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 [Corresp., Equal first author, 1, 2] , **Yu Lian** [Equal first author, 1] , **Marina Naval-Sanchez** [3] , **Inma Ponte** [4] , **Joan Enric Rodríguez-Gil** [5] , **Armand Sanchez** [1, 6] , **Alex Clop** [Corresp. 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

Corresponding Authors: Marta Gòdia, Alex Clop
Email address: marta.godia@wur.nl, alex.clop@csic.es

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. More in detail, 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 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 be both marking 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 encourage undertaking more 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.

1 **Tittle:**

2 Micrococcal nuclease sequencing of porcine sperm suggests enriched co-location between

3 retained histones and genomic regions related to semen quality and early embryo development

5 Marta Gòdia[1+#*], Yu Lian[1#], Marina Naval-Sanchez[2++], Inma Ponte[3], Joan Enric Rodríguez-Gil[4],

6 Armand Sánchez[1,5] and Alex Clop[1,6*]

8 *1. Animal Genomics Group, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-*

9 *UAB-UB, Campus UAB, Cerdanyola del Vallès, Catalonia, Spain*

10 *2. CSIRO Agriculture & Food, St. Lucia, Brisbane, QLD, Australia*

11 *3. Biochemistry and Molecular Biology Department, Biosciences Faculty, Universitat Autonòma*

12 *de Barcelona, Cerdanyola del Vallès, Catalonia, Spain*

13 *4. Department of Animal Medicine and Surgery, School of Veterinary Sciences, Universitat*

14 *Autònoma de Barcelona, Cerdanyola del Vallès, Catalonia, Spain*

15 *5. Department of Animal and Food Sciences, School of Veterinary Sciences, Universitat*

16 *Autònoma de Barcelona, Cerdanyola del Vallès, Catalonia, Spain*

17 *6. Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Catalonia, Spain*

19 #These two authors contributed equally to this work

21 +Current affiliation:

22 Animal Breeding and Genomics, Wageningen University and Research, Wageningen, The

23 Netherlands

25 [++] Current affiliation:

26 Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

28 *Corresponding authors:

29 Alex Clop: alex.clop@csic.es

30 Marta Godia: marta.godia@wur.nl

31 Abstract

32   The mammalian spermatozoon has a unique chromatin structure in which the majority of

33   histones are replaced by protamines during spermatogenesis and a small fraction of nucleosomes

34   are retained at specific locations of the genome. The sperm's chromatin structure remains

35   unresolved in most animal species, including the pig. However, mapping the genomic locations

36   of retained nucleosomes in sperm could help understanding the molecular basis of both sperm

37   development and function as well as embryo development. This information could then be useful

38   to identify molecular markers for sperm quality and fertility traits. Here, Micrococcal Nuclease

39   digestion coupled with high throughput sequencing was performed on pig sperm to map the

40   genomic location of mono- and sub-nucleosomal chromatin fractions in relation to a set of

41   diverse functional elements of the genome, some of which were related to semen quality and

42   early embryogenesis. More in detail, the investigated elements were promoters, the different

43   sections of the gene body, coding and non-coding RNAs present in the pig sperm, potential

44   transcription factor binding sites, genomic regions associated to semen quality traits and repeat

45   elements.

46   The analysis yielded 25,293 and 4,239 peaks in the mono- and sub-nucleosomal fractions,

47   covering 0.3% and 0.02% of the porcine genome, respectively. A cross-species comparison

48   revealed positional conservation of the nucleosome retention in sperm between the pig data and a

49   human dataset that found nucleosome enrichment in genomic regions of importance in

50   development. Both, gene ontology analysis of the genes mapping nearby the mono-nucleosomal

51   peaks and the identification of putative transcription factor binding motifs within the mono- and

52   the sub- nucleosomal peaks showed enrichment for processes related to sperm function and

53   embryo development. There was significant motif enrichment for Znf263, which in humans was

54 suggested to be a key regulator of genes with paternal preferential expression during early

55 embryogenesis. Moreover, enriched positional intersection was found in the genome between the

56 mono-nucleosomal peaks and both the RNAs present in pig sperm and the RNAs related to

57 sperm quality. There was no co-location between GWAS hits for semen quality in swine and the

58 nucleosomal sites. Finally, the data evidenced depletion of mono-nucleosomes in long

59 interspersed nuclear elements and enrichment of sub-nucleosomes in short interspersed repeat

60 elements.

61 These results suggest that retained nucleosomes in sperm could be both marking regulatory

62 elements or genes expressed during spermatogenesis linked to semen quality and fertility and act

63 as transcriptional guides during early embryogenesis. The results of this study encourage

64 undertaking more ambitious research using a larger number of samples to robustly assess the

65 positional relationship between histone retention in sperm and the reproductive ability of boars.

66

67

68

69

70

71   **Introduction**

72   Current research is showing evidences that besides the paternal genome, sperm also carries other

73   molecular information to the zygote. This includes a wide repertoire of RNAs, proteins and

74   epigenetic marks in the form of DNA methylation, retained nucleosomes and histone

75   modifications that can play a role in spermatogenesis, fertility, early embryo development and

76   even the transmission of inter-generational information (Krawetz, 2005; Gòdia, Swanson &

77   Krawetz, 2018; Spadafora, 20172; Hammoud et al., 2009; Oikawa et al., 2020). During

78   spermatogenesis, male germ cells undergo a series of profound morphological and functional

79   changes that conclude in mature spermatozoa. In the last stages of spermatogenesis, when the

80   cell is transcriptionally silent, genomic DNA (gDNA) is highly condensed to fit into the sperm

81   head (Ward & Coffey, 1991) and ensure genomic integrity, early embryo development and

82   ultimately, fertility (Ward, 2010). In mammals, the condensation of the spermatozoon chromatin

83   occurs by the progressive replacement of histones by protamines. However, a small fraction of

84   the sperm DNA remains organized in histones as shown in humans (5 to 15% of histones are

85   retained in the sperm chromatin) (Gatewood et al., 1987; Hammoud et al., 2009; Arpanahi et al.,

86   2009; Brykczynska et al., 2010; Castillo et al., 2014; Samans et al., 2014), cattle (13.4%)

87   (Samans et al., 2014) and mice (1 to 2%) (Balhorn, Gledhill & Wyrobek, 1977; Erkek et al.,

88   2013; Carone et al., 2014; Johnson et al., 2016). Several studies indicate that these nucleosomes

89   are distributed along the genome following a programmatic pattern with preferential retention at

90   gene promoters and developmental related loci (Hammoud et al., 2009; Arpanahi et al., 2009;

91   Brykczynska et al., 2010; Castillo et al., 2014; Jung et al., 2017; Yoshida et al., 2018). However,

92   these results are inconclusive because at least three other publications reported that sperm

93   nucleosomes are enriched at gene-poor regions or repeat elements (Samans et al., 2014; Carone

94 et al., 2014; Yamaguchi et al., 2018) while may be even depleted at promoters (Carone et al.,

95 2014; Samans et al., 2014) and exons (Samans et al., 2014). The reasons underlying these

96 discrepancies, which are not mutually exclusive, are unclear and could owe to technical

97 differences in the protocols used in each study. Carone and colleagues found that nucleosomes

98 differ in their resistance to MNase digestion and that the most fragile nucleosomes tended to map

99 near repeat elements while nucleosomes at promoters were more stable (Carone et al., 2014).

100 Thus, protocols using large amounts of MNase would artifactually result in nucleosomal

101 enrichment at promoters (Carone et al., 2014). However, other studies that used alternative

102 protocols not based in MNase and that claimed to have unbiasedly characterized all nucleosomes

103 present in the sperm chromatin, found enriched nucleosomes at promoters or gene bodies

104 (Johnson et al., 2016; Jung et al., 2017; Yoshida et al., 2018). Another study based on

105 nucleoplasmin instead of MNase to solubilize the sperm's chromatin concluded enrichment at

106 both, promoters and gene bodies as well as gene deserts (Yamaguchi et al., 2018). They

107 hypothesize that MNase based protocols do not retain all nucleosomes (Yamaguchi et al., 2018).

108 Nucleosome positioning in the genome and chromatin accessibility are critical in the regulation

109 of gene expression and the alteration of this epigenomic architecture has been linked to multiple

110 phenotypes in different tissues and cell types (Lai & Pugh, 2017). Nucleosomes in sperm may

111 either be leftovers of gene expression during spermatogenesis or also provide transcriptional

112 instructions upon fertilization for gamete recognition and early embryo development. Jung and

113 co-authors found that a proportion of histone modifications in gene promoters of the mouse

114 sperm chromatin, recapitulate the chromatin structure and transcriptional activity of late

115 spermatids (Jung et al., 2017). Also, a wealth of studies found that at least a proportion of histone

116 modifications in sperm have been shown to be maintained (van der Heijden et al., 2006; Lismer

117   et al., 2020) and even that nucleosome positioning (Ihara et al., 2014) and the location of histone

118   modifications (Oikawa et al., 2020) in sperm are related to gene expression in early embryos.

119   Thus, this information could be of great value to shed light into the catalog of genomic regions

120   and genes related to sperm biology and early embryo development and thus help identifying

121   elusive molecular markers for traits related to sperm quality and fertility.

122   In livestock, the architecture of the sperm chromatin at the genomic level has only been

123   investigated in cattle (Samans et al., 2014). In pigs (*Sus scrofa*), research has been carried to

124   study the sperm RNA (Gòdia et al., 2019; Gòdia et al., 2020a; Ablondi et al., 2021) and protein

125   (Mendonça et al., 2017) populations as well as the DNA methylation patterns (Khezri et al.,

126   2019; Pértille et al., 2021), but the sperm's chromatin structure remains unexplored.

127   This study is sustained by the hypothesis that nucleosome retention in the pig sperm is not

128   stochastic and owes to reminiscences of the spermatogenesis program and future information for

129   the embryogenic program. Our objective was to map the retained nucleosomes in a pool of two

130   pig sperm samples and evaluate their relative location in relation to different genomic features,

131   some of which were associated to sperm biology, semen quality and fertility. To fulfil this

132   purpose, pig spermatozoa were digested with micrococcal nuclease (MNase) and the resulting

133   nucleosome-associated DNAs were subjected to high throughput sequencing. The mono-

134   nucleosomal (MN) and sub-nucleosomal (SN) chromatin fractions were characterized and their

135   correlation with sperm RNA levels and co-location with genomic regions associated to semen

136   traits was assessed.

