Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa

Serap Erkek[1,2,3], Mizue Hisano[1], Ching-Yeu Liang[1,2], Mark Gill[1], Rabih Murr[1], Jürgen Dieker[4], Dirk Schübeler[1,2], Johan van der Vlag[4], Michael B. Stadler[1,3] and Antoine H.F.M. Peters[1,2]

1. Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland

2. Faculty of Sciences, University of Basel, Basel, Switzerland

3. Swiss Institute of Bioinformatics, Basel, Switzerland.

4. Department of Nephrology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Corresponding author:
Antoine Peters
Phone: +41 61 6978761
Fax: +41 61 6973976
Email: antoine.peters@fmi.ch

Keywords:
Nucleosome, spermatozoa, histone variants, paternal epigenetic inheritance, transgenerational inheritance

In mammalian spermatozoa, most but not all of the genome is densely packaged by protamines. Here we reveal the molecular logic underlying retention of nucleosomes in mouse spermatozoa that only contain one percent residual histones. Throughout the genome we observe high enrichment of nucleosomes at CpG-rich sequences that lack DNA methylation. Residing nucleosomes are largely composed of the H3.3 histone variant, and are trimethylated at lysine 4 of H3 (H3K4me3). Canonical H3.1 and H3.2 histones are also enriched at CpG-rich promoters marked by Polycomb-mediated H3K27me3, which is strongly predictive for gene repression in pre-implantation embryos. Histone variant specific nucleosome retention in sperm strongly associates with the level of nucleosome turnover in round spermatids. Our data shows evolutionary conservation of the basic principles of nucleosome retention in mouse and human sperm, supporting a model of epigenetic inheritance by nucleosomes between generations.

In mammals, fusion of two morphologically distinct gametes, oocytes and spermatozoa, leads to the formation of totipotent embryos. Acquisition of totipotency is thought to be mediated by extensive epigenetic reprogramming of parental genomes, affecting DNA methylation and histone modifications, and possibly replication timing and transcriptional activity in parental specific manners[1-4]. It is currently unclear to what extent differential reprogramming of maternal and paternal genomes is due to differences in chromatin states inherited from the oocyte and spermatozoon[4-11]. Beyond DNA methylation[1,2,6,12], it is unknown which types of parental chromatin states are maintained or reprogrammed in early embryos. If certain parental chromatin states did escape reprogramming in the early embryo, such information could constitute an "intrinsic intergenerational epigenetic program directing gene expression in the next generation[13]. If these chromatin states also escaped reprogramming during gametogenesis, the inheritance program would function transgenerationally[13]. An increasing number of studies point to inter- or transgenerational transmission of acquired phenotypic traits that are related to temporal exposure of (grand-)parents to alternative instructive environmental cues[14-18]. Mechanistically, such phenotypic changes may be related to (transient) alterations of an intrinsic inheritance program.

A role of histones and associated posttranslational modifications in maternal and paternal transmission of intrinsic or acquired epigenetic information is largely unknown[13]. In many metazoans, male germ cells undergo during their final differentiation into sperm an extensive chromatin remodeling process during which

genomic DNA becomes newly packaged into a highly condensed configuration by sperm specific proteins. In mammals, removal of histones is generally not complete[10,11,19-24]. Furthermore, remaining histones have been reported to stay associated with the paternal genome during *de novo* chromatin formation in the zygote following fertilization[9].

We and others recently showed that histones lasting in human sperm are to some extent enriched at regulatory sequences of genes[10,11]. We also demonstrated that H3K4me3- and H3K27me3-marked histones are retained at promoters of specific sets of genes in mouse spermatozoa[11]. The extent of evolutionary conservation of nucleosome retention at gene regulatory sequences in spermatozoa and the mechanistic principles of such retention are, however, unknown.

