

H4K20me3 marks distal intergenic and repetitive regions in human mature spermatozoa

Nihan Ozturk^{1,2,*}, Temuujin Dansranjavin^{1,*}, Sabrina Gies^{1,2}, Damien Calay³, Shanjid Shiplu^{1,2}, Catherine Creppe³, Jan Hendrickx³ and Undraga Schagdarsurengin^{1,2,‡}

ABSTRACT

Sperm histones represent an essential part of the paternally transmitted epigenome, but uncertainty exists about the role of those remaining in non-coding and repetitive DNA. We therefore analyzed the genome-wide distribution of the heterochromatic marker H4K20me3 in human sperm and somatic (K562) cells. To specify the function of sperm histones, we compared all H4K20me3-containing and -free loci in the sperm genome. Sperm and somatic cells possessed a very similar H4K20me3 distribution: H4K20me3 peaks occurred mostly in distal intergenic regions and repetitive gene clusters (in particular genes encoding odorant-binding factors and zinc-finger antiviral proteins). In both cell types, H4K20me3 peaks were enriched in LINEs, ERVs, satellite DNA and low complexity repeats. In contrast, H4K20me3-free nucleosomes occurred more frequently in genic regions (in particular promoters, exons, 5'-UTR and 3'-UTR) and were enriched in genes encoding developmental factors (in particular transcription activators and repressors). H4K20me3-free nucleosomes were also detected in substantial quantities in distal intergenic regions and were enriched in SINEs. Thus, evidence suggests that paternally transmitted histones may have a dual purpose: maintenance and regulation of heterochromatin and guidance towards transcription of euchromatin.

KEY WORDS: Sperm histones, Paternal epigenetic inheritance, H4K20me3, Human

INTRODUCTION

The epigenomes of mammalian sperm include a small subset of histones that escape histone-to-protamine replacement during spermiogenesis and remain in the chromatin of mature spermatozoa in nucleosome structures. Sperm histones (SHs) are of interest in terms of development, paternal epigenetic inheritance and male infertility owing to their high potential to contribute to both pre- and post-implantation embryogenesis, and to trigger intergenerational and transgenerational effects. As SHs represent an essential part of paternally transmitted epigenetic information, their precise genome-wide localization in both coding and non-coding regions is pivotal to understanding their function comprehensively.

Depending on the method used, human sperm have been reported to carry 3-15% of the histone complement of a somatic cell (Gatewood et al., 1987; Bench et al., 1996; Hammoud et al., 2009; Samans et al., 2014), whereas in mice, the nucleosome content is estimated to be <2% (Balhorn et al., 1977; Erkek et al., 2013), and in bulls 13% (Samans et al., 2014). Many studies addressing the genome-wide localization of mammalian SHs have utilized chromatin immunoprecipitation sequencing (ChIP-seq) and focused on certain individual histones, mostly H3, and their modifications (e.g. H3K27me3, H3K9me3, H3K4me2, H3K4me3 and H4K12ac) (Hammoud et al., 2009; Arpanahi et al., 2009; Brykczynska et al., 2010; Erkek et al., 2013; Teperek et al., 2016; Yoshida et al., 2018; Yamaguchi et al., 2018). These studies demonstrated that SHs are enriched primarily at CpG-rich sites in embryonic development-related gene regions, including imprinting gene clusters and promoters of developmental transcription and signaling factors. In our previous genome-wide assay, we sequenced the entire mono-nucleosome DNA fraction from human and bovine spermatozoa without immunoprecipitation and found that, in both species, a considerable portion of nucleosomes was also preserved in distal intergenic regions and associated with repetitive DNA elements (RDEs), particularly centromere repeats and retrotransposons (Samans et al., 2014). Our study and another (Carone et al., 2014), which also observed histone retention over broad gene-poor domains, raised a lively ongoing debate about the localization and function of SHs. Whether histones retain over-repeats in human sperm is still debatable, as our previous results were doubted (Royo et al., 2016). The reservations mainly concerned repetitive sequences and their bioinformatics evaluations, as some RDE-originating reads can map to multiple genome locations and lead to ambiguous data.

Orthogonal experimental approaches have provided several lines of evidence that histone retention in sperm occurs to a large extent over repeat elements. In murine spermatozoa, nucleohistone chromatin was shown to be hypersensitive to nuclease cleavage, organized in nucleosomal domains and enriched in retroposon DNA (Pittoggi et al., 1999). Moreover, immunostaining and fluorescence *in situ* hybridization studies in mouse and human spermatozoa have revealed colocalization of SHs with the repeat-enriched chromocenter (van der Heijden et al., 2006; Govin et al., 2007; Meyer-Ficca et al., 2013; van de Werken et al., 2014). A ChIP-seq study in mouse sperm reported that the majority of SHs occur in gene deserts, and that H3K9me3 was enriched in satellite repeats and H4 in distal intergenic regions (Yamaguchi et al., 2018). A recent study in mouse using assay for transposase-accessible chromatin (ATAC)-seq showed that the accessible sites in sperm chromatin are mostly enriched for gene promoters and repeats, including long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Luense et al., 2019). Notably, nanoscale liquid chromatography coupled to tandem mass

¹Clinic of Urology, Pediatric Urology and Andrology, Justus-Liebig University Giessen, 35392 Giessen, Germany. ²Working Group Epigenetics of the Urogenital System, Clinic of Urology, Pediatric Urology and Andrology, Justus-Liebig University Giessen, 35392 Giessen, Germany. ³Epigenetics and Bioinformatics Services Team, Diagenode SA, 4102 Liège, Belgium.

*These authors contributed equally to this work

‡Author for correspondence (Undraga.Schagdarsurengin@chiru.med.uni-giessen.de)

 U.S., 0000-0003-0127-4938

Handling Editor: Haruhiko Koseki

Received 31 August 2020; Accepted 30 June 2021

spectrometry revealed in mouse and human spermatozoa 80–100 different post-translational histone modifications (PTHMs), and analysis of different spermatogenic stages in mice showed the greatest fold increase from elongating spermatids to mature sperm for the heterochromatic marker H4K20me3 (Luense et al., 2016). In somatic and embryonic stem cells, H4K20me3 is known to be a key player in the epigenetic regulation of genomic integrity and is enriched at pericentric heterochromatin (PCH), telomeres and other RDE-rich genome segments (Jørgensen et al., 2013; Nelson et al., 2016). However, the extent to which and where H4K20me3 is preserved in human sperm are unknown.

Here, we analyzed H4K20me3 in human testes and its genome-wide distribution in mature motile spermatozoa. To obtain a complete and coherent picture, we also analyzed H4K20me3-free nucleosomes. Moreover, to specify H4K20me3 as part of the paternally transmitted epigenetic information, we analyzed the genome-wide distribution of H4K20me3 in human somatic K562 cells. Using immunohistochemical analysis (IHC) of testicular sections and quantitative western blot analysis (qWB) of spermatozoa, we demonstrate a continuous presence of H4K20me3 in all human spermatogenic stages up to ejaculated spermatozoa. ChIP-seq analyses of motile sperm samples from fertile donors show a strict separation between nucleosomes containing H4K20me3 and those without it in terms of genome-wide localization and enrichment, and associated genes and repeats. Importantly, however, sperm and somatic cells exhibit a similar genome-wide distribution of H4K20me3.

RESULTS

Continuous presence of H4K20me3 in human sperm

We used IHC to analyze H4K20me3 in human testis sections exhibiting normal spermatogenesis and immunofluorescence (IF) and qWB to evaluate the marker in ejaculated human sperm. We detected H4K20me3 at almost every spermatogenic stage, with moderate staining in spermatogonia, increased staining in preleptotene and leptotene spermatocytes, and strongest staining in round spermatids (Fig. 1A). However, in pachytene spermatocytes (stages I–V), the staining was very weak and limited to a few regions, and in elongating and maturing (elongated) spermatids (stages I, II, IV and V), H4K20me3 was barely detectable (Fig. 1A). Previous studies have found that H4K20me3 staining is significantly decreased in pachytene spermatocytes, and mostly appeared on the XY body in the course of sex chromosome inactivation (Metzler-Guillemain et al., 2008). Late elongating and maturing spermatids already have intensely condensed chromatin, which often leads to antigen-masking. Therefore, we analyzed H4K20me3 in ejaculated spermatozoa (motile and immotile fractions) using IF and qWB (Fig. 1B–D). We obtained clear staining in most motile sperm for all analyzed semen samples, and ascertained that H4K20me3 is preserved in human sperm (Fig. 1C). Semen samples of 13 fertile donors were further subjected to qWB (Table 1, Fig. 1D). We detected H4K20me3 in both motile and immotile sperm fractions, with apparent differences between individuals and between fractions (Fig. 1D). The relative content of H4K20me3 ranged in motile fractions from 0.12 to 1.91, and immotile from 0.26 to 3.30. In ten out of 13 donors, a higher content of H4K20me3 (up to 7-fold) was measured in immotile fractions (Fig. 1D).

Isolation and verification of intact mono-nucleosome fractions and associated DNA

Before starting the H4K20me3-ChIP-seq experiment in human sperm, we performed several tests to establish the optimum

conditions (Fig. S1). At least 2×10^6 spermatozoa were needed to detect a clear mono-nucleosomal band after the micrococcal-nuclease (MNase)-based procedure, and processing of portions (each 5×10^6 spermatozoa) was more efficient than using 2×10^7 spermatozoa at once (Fig. S1A); X-linking reduced nucleosomal DNA preparation efficiency and overall DNA amount (Fig. S1B); mild (Hisano et al., 2013) and harsh (Samans et al., 2014) ChIP-washing buffers did not result in differing ChIP-DNA amounts (Fig. S1C); MNase concentration of 30 units and treatment time of 5 min were most suitable to achieve a thorough separation of nucleosome- and protamine-associated chromatin (Fig. S1D,F) as well as a high enrichment of nucleosomal DNA in control regions [*EVX1*: H3K4me3/H3K27me3-occupied in human sperm (Brykczynska et al., 2010); *KCNQ1* and *ZNF510*: H4K20me3-occupied in HeLa and K562 cells (ChIP-seq data, Diagenode)] (Fig. S1F). Moreover, the resulting nucleosomal chromatin was well suited for ChIP in nucleosome-enriched genes, e.g. *EVX1*, *FOXD3*, *TSH2B* (*H2BC1*) and *HOXD8* (H3K4me3-occupied; Brykczynska et al., 2010) (Fig. S1G). Comparison of two H4K20me3-ChIP-seq antibodies (Abcam, ab9053; Diagenode, C15410207) using HeLa cells revealed a higher specificity for C15410207 for distinguishing H4K20me3-positive (*ZNF510*) and -negative regions (*GAPDH*) (Fig. S1I).

For the ChIP-seq experiment, we isolated nucleosomal chromatin from motile sperm samples from three donors (D31, D63 and D93) who exhibited normal semen parameters and were considered to be fertile according to the reference values given by the World Health Organization (Cooper et al., 2010) (Table 1, Fig. 2A). The dissociated nucleosomes in the soluble fraction were separated from the insoluble protamine fraction, and nucleosomal DNA samples were isolated and checked on agarose gel (Fig. 2B). We observed sharp bands at the expected size for mono-nucleosome DNA (146 bp) in all three samples (Fig. 2B), and another shorter band at approximately 120–130 bp when the intensity of the UV imager was dimmed. The shorter bands may be due to over-digestion of the DNA wrapping the end-sites of the nucleosomes. However, some histone 2 variants in mammalian sperm (e.g. H2A.Bbd in humans and H2AL2 in mice) result in shorter mono-nucleosome DNA of 120–130 bp (Syed et al., 2009; Ishibashi et al., 2010). After verification, nucleosomal fractions were subjected to H4K20me3-ChIP, and ChIP-DNA samples were tested by qPCR in selected control regions (Fig. 2C). As an input control for each sperm sample and ChIP-seq reaction, we used the entire mono-nucleosome DNA without immunoprecipitation. We also performed qWB analyses on total protein extracts isolated from motile sperm fractions of D31, D63 and D93. Among these three donors, D31 exhibited the highest content of H4K20me3 followed by D93 and D63 (Fig. 2D).