137

138   **Materials & Methods**

139 *Sample collection*

140 Two ejaculates, each from a different healthy Pietrain boar, MN_1 (16 months of age) and

141 MN_2 (9 months old), from two artificial insemination centres were obtained by specialists

142 during their routine sample collection using the gloved hand method (King & Macpherson, 1973)

143 and were immediately diluted (1:2) in commercial extender. Both ejaculates showed good semen

144 quality as evidenced by different sperm parameters. The percentages of sperm cell viability,

145 structurally altered acrosomes and morphological abnormalities were measured by staining the

146 samples with the eosin-nigrosin technique after 5 min incubation at 37°C. These analyses

147 showed 94.3% and 92.2% of viable cells for sample MN_1 and MN_2 respectively; 96.7% and

148 96.3% of cells with normal acrosomes, respectively; 2.2% and 1.1%, 3.1 and 1.1% and 2.2% and

149 4.6%, of head, neck and tail abnormalities, respectively. Semen samples were purified with

150 BoviPure™ (Nidacon) to remove potentially present somatic cells and processed as described in

151 (Gòdia et al., 2018). Briefly, a variable volume of sperm according to its concentration, with a

152 maximum of 1 billion cells and not exceeding 11 mL were layered over 3 mL of BoviPure™

153 diluted to a final ratio of 60% (v/v) with BoviDilute™ (Nidacon) and centrifuged at $300 \times g$ for

154 20 min at 20°C with slow acceleration and deceleration rates. The resulting cell pellet was

155 washed with PBS and resuspended in 1 mL of RNase-free PBS and further pelleted by

156 centrifugation. The pellets were stored at -80ºC until further use.

157 *Micrococcal nuclease digestion, library construction and sequencing*

158 Chromatin digestion was performed as previously described (Zalenskaya, Bradbury & Zalensky,

159 2000) with minor modifications. Forty million purified spermatozoa cells were pelleted by

160 centrifugation (3,500 rpm for 10 min) and washed twice with 1ml of 1x phosphate buffered

161 saline (PBS), and 1mM phenylmethylsulfonyl fluoride (PMSF). Pelleted cells were resuspended

162   with 1ml of 0.1% Lysolecithin (dissolved in 1x PBS and 1mM PMSF), incubated on ice for 10

163   min and centrifuged (3,500 rpm for 10 min). The sperm head was decondensed in 1ml of 1xPBS,

164   1mM PMSF with 10mM Dithiothreitol (DTT) and incubated on ice for 12 min. After

165   centrifugation (2,000 rpm for 10 min), nuclei were washed with 1ml of 1x PBS and 2mM DTT

166   and centrifuged again (3,500 rpm for 10 min). The pellet was resuspended in 100 μl of 1x PBS, 1

167   mM DTT and $CaCl_2$ was added to a final concentration of 0.6 mM. At this time, 5 U of MNase

168   (Sigma-Aldrich) were added and the samples were incubated for 7 min at 37°C. Digestion was

169   stopped by the addition of ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid

170   (EGTA) to a concentration of 5 mM. The resulting digested fraction corresponding to

171   nucleosome-bound chromatin (soluble) was separated from the undigested protamine-bound

172   fraction (pellet) by centrifugation at 20,000 rpm for 10 min. An aliquot of the digested DNA

173   fraction (which included both the MN and SN fractions) was treated with sodium dodecyl sulfate

174   (SDS) and proteinase K and evaluated in a 1.5% agarose 1x tris-acetate-EDTA (TAE) gel

175   electrophoresis. Another aliquot of the nuclease digested DNA was used to extract DNA with a

176   phenol-chloroform based protocol and was directly used for library prep after purification with

177   the Agencourt AMPure XP beads (Beckman Coulter). The MN and SN fractions were separated

178   after sequencing and read mapping using bioinformatics tools as described below. Purified DNA

179   was subjected to quality control including quantification with the Qubit™ DNA HS Assay kit

180   (Invitrogen) and Nanodrop (Thermo Scientific Fisher) as well as size and concentration

181   assessment with a 2100 Bioanalyzer and the Agilent DNA 1000 Kit (Agilent Technologies).

182   gDNA from the two sperm samples was also extracted as done by Hammoud and colleagues

183   (Hammoud et al., 2009). Briefly, sperm cells were thawed and pelleted at 10,000 x g for 1 min.

184   The supernatant was removed and the pellet was resuspended in 1 ml Lysis Reagent (1% Triton

185   X-100, 5 mM MgCl2, 320 mM Sucrose and 10 mM Tris pH 7.5). After centrifugation (20,000 x

186   g for 5 min), cells were resuspended again with 0.5ml Lysis Reagent two times, centrifuged and

187   the supernatant was discarded. The sperm pellet was dispersed with 305.5μl of Enzyme Master

188   Mix (1.25 mM MgCl$_2$, 1.25 mM Deferoxamine Mesylate, 12.5 mM Tris pH 8.0). Then, 1.6 μl

189   RNase A - 20 mg/ml -, 1.6 μl of 1 M DTT and 6.5 μl of 5M NaCl was added to each sample.

190   This cell suspension was kept at 37°C for 30 min. The cells were then dispersed with gentle

191   vertexing and subsequently, 40 μl of 10% SDS and 20μl of Proteinase K (20 mg/μl) were added.

192   Samples were then incubated at 50°C for 60 min. Following, 1ml 100% Isopropanol was added

193   and the tubes were inverted several times until a gel-like DNA precipitate appeared. Then, 0.6 ml

194   of NaI Solution was added (40 mM Tris pH 8.5, 20 mM EDTA and 7.6 M NaI), and the tubes

195   were inverted until DNA formed a white precipitate. The DNA pellet was washed 3 times with

196   70% EtOH with a final wash with 100% EtOH. The DNA pellet was solubilized with 300 μl of

197   H$_2$O. The two resulting gDNA samples were pooled to be used as input. Pooled gDNA were

198   sheared to obtain 100 bp long fragments using a Covaris S2 instrument (Covaris Inc), according

199   to the manufacturer's instructions. Sequencing libraries of the two MNase treated samples and

200   the input gDNAs were prepared with the PrepX™ DNA Library Kit (Takara). The libraries were

201   used as a template to generate 50 bp paired end (PE) reads in an Illumina HiSeq2500 system.

202   ***MNase-Seq data preprocessing and quality evaluation***

203   Raw sequencing data was filtered to remove low quality bases, adaptor sequences and short

204   reads (< 25 bp length) with Trimmomatic v.0.36 (Bolger, Lohse & Usadel, 2014). Trimmed

205   reads were aligned to the porcine genome (Sscrofa11.1) with Bowtie2 v.2.4.1 (Langmead &

206   Salzberg, 2012) fitting default parameters except "--very sensitive". Pearson correlation of the

207   read coverages for genomic regions along the genome was calculated using the

208   multiBamSummary file based on the two MNase samples with the tool deepTools v.3.3.2

209   (Ramírez et al., 2016) with options: "plotCorrelation --removeOutliers –skipZeros". Since the

210   correlation between the samples was very high, the reads from the two replicates (MN_1:

211   replicate A; MN_2: replicate B) were merged and processed as a pool with the aim to increase

212   sequencing depth and peak detection power and accuracy. Then, the MN and SN fractions were

213   bioinformatically separated based on the genomic distance between the two paired reads. To

214   extract the reads from the SN fraction, with a length slightly below 100 bp (see Fig. S1a), the

215   parameter "--maxins 110", which indicates that the mapped paired-end reads should not exceed

216   110 bp, was used. To obtain the reads from the MN fraction, with a length around 150 bp, the

217   parameters "--minins 111" and "--maxins 1000" were applied. Subsequently, duplicate reads of

218   these fractions and the input sample were removed with Picard-Tools MarkDuplicates v.1.56

219   (http://picard.sourceforge.net). Finally, for both fractions, peak calling was performed with

220   MACS2 v.2.1.0 (Zhang et al., 2008) with "-q 0.05 -B -g hs" and compared against the input

221   sample. Visualization of the genomic heatmaps from the MNase-Seq signals using transcription

222   start sites (TSS) as reference points was carried with the deepTools (Ramírez et al., 2016)

223   computeMatrix tool with parameters: "reference-point -b 500 -a 500 –skipZeros" and

224   plotHeatmap tool using "--refPointLabel 'TSS'".

225   ***Peak location relative to gene annotation***

226   Peaks were categorized by their position in relation to gene features as annotated in Ensembl

227   (v96) using BEDtools v.2.29.2 (Quinlan & Hall, 2010) intersect. The peaks were classified as

228   mapping to TSSs, promoters, overlapping 5' untranslated region (UTR), 3' UTR, exonic,

229   intronic and intergenic as carried in (Halstead et al., 2020).

230   In order to determine any potential positional preference or avoidance of the peaks within the

231   gene features, 1,000 iterative permutations of the peak genomic locations with the same number

232   of peaks following the same peak width distribution but with randomized genomic locations

233   were carried using BEDtools (Quinlan & Hall, 2010) shuffle. Subsequently, the randomized

234   peaks were assigned to genomic features according to the genomic position within these features.

235   These results were then contrasted against our results on real data with a permutation test that

236   calculates the proportion of permutations where the number of peaks that map at a given feature

237   deviates from the mean of the distribution more than the number of overlaps observed in the real

238   data, or in other words, are located more towards the tails of the distribution than the real data.

239   The fold change (FC) was also measured. FC was calculated by dividing the number of overlaps

240   between the real MN or SN peaks and each of the genomic features by the equivalent mean of

241   the distribution of the number of overlaps obtained with the shuffled data. These comparisons

242   that yielded $P < 0.001$ and FC $\geq$ 1.5 or FC $\leq$ 0.66, were considered enriched or depleted,

243   respectively.

244   ***Gene Ontology analysis of the genes near MNase peaks***

245   The genes mapping within or less than 5 kbp apart from the identified peaks were extracted from

246   Ensembl v96 with BEDtools (Quinlan & Hall, 2010) closestBed and were used for Gene

247   Ontology (GO) enrichment analysis. GO was carried out with Cytoscape v.3.8.2 plugin ClueGO

248   v.2.5.7 (Bindea et al., 2009) using the Cytoscape's porcine dataset and default settings. Only the

249   significant Bonferroni corrected *p*-values were considered.

