40 healthy Pietrain boars belonging to commercial artificial insemination farms, including 40 total and 35 small RNA-seq datasets previously generated by our group (NCBI's BioProject PRJNA520978) was used. These datasets provided a list of 4,120 protein coding RNAs (*Gòdia et al., 2020b*), 1,574 circular RNAs (circRNAs) mapping in pig chromosomes (*Gòdia et al., 2020a*) and 6,729 PIWI-interacting RNAs (piRNAs) (*Ablondi et al., 2021*). The averaged RNA abundances from all the samples were used for further analysis. Protein coding genes were classified according to their RNA abundances measured in Fragments per Kilobase per Million mapped reads (FPKM) as (i) absent (<1 FPKM); (ii) low abundance (≥1 to <10 FPKM); (ii) intermediate abundance (≥10 to <100 FPKM) and (iii) high abundance (≥100 FPKM). Then, the existence of positional co-location enrichment between each of these gene fractions and the MNase peaks was assessed using the Fisher Exact Test (two-tailed). Any distance below 5 kb between an RNA and a MN or SN peak was considered as co-location. Moreover, the RNA abundance of the genes that co-located with MN and SN peaks with the RNA levels of the whole set of genes annotated in the pig genome was compared by employing the Kruskal-Wallis test using RNA abundances stabilized with the log2 transformation.

The positional enrichment of circRNAs or piRNAs at MNase peaks was also studied. First, the genomic regions of the MN and SN peaks were iteratively randomized 1,000 times using BEDtools (*Quinlan & Hall, 2010*) shuffle. The overlap in the real data and the permutations with the circRNAs or piRNAs was determined with BEDtools v.2.29.2 (*Quinlan & Hall, 2010*) intersect. Then, the enrichment of the genomic coordinates of the real MNase peaks at piRNA and circRNAs was assessed by comparison with the overlap of the randomized MNase peak locations using the permutation test. The FC was also calculated by dividing the number of overlaps between the real MN or SN peaks and circRNAs or piRNAs by the mean of the distribution of the number of overlaps obtained with the shuffled data. A threshold of P-value < 0.001 and FC ≥ 1.5 (enrichment) or FC ≤ 0.66 (depletion), respectively, was set to consider co-location enrichment or depletion.

MNase profiles were also evaluated against the genomic regions showing genetic association with sperm quality traits in Pietrain pigs. These associations were identified through a Genome-Wide Association Study (GWAS) conducted by our group, which included the two samples used in this study (*Gòdia et al., 2020b*). This GWAS provided 18 regions showing genetic association with the percentage of sperm cells with head abnormalities (seven regions), neck abnormalities (four regions), percentage of abnormal acrosomes after 5 min incubation at 37 °C (two regions), the ratio of abnormal acrosomes after 5 min and 90 min of incubation (one region), percentage of motile spermatozoa after 5 min incubation (two regions) and after 90 min incubation (two regions, one shared with motility after 5 min incubation) and the percentage of sperm cells with proximal droplets (one region). These regions spanned in total, 16.5 Mbp (0.7%) of the porcine autosomal genome (*Gòdia et al., 2020b*). Again, to evaluate whether MNase peaks were enriched at GWAS hits, the genomic regions of the MN and SN peaks were iteratively randomized 1,000 times using BEDtools (*Quinlan & Hall, 2010*) shuffle and the overlap of both, the real observed data and the 1,000 randomized files with the GWAS hits was evaluated with BEDtools v.2.29.2 (*Quinlan & Hall, 2010*) intersect. This randomized overlap was

compared to the overlap of the real data using the permutation test. The FC was also calculated by dividing the number of overlaps between the real MN or SN peaks and the GWAS hits by the mean of the distribution of the number of overlaps obtained with the shuffled data. A threshold of *P*-value < 0.001 and FC = 1.5 (enrichment) or FC = 0.66 (depletion), respectively, was set to consider co-location enrichment or depletion.

Peak location relative to repeat elements

The list of repeat elements in the pig genome from RepeatMasker was downloaded from the UCSC genome browser. The same approach as described for the other genomic features was followed to evaluate enrichment or depletion of MNase peaks at the whole set of repeat elements, long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs). The genomic locations of the MN and the SN peaks were randomized 1,000 times and the overlap with the different sets of repeat elements was assessed with BEDtools v.2.29.2 (*Quinlan & Hall, 2010*) intersect. A permutation test was carried to evaluate enrichment (*P*-value < 0.001; FC ≥ 1.5) or depletion (*P*-value < 0.001; FC ≤ 0.66) of MN and SN peaks at repeat elements.

All scripts, including R, bash and software parameters used are available on Figshare: https://doi.org/10.6084/m9.figshare.21997523.v1.

RESULTS

MNase digestion, library preparation and data preprocessing

The MNase protocol yielded 95.2 and 75.6 ng of DNA for sample 1 and sample 2, respectively. The extraction of gDNA provided 1227.6 ng/ul. The MNase treatment of the pig sperm samples generated ~147 bp fragments corresponding to the MN fraction and an additional ~100 bp fragment corresponding to the SN fraction (Fig. S1).

The sequencing of the libraries yielded more than 43 million PE reads for each MNase biological replicate and 41.8 million PE reads for the input sample (Table 1). In average, 90.2% of the MNase reads and 92.7% for the input reads mapped to the porcine genome (Sscrofa11.1). The genome-wide profiles of MNase sensitivity of the two samples showed high correlation (Pearson R = 0.87) (Fig. S2) and similar mapping statistics (Table 1). Thus, the sample pooling strategy carried by Hammoud and collaborators (*Hammoud et al., 2009*) was replicated and the two samples were pooled with the aim to increase read depth and peak call sensitivity and accuracy. The final number of PE reads in each, the MN and SN fractions, was 62.1 and 7.1 million, respectively (Table 1).

Characterization of the MN and SN peaks

The analysis yielded 25,293 MN and 4,239 SN peaks (Tables S1 and S2). The MN peaks averaged 270 bp in width and covered 0.3% of the porcine genome. The SN peaks were 141 bp wide and covered 0.02% of the genome. Most MN (70.2%) and SN (94.5%) peaks were annotated in intronic and intergenic regions (Fig. 1). MN peaks were significantly enriched at promoters (*P*-value < 0.001, FC = 5.05), TSS (*P*-value < 0.001, FC = 11.26) (Fig. 2), 5′-UTR (*P*-value < 0.001, FC = 12.20), coding sequences (*P*-value < 0.001, FC = 4.16) and 3′-UTR (*P*-value < 0.001, FC = 1.56) (Table 2 and Fig. S3). This result was not replicated in

Table 1 **MNase-Seq sequencing and pre-processing metrics for the input and the two MNase biological replicates.**

	Input DNA	MNase replicate A	MNase replicate B
Sequencing reads (PE)	41,805,617	44,372,118	43,267,605
Reads mapped to Sscrofa11.1	38,332,480	39,906,605	38,303,527
% mapping Sscrofa11.1	92.7%	91.4%	89.6%
% duplicates	23.4%	9.3%	8.4%
Number of reads in the MN fraction (PE)		62,133,428	
Number of reads in the SN fraction (PE)		7,139,475	

Note:
PE, paired end reads.