Validation of ChIP-seq data and general characterization of detected peaks

We analyzed a total of three human sperm samples (D31, D63 and D93; i.e. three biological replicates) with regard to genome-wide distribution of H4K20me3. To specify the genomic distribution of H4K20me3 in human sperm, we also considered the sperm nucleosome fraction without H4K20me3. Therefore, sperm inputs were also subjected to peak calling, and input peaks cleared from overlapping H4K20me3 peaks were considered nucleosome-binding sites without H4K20me3. To identify those H4K20me3 loci, which may be important for paternal epigenetic inheritance, we compared the genome-wide distribution of H4K20me3 in human sperm with the distribution in human somatic cells. We used the K562 myelogenous leukaemia cell line. The ChIP-seq data from

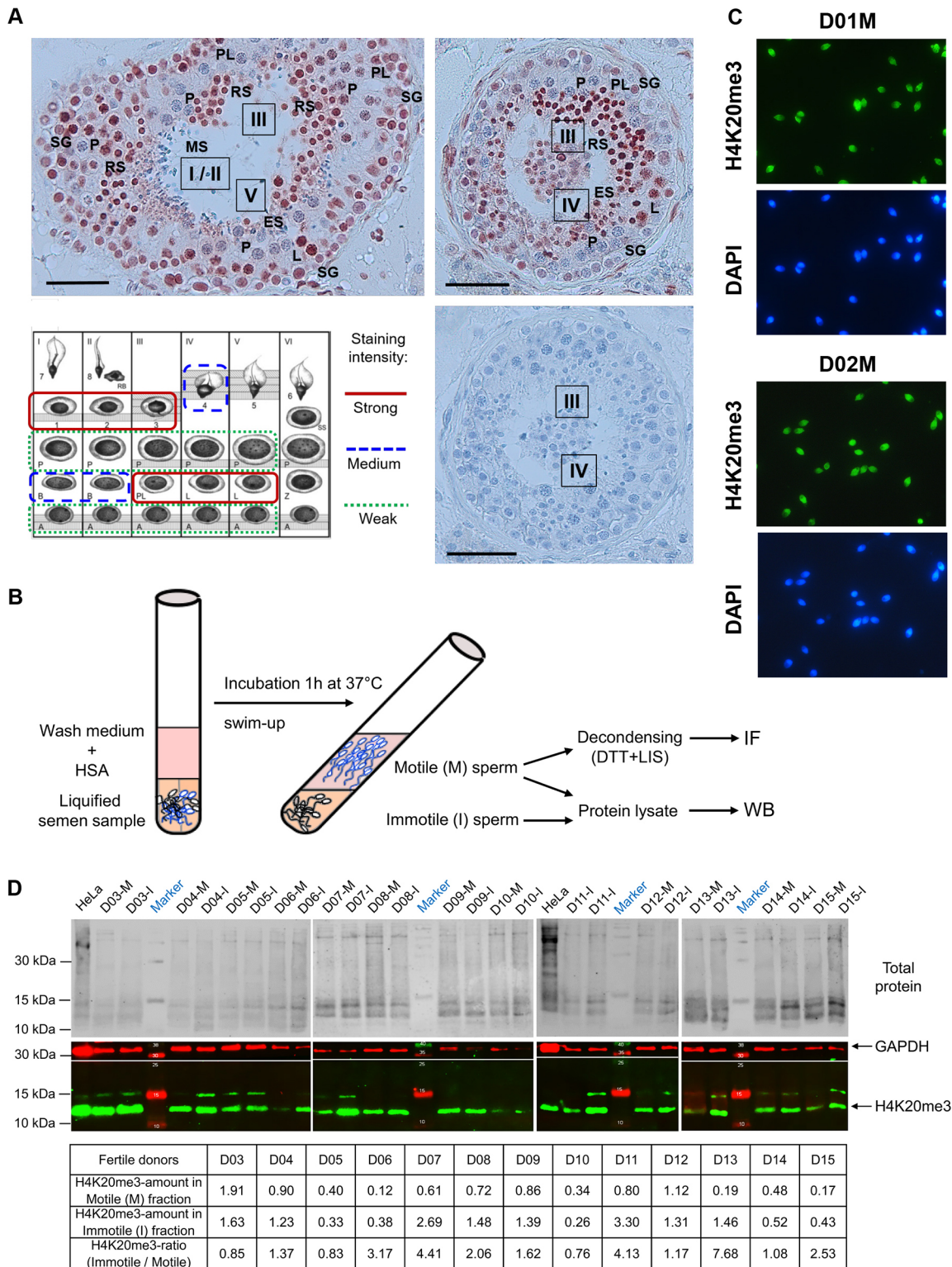


Fig. 1. See next page for legend.

K562 cells ($n=5$ technical replicates) were generated and kindly provided by Diagenode (Diagenode SA, Belgium; GSE129239).

The main ChIP-seq and peak calling statistics are given in Table S1. From all reads generated by H4K20me3-ChIP-seq (D31: 76,831,541 reads; D63: 58,675,691; D93: 72,016,300), approximately 89% in D31 and D93 and 94% in D63 were mapped efficiently. After filtering redundant H4K20me3 reads, non-redundant reads

accounted for 71.6% of efficiently mapped reads in D31, 60.7% in D63 and 61.4% in D93. Among the five K562 replicates, a mapping efficiency of 83-96% was achieved, with 61-88% non-redundant H4K20me3 reads (Table S1). Among the sperm input reads, a mapping efficiency of 67-79% was achieved, with 56-84% non-redundant reads (Table S1). Principal component analysis revealed correlation between different sperm samples, which were clearly

Fig. 1. Analysis of H4K20me3 in human testis and semen samples reveals its presence in every spermatogenic stage and in ejaculated spermatozoa. (A) Immunohistochemistry of testis sections obtained from men with obstructive azoospermia and exhibiting normal spermatogenesis revealed H4K20me3 in spermatogonia (SG, weak to medium staining), pre-leptotene (PL, strong staining), leptotene (L, strong staining), and pachytene spermatocytes (P, weak staining limited to few regions), and round and early elongating spermatids (RS and ES, strong to medium staining, respectively). No staining was detectable in late elongating (maturing) spermatids (MS). Stage VI is very rare and not assessable. Top panels show exemplary tubules representing stages I to V of the seminiferous epithelial cycle; bottom-right panel is a serial section of the tubule shown above representing the secondary antibody control; bottom-left panel is a schematic summarizing H4K20me3 levels in the different stages of human spermatogenesis. A, A type spermatogonia; B, B type spermatogonia; L, leptotene; P, pachytene primary spermatocytes; PL, pre-leptotene; SS, secondary spermatocytes; Z, zygotene. Roman numerals represent the step of spermiogenesis; RB: residual body. Scale bars: 50 μ m. (B) Schematic depicting separation and analysis of motile and immotile human sperm cells. Protein lysates from both fractions were analyzed for H4K20me3 and GAPDH by WB. Motile spermatozoa were decondensed using DTT and LIS and analyzed for H4K20me3 by IF. (C) IF revealed H4K20me3 retention in human motile spermatozoa as shown in two donor samples (D01M and D02M). Sperm DNA was stained with DAPI. (D) Quantitative WB analysis of motile (M) and immotile (I) sperm samples obtained from fertile donors (D03-D15) revealed an inter-individual variability of H4K20me3 content in motile and immotile sperm fractions. Immotile sperm fractions exhibited more frequently a higher content of H4K20me3 in comparison with corresponding motile fractions. The quantification of H4K20me3 signals was calculated using the total protein stain in individual lanes, GAPDH was used as quality control of protein isolation, and HeLa protein extract was used as a control for antibodies.

separated from K562 replicates (Fig. S2A). Differential binding analysis of ChIP-seq data revealed a sharp difference between sperm and somatic cells, as demonstrated in the correlation heatmap generated using the affinity (read counts) data (Fig. S2B). A Pearson's correlation analysis of the sequencing reads was performed in 10 kb and 500 bp window sizes (Fig. S3A,B). It showed that the three input controls from sperm samples correlated well ($R=0.78-0.90$ in 10 kb window; $0.66-0.77$ in 500 bp window) and indicated high reproducibility of the mono-nucleosome DNA preparation procedure.

The H4K20me3-IPs in three sperm samples (reads produced on immunoprecipitated DNA) showed R-values of 0.66-0.87 in the 10 kb window, and 0.51-0.64 in the 500 bp window, and were comparable to K562 technical replicates ($R=0.81-0.90$ in 10 kb window; $0.46-0.67$ in 500 bp window) (Fig. S3A,B). To verify study-to-study biases and assess whether freezing and thawing of human motile spermatozoa has a diminishing effect on nucleosome preparation, we compared our previous MNase-seq reads generated on the pooled input DNA of four fertile donors by using non-frozen swim-up spermatozoa (GSE47843; Samans et al., 2014) to input reads generated in this study for D31, D63 and D93 by using frozen-and-thawed swim-up spermatozoa. Correlation coefficients calculated by the Spearman method indicated a high correlation between previous and current input reads ($R=0.75-0.87$, 10 kb window), excluding study-to-study biases and experimental variations in the preparation of mono-nucleosomes (Fig. S3C). Further comparison of MNase-seq reads in typical nucleosome-enriched gene promoters using the integrated genome viewer (IGV) showed a similar enrichment of sequencing reads generated from non-frozen and frozen-thawed spermatozoa (Fig. S3D).

After peak calling, the number of H4K20me3 peaks was evaluated for human sperm and K562 cells. The number of input peaks was also evaluated for sperm samples (Table S1). We found considerable variation in H4K20me3 peaks among sperm samples (D31: 8156 peaks; D63: 1254 peaks; D93: 2433 peaks), and the number of peaks detected for K562 cells was between 1730 and 2799. The high number of peaks in D31 may be explained by the higher content of H4K20me3 detected in D31 sperm (Fig. 2D). Moreover, a higher amount of dsDNA was ChIP-ed in D31 than in D63 and D93, so that the library preparation started with different DNA quantities (D31: IP and input each 1000 pg; D63 and D93: IP and input each 620 pg). Accordingly, different numbers of PCR cycles were needed to obtain a sufficient number of libraries (11 for D31, and 13 for D63 and D93). As a result, there was a higher proportion of non-redundant reads in D31 that could be used for peak calling. This may also have contributed to a higher number of H4K20me3 peaks called in D31. The number of input peaks detected in sperm was 3000 for D31, 2030 for D63, and 3801 for

Table 1. Age and semen parameters of healthy donors analyzed in this study by immunocytochemistry (D01, D02), qWB (D03-D15) and ChIP-seq (D31, D63, D93)

Donor ID	Age (years)	Semen volume (ml)	Semen pH	Sperm concentration ($\times 10^6$ /ml)	Global motility (%)			
					A	B	C	D
D31	31	2	8.3	102	46	3	7	44
D63	27	1.5	8.1	373	44	23	5	28
D93	39	2.5	8.7	127	51	18	13	18
D01	29	2.4	8.4	31.2	71	16	4	9
D02	29	2.8	8.7	69.8	88	4	2	6
D03	26	3	7.9	44.6	78	7	9	6
D04	26	2.2	8.3	85.2	51	6	13	31
D05	26	2.6	9	80.4	61	11	7	21
D06	32	4.2	8.1	56.5	40	14	12	34
D07	34	2	8.6	197	18	11	6	65
D08	27	5.4	8.3	63.6	61	18	2	19
D09	35	3.9	8.7	74.2	56	9	9	26
D10	36	4.2	8.1	77.8	43	21	14	22
D11	36	8.7	8.5	35.6	60	8	4	28
D12	21	2.5	8	80.6	53	18	12	17
D13	26	4	8.3	152	84	8	1	7
D14	28	2	8.7	88.6	41	20	6	33
D15	22	2.6	7.8	44.8	72	21	6	1

The sperm samples met normal semen parameters according to the reference values given by the World Health Organization (Cooper et al., 2010). For sperm motility: A (rapidly progressive), B (slowly progressive), C (non-progressive), D (immotile).