250   ***Motif enrichment analysis at the MN and SN underlying sequences***

251   Motif enrichment analysis was carried out with the software HOMER v.4.10.0 (Heinz et al.,

252   2010) findMotifsGenome.pl function with default parameters.These default parameters include:

253   (i) random selection of pick the background genome to compare against the input sequence data

254 and using the background frequency of nucleotides in the human genome, hg19; (ii) a default

255 length of the motifs to search for is of 8-12 nucleotides long; (iii) motif search on both the

256 forward and reverse strands of the input sequences.

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

258 A comparable human sperm MNase-Seq dataset that sequenced the mono-nucleosomal fraction

259 (GSE15690) from Hammoud et al. (Hammoud et al., 2009), obtained from a pool of four donors

260 was downloaded from the Gene Expression Omnibus database. The genomic coordinates from

261 the human sperm MN peaks were liftover to Sscrofa11.1 using the UCSC liftover tool (Kuhn,

262 Haussler & Kent, 2013) with default parameters except for "-minMatch=0.1". For the

263 enrichment evaluation, the genomic location of the porcine MNase peaks were randomized 1,000

264 times using BEDtools (Quinlan & Hall, 2010) shuffle and overlapped to the human, or cattle

265 MNase peaks using BEDtools v.2.29.2 (Quinlan & Hall, 2010) intersect. Thereafter, the results

266 were contrasted to the overlap of the real data with the permutation test that calculates the

267 proportion of permutations where the number of overlaped peaks lie more toward the tails of the

268 distribution than the value observed in the real data. FC was also calculated by dividing the

269 number of overlaps between the real MN or SN peaks and human MNase peaks by the mean of

270 the distribution of the number of overlaps obtained with the shuffled data. To consider co-

271 location enrichment or depletion a threshold of *P < 0.001* and FC ≥ 1.5 or FC ≤ 0.66, was set to

272 determine enrichment and depletion, respectively.

273 The same approach was used for two cattle (GSE47843) comparable MNase-Seq datasets that

274 sequenced the mono-nucleosomal fractions from two bulls (samples s_1_1_bovine and

275 s_5_1_bovine) from the Holstein breed (Samans et al., 2014). Genome coordinates from bull

276 (busTau7) genomes were first liftover to hg19, and then from hg19 to Sscr11.

277 ***Integration with other -omics data***

278 Although mature sperm is assumed to be transcriptionally inactive, RNAs in sperm may both

279 reflect preceding events in transcriptionally active spermatogenic precursor cells during male

280 germ cell development and have a role in early development after fertilization (Gòdia, Swanson

281 & Krawetz, 2018; Ostermeier et al., 2004). Thus, the existence of a relationship between the

282 location of retained nucleosomes in sperm and the sperm transcriptome profile was

283 hypothesized. To test this hypothesis, sperm transcriptome data from 40 healthy Pietrain boars

284 belonging to commercial artificial insemination farms, including 40 total and 35 small RNA-seq

285 datasets previously generated by our group (NCBI's BioProject PRJNA520978) was used. These

286 datasets provided a list of 4,120 protein coding RNAs (Gòdia et al., 2020b), 1,574 circular RNAs

287 (circRNAs) mapping in pig chromosomes (Gòdia et al., 2020a) and 6,729 PIWI-interacting

288 RNAs (piRNAs) (Ablondi et al., 2021). The averaged RNA abundances from all the samples

289 were used for further analysis. Protein coding genes were classified according to their RNA

290 abundances measured in Fragments per Kilobase per Million mapped reads (FPKM) as (i) absent

291 (< 1 FPKM); (ii) low abundance (≥ 1 to <10 FPKM); (ii) intermediate abundance (≥ 10 to < 100

292 FPKM) and (iii) high abundance (≥ 100 FPKM). Then, the existence of positional co-location

293 enrichment between each of these gene fractions and the MNase peaks was assessed using the

294 Fisher Exact Test (two-tailed). Any distance below 5 kb between an RNA and a MN or SN peak

295 was considered as co-location. Moreover, the RNA abundance of the genes that co-located with

296 MN and SN peaks with the RNA levels of the whole set of genes annotated in the pig genome

297 was compared by employing the Kruskal-Wallis test using RNA abundances stabilized with the

298 log2 transformation.

299   The positional enrichment of circRNAs or piRNAs at MNase peaks was also studied. First, the

300   genomic regions of the MN and SN peaks were iteratively randomized 1,000 times using

301   BEDtools (Quinlan & Hall, 2010) shuffle. The overlap in the real data and the permutations with

302   the circRNAs or piRNAs was determined with BEDtools v.2.29.2 (Quinlan & Hall, 2010)

303   intersect. Then, the enrichment of the genomic coordinates of the real MNase peaks at piRNA

304   and circRNAs or piRNA sites was assessed by comparison with the overlap of the randomized

305   with the MNase peak locations using the permutation test. The FC was also calculated by

306   dividing the number of overlaps between the real MN or SN peaks and circRNAs or piRNAs by

307   the mean of the distribution of the number of overlaps obtained with the shuffled data. A

308   threshold of $P < 0.001$ and FC $\geq$ 1.5 (enrichment) or FC $\leq$ 0.66 (depletion), respectively, was set

309   to consider co-location enrichment or depletion.

310   MNase profiles were also evaluated against the genomic regions showing genetic association

311   with sperm quality traits in Pietrain pigs in a Genome-Wide Association Study (GWAS) carried

312   by our group and which included the two samples used in this study (Gòdia et al., 2020b). This

313   GWAS provided 18 regions showing genetic association with the percentage of sperm cells with

314   head abnormalities (7 regions), neck abnormalities (4 regions), percentage of abnormal

315   acrosomes after 5 minutes incubation at 37°C (2 regions), the ratio of abnormal acrosomes after

316   5 minutes and 90 minutes of incubation (1 region), percentage of motile spermatozoa after 5

317   minutes incubation (2 regions) and after 90 minutes incubation (2 regions, one shared with

318   motility after 5 minutes incubation) and the percentage of sperm cells with proximal droplets (1

319   region). These regions spanned in total, 16.5 Mbp (0.7%) of the porcine autosomal genome

320   (Gòdia et al., 2020b). Again, to evaluate whether MNase peaks were enriched at GWAS hits, the

321   genomic regions of the MN and SN peaks were iteratively randomized 1,000 times using

322   BEDtools (Quinlan & Hall, 2010) shuffle and the overlap of both, the real observed data and the

323   1,000 randomized files with the GWAS hits was evaluated with BEDtools v.2.29.2 (Quinlan &

324   Hall, 2010) intersect. This randomized overlap was compared to the overlap of the real data

325   using the permutation test. The FC was also calculated by dividing the number of overlaps

326   between the real MN or SN peaks and the GWAS hits by the mean of the distribution of the

327   number of overlaps obtained with the shuffled data. A threshold of $P < 0.001$ and $FC = 1.5$

328   (enrichment) or $FC = 0.66$ (depletion), respectively, was set to consider co-location enrichment

329   or depletion.

330   *Peak location relative to repeat elements*

331   The list of repeat elements in the pig sperm from RepeatMasker was downloaded from the

332   UCSC genome browser. The same approach as described for the other genomic features was

333   followed to evaluate enrichment or depletion of MNase peaks at the whole set of repeat

334   elements, long interspersed nuclear elements (LINEs) or short interspersed nuclear elements

335   (SINEs). The genomic locations of the MN and the SN peaks were randomized 1,000 times and

336   the overlap with the different sets of repeat elements was assessed with BEDtools v.2.29.2

337   (Quinlan & Hall, 2010) intersect. A permutation test was carried to evaluate enrichment ($P <$

338   0.001; $FC \geq 1.5$) or depletion ($P < 0.001$; $FC \leq 0.66$) of MN and SN peaks at repeat elements.

339   All scripts, including R, bash and software parameters used are available on Figshare:

340   https://doi.org/10.6084/m9.figshare.21997523.v1.

341

342   **Results**

343   *MNase digestion, library preparation and data preprocessing*

344   The MNase protocol yielded 95.2 and 75.6 ng of DNA for sample 1 and sample 2, respectively.

345   The extraction of gDNA provided 1227.6 ng/ul. The MNase treatment of the pig sperm samples

346   generated ~ 147 bp fragments corresponding to the MN fraction and an additional ~ 100 bp

347   fragment corresponding to the SN fraction (Fig. S1).

348   The sequencing of the libraries yielded more than 43 million PE reads for each MNase biological

349   replicate and 41.8 million PE reads for the input sample (Table 1). In average, 90.2% of the

350   MNase reads and 92.7% for the input reads mapped to the porcine genome (Sscrofa11.1). The

351   genome-wide profiles of MNase sensitivity of the two samples showed high correlation (Pearson

352   R = 0.87) (Fig. S2) and similar mapping statistics (Table 1). Thus, the sample pooling strategy

353   carried by Hammoud and collaborators (Hammoud et al., 2009) was replicated and the two

354   samples were pooled with the aim to increase read depth and peak call sensitivity and accuracy.

355   The final number of PE reads in each, the MN and SN fractions, was 62.1 million and 7.1

356   million, respectively (Table 1).

357   ***Characterization of the MN and SN peaks***

358   The analysis yielded 25,293 MN and 4,239 SN peaks (Tables S1 and S2). The MN peaks

359   averaged 270 bp in width and covered 0.3% of the porcine genome. The SN peaks were 141 bp

360   wide and covered 0.02% of the genome. Most MN (70.2%) and SN (94.5%) peaks were

361   annotated in intronic and intergenic regions (Fig. 1). MN peaks were significantly enriched at

362   promoters (*p*-value < 0.001, FC = 5.05), TSS (*p*-value < 0.001, FC = 11.26) (Fig. 2), 5'-UTR (*p*-

363   value < 0.001, FC = 12.20), coding sequences (*p*-value < 0.001, FC = 4.16) and 3'-UTR (*p*-

364   value < 0.001, FC = 1.56) (Table 2, Fig. S3). This result was not replicated in the SN peaks,

365   which showed mild depletion at 3'-UTR sites (p-value < 0.001, FC = 0.47) (Table 2 and Fig.