To address conservation and to dissect the molecular logic underlying nucleosome retention, we determined the genome-wide nucleosome occupancy in mouse spermatozoa that only contain 1% residual histones. We show here that combinatorial effects of sequence composition, histone variants and histone modifications uniquely determine the packaging of sperm DNA. Nucleosomes in sperm mainly localize to unmethylated CpG-rich sequences in a histone variant specific manner and are differentially modified. Comparison of histone variant profiles between post-meiotic round spermatids and sperm argues that retention of variant specific nucleosomes in sperm are linked to levels of nucleosome turnover in haploid round spermatids.

RESULTS

Nucleosomes localize at GC-rich sequences in mouse sperm

To assess the potential of paternal epigenetic inheritance by nucleosomes in mouse we first aimed to determine the location of nucleosomes in spermatozoa. We isolated motile spermatozoa from caudal epididymi and performed deep-sequencing of DNA associated with mono-nucleosomes prepared by micrococcal nuclease (MNase) digestion of sperm chromatin. Genome-wide analyses indicated an approximately 10- and 2-fold overrepresentation of nucleosomes at promoter regions and exons respectively, while nucleosomes were underrepresented at introns and repeat regions (Supplementary Fig. 1a, 1b). We observed promoter enrichment at many but not all genes (Fig. 1a). Classification of promoters according to their GC content, CpG ratio and length of CpG-rich region[25] revealed that high-CpG (HCP) and intermediate-CpG (ICP) promoters are highly and moderately enriched in nucleosomes respectively while most promoters with low CpG content (LCP) lack nucleosomes (Fig. 1b). Nucleosomal enrichment is not restricted to CGI-promoters

but is also detected at intra- and intergenic CGIs as well as within GC-rich simple repeat sequences (Fig. 1c; data not shown).

To investigate whether nucleosomal occupancy in sperm correlates with a specific sequence composition, we determined single nucleotide frequencies in 1kb windows tiled throughout the genome. While guanine and cytosine correlate positively with nucleosome occupancy genome-wide, adenine and thymine do not (Fig. 1d). We next assessed the contribution of different dinucleotides to nucleosome occupancy, independent of single nucleotide frequencies, by calculating the ratio of "observed over expected" frequencies for each dinucleotide. Remarkably, these analyses revealed that predominantly the CpG dinucleotide contributes to sequence-related nucleosomal packaging of sperm DNA (Fig. 1d), while the GpC dinucleotide has almost no contribution. The ApA and TpT dinucleotides contribute moderately.

To establish whether the observed CpG dinucleotide association reflects an intrinsic DNA sequence preference for nucleosome formation, we reanalyzed *in vitro* nucleosome reconstitution data of histone octamers assembled onto yeast genomic DNA[26]. Similar to Tillo and Hughes[27], we observed a strong contribution of guanine and cytosine to *in vitro* nucleosome formation, yet no specific contributions of either CpG nor GpC dinucleotides (Supplementary Fig. 2a, 2b). Thus, the strong association of CpG density to nucleosome retention in mouse sperm does not reflect an intrinsic preference of nucleosomes to CpG-rich DNA. Instead, it represents a novel feature of CpG islands (CGIs) in genome function executed during mouse male germ cell development[28]. Motif analysis did not reveal any specific sequence compositions, other than a strong correlation to GC composition (Supplementary Fig. 1c).

Nucleosomes localize at unmethylated CpG-rich DNA in sperm

The nucleosomal occupancy at CGIs in sperm strongly contrasts with the depletion of nucleosomes at CGI-promoters in somatic cells[29,30,31]. Indeed, we observed extensive nucleosomal depletion around TSS and a clear anti-correlation between nucleosome occupancy and CpG frequency in mouse liver[32] (Fig. 1e, 1f). In somatic cells, however, nucleosomes are not depleted at CGI-promoters repressed by Polycomb Group (PcG) proteins or by DNA methylation[33]. Therefore, to investigate whether nucleosomes are preferentially retained at CGIs that are DNA-methylated in sperm, we performed bisulfite conversion and high throughput sequencing of sperm DNA associated with nucleosomes[34]. In contrast to our expectation, methylated genomic regions are devoid of nucleosomes in sperm (Fig. 2a). We observed a similar inverse relationship using genome-wide shotgun bisulfite sequencing data from mouse