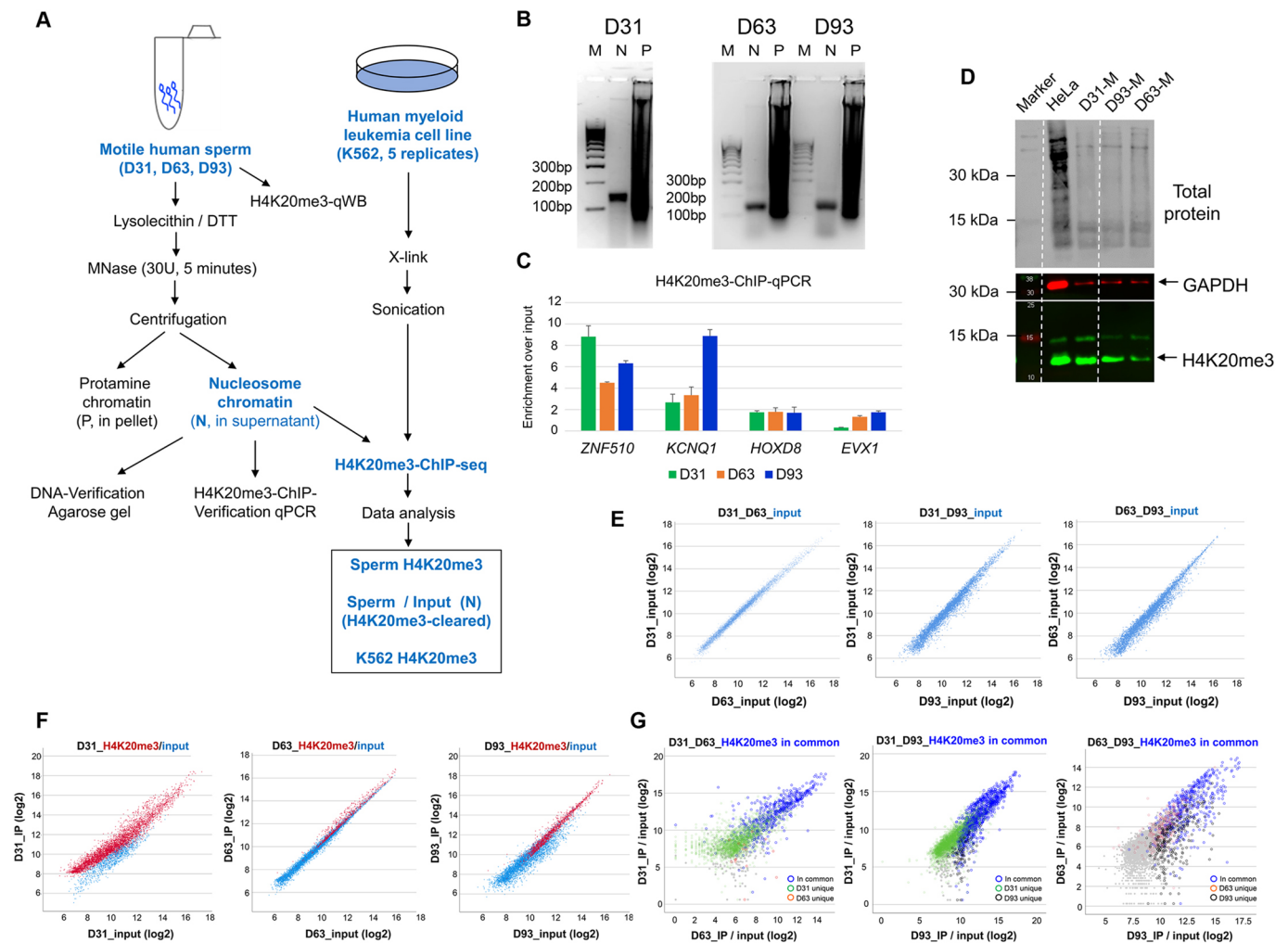


Fig. 2. Workflow of H4K20me3-ChIP-seq, verification experiments, and scatterplots of normalized sequencing reads of sperm samples. (A) Sperm nucleosome fractions from three fertile donors D31, D63 and D93 were isolated from native motile spermatozoa using an MNase-based technique and subjected to H4K20me3-ChIP and high-throughput sequencing. For K562 cells, formaldehyde-cross-linked (X) sonicated chromatin of five replicates were used for H4K20me3-ChIP-seq. Sperm input DNA (10% of total nucleosome DNA) was also subjected to sequencing and peak calling, and input peaks cleared from overlapping H4K20me3 peaks were considered as nucleosome binding sites without H4K20me3. (B) DNA isolated from the protamine (P) pellet and from a part of the nucleosome (N)-containing supernatant of D31, D63 and D93 was verified on agarose gel (M: 100 bp marker). Nucleosome-descending DNA appeared at 146 bp (mono-nucleosome size). Highly concentrated and severely degraded by DTT P-DNA appeared as a smear ranging from high to low molecular DNA. (C) H4K20me3-ChIP-DNA samples from D31, D63 and D93 were tested by qPCR in potential H4K20me3-binding sites (*ZNF510* and *KCNQ1*) were selected from ChIP-seq data in HeLa and K562 cells, Diagenode). For comparison, regions known to retain H3K4me3 and H3K27me3 in human sperm were also analyzed (*HOXD8* and *EVX1*; Bryczynska et al., 2010). (D) Total protein extracts from motile sperm fractions of D31, D63 and D93 were analyzed by qWB with regard to H4K20me3 content. D31 exhibited the highest content of H4K20me3 (relative value 1.78) followed by D93 (1.04) and D63 (0.38). The total protein amount was used as loading control for quantification, GAPDH was used as quality control of protein isolation, and HeLa protein extract was used as a control for the antibody. (E-G) All input and ChIP (IP) reads were log₂-normalized to respective read counts (library sizes), and scatterplots of input (E), ChIP (F) and ChIP/input reads (G) were generated. H4K20me3 enrichments for each sample are shown in F. H4K20me3 enrichments in common between two samples are shown in blue in G.

D93. Among input peaks occurring in at least two of three samples, 53.6% intersected with RDEs (for comparison, 93% of nucleosome input peaks intersected when read alignment was performed to every matching genome-site; Samans et al., 2014; discussed by Dansranjav and Schagdarsurengin, 2016). All input and H4K20me3 data were normalized to respective library sizes. Scatterplots of log₂-normalized read counts comparing input-to-input, ChIP-to-input and ChIP-to-ChIP are shown in Fig. 2E-G.

In sperm and K562 cells, H4K20me3-occupied DNA regions were characteristically very large, with >5000 kilobase pairs, and the same H4K20me3-enriched region often appeared in one sample as one long peak and in another sample as several short peaks (Fig. S4A-C). A direct comparison of normalized bigwig files of

H4K20me3-ChIP-seq and MNase-seq input reads from sperm samples using IGV showed that H4K20me3 peaks occurred mostly with corresponding input peaks (Fig. S4B). However, in contrast to D63 and D93, D31 exhibited numerous H4K20me3 peaks in regions lacking input peaks (Fig. S4D). These stand-alone peaks in D31 may be a result of higher ChIP efficiency with the antibody enriching for regions in a subset of spermatozoa, though the majority of cells do not carry nucleosomes at the corresponding sites and may, together with the preceding notes, explain the different ChIP-to-input scatterplot of D31 (Fig. 2F). Peaks overlapping in at least two of three sperm samples and in three of five K562 samples were considered: 7208 H4K20me3 peaks were considered for sperm samples (median size 0.036 Mbp), 2750 peaks for

H4K20me3-cleared sperm input samples (median size 0.14 Mbp) and 7240 H4K20me3 peaks for K562 samples (median size 0.027 Mbp). Next, we evaluated how much of the entire genome in human sperm is occupied by nucleosomes with and without H4K20me3. When considering the median peak size and the number of peaks identified in at least two of three sperm samples, approximately 12% of the human sperm genome was occupied by nucleosomes without H4K20me3 and 8% with H4K20me3. Remarkably, sperm nucleosomes without H4K20me3 comprised a high number of genes (10,774, with an average gene density of 28 genes/10⁶ bp), whereas those carrying H4K20me3 comprised only 1928 genes (~7 genes/10⁶ bp), similar to K562 cells, in which H4K20me3 peaks comprised 1759 genes with a similar density (~9 genes/10⁶ bp) (Table S2). Both H4K20me3-free and -containing nucleosome-binding sites in sperm included a high number of repetitive DNA elements and exhibited similar repeat densities (2844 repeats/10⁶ bp and 2368 repeats/10⁶ bp, respectively). However, remarkable differences in repeat frequencies and densities were observed when different repeat families were considered (Table S2).

Biological similarity and variability of H4K20me3 preservation in human sperm

Annotation of peaks to genomic features revealed that sperm nucleosomes with H4K20me3 occur mostly in intergenic and intronic regions (Fig. 3A). Log₂-enrichment analyses (observed versus expected at random distribution) showed a strong depletion of H4K20me3 peaks in promoters (except D63), 5'-UTR (except D63), exons, introns, 3'-UTR, and transcription termination sites (TTSs), and enrichment in distal intergenic regions (Fig. 3B). Gene ontology (GO) biological process analyses revealed that the most common H4K20me3 binding sites in human sperm were strongly enriched in genes encoding factors of 'Sensory perception of smell' (Fig. 3C,E, Fig. S4A-D). Most variable sites were enriched in genes encoding factors of xenobiotic and flavonoid metabolic processes (Fig. 3C). All ChIP-samples showed an absence of H4K20me3 in HOX gene clusters (Fig. S4E), whereas corresponding inputs (i.e. total nucleosomal DNA) exhibited peaks in HOX clusters (Fig. 3F, Fig. S4E,F). Importantly, besides gene-rich regions, sperm input peaks were also detectable in many centromeric, telomeric and intergenic regions (Fig. S4F). Log₂-enrichment analyses with regard to repeats revealed in all three sperm samples that H4K20me3 is enriched in LINEs, endogenous retroviral sequences (ERVs), low complexity repeats, satellite repeats (except D31), simple repeats (except D31) and pseudogenes, and depleted in SINEs (Fig. 3D).

Genome-wide distribution of H4K20me3-free and -containing nucleosomes in human sperm compared with K562 cells

Among the H4K20me3 peaks detected in human sperm, 4903 (68%) were also present in K562 cells (Fig. 4A). Furthermore, 2305 H4K20me3 peaks in sperm and 2337 in K562 appeared in a cell type-specific manner (i.e. exclusively in sperm and K562, respectively). Annotation of peaks to genomic features revealed that sperm nucleosomes without H4K20me3 occur more frequently in genic regions, whereas H4K20me3-containing nucleosomes occur mostly in distal intergenic regions in both sperm and K562 cells (Fig. 4B). When only common H4K20me3 peaks in sperm and K562 cells were considered, an increased percentage of those occurring in distal intergenic regions were noted (Fig. 4B). Log₂-enrichment analyses emphasized the contradictory features of H4K20me3-free and -containing nucleosomes.

Whereas H4K20me3-free nucleosomes were strongly enriched at promoters, 5'UTRs, exons, 3'UTRs and TTSs, H4K20me3-containing nucleosomes were depleted in all these regions in both sperm and K562 cells (Fig. 4C). GO term biological process analyses of sperm nucleosomes without H4K20me3 determined that they were enriched in genes playing a significant role in embryonic development (Fig. 4D, Table S3). In particular, gene promoters of developmental transcription activators and repressors were frequently enriched with H4K20me3-free nucleosomes (Fig. 4E). A total of 430 genes encoding diverse transcription regulators in human sperm contained nucleosomes in their promoters (Table S4). Among them, we detected many homeobox factors (homeobox A, B, C, D; POU, NK and SIX class homeobox), enhancer binding factors, forkhead and SRY-box proteins, Hes family transcription factors, Kruppel-like factors, nuclear receptors, several zinc finger proteins, and CCCTC-binding factor (Table S4). In comparison, H4K20me3-containing nucleosomes were enriched in both sperm and K562 cells in genes encoding odorant binding factors and olfactory receptors (Fig. 4D,E, Table S5). Notably, gene clusters that are repetitive in nature, such as olfactory receptor genes and zinc-finger genes, were enriched for H4K20me3 (Fig. S4A-D, Table S6). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of H4K20me3 peaks revealed two pathways, 'Olfactory transduction' and 'Herpes simplex virus 1 infection', to be significantly enriched in both sperm and K562 cells (Fig. 4F, Table S6).

Next, we annotated all peaks to repetitive DNA classes and worked out the main features of sperm nucleosomes with and without H4K20me3 and compared the findings to the H4K20me3 distribution in somatic cells. We found sperm nucleosome binding sites (with and without H3K20me3) to varying extents in most of the repeat classes (Table 2, Table S2). In detail, among all repeats detected in nucleosome binding sites without H4K20me3 in sperm (1,094,763 total), 33.5% were SINEs (Alu) and 12.8% LINEs (Table 2). In contrast, among all repeats detected in H4K20me3-binding sites in sperm (614,545 total), a significantly lower occurrence of SINEs (Alu) (12.7%) and a significantly higher occurrence of LINEs (23.7%) was observed (Table 2). In sperm, a significantly higher proportion of ERVs (particularly ERV1, ERVK, ERVL-MaLR), long tandem repeats (LTRs), gypsy retrotransposons, low complexity repeats, satellite and centromeric repeats was found in H4K20me3-containing nucleosomes in comparison with those without H4K20me3 (Table 2). Sperm and K562 cells exhibited similar frequencies of occurrence of different repetitive DNA elements in H4K20me3-binding sites (Table 2). These data were also supported by log₂-enrichment analyses in repeats and analyses of repeat densities in H4K20me3-free and -containing regions (Fig. 4G, Table S2). All other analyzed repeat families did not exhibit considerable differences between the groups. Furthermore, we found that H4K20me3-free nucleosome binding sites in sperm were enriched in CpG islands, whereas H4K20me3-containing nucleosomes were depleted (Fig. 4G).