366   S3). A total of 9,128 and 1,688 genes overlapped or localized less than 5 kbp apart from a MN

367   and a SN peak, respectively (Table S3). The two MN and SN fractions co-existed in 1,145 genes

368   (Table S3), which corresponds to 75% of the SN-associated genes.

369   To get an idea of the potential function of the nucleosome associated DNA, GO analysis of the

370   genes overlapping or mapping less than 5 kb away from these peaks was carried out. For the MN

371   peaks, the most significantly enriched terms were related to chemical perception, development

372   (including the nervous system, multicellular organism, tissue, embryo, etc) and cell and organ

373   morphogenesis (Table S4). The genes located in the SN fractions were enriched for less GO

374   terms and showed weaker significances and included nervous system process, cell projection,

375   regulation of GTPase activity, organelle organization and Golgi vesicle transport (Table S5). To

376   delve further into the potential functions of the nucleosome retention in sperm, the genomic

377   sequences underlying the peak regions were subjected to sequence motif enrichment analysis.

378   Enrichment for motifs from 55 plant and 57 animal (mostly mammalian: mouse and human)

379   transcription factors (TFs) was identified in the MN peaks (Table S6). The animal motifs

380   involved 15 TFs from the bHLH class (MyoD, MyoG, Myf5, NeuroG2, Olig2, Tcf21, Twist2,

381   etc) and other TFs related to embryo development and implantation (Atf4, CHOP, Erra, EBF1,

382   E2F2, Foxh1, HOXA1, HOXA2, MAFK, Pax8, IRF1, IRF3, PR, Rfx1, RORg, RUNX2, Zic,

383   Zic3). They also included TFs related to spermatogenesis or sperm function which have also

384   been related to embryo development such as CUX1, CTCFL, PAX5 and Smad2. Finally, this list

385   also contained Znf263 and FOXA1, two TFs that provide a direct link between the paternal

386   gamete and the embryo (Table S6). One study identified 514 genes with paternal preferential

387   expression during human early embryo development and proposed *ZNF263* as the strongest

388   candidate in regulating the expression of these genes (Leng et al., 2019). The pig genes that map

389   less than 500 bp apart from a MN peak harboring a Znf263 motif were identified and their the

390   overlap with the 514 human genes showing paternal preferential expression in the early embryo

391   was searched. 3,264 genes mapped less than 500 bp away from a MN peak with Znf263 binding

392   motif in pig sperm. These corresponded to 3,288 human orthologs, 92 of which (18% of 514)

393   also showed preferential paternal expression during human early embryo development (Leng et

394   al., 2019) (Table S7). This concordance nearly reached significance, with a p-value = 0.052 as

395   calculated with the hypergeometric test. The SN peaks presented enriched motifs for 28 animal

396   and 46 plant TFs. Thirteen of the animal motifs enriched in the SN peaks were also enriched in

397   the MN fraction (Table S8). The common TFs included some of the bHLH class (NeuroG2,

398   Olig2 and Twist2), as well as TFs involved in embryo development and implantation (Atf4,

399   CHOP, ERRA, Foxh1, HOXA1, HOXA2, IRF1 and Zic). The SN-specific enriched motifs

400   corresponded to several TFs related to embryo development (NFX, Duxbl, PBX1, Pitx1, SF1,

401   NFIL3, Gfi1b, HLF), meiosis (DMC1) and parent-of-origin expression driven by imprinting in

402   embryo (Zfp57).

403   A comparable human sperm MNase-Seq dataset, which included pool data from four donors

404   (Hammoud et al., 2009) and focused on MN peaks, was employed to assess the inter-species

405   conservation of the genomic location of our MNase sites. Results in human were similar to these

406   detected in our analysis in pig. 25,122 peaks were annotated in the human MN fraction, 20,641

407   of which were successfully converted to the pig genome coordinates. Twenty-six percent (5,395

408   peaks) of the human nucleosome associated sites overlapped with our porcine MN peaks. This is

409   a highly significant co-location between the two species (p-value < 0.001, FC = 25.01),

410   compared to randomization (Table 2, Fig. S3). The two cattle samples contained 2,256

411   (s_1_1_bovine) and 8,446 (s_5_1_bovine) MN peaks. Of these, 1,011 (s_1_1_bovine) and 4,186

412   (s_5_1_bovine) were successfully liftover to pig and used to determine genomic overlap. Only 7

413   (s_1_1_bovine) and 16 (s_5_1_bovine) cattle MN peaks overlapped with the porcine MN peaks

414   and no significant enrichment or depletion was observed.

415   To identify positional relationships between the nucleosome-associated regions and

416   transcriptional activity, the locations of the MN and SN peaks were compared with the repertoire

417   of RNAs that our group identified in porcine sperm (Gòdia et al., 2020b). 12,125 protein coding

418   genes were absent in sperm. On the contrary, 5,814, 3,521 and 598 genes were classified as

419   being at low, moderate and high abundance in sperm, respectively. The location of the genes that

420   are present in spermatozoa was, regardless of their abundance, highly enriched (low abundance:

421   p-value = 1.8e-72, moderate abundance: p-value = 1.5e-91, high abundance: p-value = 1.1e-11)

422   at the MN peaks when compared to the catalog of absent genes (Table 3). Similarly, the SN

423   peaks were also enriched for the low (p-value = 5.0e-17), moderate (p-value = 3.6e-47) and

424   highly abundant (p-value = 3.0e-5) genes, when compared to the set of absent genes (Table 3). In

425   line with these results, the average RNA abundance of the genes mapping to both MN (p-value =

426   4.9e-19) and SN (p-value = 4.3e-22) peaks was significantly higher than the average abundance

427   of all the genes annotated in the pig genome (Fig. 3).

428   Recently, our group annotated 6,729 sperm piRNAs in porcine sperm (Ablondi et al., 2021). The

429   piRNA regions were significantly mapping more frequently within MN peaks (p-value < 0.001,

430   FC = 4.47) with an overlap of 65 MN peaks (Table 2, Fig. S3). This enriched co-location was

431   also observed (p-value < 0.001, FC = 4.70), overlapping to 18 peaks) for the set of piRNA

432   genomic regions involving 1,355 piRNAs (Table 2, Fig. S3) which abundance in sperm

433   correlated with different sperm phenotypes including the percentage of motile sperm, the

434   percentage of morphological abnormalities or the percentage of viable cells (Ablondi et al.,

435   2021). No obvious enrichment or depletion of SN at piRNAs sites was observed (Table 2 and

436   Fig. S3). No positive or negative co-location was observed between MN or SN peaks and the

437   catalog of 1,574 sperm circRNAs and 148 circRNAs associated to sperm motility in swine

438   (Gòdia et al., 2020a) (Table 2 and Fig. S3).

439   Our group recently published a GWAS for sperm quality traits in Pietrain boars (Gòdia et al.,

440   2020b) which was used to evaluate the co-location of MNase sites with regions associated to

441   these traits that could suggest a potential link between nucleosome retention and sperm quality.

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

443   Finally, the MN peaks showed depletion at LINEs (p-value < 0.001, FC = 0.48), while the SN

444   sites presented enrichment at SINEs (p-value < 0.001, FC = 1.85) (Table 2 and Fig. S3).

445

446   **Discussion**

447   This is a descriptive study that aimed at evaluating for the first time in swine, the tendency of the

448   sperm's retained nucleosomes to co-map with different types of functional elements of the

449   genome, which could be indicative of a non-random nucleosome retention. Echoing the results

450   observed in other species (Hammoud et al., 2009; Castillo et al., 2014; Samans et al., 2014;

451   Carone et al., 2014), the digestion resulted in two DNA bands (Fig. S1). The ~ 147 bp MN band

452   corresponding to mono-nucleosomes consists of of DNA wrapped around the histone octamer

453   comprising two copies of the core histones H2A, H2B, H3, and H4 (Luger et al., 1997). A SN

454   band has been also detected in yeast (Henikoff et al., 2011), *Xenopus laevis* sperm (Oikawa et

455   al., 2020), human sperm (Castillo et al., 2014), mouse sperm (Carone et al., 2014) and cattle

456   sperm (Samans et al., 2014). The nature, composition and function of these short particles is still

457   not well defined. Proteomic analysis of SN bands in *Xenopus* sperm revealed association with

458   chromatin regulatory proteins (H3, H4, H1FX, CBX3, WDR5) and reduced amounts of H2A and

459   H2B thereby leading to the hypothesis that SN are partially unwrapped nucleosomes that have

460   lost one or the two H2A and H2B dimers (Oikawa et al., 2020). Other articles describe that at

461   least in yeast and somatic cells, these partially unwrapped nucleosomes mark TFs to their

462   cognate binding sequence, thus representing a novel signature of active chromatin as reviewed

463   by Brahma and Henikoff (Brahma & Henikoff, 2020). Another line of research suggests that SN

464   particles in sperm corresponds to DNA sequences bound to CTCF (Carone et al., 2014), which is

465   an important regulator of chromatin structure and transcriptional activity that has also been

466   related to transgenerational inheritance from sperm to the offspring (Jung et al., 2022). In our

467   study, there was no enrichment of CTCF binding sites in the SN peaks, even though enrichment

468   of CTCFL, a CTCF paralog that is specific to the male germline, was observed in the MN

469   fraction. Enrichment of putative binding sites for some TF was found, nearly half of which were

470   also enriched in the MN peaks.

471   This, together with the opposite behavior observed between MN and SN regarding their location

472   in relation to 3'UTRs and repeat elements (MN are depleted at LINES and SN are enriched at

473   SINEs) could be indicating that the SN fraction identified in our study may correspond to

474   partially unwrapped nucleosomes generated during spermiogenesis or by the MNase digestion

475   protocol.

476   The extraction of nucleosomal DNA was less efficient than in experiments carried in sperm from

477   other species (Hammoud et al., 2009; Brykczynska et al., 2010; Castillo et al., 2014; Samans et

478   al., 2014; Erkek et al., 2013), and it did not yield sufficient amount of DNA from the

479   electrophoretic bands for high throughput sequencing. Nevertheless, the total amount of digested

480   DNA was enough for sequencing and fragment size separation after read mapping using

481   bioinformatics tools. To directly sequence the MN and SN fractions, future experiments will

482   require to optimize the efficiency of the protocol or, less ideal, to process a larger number of

483   sperm cells to isolate sufficient DNA from each MNase electrophoretic band.