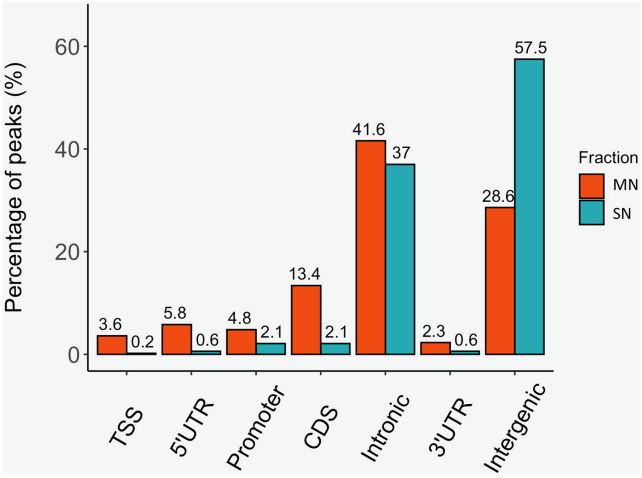

Figure 1 **Distribution of the MN and SN peaks relative to gene features in the pig genome.** Peaks were classified according to their co-location with gene features as, transcription start site (TSS), 5′ untranslated region (5′UTR), 3′UTR, promoter, coding sequence (CDS), intronic and intergenic.
Full-size DOI: 10.7717/peerj.15520/fig-1

the SN peaks, which showed mild depletion at 3'-UTR sites (P-value < 0.001, FC = 0.47) (Table 2 and Fig. S3). A total of 9,128 and 1,688 genes overlapped or localized less than 5 kbp apart from a MN and a SN peak, respectively (Table S3). The two MN and SN fractions co-existed in 1,145 genes (Table S3), which corresponds to 75% of the SN-associated genes.

To get an idea of the potential function of the nucleosome associated DNA, a GO analysis was conducted on the genes that overlapped with or were located within 5 kb of these peaks. For the MN peaks, the most significantly enriched terms were related to chemical perception, development (including the nervous system, multicellular organism, tissue, embryo, *etc.*) and cell and organ morphogenesis (Table S4). The genes located in the SN fractions were enriched for less GO terms and showed weaker significances and included nervous system process, cell projection, regulation of GTPase activity, organelle organization and Golgi vesicle transport (Table S5). To delve further into the potential functions of the nucleosome retention in sperm, the genomic sequences underlying the

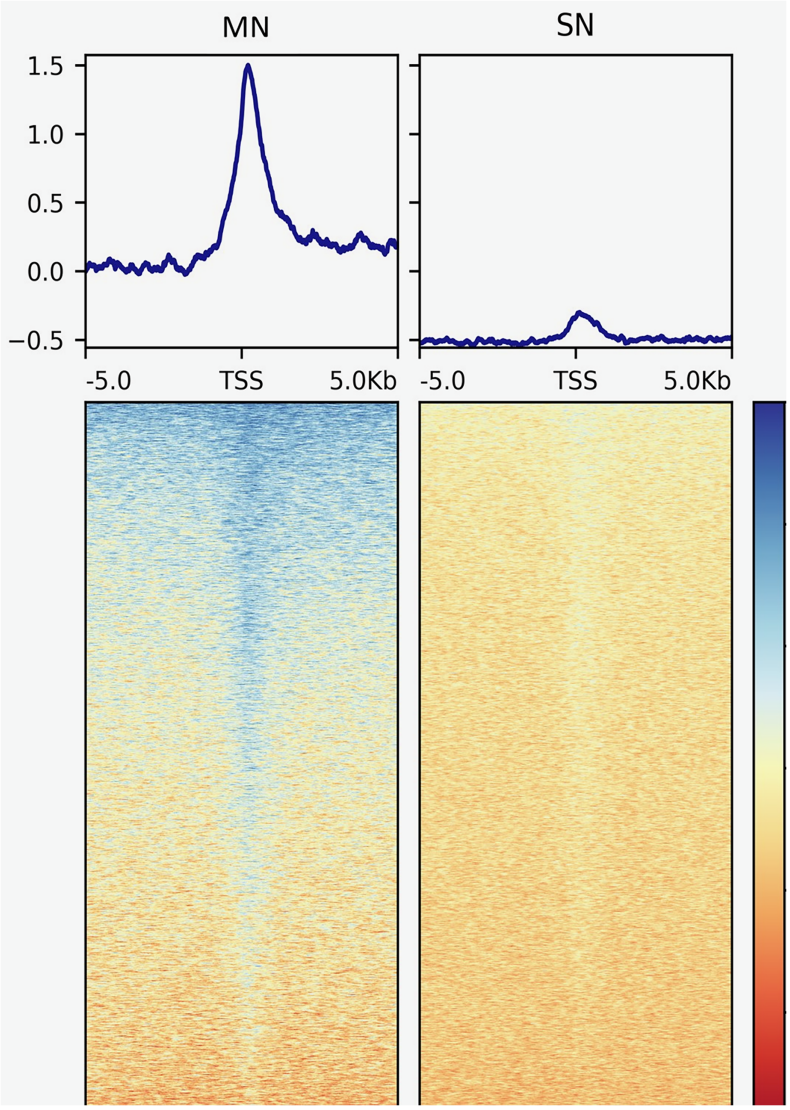

Figure 2 Genomic heatmaps depicting the normalized MNase-Seq signal centered at TSS for the MN and the SN peaks. The x axis shows the genomic location relative to the TSS. The y axis indicates the MNase-Seq signal intensity. Full-size DOI: 10.7717/peerj.15520/fig-2

peak regions were subjected to sequence motif enrichment analysis. Enrichment for motifs from 55 plant and 57 animal (mostly mammalian: mouse and human) transcription factors (TFs) was identified in the MN peaks (Table S6). The animal motifs involved 15 TFs from the bHLH class (MyoD, MyoG, Myf5, NeuroG2, Olig2, Tcf21, Twist2, *etc.*) and other TFs related to embryo development and implantation (Atf4, CHOP, Erra, EBF1, E2F2, Foxh1, HOXA1, HOXA2, MAFK, Pax8, IRF1, IRF3, PR, Rfx1, RORg, RUNX2, Zic, Zic3). They also included TFs related to spermatogenesis or sperm function which have also been related to embryo development such as CUX1, CTCFL, PAX5 and Smad2. Finally, this list also contained Znf263 and FOXA1, two TFs that provide a direct link between the paternal gamete and the embryo (Table S6). One study identified 514 genes with paternal preferential expression during human early embryo development and

Table 2 **Summary of the results of the permutation tests to evaluate the location of the MN and SN peaks in relation to different genomic features.**