Differential binding of H4K20me3 in sperm and somatic K562 cells and genome-wide distribution of unique sites

DiffBind analyses of consensus (common) peaks between sperm and K562 cells (4903 total) revealed 1733 differentially bound sites, of which 1598 showed an increased affinity in sperm (Fig. 5A,B). A comparison of H4K20me3-binding sites with an increased affinity in sperm to those with an increased affinity in K562 cells with regard to GO term biological processes revealed a significant enrichment of several categories in sperm, but not in K562

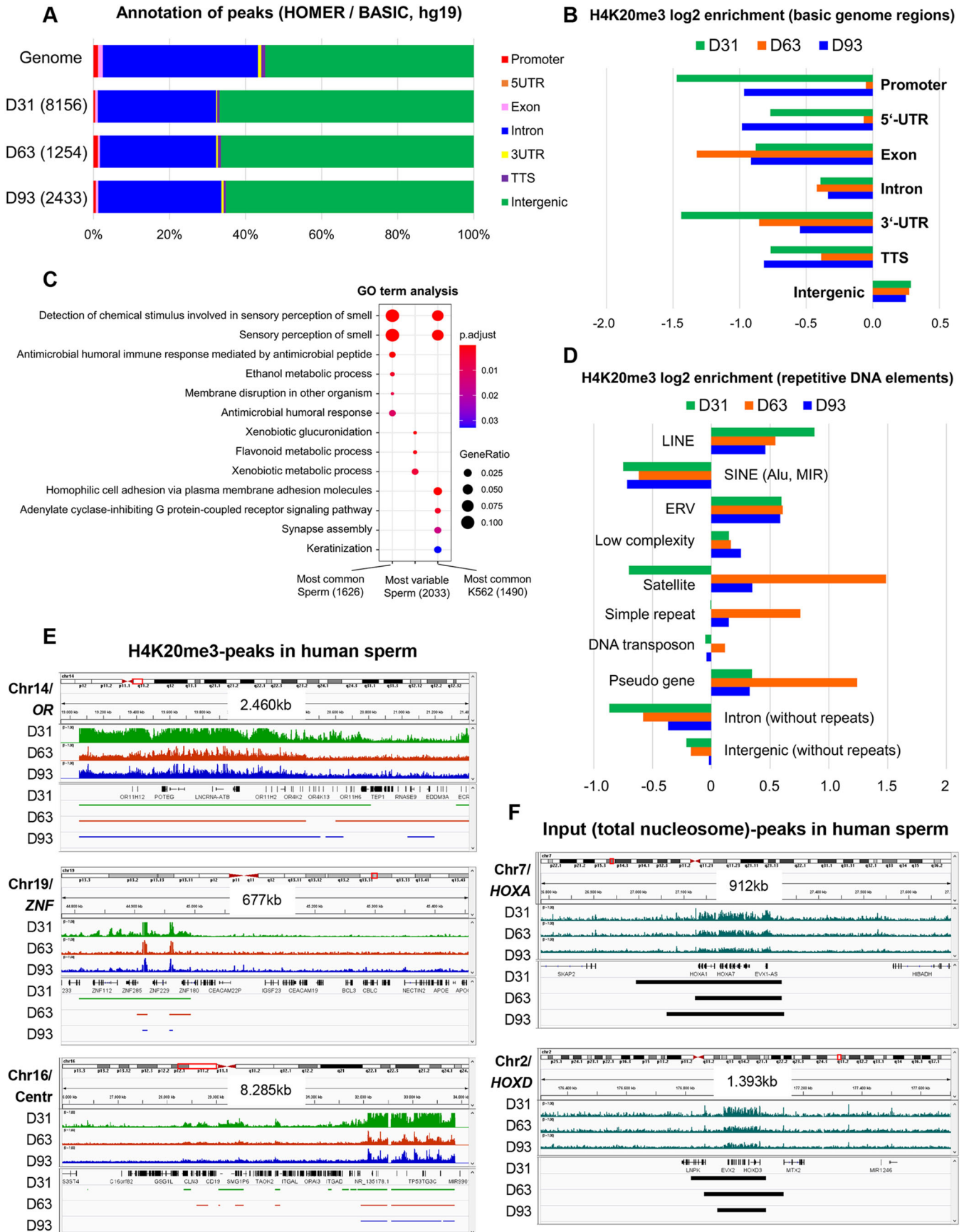


Fig. 3. See next page for legend.

Fig. 3. Biological similarities and variability of the genome-wide H4K20me3 distribution and enrichment in human sperm. (A) Annotation of peaks detected in sperm samples D31, D63 and D93 to genome regions was carried out using HOMER package (annotatePeaks.pl function, basic annotation). The number of peaks in each sample and the expected genome distribution are shown. (B) Log₂-enrichment analysis (observed versus expected) of H4K20me3 peaks in D31, D63 and D93 revealed comparable enrichments of H4K20me3 in intergenic regions and depletions in genic regions (except D63, for which depletion was found to a lesser extent in promoter and 5'-UTR regions). (C) Gene ontology (GO) term biological process analysis of the most common H4K20me3 peaks in sperm (i.e. peaks detected in all three samples) revealed a significant enrichment of genes involved in sensory perception of smell. The most variable H4K20me3 peaks (i.e. peaks detected in one of three samples) were enriched in genes involved in xenobiotic and flavonoid metabolic processes. In K562 cells, the most common H4K20me3 peaks (i.e. peaks detected in all five technical replicates) were significantly enriched in genes involved in sensory perception of smell and homophilic cell adhesion. Considered gene numbers, gene ratios and adjusted *P*-values are given. (D) Log₂-enrichment analysis (observed versus expected) of H4K20me3 peaks in D31, D63 and D93 in different repeat families showed similar enrichments in LINEs, ERVs, low complexity repeats and pseudogenes, and similar depletions in SINEs. Diverging H4K20me3-enrichment profiles were found for D31 in satellite and simple repeats. (E) IGV snapshots of H4K20me3 peaks in D31, D63 and D93 in olfactory receptor (OR) and zinc-finger (ZNF) gene clusters, and in the centromere (Centr) region. (F) IGV snapshots of sperm input peaks in D31, D63 and D93 in *HOXA* and *HOXD* gene clusters. For E and F, normalized .bigwig files (upper three rows) and .bed files (lower three rows) were used.

(Fig. 5C). Besides the genes involved in sensory perception of smell, many genes involved in antimicrobial and antiviral immune response could be detected in H4K20me3 sites with an increased affinity in sperm (Fig. 5C). Moreover, a more frequent occurrence of LINE1s and a less frequent occurrence of SINEs was detected for H4K20me3 sites with an increased affinity in sperm (Fig. 5D).

We also analyzed the genome-wide distribution of cell type-specific H4K20me3-binding sites (i.e. unique peaks occurring exclusively in sperm or K562 cells) with regard to their occurrence in different genome regions including repeats, and GO term enrichment (Figs 4B, 5E-G). Interestingly, unique H4K20me3 sites in sperm were enriched in 5'-UTR, whereas K562-specific sites were depleted (Fig. 5E). Moreover, a significant enrichment of genes involved in antimicrobial and antiviral immune response could be found in unique H4K20me3-binding sites in sperm, but not in K562 cells (Fig. 5F). A higher enrichment in LINE1s, ERVs, low complexity repeats and satellite repeats, and DNA transposons, and a depletion in SINEs was observed for unique H4K20me3-binding sites in sperm (Fig. 5G).

DISCUSSION

Several studies have provided evidence that SHs remain in distal intergenic regions and RDEs, but the biological relevance of this is still largely unknown. As RDEs occupy more than two-thirds of the entire mammalian genome and are tightly regulated during the life cycle, it is probable that SHs are involved in regulation of RDEs (Dansranjav and Schagdarsurengin, 2016). We therefore analyzed the genome-wide distribution of the heterochromatic marker H4K20me3 in human sperm and somatic (K562) cells, and compared it with that of H4K20me3-free loci in sperm. Sperm and somatic cells exhibited a similar genome-wide distribution pattern of H4K20me3, which was clearly distinguishable from H4K20me3-free nucleosomes in sperm (Fig. 6). Functional attribution of sperm nucleosomes were obtained in both transcriptional regulation of euchromatin (particularly H4K20me3-free nucleosomes enriched in

gene promoters) and maintenance and regulation of heterochromatin (particularly H4K20me3-containing nucleosomes enriched in distal intergenic regions and RDEs, with strong similarities to the distribution in somatic cells) (Fig. 6).

A recent sperm-chromatin-structure-assay (SCSA) in mice demonstrated that swim-up-sperm exhibited HDS (high DNA stainability) values of $6.1 \pm 2.1\%$ and contained immature cells with an uncompleted histone-to-protamine-replacement (Yoshida et al., 2018). As we did not perform SCSA on our samples and could not find HDS data generated on motile spermatozoa of fertile men in literature, we cannot exclude the possibility of a contamination with immature cells. However, in humans, HDS is highly negatively correlated with sperm motility, and large-scale SCSA studies in infertile men indicated HDS values of $3.3 \pm 2.5\%$ with numerous cases exhibiting even 0.1-0.8% (Gandini et al., 2004; Niu et al., 2011). On this basis, one might assume that swim-up spermatozoa of fertile men are likely to have HDS values of <1%.

In the present study, we detected enrichments of H4K20me3-free nucleosomes in euchromatic regions of the sperm genome, more precisely in gene promoters, exons, 5'-UTR, 3'-UTR and TTSs. GO enrichment analyses of H4K20me3-free promoters revealed a significant enrichment of 430 genes encoding RNA polymerase II-specific transcription activators and repressors with potential relevance in embryonic development, postnatal development, and beyond. These results largely confirmed the findings of previous studies (Hammoud et al., 2009; Arpanahi et al., 2009; Brykczynska et al., 2010; Erkek et al., 2013; Teperek et al., 2016; Yoshida et al., 2018; Yamaguchi et al., 2018) and supported the potential role of SHs in the transcriptional regulation of genes in the next generation.

Our IHC analyses demonstrated that H4K20me3 could be detected at all human spermatogenic stages. However, the strongest IHC staining was observed in spermatocytes and elongating spermatids. A recent study in mouse showed that H1t, the major linker histone variant in pachytene spermatocytes, is predominantly associated with LINEs and LTRs, and that H1t-containing domains carried repressive marks such as methylated CpGs, H3K9me3, H4K20me3 and heterochromatin-associated proteins, including HP1 β and PIWIL1 (Mahadevan et al., 2020). In late human pachynema, SUMO1, a ubiquitin-like protein involved in chromatin inactivation and transcriptional repression, has been shown to colocalize with H3K9me3, H4K20me3 and HP1 α at constitutive heterochromatin (Metzler-Guillemain et al., 2008). A study in rats demonstrated that the spermatid-specific linker histone HILS1, the expression of which overlaps with histone-to-protamine exchange, is a poor condenser of DNA and, in condensing spermatids, preferentially binds to LINE1s within intergenic and intronic regions (Mishra et al., 2018). In HILS1-bound chromatin, enrichments of H3K9me3, H4K20me3 and H4 acetylation were observed, whereas H3K4me3 and H3K27me3 were absent.

In both human sperm and K562 cells, we detected H4K20me3 enrichments in satellite repeats, and IGV analyses showed H4K20me3 peaks in centromeric and telomeric regions. In somatic and embryonic stem cells, H4K20me3 was shown to colocalize with H3K9me3 in PCH and telomeres, and to participate in maintenance of the constitutive heterochromatin (Schotta et al., 2004; Benetti et al., 2007). PCH-associated H4K20me3 is evolutionarily conserved and is crucially important for the genome as well as epigenome integrity (Jørgensen et al., 2013; Nelson et al., 2016). Silencing markers H3K9me3, HP1 and H4K20me3 were shown to be retained on PCH throughout murine primordial germ cell development (Magaraki et al., 2017). Importantly, studies in mouse zygotes showed that both