sperm (Fig. 2b)[6]. This exclusive inverse relationship is nicely illustrated at imprinting control regions (ICRs) in mouse sperm. While paternal ICRs regulating somatic expression of genes such as *H19*, *Dlk1*, *Gtl2* and *Rasgrf1*[35] are methylated and devoid of nucleosomes, ICRs controlling maternally imprinted gene clusters (e.g. of *Kcnq1ot1*, *Snrpn,* and *Peg10*) are unmethylated and contain nucleosomes (Supplementary Fig 3). Furthermore, GC-rich retro-elements like LINE1 elements that are methylated in sperm and become demethylated after fertilization[1] lack nucleosomes in sperm (data not shown). These data are compatible with a model in which DNA methylation established early during male germ cell development[36] prevents nucleosome retention during spermiogenesis.

By combining sequence characteristics of CGIs and their DNA methylation states, we found strong positive correlations between nucleosomal enrichment and the number and density of CpG dinucleotides within CGIs devoid of DNA methylation (Fig. 2c). Using a linear mathematical model, we can predict nucleosome occupancy in mouse sperm as a function of CpG dinucleotide frequency and DNA methylation level (Fig. 2d).

We and others previously showed that retained histones are not randomly distributed in human sperm, but are to some extent enriched at GC-rich regulatory elements of genes[10,11,37]. As for mouse, we observed an inverse relationship between the degree of nucleosomal occupancy and the level of DNA methylation in human sperm[38] (Supplementary Figure 2c - 2f). Thus, these analyses demonstrate that nucleosome retention at unmethylated CGIs is conserved between mouse and human spermatozoa.

Histone H3 variant specific occupancy at CGIs in mouse sperm

The unique nucleosomal organization in sperm, highly distinct from that in somatic cells[29-31] , emphasizes extensive chromatin remodeling processes occurring during the formation of spermatozoa. Given the important roles of histone variants in transcription, cellular differentiation, reproduction and development[39-41] we asked whether canonical H3.1 and H3.2 and variant H3.3 histones may serve specific functions in nucleosome eviction versus retention during spermiogenesis. We performed Western blot analysis with antibodies specific for either H3.3 (Supplementary Fig. 4a) or H3.1 and H3.2 (referred to as H3.1/H3.2 since the antibody recognizes an epitope shared by H3.1 and H3.2)[42]. Compared to proliferating embryonic stem cells (ESCs) and even to quiescent aging neurons[43], H3.3 is incorporated into chromatin of round spermatids and sperm to very high levels relative to H3.1/H3.2, suggesting an extensive and rapid replacement of

canonical histones by the H3.3 variant, presumably upon entry into meiotic prophase and/or during spermatid differentiation (Fig. 3a). In sperm, H3.3 ChIP-sequencing profiles are highly similar to nucleosomal profiles whereas H3.1/H3.2 profiles are not (Fig. 3b). Consistently, H3.3 enrichments are well predicted by the linear model, suggesting a CpG density-linked retention mechanism for H3.3 containing nucleosomes (Fig. 3c). Regions containing H3.1/H3.2 histones are in contrast systematically underestimated by the model, suggesting that retention of canonical and H3.3 variant histones may be differentially regulated.

Nucleosome turnover in round spermatids

To understand the timing and mechanisms of chromatin remodeling, we profiled the occupancy of H3.3 and H3.1/H3.2 nucleosomes and measured levels of mRNA transcripts by ChIP- and RNA-sequencing in round spermatids. In contrast to sperm, we observed a widespread reduction in H3.1/H3.2-nucleosomal occupancy around transcriptional start sites (TSS) of genes in round spermatids (Fig. 4a). We next classified gene promoters according to CpG density and RNA transcript levels of associated genes (Fig. 4b). For expressed genes, we observed eviction of H3.1/H3.2 nucleosomes around TSS of CpG-rich (≥3% CpG) and CpG-poor (<3% CpG) genes correlating well with mRNA levels of associated genes. For medium to highly expressed CpG-rich genes, we also observed clear positioning of remaining nucleosomes around TSS. These data argue for a transcription-coupled eviction of canonical histones.