	Source	Genomic feature	Real data	Min	Max	Mean	SE	*P*-val	Fold change
MN	Genome annotation	Promoter	2,793	489	624	553.01	22.38	<0.001	5.05
		TSS	914	55	112	81.21	8.97	<0.001	11.26
		5'UTR	2,070	128	205	169.74	12.99	<0.001	12.20
		CDS	3,386	734	891	814.82	28.07	<0.001	4.16
		Intron	10,527	9,956	10,458	10,168.04	79.43	<0.001	1.04
		3'UTR	576	310	431	368.88	18.99	<0.001	1.56
		Intergenic	12,356	14,466	14,931	14,735.91	76.56	<0.001	0.84
	Data specific from pig sperm	circRNA	3,198	2,679	3,023	2,847.75	48.15	<0.001	1.12
		circRNA_corr	14	5	27	15.03	3.82	0.46	0.93
		piRNAs	65	4	29	14.54	3.85	<0.001	4.47
		piRNAs_corr	18	0	11	3.83	1.92	<0.001	4.70
		GWAS	189	123	221	171.05	12.98	0.09	1.10
	RepeatMasker	RE	12,666	16,080	16,597	16,341.85	74.27	<0.001	0.78
		SINE	5,892	6,968	7,421	7,182.70	68.15	<0.001	0.82
		LINE	3,826	7,651	8,085	7,889.53	72.58	<0.001	0.48
	Other species	Human	5,395	163	265	215.71	15.23	<0.001	25.01
		s_1_1_bovine	7	3	34	17.08	4.31	0.009	0.41
		s_5_1_bovine	16	12	44	26.75	5.11	0.014	0.60
SN	Genome annotation	Promoter	104	58	119	87.71	9.37	<0.001	1.19
		TSS	8	1	19	8.18	2.93	0.57	0.98
		5'UTR	27	7	41	22.06	4.76	0.18	1.22
		CDS	88	73	145	103.62	10.06	0.06	0.85
		Intron	1,568	1,575	1,806	1,693.45	32.20	<0.001	0.93
		3'UTR	27	32	83	57.76	7.59	<0.001	0.47
		Intergenic	2,575	2,368	2,571	2,464.93	32.35	<0.001	1.04
	Data specific from pig sperm	circRNA	460	400	542	477.26	20.71	0.21	0.96
		circRNA_corr	4	0	8	2.45	1.60	0.25	1.63
		piRNAs	5	0	9	1.59	1.29	0.02	3.15
		piRNAs_corr	3	0	3	0.38	0.62	0.01	7.87
		GWAS	24	14	48	28.92	5.48	0.22	0.83
	RepeatMasker	RE	3,559	2,381	2,579	2,485.37	32.50	<0.001	1.43
		SINE	1,817	890	1,062	979.63	27.83	<0.001	1.85
		LINE	874	1,065	1,250	1,163.82	29.09	<0.001	0.75
	Other species	Human	88	17	50	31.73	5.55	<0.001	2.77
		s_1_1_bovine	1	0	9	2.68	1.61	0.249	0.37
		s_5_1_bovine	1	0	11	3.64	1.92	0.112	0.28

Note:
Min, Max, Mean, minimum, maximum and mean number of overlaps for the 1,000 simulations, respectively. SE, standard error. TSS, transcription start site. 5′UTR, 5′ untranslated region. CDS, coding sequence. 3′UTR, 3′ untranslated region. circRNA_corr, circular RNAs which abundance level in sperm correlated to semen quality traits in swine as described by *Gòdia et al. (2020a)*. piRNA_corr, piRNAs which abundance level in sperm correlated to semen quality traits in swine as described by *Ablondi et al. (2021)*. GWAS, GWAS hits for semen quality traits in pigs as described by *Gòdia et al. (2020b)*. RE, repeat elements in the pig genome. SINE, short interspersed nuclear elements in swine. LINE, long interspersed nuclear element in swine.

proposed *ZNF263* as the strongest candidate in regulating the expression of these genes (*Leng et al., 2019*). The pig genes that map less than 500 bp apart from a MN peak harboring a Znf263 motif were identified, and their overlap with the 514 human genes showing paternal preferential expression in the early embryo was determined. 3,264 genes mapped less than 500 bp away from a MN peak with Znf263 binding motif in pig sperm. These corresponded to 3,288 human orthologs, 92 of which (18% of 514) also showed preferential paternal expression during human early embryo development (*Leng et al., 2019*) (Table S7). This concordance almost reached significance, with a *P*-value of 0.052 as calculated using the hypergeometric test. The SN peaks presented enriched motifs for 28 animal and 46 plant TFs. Thirteen of the animal motifs enriched in the SN peaks were also enriched in the MN fraction (Table S8). The common TFs included some of the bHLH class (NeuroG2, Olig2 and Twist2), as well as TFs involved in embryo development and implantation (Atf4, CHOP, ERRA, Foxh1, HOXA1, HOXA2, IRF1 and Zic).
The SN-specific enriched motifs corresponded to several TFs related to embryo development (NFX, Duxbl, PBX1, Pitx1, SF1, NFIL3, Gfi1b, HLF), meiosis (DMC1) and parent-of-origin expression driven by imprinting in embryo (Zfp57).

A comparable human sperm MNase-Seq dataset, which included pool data from four donors (*Hammoud et al., 2009*) and focused on MN peaks, was employed to assess the inter-species conservation of the genomic location of our MNase sites. Results in human were similar to these detected in our analysis in pig. 25,122 peaks were annotated in the human MN fraction, 20,641 of which were successfully converted to the pig genome coordinates. Twenty-six percent (5,395 peaks) of the human nucleosome associated sites overlapped with our porcine MN peaks. This is a highly significant co-location between the two species (*P*-value < 0.001, FC = 25.01), compared to randomization (Table 2 and Fig. S3). The two cattle samples contained 2,256 (s_1_1_bovine) and 8,446 (s_5_1_bovine) MN peaks. Of these, 1,011 (s_1_1_bovine) and 4,186 (s_5_1_bovine) were successfully liftover to pig and used to determine genomic overlap. Only 7 (s_1_1_bovine) and 16 (s_5_1_bovine) cattle MN peaks overlapped with the porcine MN peaks and no significant enrichment or depletion was observed.

To identify positional relationships between the nucleosome-associated regions and transcriptional activity, the locations of the MN and SN peaks were compared with the repertoire of RNAs that our group identified in porcine sperm (*Gòdia et al., 2020b*). A total of 12,125 protein coding genes were absent in sperm. On the contrary, 5,814, 3,521 and 598 genes were classified as being at low, moderate and high abundance in sperm, respectively. The location of the genes that are present in spermatozoa was, regardless of their abundance, highly enriched (low abundance: *P*-value = 1.8E−72, moderate abundance: *P*value = 1.5E−91, high abundance: P-value = 1.1E−11) within or near the MN peaks when compared to the catalog of absent genes (Table 3). Similarly, the SN peaks were also enriched for the low (*P*-value = 5.0E−17), moderate (*P*-value = 3.6E−47) and highly abundant (*P*value = 3.0E−5) genes, when compared to the set of absent genes (Table 3). In line with these results, the average RNA abundance of the genes mapping to both MN (*P*value = 4.9E−19) and SN (*P*-value = 4.3E−22) peaks was significantly higher than the average abundance of all the genes annotated in the pig genome (Fig. 3).