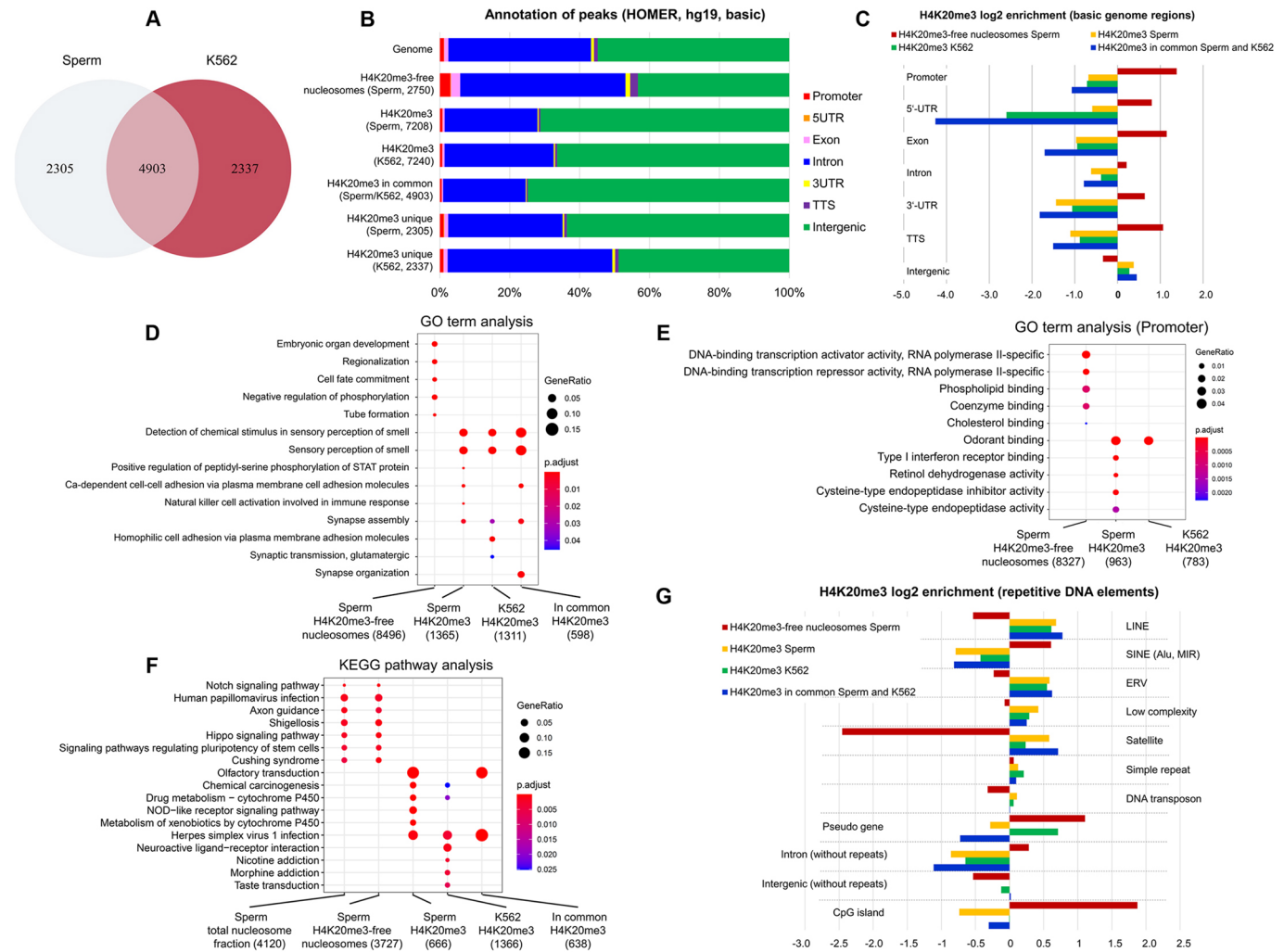


Fig. 4. Genome-wide distribution and enrichment of sperm nucleosomes (with and without H4K20me3) and H4K20me3 in somatic K562 cells. (A) Venn diagram showing the number of common H4K20me3 sites between sperm and K562 cells and the number unique for each group. (B) Annotation of H4K20me3 and H4K20me3-cleared input peaks in sperm (detected in at least two of three samples), H4K20me3 peaks in K562 cells (detected in at least three of five replicates), and common and unique H4K20me3 peaks in sperm and K562. In comparison with expected (i.e. random) distribution, H4K20me3 peaks occurred more frequently in intergenic regions, and H4K20me3-cleared peaks (i.e. sperm nucleosomes without H4K20me3) more frequently in genic regions, in particular promoters and exons. (C) Log₂-enrichment analysis of peaks (observed versus expected) in different genome regions. H4K20me3-free nucleosomes were enriched in genic regions (in particular promoters, exons, 5'- and 3'-UTRs, and TTSs). H4K20me3-containing nucleosomes in sperm and K562 were depleted in all genic regions and enriched in intergenic. (D) GO term biological process enrichment analysis of sperm input peaks (H4K20me3-cleared) and H4K20me3 peaks in sperm and K562 cells (individual and in common). H4K20me3-free nucleosomes in sperm were enriched in genes encoding factors of embryonic organ development, regionalization and cell fate commitment. H4K20me3 was mostly enriched in genes encoding odorant binding factors in sperm and K562. Considered gene numbers, gene ratios and adjusted *P*-values are given. (E) GO term molecular function enrichment analysis of peaks detected in gene promoters. In sperm, H4K20me3-free nucleosomes were significantly enriched in promoters of genes encoding RNA polymerase II-specific DNA-binding transcription activators and repressors as well as phospholipid and co-enzyme binding factors. In sperm, H4K20me3 was enriched in promoters of odorant and interferon receptor binding factors, and factors involved in retinol dehydrogenase and cysteine-type endopeptidase activity. In K562, H4K20me3 was enriched only in promoters of odorant binding factors. (F) KEGG pathway analysis of sperm input peaks (total and H4K20me3-cleared) and H4K20me3 peaks in sperm and K562 cells. In sperm and K562, H4K20me3 was enriched in genes involved in pathways 'Olfactory transduction' and 'Herpes simplex virus 1 infection'. In sperm, H4K20me3 was also enriched in genes involved in 'Chemical carcinogenesis', 'Drug metabolism' and several others. Sperm nucleosomes without H4K20me3 were enriched in genes involved in 'Signaling pathways regulating pluripotency of stem cells', 'Notch signaling pathway' and several others. (G) Log₂-enrichment analysis (observed versus expected) of peaks in repeats revealed enrichments of H4K20me3 in LINES, ERVs, low complexity repeats and satellite repeats, and depletions in SINEs. H4K20me3-free nucleosomes in sperm were depleted in all repeats except SINEs. A depletion of H4K20me3 in CpG islands was found in sperm, but not K562 cells. H4K20me3-free nucleosomes in sperm were enriched in CpG islands.

H3K9me3 and H4K20me3 were exclusively detected on the maternal pronucleus (Eid et al., 2016; Zeng et al., 2019), whereas in human embryos H3K9me3 and H4K20me3 were paternally transmitted to the oocyte and contributed to the formation of the paternal constitutive heterochromatin, which was subsequently propagated over embryonic cleavage divisions (van de Werken et al., 2014).

Next, we found that sperm nucleosomes with as well as without H4K20me3 occurred frequently in intronic and intergenic regions and were specifically enriched in retrotransposons. Interestingly, sperm nucleosomes with H4K20me3 occurred twice as frequently in LINE1s and ERVs (e.g. ERV1, ERVK, ERVL and ERVLMaLR), whereas H4K20me3-free nucleosomes occurred three

Table 2. Distribution of repetitive DNA elements in genomic regions occupied with H4K20me3 and H4K20me3-free nucleosomes in human sperm, and with H4K20me3 in K562 cells

Repeat families	Sperm H4K20me3-free nucleosomes		Sperm H4K20me3		K562 H4K20me3		In common H4K20me3	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Retrotransposons								
LINE1**	140,396	12.824	145,619	23.695	121,804	23.543	94,162	24.943
LINE2	91,709	8.377	47,805	7.779	39,418	7.619	28,073	7.436
SINE (Alu)**	367,239	33.545	77,804	12.660	76,183	14.725	44,351	11.748
SINE (MIR)	138,151	12.619	50,563	8.228	43,199	8.350	29,370	7.780
ERV1**	29,614	2.705	30,725	5.000	25,785	4.984	19,454	5.153
ERVK**	1635	0.149	2329	0.379	2053	0.397	1556	0.412
ERVL**	23,898	2.183	27,391	4.457	21,811	4.216	17,358	4.598
ERVL-MaLR**	55,806	5.098	56,556	9.203	44,690	8.638	35,561	9.420
LTR**	242	0.022	360	0.059	292	0.056	206	0.055
Gypsy**	2858	0.261	2507	0.408	1887	0.365	1424	0.377
RTE	2994	0.273	2181	0.355	1797	0.347	1289	0.341
DNA transposons								
hAT	3459	0.316	1431	0.233	1258	0.243	819	0.217
hAT-Black-jack	3652	0.334	2220	0.361	1858	0.359	1359	0.360
hAT-Charlie	50,998	4.658	22,737	3.700	19,152	3.702	12,945	3.429
hAT-Tip100	6745	0.616	3304	0.538	2740	0.530	1979	0.524
TcMar	862	0.079	527	0.086	405	0.078	282	0.075
TcMar-Mariner	2311	0.211	2162	0.352	1789	0.346	1345	0.356
TcMar-Tc2	1444	0.132	832	0.135	681	0.132	464	0.123
TcMar-Tigger	17,281	1.579	13,390	2.179	11,101	2.146	8073	2.138
PiggyBac	354	0.032	307	0.050	273	0.053	192	0.051
Non-transposable repeats								
Low complexity**	54,189	4.950	61,306	9.976	47,270	9.137	38,554	10.213
Simple repeats	80,736	7.375	51,882	8.442	43,040	8.319	32,163	8.520
Satellites**	538	0.049	1029	0.167	847	0.164	794	0.210
Centromeric repeats**	16	0.001	321	0.052	306	0.059	293	0.078
CR1	11,422	1.043	5829	0.949	4894	0.946	3403	0.901
Other	3140	0.287	2189	0.356	1777	0.343	1330	0.352
ncRNA								
RNA	115	0.011	80	0.013	64	0.012	45	0.012
rRNA	338	0.031	208	0.034	212	0.041	153	0.041
scRNA**	431	0.039	88	0.014	69	0.013	43	0.011
snRNA	1040	0.095	388	0.063	310	0.060	201	0.053
srpRNA**	492	0.045	92	0.015	87	0.017	51	0.014
tRNA	658	0.060	383	0.062	317	0.061	219	0.058
Total repeats	1,094,763	100%	614,545	100%	517,369	100%	377,511	100%

***P*<0.001 (Pearson's χ^2 test: repeat frequencies in H4K20me3-free nucleosome regions versus H4K20me3-containing regions). Repeats showing a significant difference in percentage distribution are marked in bold.

times more frequently in SINEs (particularly Alu). Remarkably, LINE1s and SINEs constitute at least one-third of human genomic DNA, and both are involved in transcriptional regulation. Thousands of LINEs reside within genes, and these sequences are conserved and regulate the expression of their host genes (Wanichnopparat et al., 2013). Cells use LINE1s within gene bodies as cis-regulatory elements to modulate gene expression, and they are involved in the regulation of many biological processes, including embryogenesis, cell differentiation, and cellular responses to external stimuli (Wanichnopparat et al., 2013). Alu elements make up the greatest part of SINEs and those residing within 3'-UTR can influence mRNA export from the nucleus to the cytoplasm, mRNA translation and mRNA decay via proteins, and thereby may be involved in post-transcriptional gene regulation (Maquat, 2020). Colocalization of H4K20me3 with activating histone modifications at transcriptionally dynamic regions was recently shown in embryonic stem cells (Xu and Kidder, 2018). Bivalent domains consisting of H3K4me3/H4K20me3 were predominantly located in intergenic regions and near TSSs of active genes, whereas H3K36me3/H4K20me3 were located in intergenic regions and within gene body regions of active genes, and

both types of domains were enriched with LINEs (Xu and Kidder, 2018). Therefore, we hypothesize that paternally transmitted nucleosomes in certain retrotransposons may be required for activation of the embryonic genome.

The function of SHs and RDEs with regard to early embryogenesis is still largely unexplored. A recent study compared the accessible regions between human sperm and zygotes and found that the majority of accessible chromatin features in sperm were distributed in gene-poor regions, showing an opposite pattern to zygotes and all other pre-implantation stages (Liu et al., 2019). They concluded that this unique feature of sperm chromatin in gene-poor regions may be important for the access of maternal transcription factors to completely reprogram the paternal genome, in which the gene-rich regions are largely protected by protamines (Liu et al., 2019). In early mouse embryos, LINE1 activation after fertilization was shown to regulate global chromatin accessibility and to be an integral part of the developmental program (Jachowicz et al., 2017). In line with this study, activation of MERVL and ERVL-MaLR retrotransposons was associated with cleavage-stage embryos, and ERVL-MaLR and ERVK elements have been suggested to drive gene expression in oocytes and early mouse embryos (Peaston et al.,

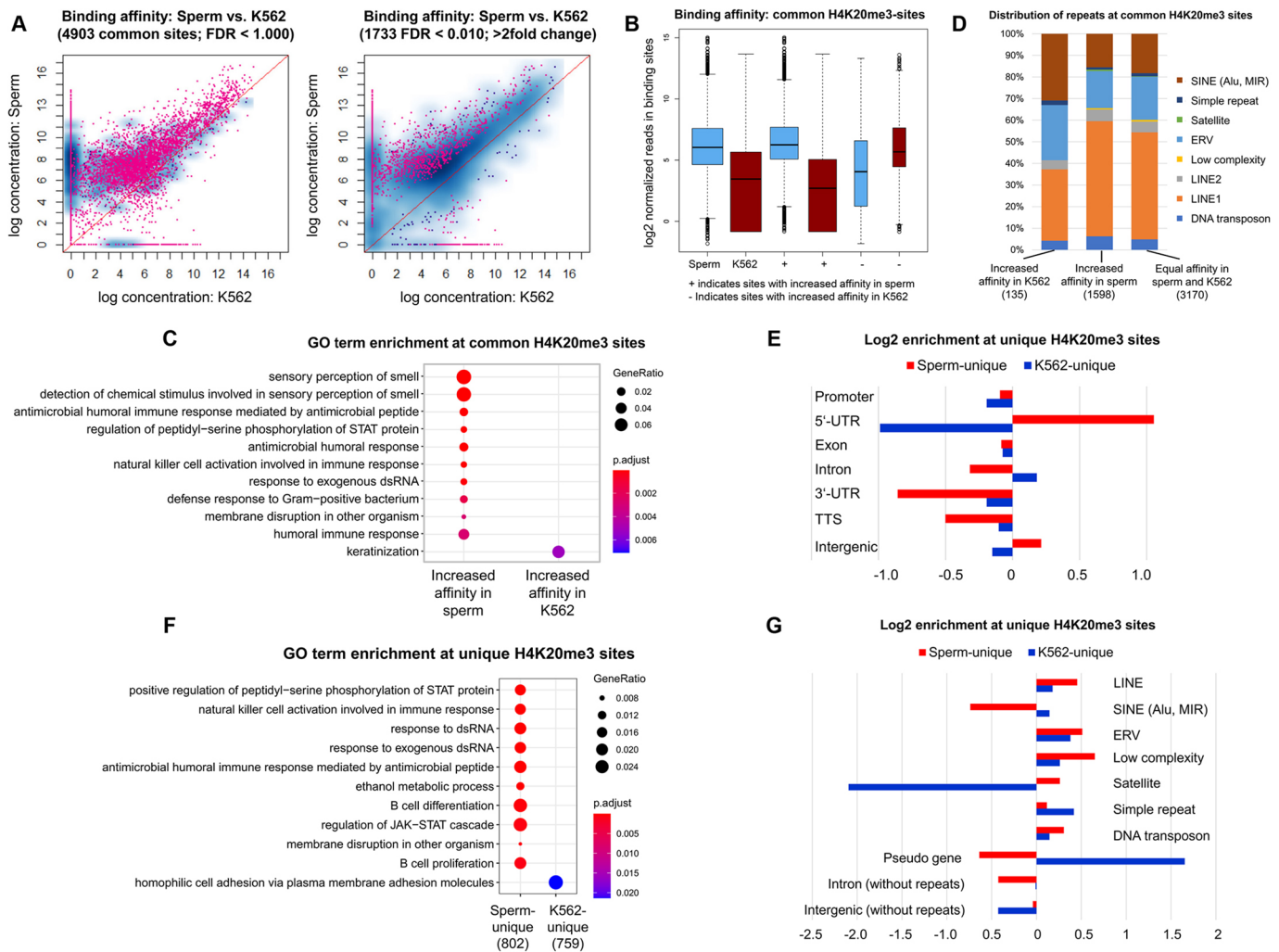


Fig. 5. Differential binding analysis of H4K20me3 in sperm and somatic K562 cells and genome-wide distribution of unique H4K20me3 sites. (A-D) Analysis of consensus H4K20me3 sites in sperm and K562 cells using DiffBind. Sample-to-sample plots show all common (A, left) and differentially bound (A, right) sites between sperm and K562. Log fold change >2 and FDR <0.01 was considered as significant. Dots above the center line represent a gain of binding affinity, and below the center line a loss of binding affinity in sperm versus K562. Box plots show the distribution of reads over all bound common H4K20me3 sites (B; horizontal line indicates median). H4K20me3 peaks with an increased affinity in sperm were enriched in genes encoding olfactory signal transducers and factors involved in immune response (C), and occurred more frequently in LINEs and less frequently in SINEs in comparison with H4K20me3 sites with an increased affinity in K562 (D). (E-G) Analysis of unique (i.e. cell type-specific) H4K20me3-binding sites in sperm (2305) and K562 (2337). Log₂-enrichment analysis in different genome regions (observed versus expected) revealed considerable differences at 5'-UTR, intronic and intergenic regions (E). In contrast to K562, sperm-unique H4K20me3 binding sites were enriched for several GO categories, in particular in those related to antimicrobial and antiviral immune responses (F), and occurred more frequently in LINEs, ERVs, low complexity repeats, satellite repeats and DNA transposons (G).