For non-expressed genes we measured low levels of eviction of canonical histones at CpG-rich TSS regions but not at CpG-poor TSS regions (Fig. 4b). This finding is consistent with studies reporting nucleosome depletion around silent CGI promoters in somatic cells[30-32]. Depletion of H3.1/H3.2 nucleosomes around TSS in spermatids is more pronounced than that of H3.2-HA tagged nucleosomes in ESCs (Fig. 4b; Supplementary Fig. 4b)[44]. Possibly, this is due to progressive loss of canonical histones during transcription in post-replicative germ cells.

For H3.3 nucleosomes, we also measured some depletion around TSS that was more pronounced downstream of TSS at medium and highly expressed genes (Fig. 4a, 4b). Comparison of H3.3 to H3.1/H3.2 occupancy levels suggests extensive transcription-coupled eviction of canonical histones and subsequent replacement by H3.3 nucleosomes in round spermatids. We interpret the ratio between H3.3 over H3.1/H3.2 as a surrogate measure for nucleosome turnover.

Control of H3 variant specific occupancy at CGIs in sperm

To understand the relationship between histone variant specific nucleosome turnover in round spermatids and retention in sperm, we compared the level of occupancy for both variants in both cell types, at regions with nucleosomal enrichments in spermatozoa. We observed that regions that are strongly and intermediately enriched for H3.3-containing nucleosomes in sperm are actually depleted of such nucleosomes in round spermatids, suggesting dynamic redistribution *in cis* or *de novo* incorporation of H3.3 nucleosomes later during spermatid differentiation, e.g. in late round or elongating spermatids (Fig. 5a). In contrast, H3.1/H3.2 nucleosomes are predominantly detected at weak nucleosomal peak regions in spermatozoa. Furthermore, such local H3.1/H3.2 enrichments in sperm highly resemble the ones in round spermatids suggesting that H3.1/H3.2 nucleosomes retained in sperm largely reflect reduced turnover of canonical H3.1/H3.2 histones inspermatids (Fig. 5a).

We next assessed the connection between CpG density (Fig. 3) and nucleosome turnover in spermatids (Fig. 4) in relation to histone variant specific nucleosome retention at promoter regions of genes in sperm (Fig. 5b). CGI promoters (with ≥3% CpG) that undergo intermediate to high levels of nucleosomal turnover in round spermatids harbor high levels of H3.3 in sperm. In contrast, non-CGI promoters (<3% CpG) are subjected to low to intermediate levels of nucleosome turnover and are relatively enriched for H3.1/H3.2. Finally, a group of CGI promoters is enriched for both H3.1/H3.2 and H3.3 (Fig. 5b). These promoters are generally characterized by intermediate turnover levels in spermatids. Together, these data show that CpG density and the extent of turnover in spermatids strongly relate to the identity of histones retained in sperm.

H3K27me3 associates with H3.1/H3.2 retention in sperm

To study whether histone modification states may affect nucleosome dynamics during spermiogenesis, we performed ChIP-sequencing for H3K4me3 and H3K27me3, two modifications that are associated with CGIs in somatic cells. We measured comparable enrichments around TSS for both modifications in round spermatids and sperm (Supplementary Fig. 5a) indicating propagation of the modification state during spermiogenesis. CGI promoters (with ≥3% CpG) containing H3.3 nucleosomes are generally marked by H3K4me3 in sperm (Fig. 6a). A fraction of CGIs with intermediate H3K4me3 levels are strongly positive for H3K27me3, indicating the presence of bivalent promoters in sperm (Fig. 6b; cluster 4 in Supplementary Fig. 5b). Importantly, such H3K4me3/H3K27me3 double marked CGI promoters also show enrichment for H3.1/H3.2 histones in sperm (Fig. 6a, 6b). These data suggest the presence of bivalent promoters in sperm that contain a

mixture of H3.3 and H3.1/H3.2 nucleosomes. The data further suggest that Polycomb proteins and/or PRC2-mediated H3K27me3 suppress, at least in part, the default eviction of H3.1/H3.2 histones at CGIs in round spermatids and consequently promote the retention of pre-existing canonical histones during chromatin remodeling in elongating spermatids. In accordance, CGI promoters with low-intermediate levels of nucleosome turnover in spermatids are H3K27me3 positive in spermatids and in sperm (Supplementary Fig. 5d).