Table 3 **Distribution of the protein coding genes within the MN and SN fractions, according to their RNA abundance in sperm.**

		Genes within the MN peaks			Genes within the SN peaks		
		Less than 5 kb from a MN peak	More than 5 kb from a MN peak	*P*-value	Less than 5 kb from a SN peak	More than 5 kb from a SN peak	*P*-value
Gene's RNA levels in sperm	Not present	4,083	8,042		644	11,481	
	Low abundance	2,774	3,040	1.80E−72	504	5,310	5.00E−17
	Intermediate abundance	1,857	1,664	1.50E−91	455	3,066	3.60E−47
	High abundance	296	302	1.10E−11	60	538	3.00E−05

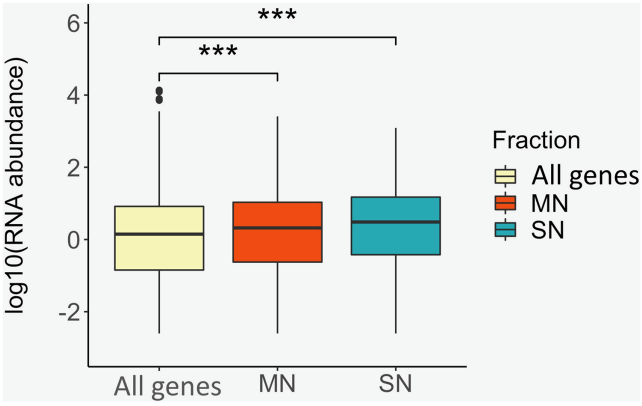

Figure 3 **Box plots showing the average RNA abundance of (i) all genes present in the genome (yellow), (ii) the genes present in the MN fraction (red) and (iii) the genes present in the SN fraction (blue).** ***: *P*-value ≤ 0.001. Full-size DOI: 10.7717/peerj.15520/fig-3

Recently, our group annotated 6,729 sperm piRNAs in porcine sperm (*Ablondi et al., 2021*). The piRNA regions showed a significantly higher frequency of mapping within MN peaks (*P*-value < 0.001, FC = 4.47) and exhibited an overlap of 65 MN peaks (Table 2 and Fig. S3). A similar enriched co-location was also observed ((*P*-value < 0.001, FC = 4.70, overlap with 18 MN peaks) for the set of piRNA genomic regions involving 1,355 piRNAs (Table 2 and Fig. S3), which abundance in sperm was also found to correlate with various sperm phenotypes. These phenotypes include the percentage of motile sperm, the percentage of morphological abnormalities or the percentage of viable cells (*Ablondi et al., 2021*). No obvious enrichment or depletion of SN at piRNAs sites was observed (Table 2 and Fig. S3). No positive or negative co-location was observed between MN or SN peaks and the catalog of 1,574 sperm circRNAs and 148 circRNAs associated to sperm motility in swine (*Gòdia et al., 2020a*) (Table 2 and Fig. S3).

Our group recently published a GWAS for sperm quality traits in Pietrain boars (*Gòdia et al., 2020b*) which was used to evaluate the co-location of MNase sites with regions associated to these traits that could suggest a potential link between nucleosome retention

and sperm quality. No significant co-location between MN or SN and the GWAS hits was found (Table 2 and Fig. S3).

Finally, the MN peaks showed depletion at LINEs (*P*-value < 0.001, FC = 0.48), while the SN sites presented enrichment at SINEs (*P*-value < 0.001, FC = 1.85) (Table 2 and Fig. S3).

DISCUSSION

This is a descriptive study that aimed at evaluating for the first time in swine, the tendency of the sperm's retained nucleosomes to co-map with different types of functional elements of the genome, which could be indicative of a non-random nucleosome retention. Echoing the results observed in other species (*Hammoud et al., 2009*; *Castillo et al., 2014*; *Samans et al., 2014*; *Carone et al., 2014*), the digestion resulted in two DNA bands (Fig. S1). The ~147 bp MN band corresponding to mono-nucleosomes consists of of DNA wrapped around the histone octamer comprising two copies of the core histones H2A, H2B, H3, and H4 (*Luger et al., 1997*). A SN band has been also detected in yeast (*Henikoff et al., 2011*), *Xenopus laevis* sperm (*Oikawa et al., 2020*), human sperm (*Castillo et al., 2014*), mouse sperm (*Carone et al., 2014*) and cattle sperm (*Samans et al., 2014*). The nature, composition and function of these short particles is still not well defined. Proteomic analysis of SN bands in *Xenopus* sperm revealed association with chromatin regulatory proteins (H3, H4, H1FX, CBX3, WDR5) and reduced amounts of H2A and H2B thereby leading to the hypothesis that SN are partially unwrapped nucleosomes that have lost one or the two H2A and H2B dimers (*Oikawa et al., 2020*). Other articles describe that at least in yeast and somatic cells, these partially unwrapped nucleosomes mark TFs to their cognate binding sequence, thus representing a novel signature of active chromatin as reviewed by *Brahma & Henikoff (2020)*. Another line of research suggests that SN particles in sperm corresponds to DNA sequences bound to CTCF (*Carone et al., 2014*), which is an important regulator of chromatin structure and transcriptional activity that has also been related to transgenerational inheritance from sperm to the offspring (*Jung et al., 2022*). In our study, there was no enrichment of CTCF binding sites in the SN peaks, even though enrichment of CTCFL, a CTCF paralog that is specific to the male germline, was observed in the MN fraction. Enrichment of putative binding sites for some TF was found, nearly half of which were also enriched in the MN peaks.

This, together with the opposite behavior observed between MN and SN regarding their location in relation to 3'UTRs and repeat elements (MN are depleted at LINEs and SN are enriched at SINEs) could be indicating that the SN fraction identified in our study may correspond to partially unwrapped nucleosomes generated during spermiogenesis or by the MNase digestion protocol.

The extraction of nucleosomal DNA was less efficient than in experiments carried in sperm from other species (*Hammoud et al., 2009*; *Brykczynska et al., 2010*; *Castillo et al., 2014*; *Samans et al., 2014*; *Erkek et al., 2013*), and it did not yield sufficient amount of DNA from the electrophoretic bands for high throughput sequencing. Nevertheless, the total amount of digested DNA was enough for sequencing and fragment size separation after read mapping using bioinformatics tools. To directly sequence the MN and SN fractions, future experiments will require to optimize the efficiency of the protocol or, less ideal, to

process a larger number of sperm cells to isolate sufficient DNA from each MNase electrophoretic band.