2004; Veselovska et al., 2015; Whiddon et al., 2017). Existing studies in human preimplantation embryos have reported stage-specific expression of more than 1000 repeat elements, mainly retrotransposons, in the course of embryonic genome activation (Yandim and Karakulah, 2019). ATAC-seq has shown that accessible chromatin in human preimplantation embryos is widely shaped by transposable elements and overlaps extensively with putative cis-regulatory sequences (Wu et al., 2016). Accessible chromatin regions found in early human embryos resided in distal regions that overlap with the hypomethylated domains of DNA in human oocytes and were enriched for transcription factor-binding sites (Wu et al., 2018). A potential link between the expression of repeats and nearby genes was also emphasized for SVA (SINE-VNTR-Alu) in early human embryos (Yandim and Karakulah, 2019).

Interestingly, we detected a large proportion of H4K20me3-binding sites in human sperm and somatic cells in genic (i.e.

euchromatic) regions, mostly in introns, but also in repetitive gene clusters. Our GO enrichment analyses showed that both human sperm and somatic cells were highly enriched in H4K20me3 in genes encoding odorant receptors and their promoters. This finding, especially concerning sperm cells, is intriguing because the parental olfactory experience in mice has been shown to influence behavior and neural structure in subsequent generations (Dias and Ressler, 2014; Rando, 2016). For example, when one of two odorants was introduced to male mice and paired with foot shock, the offspring exhibited increased sensitivity to this specific odorant, accompanied by an increase in the number of relevant olfactory receptor-positive neurons. This study suggested the possibility that paternal olfactory experience is inherited by offspring and transmitted in a DNA-independent manner (Dias and Ressler, 2014; Rando, 2016). The possibility of a paternally inherited sensitivity to odorants is also suspected in insects, such as *Drosophila melanogaster* flies (Williams, 2016) and *Bicyclus anynana* butterflies (Gowri et al.,

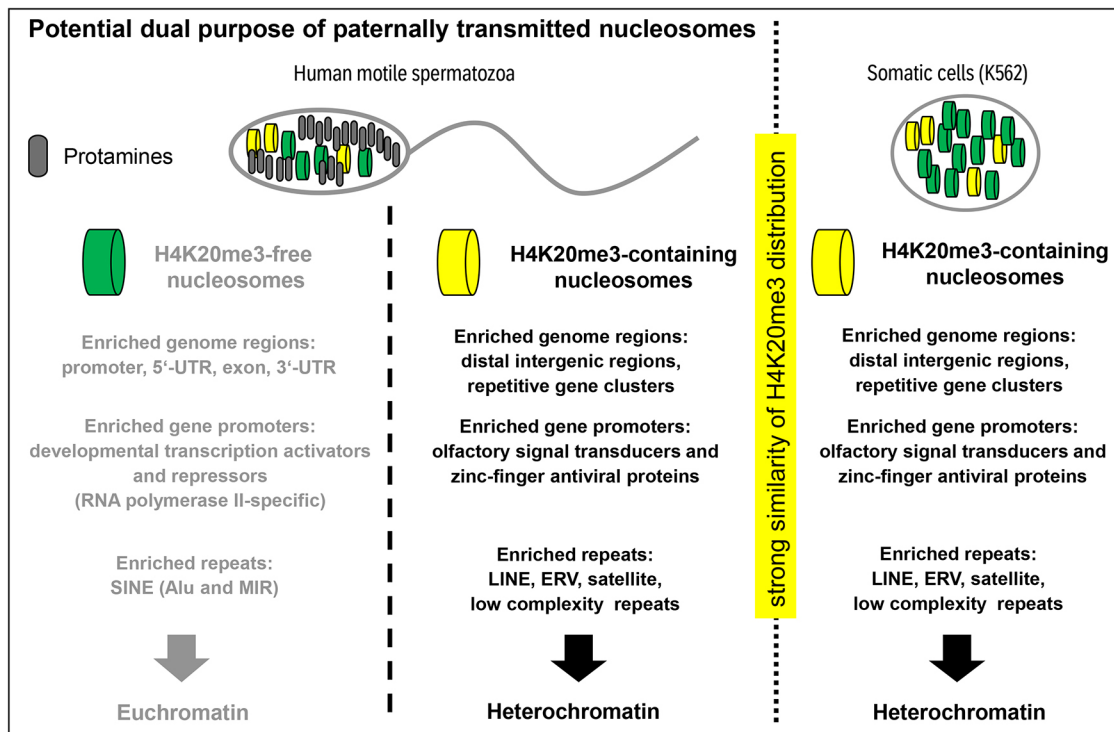


Fig. 6. Summary of study findings emphasizing the potential dual purpose of preserved nucleosomes in human sperm. In human sperm, nucleosomes without H4K20me3 were enriched in euchromatin, in particular gene promoters of developmental transcription activators and repressors, and frequently found in distal intergenic regions, especially in SINEs. In contrast, sperm nucleosomes carrying H4K20me3 occurred mostly in gene-poor distal intergenic regions and repetitive gene clusters encoding olfactory signal transducers and zinc-finger antiviral proteins, and were enriched in LINEs, ERVs, satellite repeats and low complexity repeats. A strong similarity between sperm and K562 cells was determined with regard to the overall genome-wide distribution of H4K20me3 as well as its enrichment in certain repeat families and genes. Thus, a dual purpose of paternally transmitted histones may be well expected: maintenance and regulation of heterochromatin and guidance towards transcription of euchromatin.

2019). Furthermore, our KEGG pathway analyses showed that both human sperm and somatic cells were highly enriched in H4K20me3 in genes encoding zinc-finger antiviral proteins. Remarkably, among sperm-unique H4K20me3-binding sites (i.e. sites present in sperm, but not in somatic K562 cells) we found a high enrichment of genes involved in antimicrobial and antiviral immune response.

Taken together, our results suggest a potential dual purpose of preserved nucleosomes in human sperm, which is attributable to regulation of both euchromatin and heterochromatin. Our study contributes to a better understanding of the function of paternally transmitted histones in humans and clarifies the still controversial issue of whether sperm nucleosomes are retained in gene-poor regions and repeats.

MATERIALS AND METHODS

Collection of human semen samples and swim-up procedures

The study included collection and analysis of semen samples from healthy donors and was approved by the Ethics Commission of the Medical Faculty of the Justus-Liebig-University Giessen (approval from 15 December 2010 in the frame of the Clinical Research Unit KFO181/Period 2 'Mechanisms of male factor infertility', confirmed on 17 December 2014). All study participants gave their written informed consent. Human semen samples were obtained from healthy men after sexual abstinence for 3–5 days at the Department of Urology, Pediatric Urology and Andrology, Justus-Liebig-University Giessen. Ejaculates were liquefied at 37°C for 30 min and individual semen parameters analyzed. According to WHO reference values (Cooper et al., 2010), all donors analyzed in this study had normal spermogram values and were considered as fertile (Table 1). The motile sperm fraction was separated from the immotile fraction using the direct swim-up technique. Briefly, 1 ml of semen from each donor was placed at

the bottom of a test tube and overlaid with 1.2 ml of semen wash medium containing freshly prepared human tubal fluid medium complemented with 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 1% human serum albumin (HSA, Biotest). The tubes were placed at a 45° angle to increase the interaction surface and incubated for 1 h at 37°C. The motile sperm fraction was collected carefully from the upper phase. The immotile sperm fraction was treated with a somatic cell lysis buffer [0.1% sodium dodecyl sulfate (SDS) and 0.5% Triton X-100] for 30 min on ice in order to eliminate somatic cells (Luense et al., 2016). Subsequently, the motile and immotile sperm fractions were centrifuged at 300 g for 10 min and washed twice with PBS. The purified motile and immotile sperm fractions were stored at –80°C until nucleosome preparation and ChIP.

Preparation and validation of nucleosome DNA from human sperm samples

Before starting the actual H4K20me3-ChIP-seq experiment, we performed several tests to determine the following: the number of motile spermatozoa required for isolation of a sufficient amount of nucleosomal DNA; the efficiency of native (N)ChIP versus cross-link (X)ChIP in sperm; the suitability of different ChIP washing buffers; the efficiency of MNase-treatment conditions in separation of nucleosome- and protamine-DNA; the impact of MNase concentration and treatment time on subsequent ChIP in euchromatic versus heterochromatic genome regions in sperm; the suitability of the resulting nucleosomal chromatin for ChIP in previously described nucleosome-occupied regions; and the specificity of ChIP-seq antibodies (Fig. S1, Table S7).

After positive testing, nucleosome fractions were prepared from human sperm samples as described in published protocols (Hammoud et al., 2009; Samans et al., 2014) with some minor modifications. Motile sperm cells (5×10^6) were diluted in PBS containing 0.1% lyssolecithin (Sigma-Aldrich) and protease inhibitors, and incubated for 15 min on ice. The sperm cells

were centrifuged at 2500 *g* for 5 min, washed with PBS, and re-centrifuged. The sperm pellets were then resuspended in PBS containing 20 mM dithiothreitol (DTT) and protease inhibitors and incubated at 37°C for 30 min. Sequential treatment with lysolecithin and DTT was performed in order to achieve more effective reduction of the disulfide bonds between protamines and to improve solubilization of the sperm chromatin. Subsequently, the samples were centrifuged at 2500 *g* for 5 min, washed with PBS, and re-centrifuged. Finally, the sperm pellets were resuspended in PBS containing 2 mM CaCl₂ and protease inhibitors, and the released chromatin digested with 30 units MNase for 5 min. MNase was stopped by the addition of 5 mM ethylenediamine tetra-acetic acid (EDTA) and the samples placed briefly on ice, and then centrifuged at 17,000 *g* at 4°C to separate the nucleosome fraction (in supernatant) from the protamine fraction (in pellet).

The effectiveness of the preparation of nucleosome- and protamine-associated chromatin was verified on the basis of protein and DNA isolated from each fraction and analyzed by WB (Fig. S1D), agarose gel electrophoresis (Fig. 2B, Fig. S1B), and qPCR (Fig. S1E, Table S7). For WB, H4K20me3 was detected with rabbit anti-histone H4 (tri methyl K20) antibody (ab9053, Abcam; 1 µg/ml) and IRDye 800CW goat anti-rabbit IgG secondary antibody (926-32211, 1:5000, LI-COR), and H3K27me3 was detected with mouse anti-histone H3 (tri methyl K27) antibody (ab6002, Abcam; 2.5 µg/ml) and IRDye 680RD goat anti-mouse IgG secondary antibody (926-68070, 1:5000, LI-COR) (detailed WB protocol is given in ‘Protein extraction from human sperm and qWB analysis of H4K20me3’ section). For verification of DNA, the nucleosome and protamine fractions were first treated with RNase A (Thermo Fisher Scientific) at 37°C for 30 min and then with proteinase K (Carl Roth) at 56°C for 3.5 h. DNA was isolated by a standard phenol-chloroform procedure and precipitated overnight at -20°C with 2 volumes absolute ethanol, 1/10 volume sodium acetate (pH 5.2) and 20 µg glycogen. The purified DNA samples were analyzed on a 2.5% agarose gel and visualized using the Odyssey Fc imaging system.