Determinants of nucleosome retention in sperm

On the basis of sequence composition and occupancy levels of nucleosomes, histone variants and histone modifications at gene promoters as well as expression states, we can classify genes into five different clusters (Fig. 6c) that correlate well with different gene functions in cellular homeostasis (clusters 2 and 3), germ cell and embryonic development (clusters 1 and 4 respectively), and stimulus perception and host defense (cluster 5) (Supplementary Table 1; Supplementary Fig. 5b, 6a - 6j).

To quantify the extent by which expression and the different chromatin characteristics such as histone variants and modifications measured in round spermatids, as well as CpG density contribute to nucleosome occupancy in sperm, we performed a variance partitioning analysis for promoter regions (Fig. 6d). Combining all of these variables, a total of 79.4% and 70% of the variance in H3.3 and H3.1/H3.2 occupancies in sperm can be explained.

For H3.3 occupancy in sperm, CpG density of promoters has, as expected, the highest unique contribution while H3.3 occupancy in spermatids has a small unique contribution (Fig. 6d; clusters 1-4 in Fig. 6c). Notably, 78% of the variance is explained when CpG density and H3.3 occupancy in spermatids are used together as the only variables in the partitioning analysis. These data suggest that extensive H3.1/H3.2 turnover and ensuing H3.3 deposition at CGIs in round spermatids contribute to H3.3 retention at such promoters in sperm.

In contrast, H3.1/H3.2 enrichments in sperm mostly relate to H3.1/H3.2 enrichments in round spermatids (Fig. 6d; cluster 5 in Fig. 6c). Moreover, CGI promoters marked by H3K27me3 in round spermatids preferentially retain H3.1/H3.2 in sperm (Fig. 6d; cluster 4 in Fig. 6c). When taking H3.1/H3.2 and H3K27me3 enrichments in round spermatids as the only variables, 68% of the total of 70% variance is explained. These quantifications argue that low levels of nucleosome turnover at H3K27me3-marked promoters in spermatids substantially contributes to H3.1/H3.2 retention in sperm.

When performing the variance partitioning analysis genome-wide at 1kb windows that do not intersect with TSS regions and any other CGIs we observed that only 30% to 42% of variance for H3.3 and H3.1/H3.2 respectively is explained by CpG density and chromatin characteristics measured in round spermatids (Supplementary Fig. 5e). Nonetheless, occupancy levels of histone variants in sperm relate well to the occupancy of the corresponding variants in spermatids. This relationship supports a model of nucleosome retention without major remodeling *in cis* during spermatid maturation and protamine incorporation.

H3K27me3 associates with gene repression in early embryos

To assess the potential of nucleosomes and associated modifications retained in sperm for regulating transcription in the next generation, we analyzed the expression of genes belonging to the five different clusters shown in Fig. 6c in oocytes and in pre-implantation embryos[45,11]. We observe that housekeeping genes in cluster 2 are significantly more likely to be *de novo* transcribed (48.7% vs. 39.6%; Fig. 7a) in early embryos as well as being expressed in oocytes (48.8% vs. 40.4%; Fig. 7b) than genes belonging to cluster 1 enriched for germ line functions. Intriguingly, H3K4me3 is more enriched around TSS of cluster 1 genes than cluster 2 genes in sperm (Fig. 6c; Supplementary Fig. 5b). Analogously, genes of clusters 3 and 4 have similar H3K4me3 enrichments in sperm (Supplementary Fig. 5b), yet display significantly different expression states in early embryos (Fig. 7a, b). These data argue for a rather limited potential, if any, of H3K4 tri-methylated nucleosomes in sperm to predetermine transcription in early embryos (Fig. 7a). This may relate to the prevalent H3K4 tri-methylation at CGIs in spermatids, ESCs and during somatic differentiation that is independent of their transcriptional status[28] (Supplementary Fig. 5c).