484   The nucleosome-associated DNA spanned ~ 0.3% of the porcine sperm's genome, which

485   indicates a nearly fifteen, three and 45 -fold decrease in nucleosome retention when compared to

486   what has been described in human (Hammoud et al., 2009; Arpanahi et al., 2009; Gatewood et

487   al., 1987; Brykczynska et al., 2010; Castillo et al., 2014), mouse (Balhorn, Gledhill & Wyrobek,

488   1977; Erkek et al., 2013; Johnson et al., 2016), and cattle (Samans et al., 2014), respectively.

489   This difference could have technical causes as previously described by Carone and co-authors

490   regarding the existence of fragile and stable nucleosomes differing in their susceptibility to

491   MNase digestion and the fact that the extent of MNase digestion can impact on the number of

492   nucleosomes remaining in the sample (Carone et al., 2014). However, the existence of

493   interspecies variability cannot be ruled out, which could be partially driven by differences in the

494   protamine amino acid content resulting in changes on the extent or arrangement of the protamine

495   disulfide bonding (Perreault et al., 1988).

496   An increasing body of evidence in other animal species points towards a programmatic

497   nucleosome retention in the sperm genome (Hammoud et al., 2009; Castillo et al., 2014). In our

498   study, the largest abundance of MN and SN peaks in intronic and intergenic regions (Fig. 1), is

499   not unexpected since these gene features cover the vast majority of the genomic space. As a

500   matter of fact, MN enrichment at promoters and exons (Table 2, Fig. S3), provides the first

501   indication that nucleosome retention in sperm may not be stochastic and may relate to the

502   genome's functional activity. This hypothesis is further supported by the results obtained by

503   comparing the genomic location of the MN peaks in porcine sperm with that in human; the gene

504   annotation of the pig genome; the prediction of TF binding motifs and our own data on the pig

505   sperm transcriptome.

506   The concordance with previous findings in other species regarding the predominant location of

507   nucleosome peaks in gene promoters (Hammoud et al., 2009) and gene bodies (Yamaguchi et al.,

508   2018) as well as the overlap with their human synthenic regions (Hammoud et al., 2009),

509   indicates a degree of inter-species conservation in gene regulation, normally associated to the

510   maintenance of important functions. Despite the significant overlap with the human MN map,

511   still 73% of the human MN did not overlap to any of the porcine counterparts. This lack of full

512   overlap could be attributed to both technical and biological reasons. As previously mentioned,

513   the extend of MNase digestion could lead to differences in the MN profiles obtained by MNase-

514   Seq (Carone et al., 2014). Furthermore, cannot even rule out the possibility that the porcine and

515   human sperm have different sensitivities to the MNase digestion. The biological factors that

516   could lead to the inter-species differences in the MN retention include the existence of a certain

517   degree of random retention of nucleosomes due to their inefficient removal during

518   spermiogenesis, differences in the molecular mechanisms underlying spermatogenesis (Murat et

519   al., 2023) or embryogenesis (Lu et al., 2021), or genetic redundancy (Nowak et al., 1997).

520   Likewise, the lack of overlap with the cattle dataset might be attributed to differences in the

521   protocol but also to the small number of MN peaks detected in the two cattle samples and the

522   poor performance of the liftover from cattle to pig. The poor success of the liftover might have

523   been caused by the double-step liftover from cattle to human and from human to pig which was

524   needed as not direct liftover between cattle and pig in the UCSC liftover tool.

525   A larger number of strongly enriched GO terms for the catalog of genes mapping nearby the MN

526   when compared to the SN fractions was observed (Tables S4 and S5). This could be in part due

527   to both, the low number of observed SN peaks when compared to MN sites, and the fact that the

528   percentage of intergenic SN locations (57.5%) doubled that of intergenic MN peaks (28.6%).

529   Hence, the GO analysis run for SN included a lower number of genes, and consequently, yielded

530   a small number of enriched GO terms. The most enriched terms for the MN fraction included

531   sensory perception, which is important for sperm homeostasis (e.g., *SOD2*; (Malivindi et al.,

532   2018)), capacitation (e.g., *TRPV1*; (Osycka-Salut et al., 2020)) and chemotaxis guidance to the

533   egg for fertilization (e.g., *CA6*; (Boué et al., 1995)) and cell differentiation, development and

534   morphogenesis, all with obvious links to embryogenesis (Table S4).

535   The search for TF motifs at MN peaks also indicated that MN peaks may carry instructions for

536   embryo development. First, enriched motifs for several TFs related to embryo development were

537   identified (Table S6). Some of these genes point toward a role of the paternal gamete in

538   regulating embryo development. This is the case for *ATF4* (Puscheck et al., 2015) and *CHOP*

539   (Ali et al., 2018), which are relevant in the stress response of the mouse embryo at key

540   embryogenesis steps such as zygotic genome activation. FOXA1 is bound, in the chromatin of

541   human sperm, to genomic regions corresponding to enhancer elements in embryonic stem cells

542   and in several cell types of the embryo, which could suggest that the position of this TF in sperm

543   may guide the location of relevant enhancers for embryo development (Jung et al., 2019). The

544   identification of enriched Znf263 motifs is another relevant finding. *ZNF263* has been proposed

545   as a master regulator of the genes showing paternal preferential expression in the early

546   developing human embryo (Leng et al., 2019). The nearly significant enrichment (p-*value* =

547   0.052), of genes with paternal preferential expression in the human early embryo and the pig

548   orthologs mapping near a MN peak with predicted ZNF263 binding motif in sperm (Table S7)

549   may be indicating an epigenetic mark in sperm triggering gene expression of specific genes upon

550   fertilization. All in all, this data is in line with the hypothesis that nucleosome retention in sperm

551   has guiding relevance in the early embryo development.

552   Although mature sperm is assumed to be transcriptionally silent, it carries a wide repertoire of

553   RNAs related to spermatogenesis, fertilization, embryo development and offspring phenotype

554   [reviewed in: (Gòdia, Swanson & Krawetz, 2018)], a large proportion of which were transcribed

555   in transcriptionally active spermatogenic cells during previous stages of spermatogenesis. Our

556   group has generated sperm RNA-Seq (Gòdia et al., 2020a; Ablondi et al., 2021; Gòdia et al.,

557   2020b) and GWAS data for semen quality traits (Gòdia et al., 2020b) from porcine samples. The

558   RNA-Seq and GWAS studies included 40 and 276 samples, respectively. Thirty-five of the 40

559   pigs that provided sperm for RNA-Seq also contributed data for the GWAS. The 2 boars used in

560   the MNase study, also Pietrain, participated in the GWAS but not in the RNA-Seq. The

561   regulation of transcription is modulated by nucleosome occupancy and the accessibility of the

562   genome to the transcriptional machinery (Henikoff, Furuyama & Ahmad, 2004). Our results

563   show that the genes mapping within or nearby retained nucleosomes tend to display larger RNA

564   abundance than the full set of genes annotated in the porcine genome (Fig. 3). These results also

565   indicate that the genes that contributed detectable RNA levels by RNA-Seq in sperm, tend to

566   map to nucleosome-retained loci in both the MN and SN fractions. Not only protein coding but

567   also a family of regulatory RNAs, the piRNAs, were also significantly enriched at MN and SN

568   peaks. Again, this data supports the notion that nucleosome occupancy is key in modulating gene

569   expression. Moreover, the enriched co-location between the MN and SN sites and the catalog of

570   piRNA regions (Ablondi et al., 2021) which abundance correlated with sperm quality phenotypes

571   further suggests that this regulation during spermatogenesis may have phenotypic consequences

572   on semen quality and reproduction. Thus, the nucleosomes and sub-nucleosomes associated to

573   these RNAs may be leftovers from spermatogenesis and might provide useful information for the

574   identification of markers of abnormal spermatogenesis and sperm quality.

575   In light of these results, and considering that nucleosome positioning can modulate gene

576   expression (Bai & Morozov, 2010; Lorch, LaPointe & Kornberg, 1987; Li, Carey & Workman,

577   2007), it was hypothesized that sperm-retained nucleosomes could encompass DNA variants

578   altering elements such as promoters or enhancers regulating the expression of genes that played a

579   role during spermatogenesis. In such case, some degree of co-location between MNase peaks and

580   GWAS hits should be observed. However, this trend between the MN and SN peaks with the

581   GWAS hits for semen quality was not detected (Table 2 and Fig. S3) (Gòdia et al., 2020b).

582   The reasons, whether biological or technical, underlying the lack of a co-location trend with

583   GWAS hits for semen quality and the catalog of sperm circRNAs cannot be determined. For the

584   GWAS, one possibility is that because the GWAS hits are in fact, highlighting genomic regions

585   in linkage disequilibrium with the causal functional variant, the GWAS region is expected to be

586   much larger than this causal variant and probably most of the GWAS regions do not correspond

587   to functional elements of the genome (e.g, promoters, enhancers or genes) that could be

588   overlapping with nucleosomes. Regarding the circRNAs, although the FC did not reach our

589   threshold, none of the random permutations reached the same number of overlaps between MN

590   peaks and the catalog of circRNAs indicating that, although modest, there might be a trend of co-

591   location between circRNAs and MN peaks (Table 2).

592   Noteworthy, our results are not in disagreement with the hypothesis that sperm nucleosomes

593   mostly map at gene-poor regions and repeat elements (Carone et al., 2014; Samans et al., 2014;

594   Oikawa et al., 2020). The depletion of MN and enrichment of SN at LINEs and SINEs,

595   respectively, might be indicating that nucleosomes in these repeat elements are more susceptible

596   to MNase digestion as suggested by Carone and colleagues (Carone et al., 2014).

597   Four technical considerations in relation to the results of this study need to be taken into

598   consideration. The first consideration is that the presence of somatic cells, which contain large

599   number of histones in their chromatin, in the sperm samples could seriously mask the

600   nucleosome profiles of the sperm cells. Confidence can be placed on the absence of somatic cells

601   in our samples for two reasons. First, commercial pig sperm doses rarely present somatic cell

602   contamination. Second, the sperm samples were subjected to Bovipure^{TM} purification which

603   removes somatic cells. Moreover, the number of MN peaks identified in our study is close to the

604   number of MN peaks found by Hammoud and colleagues in human (Hammoud et al., 2009),

605   thereby indicating that the MNase profiles obtained in our study are not influenced by somatic

606   cells. The second consideration is the fact that our chromatin extraction protocol yielded lower

607   amount of DNA than what has been obtained in other species described in different studies,

608   which could be indicating that our conditions were not fully optimized for pigs. The third

609   consideration is that ChIP-Seq, either for H3 or for different histone modifications related to

610   specific genome activity would have been more precise and informative than MNase-Seq.