The nucleosome-associated DNA spanned ~0.3% of the porcine sperm's genome, which indicates a nearly fifteen, three and 45 -fold decrease in nucleosome retention when compared to what has been described in human (*Hammoud et al., 2009*; *Arpanahi et al., 2009*; *Gatewood et al., 1987*; *Brykczynska et al., 2010*; *Castillo et al., 2014*), mouse (*Balhorn, Gledhill & Wyrobek, 1977*; *Erkek et al., 2013*; *Johnson et al., 2016*), and cattle (*Samans et al., 2014*), respectively. This difference could have technical causes as previously described by Carone and co-authors regarding the existence of fragile and stable nucleosomes differing in their susceptibility to MNase digestion and the fact that the extent of MNase digestion can impact on the number of nucleosomes remaining in the sample (*Carone et al., 2014*). However, the existence of interspecies variability cannot be ruled out, which could be partially driven by differences in the protamine amino acid content resulting in changes on the extent or arrangement of the protamine disulfide bonding (*Perreault et al., 1988*).

An increasing body of evidence in other animal species points towards a programmatic nucleosome retention in the sperm genome (*Hammoud et al., 2009*; *Castillo et al., 2014*). In our study, the largest abundance of MN and SN peaks in intronic and intergenic regions (Fig. 1), is not unexpected since these gene features cover the vast majority of the genomic space. As a matter of fact, MN enrichment at promoters and exons (Table 2 and Fig. S3), provides the first indication that nucleosome retention in sperm may not be stochastic and may relate to the genome's functional activity. This hypothesis is further supported by the results obtained by comparing the genomic location of the MN peaks in porcine sperm with that in human; the gene annotation of the pig genome; the prediction of TF binding motifs and our own data on the pig sperm transcriptome.

The concordance with previous findings in other species regarding the predominant location of nucleosome peaks in gene promoters (*Hammoud et al., 2009*) and gene bodies (*Yamaguchi et al., 2018*) as well as the overlap with their human synthenic regions (*Hammoud et al., 2009*), indicates a degree of inter-species conservation in gene regulation, normally associated to the maintenance of important functions. Despite the significant overlap with the human MN map, still 73% of the human MN did not overlap to any of the porcine counterparts. This lack of full overlap could be attributed to both technical and biological reasons. As previously mentioned, the extend of MNase digestion could lead to differences in the MN profiles obtained by MNase-Seq (*Carone et al., 2014*). Furthermore, the possibility that porcine and human sperm have different sensitivities to the MNase digestion cannot be ruled out. The biological factors that could lead to the inter-species differences in the MN retention include the existence of a certain degree of random retention of nucleosomes due to their inefficient removal during spermiogenesis, differences in the molecular mechanisms underlying spermatogenesis (*Murat et al., 2023*) or embryogenesis (*Lu et al., 2021*), or genetic redundancy (*Nowak et al., 1997*). Likewise, the lack of overlap with the cattle dataset might be attributed to differences in the protocol but also to the small number of MN peaks detected in the two cattle samples and the poor performance of the liftover from cattle to pig. The poor success of the liftover might have been caused by the double-step liftover from cattle to human and from human to pig

which was necessary due to the lack of a direct liftover between cattle and pig in the UCSC liftover tool.

A larger number of strongly enriched GO terms for the catalog of genes mapping nearby the MN when compared to the SN fractions was observed (Tables S4 and S5). This could be in part due to both, the low number of observed SN peaks when compared to MN sites, and the fact that the percentage of intergenic SN locations (57.5%) doubled that of intergenic MN peaks (28.6%). Hence, the GO analysis run for SN included a lower number of genes, and consequently, yielded a small number of enriched GO terms. The most enriched terms for the MN fraction included sensory perception, which is important for sperm homeostasis (*e.g.*, *SOD2*; (*Malivindi et al., 2018*)), capacitation (*e.g.*, *TRPV1*; (*Osycka-Salut et al., 2020*)) and chemotaxis guidance to the egg for fertilization (*e.g.*, *CA6*; (*Boué et al., 1995*)) and cell differentiation, development and morphogenesis, all with obvious links to embryogenesis (Table S4).

The search for TF motifs at MN peaks also indicated that MN peaks may carry instructions for embryo development. First, enriched motifs for several TFs related to embryo development were identified (Table S6). Some of these genes point toward a role of the paternal gamete in regulating embryo development. This is the case for *ATF4* (*Puscheck et al., 2015*) and *CHOP* (*Ali et al., 2018*), which are relevant in the stress response of the mouse embryo at key embryogenesis steps such as zygotic genome activation. FOXA1 is bound, in the chromatin of human sperm, to genomic regions corresponding to enhancer elements in embryonic stem cells and in several cell types of the embryo, which could suggest that the position of this TF in sperm may guide the location of relevant enhancers for embryo development (*Jung et al., 2019*). The identification of enriched Znf263 motifs is another relevant finding. *ZNF263* has been proposed as a master regulator of the genes showing paternal preferential expression in the early developing human embryo (*Leng et al., 2019*). The nearly significant enrichment (*Pvalue* = 0.052), of genes exhibiting paternal preferential expression in the human early embryo that correspond to pig orthologs mapping near a MN peak with predicted ZNF263 binding motif in sperm (Table S7) suggest that these MN peaks may be epigenetic marks in sperm that trigger the expression of specific genes upon fertilization. All in all, this data is in line with the hypothesis that nucleosome retention in sperm has guiding relevance in the early embryo development.

Although mature sperm is assumed to be transcriptionally silent, it carries a wide repertoire of RNAs related to spermatogenesis, fertilization, embryo development and offspring phenotype (reviewed in: (*Gòdia, Swanson & Krawetz, 2018*)), a large proportion of which were transcribed in transcriptionally active spermatogenic cells during previous stages of spermatogenesis. Our group has generated sperm RNA-Seq (*Gòdia et al., 2020a*; *Ablondi et al., 2021*; *Gòdia et al., 2020b*) and GWAS data for semen quality traits (*Gòdia et al., 2020b*) from porcine samples. The RNA-Seq and GWAS studies included 40 and 276 samples, respectively. Thirty-five of the 40 pigs that provided sperm for RNA-Seq also contributed data for the GWAS. The two boars used in the MNase study, also Pietrain, participated in the GWAS but not in the RNA-Seq. The regulation of transcription is modulated by nucleosome occupancy and the accessibility of the genome to the

transcriptional machinery (*Henikoff, Furuyama & Ahmad, 2004*). Our results show that the genes mapping within or nearby retained nucleosomes tend to display larger RNA abundance than the full set of genes annotated in the porcine genome (Fig. 3). These results also indicate that the genes that contributed detectable RNA levels by RNA-Seq in sperm, tend to map to nucleosome-retained loci in both the MN and SN fractions. Not only protein coding but also a family of regulatory RNAs, the piRNAs, were also significantly enriched at MN and SN peaks. Again, this data supports the notion that nucleosome occupancy is key in modulating gene expression. Moreover, the enriched co-location between the MN and SN sites and the catalog of piRNA regions (*Ablondi et al., 2021*) which abundance correlated with sperm quality phenotypes further suggests that this regulation during spermatogenesis may have phenotypic consequences on semen quality and reproduction. Thus, the nucleosomes and sub-nucleosomes associated to these RNAs may be leftovers from spermatogenesis and might provide useful information for the identification of markers of abnormal spermatogenesis and sperm quality.