In contrast, only ~16% of CGI promoters marked by H3K27me3 (and H3K4me3) in sperm (cluster 4) are expressed in pre-implantation embryos (Fig. 7a). Moreover, many Polycomb target genes in sperm are similarly modified by H3K27me3 in ESCs (Supplementary Fig. 5c). These data support a model of H3K27me3 mediating epigenetic inheritance of transcriptional repression between generations.

DISCUSSION

The role of histones and associated posttranslational modifications in maternal and paternal transmission of epigenetic information is currently unknown. Here we describe a systematic genome-wide characterization of chromatin states in mouse spermatids and spermatozoa. We show that the one percent of histones retained in

sperm of mice[11] are strongly enriched at CGIs that are not methylated at the underlying sequence. Likewise, we demonstrate that in human sperm the CGIs to which the 10-15 percent of residual histones are somewhat enriched[10,11,37] are also unmethylated. Since CGIs are frequently associated with gene promoter function, the evolutionary conserved presence of modified nucleosomes at unmethylated CGIs in sperm of mammals suggest a central role for CGIs and retained nucleosomes in paternal intrinsic epigenetic inheritance between generations[13].

Our data show that CGIs in mouse sperm generally contain the variant H3.3 protein (cluster 1-4 in Fig. 6c) while canonical H3.1/H3.2 proteins are only present at some CGIs (cluster 4 in Fig. 6c). Comparative analysis of chromatin states in round spermatids and sperm strongly suggests that the level of nucleosome turnover in round spermatids determines the type of H3 histone retained at CGIs in sperm.

In round spermatids, we measured an extensive eviction of canonical nucleosomes around TSS of genes and replacement by H3.3 containing nucleosomes. The extent of nucleosome turnover positively correlates with the level of transcriptional activity of associated gene promoters, as in somatic cells[32,44]. CGIs in round spermatids also show transcription-independent nucleosome turnover, as observed in somatic cells[30-32], possibly reflecting dynamic competition between nucleosomes and transcription factors for CGI binding. Notably, the overall extent of H3.1/H3.2 to H3.3 replacement around TSS is more pronounced in post-mitotic round spermatids than e.g. in replicating ESCs. These data argue that male germ cells undergo extensive remodeling of their chromatin during the approximately two weeks following their entry into meiosis and subsequent differentiation as haploid spermatids.

In sperm, H3.3 is enriched at most CGIs, as a reflection of turnover in spermatids. In contrast, H3.1/H3.2 is only present at CGIs with low-intermediate nucleosome turnover in spermatids and that are marked by H3K27me3 in round spermatids and sperm. These findings suggest that PRC2 proteins directly or indirectly via H3K27me3 inhibit nucleosome turnover in round spermatids, thereby promoting H3.1/H3.2 retention in sperm. At non-CGI promoter genes (cluster 5 in Fig. 6c), we observed only minor enrichment of H3.1/H3.2 nucleosomes around TSS of some genes, supporting the notion of poor nucleosome retention, if any, at non-CGI promoters.

Currently, the mechanisms driving nucleosome retention versus eviction during spermiogenesis are unknown. Our findings support a model in which H3.3 nucleosomes present at CGIs in sperm become stably incorporated into chromatin and marked by H3K4me3 in late round spermatids in response to a global cessation

of histone turnover and transcription (Fig. 7c). Reduced nucleosome turnover, as observed at H3K27me3-marked CGIs in spermatids, would promote retention of canonical H3.1/H3.2 in sperm. This model entails that CpG-rich DNA would somehow resist loading of transition proteins and protamines in elongating spermatids, thereby enabling nucleosome retention at CGIs as measured in sperm. Resistance to loading could be mediated by CGI-binding proteins binding to unmethylated DNA and protecting nucleosomes locally from eviction. Alternatively, it could reflect a reduced intrinsic affinity of protamines for CG-rich DNA. A variation on this model is that transcription/chromatin factors and H3.3 nucleosomes would continue to compete for binding to CGIs during the histone-to-protamine exchange process in elongating spermatids. This dynamic process may block protamine incorporation.