611   However, the amount of DNA obtained after the MNase digestion was limited and the

612   immunoprecipitation would most likely have resulted in insufficient amounts of DNA for high

613   throughput sequencing. Nevertheless, the MNase protocol has provided the required information

614   to suggest that nucleosome retention in sperm may contain relevant information in relation to

615   spermatogenesis, semen quality and embryo development. The fourth consideration relates to the

616   fact that the results are based on a single pool of two samples and consequently, the robustness of

617   the location of each particular nucleosomal site cannot be assessed. Nonetheless, the aim of this

618   work was not to robustly map the specific location of nucleosomes but to evaluate whether the

619   retained histones tend to co-map within functional features of the genome, which would indicate

620   that nucleosome retention is not random but programmatic. The statistical confidence of our

621   study was assessed within a sample, by testing the enrichment of positional overlap between the

622   identified nucleosomal sites and the different functional features of the genome against the null

623   basis of random location. Remarkably, all but two of the comparisons showed statistically

624   significant co-location thereby indicating, that nucleosome retention in pig sperm is not

625   stochastic and may owe to a reminiscent gametogenesis program and a future embryogenesis

626   program. Still, since only one pool of two samples was evaluated, the results should be

627   considered as preliminary and future research will be required to robustly establish the functional

628   relevance of nucleosome retention in the sperm genome in swine.

629

630   **Conclusions**

631   The results of our analyses show than the retained nucleosomes in the porcine sperm chromatin

632   tend to overlap with different functional features of the genome such as promoters, exons, genes

633   expressed in sperm and associated to semen quality, putative TFBS for TFs related to

634   reproduction and embryogenesis. In conclusion, the findings suggest that nucleosome retention

635   in mature spermatozoa is related to transcriptional regulation during spermatogenesis and is also

636   an instructional contributor to early embryo development. Hence, interrogating the nucleosome

637   occupancy in the sperm chromatin could help elucidating the biological basis of sperm quality

638   and early embryo development and perhaps, could even assist in the search for biomarkers for

639   these sets of traits.

640

641 **Acknowledgements**

642 We would like to thank Sue Hammoud and all the members of Hammoud lab for their support.

643 We also would like to thank Sam Balasch (Grup Gepork S.A., Catalonia) for supplying the

644 sperm samples.

645

646 **References**

647 Ablondi M, Gòdia M, Rodriguez-Gil JE, Sánchez A, Clop A. 2021. Characterisation of sperm

648   piRNAs and their correlation with semen quality traits in swine. *Animal Genetics* 52:114–

649   120. DOI: 10.1111/age.13022.

650 Ali I, Liu HX, Zhong-Shu L, Dong-Xue M, Xu L, Shah SZA, Ullah O, Nan-Zhu F. 2018.

651   Reduced glutathione alleviates tunicamycin-induced endoplasmic reticulum stress in

652   mouse preimplantation embryos. *Journal of Reproduction and Development* 64:15–24.

653   DOI: 10.1262/jrd.2017-055.

654 Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K,

655   Tedder P, Miller D. 2009. Endonuclease-sensitive regions of human spermatozoal

656   chromatin are highly enriched in promoter and CTCF binding sequences. *Genome*

657   *Research* 19:1338–1349. DOI: 10.1101/gr.094953.109.

658 Bai L, Morozov A. 2010. Gene regulation by nucleosome positioning. *Trends in Genetics*

659   26:476–483. DOI: 10.1016/j.tig.2010.08.003.

660 Balhorn R, Gledhill BL, Wyrobek AJ. 1977. Mouse sperm chromatin proteins: quantitative

661   isolation and partial characterization. *Biochemistry* 16:4074–4080. DOI:

662   10.1021/bi00637a021.

663   Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès

664   F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally

665   grouped gene ontology and pathway annotation networks. *Bioinformatics* 25:1091–1093.

666   DOI: 10.1093/bioinformatics/btp101.

667   Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence

668   data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

669   Boué F, Duquenne C, Lassalle B, Lefèvre A, Finaz C. 1995. FLB1, a Human Protein of

670   Epididymal Origin that is Involved in the Sperm-Oocyte Recognition Process. *Biology of*

671   *Reproduction* 52:267–278. DOI: 10.1095/biolreprod52.2.267.

672   Brahma S, Henikoff S. 2020. Epigenome Regulation by Dynamic Nucleosome Unwrapping.

673   *Trends in Biochemical Sciences* 45:13–26. DOI: 10.1016/j.tibs.2019.09.003.

674   Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schubeler D,

675   Stadler MB, Peters AH. 2010. Repressive and active histone methylation mark distinct

676   promoters in human and mouse spermatozoa. *Nature Structural & Molecular Biology*

677   17:679–687. DOI: 10.1038/nsmb.1821.

678   Carone BR, Hung JH, Hainer SJ, Chou M te, Carone DM, Weng Z, Fazzio TG, Rando OJ. 2014.

679   High-resolution mapping of chromatin packaging in mouse embryonic stem cells and

680   sperm. *Developmental Cell* 30:11–22. DOI: 10.1016/j.devcel.2014.05.024.

681   Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballesca JL, Oliva R. 2014. Genomic

682   and proteomic dissection and characterization of the human sperm chromatin. *Molecular*

683   *Human Reproduction* 20:1041–1053. DOI: 10.1093/molehr/gau079.

684 Erkek S, Hisano M, Liang C-Y, Gill M, Murr R, Dieker J, Schübeler D, Vlag J van der, Stadler

685   MB, Peters AHFM. 2013. Molecular determinants of nucleosome retention at CpG-rich

686   sequences in mouse spermatozoa. *Nature Structural & Molecular Biology* 20:868–875.

687   DOI: 10.1038/nsmb.2599.

688 Gatewood JM, Cook GR, Balhorn R, Bradbury EM, Schmid CW. 1987. Sequence-Specific

689   Packaging of DNA in Human Sperm Chromatin. *Science* 236:962–964. DOI:

690   10.1126/science.3576213.

691 Gòdia M, Mayer FQ, Nafissi J, Castelló A, Rodríguez-Gil JE, Sánchez A, Clop A. 2018. A

692   technical assessment of the porcine ejaculated spermatozoa for a sperm-specific RNA-seq

693   analysis. *Systems Biology in Reproductive Medicine* 64:291–303. DOI:

694   10.1080/19396368.2018.1464610.

695 Gòdia M, Swanson G, Krawetz SA. 2018. A history of why fathers' RNA matters. *Biology of*

696   *Reproduction* 99:147–159. DOI: 10.1093/biolre/ioy007.

697 Gòdia M, Estill M, Castelló A, Balasch S, Rodríguez-Gil JE, Krawetz SA, Sánchez A, Clop A.

698   2019. A RNA-Seq Analysis to Describe the Boar Sperm Transcriptome and Its Seasonal

699   Changes. *Frontiers in Genetics* 10:299. DOI: 10.3389/fgene.2019.00299.

700 Gòdia M, Castelló A, Rocco M, Cabrera B, Rodríguez-Gil JE, Balasch S, Lewis C, Sánchez A,

701   Clop A. 2020a. Identification of circular RNAs in porcine sperm and evaluation of their

702   relation to sperm motility. *Scientific Reports* 10:1–11. DOI: 10.1038/s41598-020-64711-

703   z.

704 Gòdia M, Reverter A, González-Prendes R, Ramayo-Caldas Y, Castelló A, Rodríguez-Gil JE,

705   Sánchez A, Clop A. 2020b. A systems biology framework integrating GWAS and RNA-

706   seq to shed light on the molecular basis of sperm quality in swine. *Genetics Selection*

707   *Evolution* 52:1–21. DOI: 10.1186/s12711-020-00592-0.

708   Halstead MM, Kern C, Saelao P, Wang Y, Chanthavixay G, Medrano JF, van Eenennaam AL,

709   Korf I, Tuggle CK, Ernst CW, Zhou H, Ross PJ. 2020. A comparative analysis of

710   chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics* 21:698. DOI:

711   10.1186/s12864-020-07078-9.

712   Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. 2009. Distinctive chromatin

713   in human sperm packages genes for embryo development. *Nature* 460:473–478. DOI:

714   10.1038/nature08162.

715   van der Heijden GW, Derijck AAHA, Ramos L, Giele M, van der Vlag J, de Boer P. 2006.

716   Transmission of modified nucleosomes from the mouse male germline to the zygote and

717   subsequent remodeling of paternal chromatin. *Developmental Biology* 298:458–469.

718   DOI: 10.1016/j.ydbio.2006.06.051.

719   Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass

720   CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-

721   regulatory elements required for macrophage and B cell identities. *Molecular Cell*

722   38:576–589. DOI: 10.1016/j.molcel.2010.05.004.

723   Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome

724   characterization at single base-pair resolution. *Proceedings of the National Academy of*

725   *Sciences of the United States of America* 108:18318–18323. DOI:

726   10.1073/pnas.1110731108.

727   Henikoff S, Furuyama T, Ahmad K. 2004. Histone variants, nucleosome assembly and

728   epigenetic inheritance. *Trends in Genetics* 20:320–326. DOI: 10.1016/j.tig.2004.05.004.

729   Ihara M, Meyer-Ficca ML, Leu NA, Rao S, Li F, Gregory BD, Zalenskaya IA, Schultz RM,

730   Meyer RG. 2014. Paternal Poly (ADP-ribose) Metabolism Modulates Retention of

731   Inheritable Sperm Histones and Early Embryonic Gene Expression. *PLoS Genetics*

732   10:e1004317. DOI: 10.1371/journal.pgen.1004317.

733   Johnson GD, Jodar M, Pique-Regi R, Krawetz SA. 2016. Nuclease Footprints in Sperm Project

734   Past and Future Chromatin Regulatory Events. *Scientific Reports* 6:25864. DOI:

735   10.1038/srep25864.