In light of these results, and considering that nucleosome positioning can modulate gene expression (*Bai & Morozov, 2010*; *Lorch, LaPointe & Kornberg, 1987*; *Li, Carey & Workman, 2007*), it was hypothesized that sperm-retained nucleosomes could encompass DNA variants altering elements such as promoters or enhancers regulating the expression of genes that played a role during spermatogenesis. In such case, some degree of co-location between MNase peaks and GWAS hits should be observed. However, this trend between the MN and SN peaks with the GWAS hits for semen quality was not detected (Table 2 and Fig. S3) (*Gòdia et al., 2020b*).

The reasons, whether biological or technical, underlying the lack of a co-location trend with GWAS hits for semen quality and the catalog of sperm circRNAs cannot be determined. For the GWAS, one possibility is that because the GWAS hits are in fact, highlighting genomic regions in linkage disequilibrium with the causal functional variant, the GWAS region is expected to be much larger than this causal variant and probably most of the GWAS regions do not correspond to functional elements of the genome (*e.g.*, promoters, enhancers or genes) that could be overlapping with nucleosomes. Regarding the circRNAs, although the FC did not reach our threshold, none of the random permutations reached the same number of overlaps between MN peaks and the catalog of circRNAs indicating that, although modest, there might be a trend of co-location between circRNAs and MN peaks (Table 2).

Noteworthy, our results are not in disagreement with the hypothesis that sperm nucleosomes mostly map at gene-poor regions and repeat elements (*Carone et al., 2014*; *Samans et al., 2014*; *Oikawa et al., 2020*). The depletion of MN and enrichment of SN at LINEs and SINEs, respectively, might be indicating that nucleosomes in these repeat elements are more susceptible to MNase digestion as suggested by Carone and colleagues (*Carone et al., 2014*).

Four technical considerations in relation to the results of this study need to be taken into consideration. The first consideration is that the presence of somatic cells, which contain large number of histones in their chromatin, in the sperm samples could seriously mask the nucleosome profiles of the sperm cells. Confidence can be placed on the absence of

somatic cells in our samples for two reasons. First, commercial pig sperm doses rarely present somatic cell contamination. Second, the sperm samples were subjected to Bovipure™ purification which removes somatic cells. Moreover, the number of MN peaks identified in our study is close to the number of MN peaks found by *Hammoud et al. (2009)* in human, thereby indicating that the MNase profiles obtained in our study are not influenced by somatic cells. The second consideration is the fact that our chromatin extraction protocol yielded lower amount of DNA than what has been obtained in other species described in different studies, which could be indicating that our conditions were not fully optimized for pigs. The third consideration is that ChIP-Seq, either for H3 or for different histone modifications related to specific genome activity would have been more precise and informative than MNase-Seq. However, the amount of DNA obtained after the MNase digestion was limited and the immunoprecipitation would most likely have resulted in insufficient amounts of DNA for high throughput sequencing. Nevertheless, the MNase protocol has provided the required information to suggest that nucleosome retention in sperm may contain relevant information in relation to spermatogenesis, semen quality and embryo development. The fourth consideration relates to the fact that the results are based on a single pool of two samples and consequently, the robustness of the location of each particular nucleosomal site cannot be assessed. Nonetheless, the aim of this work was not to robustly map the specific location of nucleosomes but to evaluate whether the retained histones tend to co-map within functional features of the genome, which would indicate that nucleosome retention is not random but programmatic. The statistical confidence of our study was assessed within a sample, by testing the enrichment of positional overlap between the identified nucleosomal sites and the different functional features of the genome against the null basis of random location. Remarkably, all but two of the comparisons showed statistically significant co-location thereby indicating, that nucleosome retention in pig sperm is not stochastic and may owe to a reminiscent gametogenesis program and a future embryogenesis program. Still, since only one pool of two samples was evaluated, the results should be considered as preliminary and future research will be required to robustly establish the functional relevance of nucleosome retention in the sperm genome in swine.

CONCLUSIONS

The results of our analyses show than the retained nucleosomes in the porcine sperm chromatin tend to overlap with different functional features of the genome such as promoters, exons, genes expressed in sperm and associated to semen quality, putative TFBS for TFs related to reproduction and embryogenesis. In conclusion, the findings suggest that nucleosome retention in mature spermatozoa is related to transcriptional regulation during spermatogenesis and is also an instructional contributor to early embryo development. Hence, interrogating the nucleosome occupancy in the sperm chromatin could help elucidating the biological basis of sperm quality and early embryo development and perhaps, could even assist in the search for biomarkers for these sets of traits.

ACKNOWLEDGEMENTS

We would like to thank Sue Hammoud and all the members of Hammoud lab for their support. We also would like to thank Sam Balasch (Grup Gepork S.A., Catalonia) for supplying the sperm samples.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under grant AGL2013-44978-R and grant AGL2017-86946-R and by the CERCA Programme/Generalitat de Catalunya. AGL2017-86946-R was also funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERDF). We had financial support from the Agency for Management of University and Research Grants (AGAUR) of the Generalitat de Catalunya (Grant Numbers 2014 SGR 1528 and 2017 SGR 1060) and from the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centers of Excellence in R&D SEV-2015-0533 and CEX2019-000902-S. Marta Gòdia had a Ph.D. studentship from MINECO (Grant Number BES-2014-070560) and Short-Stay fellowships from MINECO at SSH group (EEBB-I-16-11528) and at CSIRO (EEBB-I-18-12860). The publication of this work was supported by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Spanish Ministry of Economy and Competitiveness (MINECO): AGL2013-44978-R, AGL2017-86946-R.
CERCA Programme/Generalitat de Catalunya.
Spanish State Research Agency (AEI) and the European Regional Development Fund (ERDF): AGL2017-86946-R.
Agency for Management of University and Research Grants (AGAUR) of the Generalitat de Catalunya: 2014 SGR 1528 and 2017 SGR 1060.
Spanish Ministry of Science and Innovation.
Severo Ochoa Programme for Centers of Excellence in R&D: SEV-2015-0533 and CEX2019-000902-S.
MINECO: BES-2014-070560.
Short-Stay fellowships from MINECO at SSH group: EEBB-I-16-11528.
CSIRO: EEBB-I-18-12860.
CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Marta Gòdia conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yu Lian analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Marina Naval-Sanchez conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Inma Ponte performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Joan Enric Rodríguez-Gil performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Armand Sanchez conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Alex Clop conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at the NCBI's Short Read Archive: SRR14117448, SRR14117447, SRR14117446.

The code is available at Figshare: Gòdia, M (2023): Commandlines_MNase_seq.sh. figshare. Online resource. https://doi.org/10.6084/m9.figshare.21997523.v1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.15520#supplemental-information.