Native chromatin immunoprecipitation (N-ChIP) of H4K20me3 in human sperm

Nucleosome-associated-chromatin samples isolated from 2×10⁷ motile sperm cells for each donor (D31, D63, D93) were diluted 1:4 with dilution buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.1% Triton X-100). Next, 10% of each diluted chromatin was spared for use as the input DNA control and the rest incubated with 2 µg of H4K20me3 antibody (ChIP-seq grade, C15410207, Diagenode) overnight at 4°C under gentle rotation. The antibody-antigen complexes were captured with Dynabeads protein A (Thermo Fisher Scientific) and the beads washed once with each of the N-ChIP wash buffers: buffer A (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 75 mM NaCl), buffer B1 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 125 mM NaCl, 0.04% Triton X-100) and buffer B2 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 125 mM NaCl, 0.01% Triton X-100). After the washing steps, the beads were resuspended in 200 µl of freshly prepared elution buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS, 0.1 M NaHCO₃), incubated at room temperature for 15 min with rotation, and vortexed gently for 15 s. The elution step was repeated once more, and the pooled eluates together with the input controls were subjected to RNase A (Thermo Fisher Scientific) treatment at 37°C for 30 min and proteinase K (Carl Roth) treatment at 56°C for 3.5 h. DNA was purified by a standard phenol-chloroform procedure and precipitated with 2 volumes absolute ethanol, 1/10 volume sodium acetate (pH 5.2) and 6 µg glycogen overnight at -20°C. The eluted ChIP-DNA was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and tested by qPCR in potential H4K20me3-positive regions (Table S7).

Cross-link ChIP of H4K20me3 in somatic K562 cells

The human myeloid leukemia cell line K562 was authenticated and tested for contamination by Diagenode. Cross-link ChIP was performed in five technical replicates using the iDeal ChIP-seq kit for histones (Diagenode). Briefly, for each replicate, K562 cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Chromatin was sheared using a Bioruptor Pico sonication device (Diagenode) to an average fragment length of 150-200 bp as assessed by the High Sensitivity NGS

Fragment Analysis Kit (DNF-474) on a Fragment Analyzer (Advanced Analytical Technologies). ChIP was performed manually following the protocol of the aforementioned kit. Chromatin corresponding to 1% was set apart as the input control. Chromatin corresponding to 0.1 million cells was immunoprecipitated overnight at 4°C under gentle rotation by adding 0.5 µg of the same H4K20me3 antibody used for human sperm samples (Diagenode). The antibody-antigen complexes were captured with protein A-coated magnetic beads and the beads washed with the ChIP wash buffers provided in the kit. After reverse cross-linking, the eluted DNA was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

Library preparation, ChIP-seq, and data processing

Libraries were prepared using the IP-Star Compact Automated System (Diagenode) from input and H4K20me3-ChIP DNA (starting amount 1 ng for D31 and K562, and each 620 pg for D63 and D93) using the MicroPlex Library Preparation Kit v2 (12 indices) (Diagenode). Library amplification (11 PCR cycles for D31 and K562, and 13 PCR cycles each for D63 and D93) was assessed using the High Sensitivity NGS Fragment Analysis Kit (DNF-474) on a Fragment Analyzer (Advanced Analytical Technologies). Libraries were then double size-selected and purified using Agencourt AMPure XP (Beckman Coulter) and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Finally, the fragment sizes were analyzed by the High Sensitivity DNA Analysis Kit on a 2100 Bioanalyzer system (Agilent). High-throughput sequencing was performed using an Illumina NextSeq 500 system in single-end mode (50 bp, single end), as previously described in studies on heterochromatic histone marks such as H4K20me3 (Nelson et al., 2016; Zhang et al., 2020), by running NextSeq Control Software 2.1.0.31.

Quality control for the sequencing reads was performed using FastQC (Andrews, 2010). Sequencing reads were trimmed to remove low-quality score reads and then aligned to the human reference genome hg19 using BWA software v.0.7.5a. The main command line to generate the bam alignment was: `bwa aln -t 4 ${bwa_reference} ${sample_id}.fq.gz > ${sample_id}.sai; bwa samse ${bwa_reference} ${sample_id}.sai ${sample_id}.fq.gz > ${sample_id}.sam; samtools sort -@ 4 -O bam -o ${sample_id}.bam ${sample_id}.sam`. When a read mapped to multiple locations with the same scoring alignment, the mapping of the read to only one of these positions was chosen randomly and considered in peak calling. Once all reads were aligned, the PCR duplicates were removed using SAMtools v.1.3.1, and the regions blacklisted by ENCODE were filtered out. Total and mapped read numbers are given in Table S1. Alignment coordinates were converted to BED format using bedTools v.2.17 (Ramirez et al., 2016) and peak calling performed using SICER v.1.1 (Zang et al., 2009) with adjusted parameters for H4K20me3, i.e. for broad marks (window size: 3000 bp; gap size: 21,000 bp; FDR: q-value 0.3). H4K20me3 peaks were characteristically very large, ranging from a few kilobase pairs to some megabase pairs and spanning both genic and intergenic regions. We also tested a second set of parameters for peak calling (window size: 1000 bp; gap size: 3000 bp; FDR: q-value 0.1), whereby 3- to 5-times more peaks were generated after peak calling using SICER. However, we observed that these peaks were actually found in the same regions as the peaks detected by the first set of parameters, but split into several smaller peaks. Thus, for further data analyses, we decided to take into consideration the peaks generated by the first set of parameters. Peak calling on the three input samples alone was performed using the corresponding SICER functionality (=SICER-rb).

Annotations in genome regions (basic and detailed) and log₂-enrichment analyses (observed versus expected) were performed using annotatePeaks.pl from HOMER software (Heinz et al., 2010). GO term analyses were conducted with the R package ChIPpeakAnno (Zhu et al., 2010) and visualization was performed with ChIPseeker (Yu et al., 2015). The bigwig files were obtained using the bamCoverage tool of deepTools v.2.26.0. IGV v.2.4.16 (Thorvaldsdóttir et al., 2013) was used to visualize peaks on particular genomic regions. A correlation heatmap was generated using the affinity binding peaks profile (read counts) data, and the principal component analysis graphs were generated by the R package DiffBind v.2.2.12 (Stark and Brown, 2019). Differential affinity binding analyses

were performed with DiffBind. Additional bioinformatics analyses were performed using the Galaxy server for Computational Genomics hosted and maintained by the Department of Bioinformatics and Systems Biology, JLU Giessen (<https://www.computational.bio.uni-giessen.de/galaxy/>). Data reproducibility was evaluated by calculating ChIP read intensities every 10 kb and 500 bp of the human genome regions using deepTools v.3.1.2 with the multiBamSummary tool. Pearson's and Spearman's correlation coefficients were visualized as a heatmap using deepTools with the plotCorrelation tool.

Immunohistochemical analysis of H4K20me3 in human testes

Human testis tissue samples were obtained from patients with obstructive azoospermia but exhibiting normal spermatogenesis (Clinic of Urology, Pediatric Urology and Andrology, JLU Giessen). The samples were fixed in formalin, embedded in paraffin, and mounted on microscope slides (R. Langenbrinck). For IHC, testis slides were deparaffinized in xylol, hydrated through an alcohol series of decreasing concentration (100%, 96% and 70% ethanol), and boiled for 20 min in citrate buffer for antigen retrieval. Slides were cooled for 30 min to room temperature, and peroxidase activity was inhibited by adding an ice-cold 3% hydrogen peroxide solution in methanol. Slides were washed in Tris-HCl buffer (0.1 M, pH 7.4), blocked for 20 min in 5% bovine serum albumin (BSA) solution in Tris-HCl, and incubated with H4K20me3 antibody (ab9053, Abcam; 5 µg/ml) overnight at 4°C. Blocking buffer without primary antibody was used as the negative control. After washing with Tris-HCl buffer, the slides were incubated with secondary antibody (goat anti-rabbit IgG, Dako, 1:200) at room temperature for 1 h and washed several times with Tris-HCl. Slides were developed for 10-20 min using the ABC-Peroxidase Staining Kit (Thermo Fisher Scientific), and development was stopped by washing in ddH₂O. The nuclei were counterstained for 5 s in Mayer's Hematoxylin (Merck). Slides were mounted with Dako Faramount aqueous mounting medium (Agilent Technologies) and covered with glass plates before microscopic examination.

Immunofluorescence analysis of H4K20me3 in human sperm cells

Motile human sperm cells isolated by the swim-up procedure were scratched out and dried on slides, decondensed by treating first with 10 mM DTT for 10 min, and then treated with lithium diiodo salicylate (LIS) solution (10 mM LIS in 1 mM DTT-Tris-HCl buffer) for 2 h. Sperm cells were fixed with 4% paraformaldehyde (PFA, in PBS, pH 7.4) and blocked for 15 min with blocking solution (0.1% BSA and 2% Triton X-100 in PBS). H4K20me3 antibody (ab9053, Abcam; 5 µg/ml in blocking buffer) was added and incubated overnight at 4°C. Slides were washed in Tris-HCl buffer and incubated with secondary antibody (goat anti-rabbit IgG, Alexa Fluor 488, Invitrogen, 35552) at a concentration of 4 µg/ml in PBS at room temperature for 1 h. The nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI, blue). Slides were mounted with Dako Faramount aqueous mounting medium (Agilent Technologies) and covered with glass plates before microscopic examination.

Protein extraction from human sperm and qWB analysis of H4K20me3

Whole protein extracts from motile and immotile sperm fractions were analyzed by WB. Prior to protein lysate preparation, 2×10^7 sperm cells from each fraction were incubated in somatic cell lysis buffer (0.1% SDS and 0.5% Triton X-100 in ddH₂O) for 30 min on ice, washed with PBS, and pelleted by centrifugation. Sperm cells were resuspended in Tris-urea buffer (8 M urea, 50 mM Tris-HCl pH 8.0, 105 mM NaCl, 1 mM PMSF, 10 mM DTT, 2% SDS, and proteinase inhibitors), incubated at room temperature for 30 min under strong vortexing, and sonicated for three cycles at 30% amplitude for 10 s (Vibra Cell 75022 ultrasonic processor, 130 W). Protein lysates were cleared by centrifugation at 16,430 g for 10 min at room temperature, and protein concentrations measured by nanodrop (NanoDrop ND-1000, Thermo Fisher Scientific).

Sperm protein lysate (25 µg) was mixed with Laemmli loading buffer, incubated at room temperature for 5 min, and separated on 15%

polyacrylamide gels. Semi-dry transfer (Trans-Blot SD Semi-dry transfer cell, Bio-Rad) was performed for 30 min at 150 mA (maximum 25 V) on polyvinylidene-difluoride membranes (PVDF, 0.45 µm pore size, Immobilon-FL, Merck-Millipore). The whole protein content was analyzed after transfer to the membrane using the REVERT Total Protein Stain Kit (LI-COR Biosciences). Staining was documented using the Odyssey Fc imaging system (LI-COR Biosciences). Blocking was performed using Odyssey blocking buffer (1:3 diluted in PBS) for 1 h at room temperature. The membranes were incubated overnight at 4°C with H4K20me3 (ab9053, Abcam; 1 µg/ml) and GAPDH (ab9485, Abcam; 1:2500) antibodies diluted in Odyssey blocking buffer (1:3 diluted in PBS, plus 0.1% Tween-20). Membranes were washed in PBS-Tween-20 buffer (0.1% Tween-20 in PBS) and incubated with secondary antibodies (goat anti-rabbit IgG, conjugated with IRDye 800CW for H4K20me3 detection and IRDye 680LT for GAPDH; 1:10,000) for 1 h at room temperature. Secondary antibodies were diluted in Odyssey blocking buffer (1:3 diluted in PBS, plus 0.1% Tween-20 and 0.01% SDS). After washing in PBS-Tween-20 buffer, the membranes were rinsed with PBS and fluorescence signals detected with the Odyssey Fc imaging system (2 min acquisition). The quantification of individual H4K20me3 signals on WB was calculated using the total protein stain in individual lanes (Fig. S1J).

Acknowledgements

We are grateful to H.-C. Schuppe, T. Bloch, and K. Wilhelm for excellent support in the acquisition and analysis of semen samples from healthy donors, and B. Fröhlich and A. C. Fröbisch for IF analyses.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: T.D., U.S.; Methodology: N.O., S.G., S.S., C.C., J.H.; Software: T.D., D.C.; Validation: T.D., N.O., S.G., D.C., S.S., C.C., J.H., U.S.; Formal analysis: T.D., N.O., S.G., D.C., C.C., J.H.; Investigation: T.D., N.O., S.G.; Resources: T.D., D.C., C.C., J.H., U.S.; Data curation: T.D., N.O., S.G., D.C., U.S.; Writing - original draft: N.O., T.D., U.S.; Writing - review & editing: T.D., U.S.; Visualization: T.D., N.O., S.G., U.S.; Supervision: U.S.; Project administration: U.S.; Funding acquisition: U.S.

Funding

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft; SCHA1531/2-1).

Data availability

H4K20me3 ChIP-seq data are deposited in the Gene Expression Omnibus under accession number GSE129239.