736   Jung YH, Kremsky I, Gold HB, Rowley MJ, Punyawai K, Buonanotte A, Lyu X, Bixler BJ,

737   Chan AWS, Corces VG. 2019. Maintenance of CTCF- and Transcription Factor-

738   Mediated Interactions from the Gametes to the Early Mouse Embryo. *Molecular Cell*

739   75:154-171.e5. DOI: 10.1016/j.molcel.2019.04.014.

740   Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, Corces VG. 2017. Chromatin

741   States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell*

742   *Reports* 18:1366–1382. DOI: 10.1016/j.celrep.2017.01.034.

743   Jung YH, Wang HL v., Ruiz D, Bixler BJ, Linsenbaum H, Xiang JF, Forestier S, Shafik AM, Jin

744   P, Corces VG. 2022. Recruitment of CTCF to an Fto enhancer is responsible for

745   transgenerational inheritance of BPA-induced obesity. *Proceedings of the National*

746   *Academy of Sciences of the United States of America* 119:e2214988119. DOI:

747   10.1073/pnas.2214988119.

748   Khezri A, Narud B, Stenseth E-B, Johannisson A, Myromslien FD, Gaustad AH, Wilson RC,

749   Lyle R, Morrell JM, Kommisrud E, Ahmad R. 2019. DNA methylation patterns vary in

750   boar sperm cells with different levels of DNA fragmentation. *BMC Genomics* 20:897.

751   DOI: 10.1186/s12864-019-6307-8.

752   King GJ, Macpherson JW. 1973. A Comparison of Two Methods for Boar Semen Collection.

753   *Journal of Animal Science* 36:563–565. DOI: 10.2527/jas1973.363563x.

754   Krawetz SA. 2005. Paternal contribution: new insights and future challenges. *Nature Reviews*

755   *Genetics* 6:633–642. DOI: 10.1038/nrg1654.

756   Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools.

757   *Briefings in Bioinformatics* 14:144–161. DOI: 10.1093/bib/bbs038.

758   Lai WKM, Pugh BF. 2017. Understanding nucleosome dynamics and their links to gene

759   expression and DNA replication. *Nature Reviews Molecular Cell Biology* 18:548–562.

760   DOI: 10.1038/nrm.2017.47.

761   Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*

762   9:357–359. DOI: 10.1038/nmeth.1923.

763   Leng L, Sun J, Huang J, Gong F, Yang L, Zhang S, Yuan X, Fang F, Xu X, Luo Y, Bolund L,

764   Peters BA, Lu G, Jiang T, Xu F, Lin G. 2019. Single-Cell Transcriptome Analysis of

765   Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation

766   Development. *Cell Stem Cell* 25:697-712.e6. DOI: 10.1016/j.stem.2019.09.004.

767   Li B, Carey M, Workman JL. 2007. The Role of Chromatin during Transcription. *Cell* 128:707–

768   719. DOI: 10.1016/j.cell.2007.01.015.

769   Lismer A, Siklenka K, Lafleur C, Dumeaux V, Kimmins S. 2020. Sperm histone H3 lysine 4

770   trimethylation is altered in a genetic mouse model of transgenerational epigenetic

771   inheritance. *Nucleic Acids Research* 48:11380–11393. DOI: 10.1093/nar/gkaa712.

772   Lorch Y, LaPointe JW, Kornberg RD. 1987. Nucleosomes inhibit the initiation of transcription

773   but allow chain elongation with the displacement of histones. *Cell* 49:203–210. DOI:

774   10.1016/0092-8674(87)90561-7.

775   Lu X, Zhang Y, Wang L, Wang L, Wang H, Xu Q, Xiang Y, Chen C, Kong F, Xia W, Lin Z, Ma

776   S, Liu L, Wang X, Ni H, Li W, Guo Y, Xie W. 2021. Evolutionary epigenomic analyses

777   in mammalian early embryos reveal species-specific innovations and conserved

778   principles of imprinting. *Science Advances* 7:eabi6178. DOI: 10.1126/sciadv.abi6178.

779   Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the

780   nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260. DOI: 10.1038/38444.

781   Malivindi R, Rago V, de Rose D, Gervasi MC, Cione E, Russo G, Santoro M, Aquila S. 2018.

782   Influence of all-trans retinoic acid on sperm metabolism and oxidative stress: Its

783   involvement in the physiopathology of varicocele-associated male infertility. *Journal of*

784   *Cellular Physiology* 233:9526–9537. DOI: 10.1002/jcp.26872.

785   Mendonça GA, Morandi R, Souza ET, Gaggini TS, Silva-Mendonça MCA, Antunes RC, Beletti

786   ME. 2017. Isolation and identification of proteins from swine sperm chromatin and

787   nuclear matrix. *Animal Reproduction* 14:418–428. DOI: 10.21451/1984-3143-AR816.

788   Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt

789   J, Schneider C, Mößinger K, Brüning T, Lamanna F, Belles MR, Conrad C, Kondova I,

790   Bontrop R, Behr R, Khaitovich P, Pääbo S, Marques-Bonet T, Grützner F, Almstrup K,

791   Schierup MH, Kaessmann H. 2023. The molecular evolution of spermatogenesis across

792   mammals. *Nature* 613:308–316. DOI: 10.1038/s41586-022-05547-7.

793   Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J. 1997. Evolution of genetic redundancy.

794   *Nature* 388:167–171. DOI: 10.1038/40618.

795   Oikawa M, Simeone A, Hormanseder E, Teperek M, Gaggioli V, O'Doherty A, Falk E, Sporniak

796   M, D'Santos C, Franklin VNR, Kishore K, Bradshaw CR, Keane D, Freour T, David L,

797   Grzybowski AT, Ruthenburg AJ, Gurdon J, Jullien J. 2020. Epigenetic homogeneity in

798   histone methylation underlies sperm programming for embryonic transcription. *Nature*

799   *Communications* 11:3491. DOI: 10.1038/s41467-020-17238-w.

800   Osycka-Salut CE, Martínez-León E, Gervasi MG, Castellano L, Davio C, Chiarante N, Franchi

801   AM, Ribeiro ML, Díaz ES, Perez-Martinez S. 2020. Fibronectin induces capacitation-

802   associated events through the endocannabinoid system in bull sperm. *Theriogenology*

803   153:91–101. DOI: 10.1016/j.theriogenology.2020.04.031.

804   Perreault SD, Barbee RR, Elstein KH, Zucker RM, Keefer CL. 1988. Interspecies Differences in

805   the Stability of Mammalian Sperm Nuclei Assessed in Vivo by Sperm Microinjection and

806   in Vitro by Flow Cytometry. *Biology of Reproduction* 39:157–167. DOI:

807   10.1095/biolreprod39.1.157.

808   Puscheck EE, Awonuga AO, Yang Y, Jiang Z, Rappolee DA. 2015. Molecular Biology of the

809   Stress Response in the Early Embryo and its Stem Cells. *Advances in Experimental*

810   *Medicine and Biology* 843:77–128. DOI: 10.1007/978-1-4939-2480-6_4.

811   Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic

812   features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

813   Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F,

814   Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data

815   analysis. *Nucleic Acids Research* 44:W160–W165. DOI: 10.1093/nar/gkw257.

816   Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, Wolf E, Steger K,

817   Dansranjavin T, Schagdarsurengin U. 2014. Uniformity of Nucleosome Preservation

818   Pattern in Mammalian Sperm and Its Connection to Repetitive DNA Elements.

819   *Developmental Cell* 30:23–35. DOI: 10.1016/j.devcel.2014.05.023.

820   Spadafora C. 2017. Sperm-mediated transgenerational inheritance. *Frontiers in Microbiology*

821   8:2401. DOI: 10.3389/fmicb.2017.02401.

822   Ward WS. 2010. Function of sperm chromatin structural elements in fertilization and

823   development. *Molecular Human Reproduction* 16:30–36. DOI: 10.1093/molehr/gap080.

824   Ward WS, Coffey DS. 1991. DNA Packaging and Organization in Mammalian Spermatozoa:

825   Comparison with Somatic Cell. *Biology of Reproduction* 44:569–574. DOI:

826   10.1095/biolreprod44.4.569.

827   Yamaguchi K, Hada M, Fukuda Y, Inoue E, Makino Y, Katou Y, Shirahige K, Okada Y. 2018.

828   Re-evaluating the Localization of Sperm-Retained Histones Revealed the Modification-

829   Dependent Accumulation in Specific Genome Regions. *Cell Reports* 23:3920–3932.

830   DOI: 10.1016/j.celrep.2018.05.094.

831   Yoshida K, Muratani M, Araki H, Miura F, Suzuki T, Dohmae N, Katou Y, Shirahige K, Ito T,

832   Ishii S. 2018. Mapping of histone-binding sites in histone replacement-completed

833   spermatozoa. *Nature Communications* 9:3885. DOI: 10.1038/s41467-018-06243-9.

834 Zalenskaya IA, Bradbury EM, Zalensky AO. 2000. Chromatin structure of telomere domain in

835   human sperm. *Biochemical and Biophysical Research Communications* 279:213–218.

836   DOI: 10.1006/bbrc.2000.3917.

837 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM,

838   Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome*

839   *Biology* 9:R137. DOI: 10.1186/gb-2008-9-9-r137.

840

841 **Declarations**

842 **Ethics approval and consent to participate**

843 The ejaculates obtained from pigs were privately owned for non-research purposes. The owners

844 provided consent for the use of these samples for research. Specialized professionals at the farm

845 obtained all the ejaculates following standard routine monitoring procedures and relevant

846 guidelines. No animal experiment has been performed in the scope of this research.

847 **Consent for publication**

848 Not applicable.

849 **Availability of data and material**

850 The datasets generated and analysed during the current study are available at NCBI's SRA under

851 accession numbers SRR14117448, SRR14117447, SRR14117446. All scripts, including bash

852 and software parameters used are available on Figshare

853 (https://doi.org/10.6084/m9.figshare.21997523.v1).