REFERENCES

Ablondi M, Gòdia M, Rodriguez-Gil JE, Sánchez A, Clop A. 2021. Characterisation of sperm piRNAs and their correlation with semen quality traits in swine. *Animal Genetics* **52(1)**:114–120 DOI 10.1111/age.13022.

Ali I, Liu HX, Zhong-Shu L, Dong-Xue M, Xu L, Shah SZA, Ullah O, Nan-Zhu F. 2018. Reduced glutathione alleviates tunicamycin-induced endoplasmic reticulum stress in mouse preimplantation embryos. *Journal of Reproduction and Development* **64(1)**:15–24 DOI 10.1262/jrd.2017-055.

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 Research* **19(8)**:1338–1349 DOI 10.1101/gr.094953.109.

Bai L, Morozov A. 2010. Gene regulation by nucleosome positioning. *Trends in Genetics* **26(11)**:476–483 DOI 10.1016/j.tig.2010.08.003.

Balhorn R, Gledhill BL, Wyrobek AJ. 1977. Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16(18)**:4074–4080 DOI 10.1021/bi00637a021.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J. 2009. ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25(8)**:1091–1093 DOI 10.1093/bioinformatics/btp101.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Boué F, Duquenne C, Lassalle B, Lefèvre A, Finaz C. 1995. FLB1, a human protein of epididymal origin that is involved in the sperm-oocyte recognition process. *Biology of Reproduction* **52(2)**:267–278 DOI 10.1095/biolreprod52.2.267.

Brahma S, Henikoff S. 2020. Epigenome regulation by dynamic nucleosome unwrapping. *Trends in Biochemical Sciences* **45(1)**:13–26 DOI 10.1016/j.tibs.2019.09.003.

Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schubeler D, Stadler MB, Peters AH. 2010. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature Structural & Molecular Biology* **17(6)**:679–687 DOI 10.1038/nsmb.1821.

Carone B R, Hung JH, Hainer S J, Chou M-T, Carone DM, Weng Z, Fazzio TG, Rando OJ. 2014. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental Cell* **30(1)**:11–22 DOI 10.1016/j.devcel.2014.05.024.

Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballesca JL, Oliva R. 2014. Genomic and proteomic dissection and characterization of the human sperm chromatin. *Molecular Human Reproduction* **20(11)**:1041–1053 DOI 10.1093/molehr/gau079.

Erkek S, Hisano M, Liang C-Y, Gill M, Murr R, Dieker J, Schübeler D, van der Vlag J, Stadler MB, Peters AHFM. 2013. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature Structural & Molecular Biology* **20(7)**:868–875 DOI 10.1038/nsmb.2599.

Gatewood JM, Cook GR, Balhorn R, Bradbury EM, Schmid CW. 1987. Sequence-specific packaging of DNA in human sperm chromatin. *Science* **236(4804)**:962–964 DOI 10.1126/science.3576213.

Gòdia M, Castelló A, Rocco M, Cabrera B, Rodríguez-Gil JE, Balasch S, Lewis C, Sánchez A, Clop A. 2020a. Identification of circular RNAs in porcine sperm and evaluation of their relation to sperm motility. *Scientific Reports* **10(1)**:1–11 DOI 10.1038/s41598-020-64711-z.

Gòdia M, Estill M, Castelló A, Balasch S, Rodríguez-Gil JE, Krawetz SA, Sánchez A, Clop A. 2019. A RNA-Seq analysis to describe the boar sperm transcriptome and its seasonal changes. *Frontiers in Genetics* **10**:299 DOI 10.3389/fgene.2019.00299.

Gòdia M, Mayer FQ, Nafissi J, Castelló A, Rodríguez-Gil JE, Sánchez A, Clop A. 2018. A technical assessment of the porcine ejaculated spermatozoa for a sperm-specific RNA-seq analysis. *Systems Biology in Reproductive Medicine* **64(4)**:291–303 DOI 10.1080/19396368.2018.1464610.

Gòdia M, Reverter A, González-Prendes R, Ramayo-Caldas Y, Castelló A, Rodríguez-Gil JE, Sánchez A, Clop A. 2020b. A systems biology framework integrating GWAS and RNA-seq to shed light on the molecular basis of sperm quality in swine. *Genetics Selection Evolution* **52(1)**:1–21 DOI 10.1186/s12711-020-00592-0.

Gòdia M, Swanson G, Krawetz SA. 2018. A history of why fathers' RNA matters. *Biology of Reproduction* **99(1)**:147–159 DOI 10.1093/biolre/ioy007.

Halstead MM, Kern C, Saelao P, Wang Y, Chanthavixay G, Medrano JF, van Eenennaam AL, Korf I, Tuggle CK, Ernst CW, Zhou H, Ross PJ. 2020. A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics* **21(1)**:698 DOI 10.1186/s12864-020-07078-9.

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(7254)**: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. *Molecular Cell* **38(4)**:576–589 DOI 10.1016/j.molcel.2010.05.004.

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **108(45)**:18318–18323 DOI 10.1073/pnas.1110731108.

Henikoff S, Furuyama T, Ahmad K. 2004. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics* **20(7)**:320–326 DOI 10.1016/j.tig.2004.05.004.

Ihara M, Meyer-Ficca ML, Leu NA, Rao S, Li F, Gregory BD, Zalenskaya IA, Schultz RM, Meyer RG. 2014. Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression. *PLOS Genetics* **10(5)**:e1004317 DOI 10.1371/journal.pgen.1004317.

Johnson GD, Jodar M, Pique-Regi R, Krawetz SA. 2016. Nuclease footprints in sperm project past and future chromatin regulatory events. *Scientific Reports* **6(1)**:25864 DOI 10.1038/srep25864.

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(1)**:154.e5–171.e5 DOI 10.1016/j.molcel.2019.04.014.

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(6)**:1366–1382 DOI 10.1016/j.celrep.2017.01.034.

Jung YH, Wang H-LV, Ruiz D, Bixler BJ, Linsenbaum H, Xiang J-F, Forestier S, Shafik AM, Jin P, Corces VG. 2022. Recruitment of CTCF to an Fto enhancer is responsible for transgenerational inheritance of BPA-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America* **119(50)**:e2214988119 DOI 10.1073/pnas.2214988119.

Khezri A, Narud B, Stenseth E-B, Johannisson A, Myromslien FD, Gaustad AH, Wilson RC, Lyle R, Morrell JM, Kommisrud E, Ahmad R. 2019. DNA methylation patterns vary in boar sperm cells with different levels of DNA fragmentation. *BMC Genomics* **20(1)**:897 DOI 10.1186/s12864-019-6307-8.

King GJ, Macpherson JW. 1973. A comparison of two methods for boar semen collection. *Journal of Animal Science* **36(3)**:563–565 DOI 10.2527/jas1973.363563x.