In comparison to mouse sperm, ~10-fold more nucleosomes are retained in human sperm. While using the same chromatin preparation and high-throughput sequencing procedures, we observed a ~2.5-fold lower contribution of CpG dinucleotides to nucleosome occupancy in human versus mouse spermatozoa (0.28 versus 0.71 Pearson correlation coefficient as shown in Supplementary Fig. 2c[11]; Fig. 1d). These data may therefore suggest that the eviction of nucleosomes at CpG-poor regions in the genome is less efficient during spermiogenesis in human than in mouse.

Bisulfite sequencing analysis of genomic DNA of mouse and human sperm revealed an inverse correlation between nucleosome occupancy and DNA methylation. These data are compatible with a model in which DNA methylation established early during male germ cell development[36], e.g. at paternal ICRs (Supplementary Fig. 3), prevents directly or indirectly nucleosome retention during spermiogenesis (Fig. 7c). Such a mechanism would preclude transmission of chromatin states associated with methylated DNA in immature male germ cells The differential reprogramming of DNA methylation in zygotes that were generated by micro-insemination of round spermatids versus mature spermatozoa[7] may thus indicate the presence of specific chromatin states with methylated DNA e.g. at repetitive sequences in round spermatids that are lost in sperm.

Recently, Nakamura and colleagues reported enrichment of H3K9 dimethylation at the ICRs of *H19* and *Rasgrf1*[5]. Though we observed minor enrichment for H3K27me3 at these regions, we failed to detect any noteworthy nucleosomal occupancy (compare enrichments in Supplementary Figs 3, 6). In contrast, unmethylated maternal ICRs contain nucleosomes marked by H3K4me3 and/or H3K27me3 (see also[46]). More generally, we observed higher enrichments for

modified histones than for core histones as well as extensive enrichments for histone modifications adjacent to relatively narrow peaks of nucleosomes (Supplementary Figs 3, 6). Technically, differential enrichment is likely due to higher sensitivity of antibodies for modified histones and overall lower abundance of modified histones. Biologically, enrichment for histone modifications in absence of nucleosomes may reflect retention of nucleosomes in only a low percentage of spermatozoa, possibly leading to variegated paternal transmission. These findings warrant caution to the interpretation of enrichment values for modified histones in sperm in cases where occupancy data of corresponding nucleosomes is absent.

Our study demonstrates that largely the same genes are marked by H3K27me3 in round spermatids and in sperm. In germinal vesicle (GV) oocytes deficient for *Ring1* and *Rnf2*, two key components of the PRC1 complex, 62% of the up-regulated genes are marked by H3K27me3 in mouse sperm while only 35% of the unaffected genes are PRC2 targets in sperm[47]. Notably, about 85% of PRC2 targets in sperm remain repressed during pre-implantation development. Correspondingly, ESCs only contain a slightly reduced number of PRC2 targets (Supplementary Fig. 5c). Promoters of several pluripotency factors like *Oct3/4*, *Sox*, *Esrrb*, and *Klf5* contain H3K27me3-marked nucleosomes in sperm while *Nanog* is DNA methylated and essentially devoid of nucleosomes. *Klf4* is robustly labeled with H3K4me3-marked nucleosomes and weakly with H3K27me3 (Supplementary Fig. 6k - 6p). Interestingly, repression of *Sox2* and *Klf4* in GV oocytes is dependent on *Ring1*/*Rnf2* function[47]. Together, these data suggest that Polycomb may mediate gene repression in the male as in the female germ line. While principally hypothesizing paternal transmission of modified nucleosomes, epigenetic reprogramming of some H3K27me3-marked genes such as certain pluripotency factors would be required to occur in early pre-implantation embryos to support their development. The majority of Polycomb targets, however, remain repressed in early embryos and would not need to be reprogrammed, consistent with a model of intergenerational or possibly transgenerational inheritance of an intrinsic epigenetic memory program.