854 **Competing interests**

855   The authors declare that they have no competing interests.

856   **Funding**

857   This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO)

858   under grant AGL2013-44978-R and grant AGL2017-86946-R and by the CERCA

859   Programme/Generalitat de Catalunya. AGL2017-86946-R was also funded by the Spanish State

860   Research Agency (AEI) and the European Regional Development Fund (ERDF). We thank the

861   Agency for Management of University and Research Grants (AGAUR) of the Generalitat de

862   Catalunya (Grant Numbers 2014 SGR 1528 and 2017 SGR 1060). We also acknowledge the

863   support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence

864   Severo Ochoa 2016–2019 (Grant Number SEV-2015-0533) grant awarded to the Centre for

865   Research in Agricultural Genomics (CRAG). MG acknowledged a Ph.D. studentship from

866   MINECO (Grant Number BES-2014-070560) and a Short-Stay fellowship from MINECO

867   (EEBB-I-16-11528) at SSH lab.

868

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.

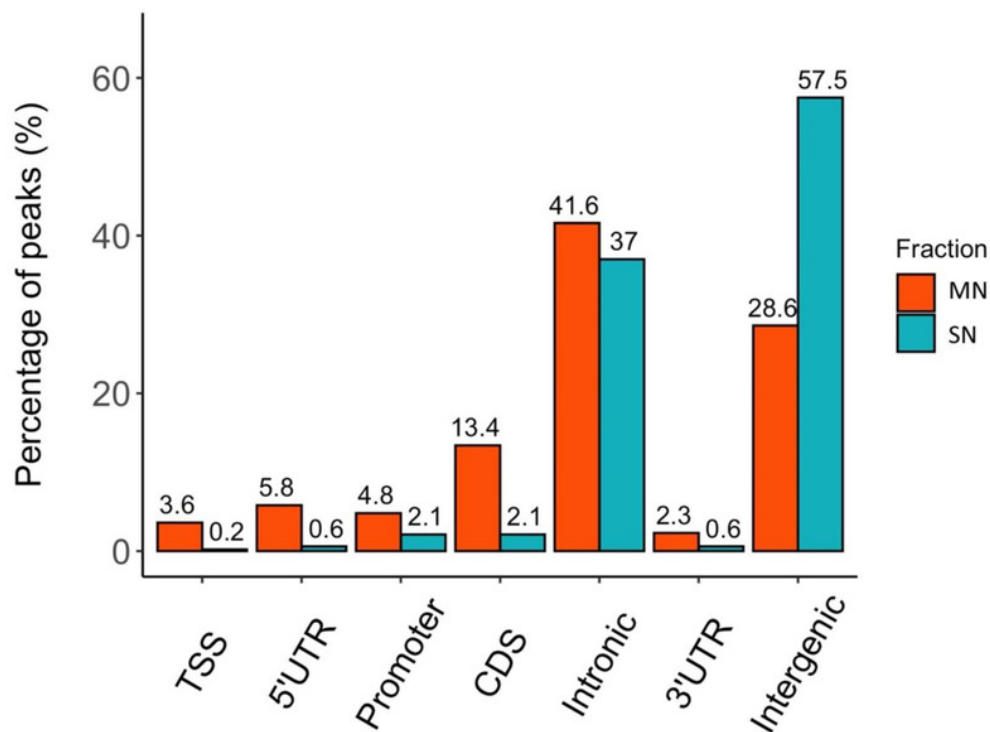



Figure 1. Distribution of the MN and SN peaks relative to gene features in the pig genome.

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.

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.

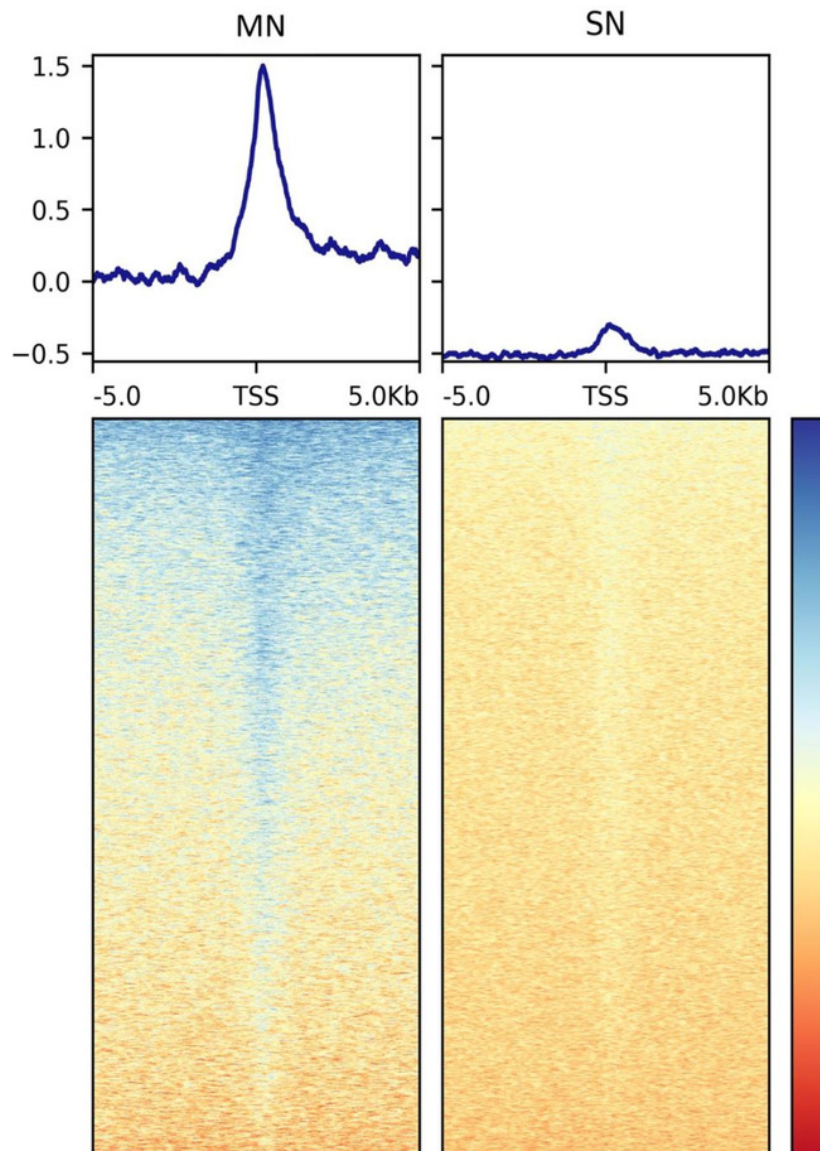



Figure 2. Genomic heatmaps depicting the normalized MNase-Seq signal centered at TSS for the MN and the SN peaks.

The x axis shows the genomic location relative to the TSS. The y axis indicates the MNase-Seq signal intensity.

Figure 3

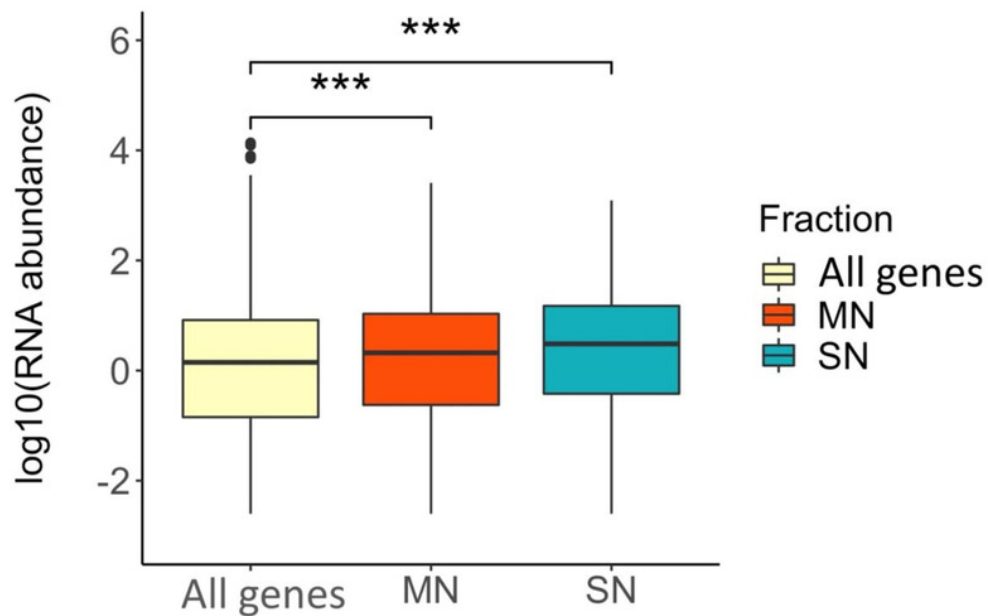
Figure 3. Box plots showing the average RNA abundance of (i) all genes present in the genome (yellow), (ii) the genes present in the MN fraction (red) and (iii) the genes present in the SN fraction (blue).

***: p-value ≤ 0.001.

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.

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

PE: paired End reads

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

2 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	

3 PE: Paired End reads

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

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 Godia 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 Godia et al. (2020b). RE: repeat elements in the pig genome. SINE: short interspersed nuclear elements in swine. LINE: long interspersed nuclear element in swine.

1   **Table 3.** Summary of the results of the permutation tests to evaluate the location of the MN and SN peaks in relation to different

2   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
		RE	12,666	16,080	16,597	16,341.85	74.27	< 0.001	0.78
	RepeatMasker	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

		Min	Max	Mean	SE	p	
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

4 Min, Max, Mean: minimum, maximum and mean number of overlaps for the 1,000 simulations, respectively. SE: standard error. TSS:
5 transcription start site. 5'UTR: 5' untranslated region. CDS: coding sequence. 3'UTR: 3' untranslated region. circRNA_corr: circular
6 RNAs which abundance level in sperm correlated to semen quality traits in swine as described by Godia et al. (2020a). piRNA_corr:
7 piRNAs which abundance level in sperm correlated to semen quality traits in swine as described by Ablondi et al. (2021). GWAS:
8 GWAS hits for semen quality traits in pigs as described by Godia et al. (2020b). RE: repeat elements in the pig genome. SINE: short
9 interspersed nuclear elements in swine. LINE: long interspersed nuclear element in swine.
10

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

1 **Table 2.** 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		
		Within MN peaks	**Not present**	*p*-value	**Within SN peaks**	**Not present**	*p*-value
Gene's RNA levels in sperm	Not present	4083	8042		644	11481	
	Low abundance	2774	3040	1.80E-72	504	5310	5.00E-17
	Intermediate abundance	1857	1664	1.50E-91	455	3066	3.60E-47
	High abundance	296	302	1.10E-11	60	538	3.00E-05