Krawetz SA. 2005. Paternal contribution: new insights and future challenges. *Nature Reviews Genetics* **6(8)**:633–642 DOI 10.1038/nrg1654.

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Briefings in Bioinformatics* **14(2)**:144–161 DOI 10.1093/bib/bbs038.

Lai WKM, Pugh BF. 2017. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature Reviews Molecular Cell Biology* **18(9)**:548–562 DOI 10.1038/nrm.2017.47.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9(4)**:357–359 DOI 10.1038/nmeth.1923.

Leng L, Sun J, Huang J, Gong F, Yang L, Zhang S, Yuan X, Fang F, Xu X, Luo Y, Bolund L, Peters BA, Lu G, Jiang T, Xu F, Lin G. 2019. Single-cell transcriptome analysis of uniparental embryos reveals parent-of-origin effects on human preimplantation development. *Cell Stem Cell* **25(5)**:697.e6–712.e6 DOI 10.1016/j.stem.2019.09.004.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128(4)**:707–719 DOI 10.1016/j.cell.2007.01.015.

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 Research* **48(20)**:11380–11393 DOI 10.1093/nar/gkaa712.

Lorch Y, LaPointe JW, Kornberg RD. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49(2)**:203–210 DOI 10.1016/0092-8674(87)90561-7.

Lu X, Zhang Y, Wang L, Wang L, Wang H, Xu Q, Xiang Y, Chen C, Kong F, Xia W, Lin Z, Ma S, Liu L, Wang X, Ni H, Li W, Guo Y, Xie W. 2021. Evolutionary epigenomic analyses in mammalian early embryos reveal species-specific innovations and conserved principles of imprinting. *Science Advances* **7(48)**:eabi6178 DOI 10.1126/sciadv.abi6178.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389(6648)**:251–260 DOI 10.1038/38444.

Malivindi R, Rago V, de Rose D, Gervasi MC, Cione E, Russo G, Santoro M, Aquila S. 2018. Influence of all-trans retinoic acid on sperm metabolism and oxidative stress: its involvement in the physiopathology of varicocele-associated male infertility. *Journal of Cellular Physiology* **233(12)**:9526–9537 DOI 10.1002/jcp.26872.

Mendonça GA, Morandi R, Souza ET, Gaggini TS, Silva-Mendonça MCA, Antunes RC, Beletti ME. 2017. Isolation and identification of proteins from swine sperm chromatin and nuclear matrix. *Animal Reproduction* **14(2)**:418–428 DOI 10.21451/1984-3143-AR816.

Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mößinger K, Brüning T, Lamanna F, Belles MR, Conrad C, Kondova I, Bontrop R, Behr R, Khaitovich P, Pääbo S, Marques-Bonet T, Grützner F, Almstrup K, Schierup MH, Kaessmann H. 2023. The molecular evolution of spermatogenesis across mammals. *Nature* **613(7943)**:308–316 DOI 10.1038/s41586-022-05547-7.

Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J. 1997. Evolution of genetic redundancy. *Nature* **388(6638)**:167–171 DOI 10.1038/40618.

Oikawa M, Simeone A, Hormanseder E, Teperek M, Gaggioli V, O'Doherty A, Falk E, Sporniak M, D'Santos C, Franklin VNR, Kishore K, Bradshaw CR, Keane D, Freour T, David L, Grzybowski AT, Ruthenburg AJ, Gurdon J, Jullien J. 2020. Epigenetic homogeneity in histone methylation underlies sperm programming for embryonic transcription. *Nature Communications* **11(1)**:3491 DOI 10.1038/s41467-020-17238-w.

Ostermeier GC, Miller D, Huntriss JD, Diamond MD, Krawetz SA. 2004. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429(6988)**:154 DOI 10.1038/429154a.

Osycka-Salut CE, Martínez-León E, Gervasi MG, Castellano L, Davio C, Chiarante N, Franchi AM, Ribeiro ML, Díaz ES, Perez-Martinez S. 2020. Fibronectin induces capacitation-associated events through the endocannabinoid system in bull sperm. *Theriogenology* **153**:91–101 DOI 10.1016/j.theriogenology.2020.04.031.

Pértille F, Alvarez-Rodriguez M, da Silva AN, Barranco I, Roca J, Guerrero-Bosagna C, Rodriguez-Martinez H. 2021. Sperm methylome profiling can discern fertility levels in the

porcine biomedical model. *International Journal of Molecular Sciences* **22(5)**:2679 DOI 10.3390/ijms22052679.

Perreault SD, Barbee RR, Elstein KH, Zucker RM, Keefer CL. 1988. Interspecies differences in the stability of mammalian sperm nuclei assessed in vivo by sperm microinjection and in vitro by flow cytometry. *Biology of Reproduction* **39(1)**:157–167 DOI 10.1095/biolreprod39.1.157.

Puscheck EE, Awonuga AO, Yang Y, Jiang Z, Rappolee DA. 2015. Molecular biology of the stress response in the early embryo and its stem cells. *Advances in Experimental Medicine and Biology* **843**:77–128 DOI 10.1007/978-1-4939-2480-6.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26(6)**:841–842 DOI 10.1093/bioinformatics/btq033.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44(W1)**:W160–W165 DOI 10.1093/nar/gkw257.

Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, Wolf E, Steger K, Dansranjavin T, Schagdarsurengin U. 2014. Uniformity of nucleosome preservation pattern in mammalian sperm and its connection to repetitive DNA elements. *Developmental Cell* **30(1)**:23–35 DOI 10.1016/j.devcel.2014.05.023.

Spadafora C. 2017. Sperm-mediated transgenerational inheritance. *Frontiers in Microbiology* **8**:2401 DOI 10.3389/fmicb.2017.02401.

van der Heijden GW, Derijck AAHA, Ramos L, Giele M, van der Vlag J, de Boer P. 2006. Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Developmental Biology* **298(2)**:458–469 DOI 10.1016/j.ydbio.2006.06.051.

Ward WS. 2010. Function of sperm chromatin structural elements in fertilization and development. *Molecular Human Reproduction* **16(1)**:30–36 DOI 10.1093/molehr/gap080.

Ward WS, Coffey DS. 1991. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cell. *Biology of Reproduction* **44(4)**:569–574 DOI 10.1095/biolreprod44.4.569.

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 Reports* **23(13)**:3920–3932 DOI 10.1016/j.celrep.2018.05.094.

Yoshida K, Muratani M, Araki H, Miura F, Suzuki T, Dohmae N, Katou Y, Shirahige K, Ito T, Ishii S. 2018. Mapping of histone-binding sites in histone replacement-completed spermatozoa. *Nature Communications* **9(1)**:3885 DOI 10.1038/s41467-018-06243-9.

Zalenskaya IA, Bradbury EM, Zalensky AO. 2000. Chromatin structure of telomere domain in human sperm. *Biochemical and Biophysical Research Communications* **279(1)**:213–218 DOI 10.1006/bbrc.2000.3917.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9(9)**:R137 DOI 10.1186/gb-2008-9-9-r137.