References

- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arpanahi, A., Brinkworth, M., Iles, D., Krawetz, S. A., Paradowska, A., Platts, A. E., Saida, M., Steger, K., Tedder, P. and Miller, D. (2009). Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res.* **19**, 1338-1349. doi:10.1101/gr.094953.109
- Balhorn, R., Gledhill, B. L. and Wyrobek, A. J. (1977). Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16**, 4074-4080. doi:10.1021/bi00637a021
- Bench, G. S., Friz, A. M., Corzett, M. H., Morse, D. H. and Balhorn, R. (1996). DNA and total protamine masses in individual sperm from fertile mammalian subjects. *Cytometry* **23**, 263-271. doi:10.1002/(SICI)1097-0320(19960401)23:4<263::AID-CYTO1>3.0.CO;2-I
- Benetti, R., Gonzalo, S., Jaco, I., Schotta, G., Klatt, P., Jenuwein, T. and Blasco, M. A. (2007). Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J. Cell Biol.* **178**, 925-936. doi:10.1083/jcb.200703081
- Bryckzynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E. J., Roloff, T. C., Beisel, C., Schübeler, D., Stadler, M. B. and Peters, A. H. F. M. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* **17**, 679-687. doi:10.1038/nsmb.1821
- Carone, B. R., Hung, J.-H., Hainer, S. J., Chou, M.-T., Carone, D. M., Weng, Z., Fazio, T. G. and Rando, O. J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev. Cell* **30**, 11-22. doi:10.1016/j.devcel.2014.05.024

- Cooper, T. G., Noonan, E., von Eckardstein, S., Auger, J., Baker, H. W. G., Behre, H. M., Haugen, T. B., Kruger, T., Wang, C., Mbizvo, M. T. et al. (2010). World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update* **16**, 231-245. doi:10.1093/humupd/dmp048
- Dansranjav, T. and Schagdarsurengin, U. (2016). The rationale of the inevitable, or why is the consideration of repetitive dna elements indispensable in studies of sperm nucleosomes. *Dev. Cell* **37**, 13-14. doi:10.1016/j.devcel.2016.03.020
- Dias, B. G. and Ressler, K. J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* **17**, 89-96. doi:10.1038/nn.3594
- Eid, A., Rodriguez-Terrones, D., Burton, A. and Torres-Padilla, M.-E. (2016). SUV4-20 activity in the preimplantation mouse embryo controls timely replication. *Genes Dev.* **30**, 2513-2526. doi:10.1101/gad.288969.116
- Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M. B. and Peters, A. H. F. M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* **20**, 868-875. doi:10.1038/nsmb.2599
- Gandini, L., Lombardo, F., Paoli, D., Caruso, F., Eleuteri, P., Leter, G., Ciriminna, R., Culasso, F., Dondero, F., Lenzi, A. et al. (2004). Full-term pregnancies achieved with ICSI despite high levels of sperm chromatin damage. *Hum. Reprod.* **19**, 1409-1417. doi:10.1093/humrep/deh233
- Gatewood, J. M., Cook, G. R., Balhorn, R., Bradbury, E. M. and Schmid, C. W. (1987). Sequence-specific packaging of DNA in human sperm chromatin. *Science* **236**, 962-964. doi:10.1126/science.3576213
- Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thévenon, J., Catena, R., Davidson, I., Garin, J., Khochbin, S. et al. (2007). Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J. Cell Biol.* **176**, 283-294. doi:10.1083/jcb.200604141
- Gowri, V., Dion, E., Viswanath, A., Piel, F. M. and Monteiro, A. (2019). Transgenerational inheritance of learned preferences for novel host plant odors in *Bicyclus anynana* butterflies. *Evolution* **73**, 2401-2414. doi:10.1111/evo.13861
- Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T. and Cairns, B. R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478. doi:10.1038/nature08162
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589. doi:10.1016/j.molcel.2010.05.004
- Hisano, M., Erkek, S., Dessus-Babus, S., Ramos, L., Stadler, M. B. and Peters, A. H. F. M. (2013). Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nat. Protoc.* **8**, 2449-2470. doi:10.1038/nprot.2013.145
- Ishibashi, T., Li, A., Eirín-López, J. M., Zhao, M., Missiaen, K., Abbott, D. W., Meistrich, M., Hendzel, M. J. and Ausió, J. (2010). H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res.* **38**, 1780-1789. doi:10.1093/nar/gkp1129
- Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J. and Torres-Padilla, M.-E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502-1510. doi:10.1038/ng.3945
- Jørgensen, S., Schotta, G. and Sørensen, C. S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res.* **41**, 2797-2806. doi:10.1093/nar/gkt012
- Liu, L., Leng, L., Liu, C., Lu, C., Yuan, Y., Wu, L., Gong, F., Zhang, S., Wei, X., Wang, M. et al. (2019). An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat. Commun.* **10**, 364. doi:10.1038/s41467-018-08244-0
- Luense, L. J., Wang, X., Schon, S. B., Weller, A. H., Lin Shiao, E., Bryant, J. M., Bartolomei, M. S., Coutifaris, C., Garcia, B. A. and Berger, S. L. (2016). Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. *Epigenet. Chromatin.* **9**, 24. doi:10.1186/s13072-016-0072-6
- Luense, L. J., Donahue, G., Lin-Shiao, E., Rangel, R., Weller, A. H., Bartolomei, M. S. and Berger, S. L. (2019). Gcn5-mediated histone acetylation governs nucleosome dynamics in spermiogenesis. *Dev. Cell* **51**, 745-758.e6. doi:10.1016/j.devcel.2019.10.024
- Magaraki, A., van der Heijden, G., Sleddens-Linkels, E., Magarakis, L., van Cappellen, W. A., Peters, A. H. F. M., Gribnau, J., Baarends, W. M. and Eijpe, M. (2017). Silencing markers are retained on pericentric heterochromatin during murine primordial germ cell development. *Epigenet. Chromatin.* **10**, 11. doi:10.1186/s13072-017-0119-3
- Mahadevan, I. A., Kumar, S. and Rao, M. R. S. (2020). Linker histone variant H1t is closely associated with repressed repeat-element chromatin domains in pachytene spermatocytes. *Epigenetics Chromatin* **13**, 9. doi:10.1186/s13072-020-00335-x
- Maquat, L. E. (2020). Short interspersed nuclear element (SINE)-mediated post-transcriptional effects on human and mouse gene expression: SINE-UP for active duty. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190344. doi:10.1098/rstb.2019.0344
- Metzler-Guillemain, C., Depetris, D., Luciani, J. J., Mignon-Ravix, C., Mitchell, M. J. and Mattei, M.-G. (2008). In human pachytene spermatocytes, SUMO protein is restricted to the constitutive heterochromatin. *Chromosome Res.* **16**, 761-782. doi:10.1007/s10577-008-1225-7
- Meyer-Ficca, M. L., Lonchar, J. D., Ihara, M., Bader, J. J. and Meyer, R. G. (2013). Alteration of poly(ADP-ribose) metabolism affects murine sperm nuclear architecture by impairing pericentric heterochromatin condensation. *Chromosoma* **122**, 319-335. doi:10.1007/s00412-013-0416-y
- Mishra, L. N., Shalini, V., Gupta, N., Ghosh, K., Suthar, N., Bhaduri, U. and Rao, M. R. S. (2018). Spermatid-specific linker histone HILS1 is a poor condenser of DNA and chromatin and preferentially associates with LINE-1 elements. *Epigenet. Chromatin* **11**, 43. doi:10.1186/s13072-018-0214-0
- Nelson, D. M., Jaber-Hijazi, F., Cole, J. J., Robertson, N. A., Pawlikowski, J. S., Norris, K. T., Criscione, S. W., Pchelintsev, N. A., Piscitello, D., Stong, N. et al. (2016). Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through preservation of genetic and epigenetic stability. *Genome Biol.* **17**, 158. doi:10.1186/s13059-016-1017-x
- Niu, Z.-H., Shi, H.-J., Zhang, H.-Q., Zhang, A.-J., Sun, Y.-J. and Feng, Y. (2011). Sperm chromatin structure assay results after swim-up are related only to embryo quality but not to fertilization and pregnancy rates following IVF. *Asian J. Androl.* **13**, 862-866. doi:10.1038/aja.2011.77
- Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D. and Knowles, B. B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597-606. doi:10.1016/j.devcel.2004.09.004
- Pittoggi, C., Renzi, L., Zaccagnini, G., Cimini, D., Degrassi, F., Giordano, R., Magnano, A. R., Lorenzini, R., Lavia, P. and Spadafora, C. (1999). A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. *J. Cell Sci.* **112**, 3537-3548. doi:10.1242/jcs.112.20.3537
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-W165. doi:10.1093/nar/gkw257
- Rando, O. J. (2016). Intergenerational transfer of epigenetic information in sperm. *Cold Spring Harb. Perspect. Med.* **6**, a022988. doi:10.1101/cshperspect.a022988
- Royo, H., Stadler, M. B. and Peters, A. H. F. M. (2016). Alternative computational analysis shows no evidence for nucleosome enrichment at repetitive sequences in mammalian spermatozoa. *Dev. Cell* **37**, 98-104. doi:10.1016/j.devcel.2016.03.010
- Samans, B., Yang, Y., Krebs, S., Sarode, G. V., Blum, H., Reichenbach, M., Wolf, E., Steger, K., Dansranjav, T. and Schagdarsurengin, U. (2014). Uniformity of nucleosome preservation pattern in mammalian sperm and its connection to repetitive DNA elements. *Dev. Cell* **30**, 23-35. doi:10.1016/j.devcel.2014.05.023
- Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D. and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251-1262. doi:10.1101/gad.300704
- Stark, R. and Brown, G. (2019). DiffBind: Differential binding analysis of ChIP-seq peak data. <https://www.cruk.cam.ac.uk/core-facilities/bioinformatics-core/software/diffbind>.
- Syed, S. H., Boulard, M., Shukla, M. S., Gautier, T., Travers, A., Bednar, J., Faivre-Moskalenko, C., Dimitrov, S. and Angelov, D. (2009). The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome. *Nucleic Acids Res.* **37**, 4684-4695. doi:10.1093/nar/gkp473
- Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G. E., Erkek, S., Kwon, T., Marcotte, E. M., Zegerman, P., Bradshaw, C. R. et al. (2016). Sperm is epigenetically programmed to regulate gene transcription in embryos. *Genome Res.* **26**, 1034-1046. doi:10.1101/gr.201541.115
- Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178-192. doi:10.1093/bib/bbs017
- van de Werken, C., van der Heijden, G. W., Eleveld, C., Teeuwssen, M., Albert, M., Baarends, W. M., Laven, J. S. E., Peters, A. H. F. M. and Baart, E. B. (2014). Paternal heterochromatin formation in human embryos is H3K9/HP1 directed and primed by sperm-derived histone modifications. *Nat. Commun.* **5**, 5868. doi:10.1038/ncomms6868
- van der Heijden, G. W., Derijck, A. A. H. A., Ramos, L., Giele, M., van der Vlag, J. and de Boer, P. (2006). Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Dev. Biol.* **298**, 458-469. doi:10.1016/j.ydbio.2006.06.051
- Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S.-I., Andrews, S. and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16**, 209. doi:10.1186/s13059-015-0769-z
- Wanichnopparat, W., Suwanwongse, K., Pin-On, P., Apornawan, C. and Mutirangura, A. (2013). Genes associated with the cis-regulatory functions of

- intragenic LINE-1 elements. *BMC Genomics* **14**, 205. doi:10.1186/1471-2164-14-205
- Whiddon, J. L., Langford, A. T., Wong, C.-J., Zhong, J. W. and Tapscott, S. J.** (2017). Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935-940. doi:10.1038/ng.3846
- Williams, Z. M.** (2016). Transgenerational influence of sensorimotor training on offspring behavior and its neural basis in *Drosophila*. *Neurobiol. Learn. Mem.* **131**, 166-175. doi:10.1016/j.nlm.2016.03.023
- Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W. et al.** (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657. doi:10.1038/nature18606
- Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S. et al.** (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256-260. doi:10.1038/s41586-018-0080-8
- Xu, J. and Kidder, B. L.** (2018). H4K20me3 co-localizes with activating histone modifications at transcriptionally dynamic regions in embryonic stem cells. *BMC Genomics* **19**, 514. doi:10.1186/s12864-018-4886-4
- Yamaguchi, K., Hada, M., Fukuda, Y., Inoue, E., Makino, Y., Katou, Y., Shirahige, K. and Okada, Y.** (2018). Re-evaluating the localization of sperm-retained histones revealed the modification-dependent accumulation in specific genome regions. *Cell Rep.* **23**, 3920-3932. doi:10.1016/j.celrep.2018.05.094
- Yandim, C. and Karakulah, G.** (2019). Expression dynamics of repetitive DNA in early human embryonic development. *BMC Genomics* **20**, 439. doi:10.1186/s12864-019-5803-1
- Yoshida, K., Muratani, M., Araki, H., Miura, F., Suzuki, T., Dohmae, N., Katou, Y., Shirahige, K., Ito, T. and Ishii, S.** (2018). Mapping of histone-binding sites in histone replacement-completed spermatozoa. *Nat. Commun.* **9**, 3885. doi:10.1038/s41467-018-06243-9
- Yu, G., Wang, L.-G. and He, Q.-Y.** (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383. doi:10.1093/bioinformatics/btv145
- Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K. and Peng, W.** (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958. doi:10.1093/bioinformatics/btp340
- Zeng, T.-B., Han, L., Pierce, N., Pfeifer, G. P. and Szabó, P. E.** (2019). EHMT2 and SETDB1 protect the maternal pronucleus from 5mC oxidation. *Proc. Natl. Acad. Sci. USA* **116**, 10834-10841. doi:10.1073/pnas.1819946116
- Zhang, Q., Thakur, C., Fu, Y., Bi, Z., Wadgaonkar, P., Xu, L., Liu, Z., Liu, W., Wang, J., Kidder, B. L. et al.** (2020). Mdig promotes oncogenic gene expression through antagonizing repressive histone methylation markers. *Theranostics* **10**, 602-614. doi:10.7150/thno.36220
- Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S. and Green, M. R.** (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237. doi:10.1186/1471-2105-11-237