Accession codes

NCBI Gene Expression Omnibus: Data have been deposited with accession code GSE42629.

Acknowledgements

We gratefully thank Sophie Dessus-Babus and Tim Roloff (FMI functional genomics group), Ina Nissen (Laboratory for Quantitative Genomics, D-BSSE, Basel), Lukas

Burger (FMI bioinformatics group), Hubertus Kohler (FMI FACS facility) and the FMI animal facility for excellent assistance. We thank members of the Peters laboratory for fruitful discussions. S.E. is supported as a recipient of a Boehringer Ingelheim Fond fellowship. M.G. and R.M. are supported by EMBO Long Term fellowships (ALTF 253-2011; ALTF 600-2008). Research in the Peters and Schübeler labs is supported by the Novartis Research Foundation and the Swiss initiative in Systems Biology (Cell Plasticity - Systems Biology of Cell Differentation). The Peters lab further acknowledges support from the Swiss National Science Foundation (31003A_125386 and National Research Programme NRP63 - Stem Cells and Regenerative Medicine), the Japanese Swiss Science and Technology Cooperation Program, the FP7 Marie Curie Initial Training Network "Nucleosome4D" and the EMBO Young Investigator Program.

Author contributions

S.E., M.H. and A.H.F.M.P. conceived and designed the experiments. S.E., M.H., C.-Y.L., M.G. performed experiments. J.D. and J.v.d.V. provided antibodies. R.M. and D.S. performed and supervised bisulfite sequencing experiments, respectively. M.B.S. provided bioinformatics training and support. S.E., M.H., C.-Y.L., M.G., M.B.S. and A.H.F.M.P. analyzed the data. S.E. and A.H.F.M.P. prepared the manuscript.

References

1. Smith, Z.D. et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-44 (2012).
2. Gu, T.P. et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-10 (2011).
3. Aoki, F., Worrad, D.M. & Schultz, R.M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Developmental biology* **181**, 296-307 (1997).
4. Puschendorf, M. et al. PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40**, 411-20 (2008).
5. Nakamura, T. et al. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415-9 (2012).
6. Kobayashi, H. et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**, e1002440 (2012).
7. Kishigami, S. et al. Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. *Developmental biology* **289**, 195-205 (2006).
8. Bui, H.T. et al. Essential role of paternal chromatin in the regulation of transcriptional activity during mouse preimplantation development. *Reproduction* **141**, 67-77 (2011).
9. van der Heijden, G.W. et al. Sperm-derived histones contribute to zygotic chromatin in humans. *BMC developmental biology* **8**, 34 (2008).

10. Hammoud, S.S. et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-8 (2009).
11. Brykczynska, U. et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* **17**, 679-87 (2010).
12. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501-2 (2000).
13. Gill, M.E., Erkek, S. & Peters, A.H. Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Current opinion in cell biology* **24**, 387-96 (2012).
14. Heijmans, B.T. et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17046-9 (2008).
15. Kaminen-Ahola, N. et al. Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model. *PLoS genetics* **6**, e1000811 (2010).
16. Anway, M.D., Memon, M.A., Uzumcu, M. & Skinner, M.K. Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *Journal of andrology* **27**, 868-79 (2006).
17. Carone, B.R. et al. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-96 (2010).
18. Zeybel, M. et al. Multigenerational epigenetic adaptation of the hepatic wound-healing response. *Nature medicine* **18**, 1369-77 (2012).
19. Balhorn, R., Gledhill, B.L. & Wyrobek, A.J. Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16**, 4074-80 (1977).
20. Gatewood, J.M., Cook, G.R., Balhorn, R., Bradbury, E.M. & Schmid, C.W. Sequence-specific packaging of DNA in human sperm chromatin. *Science* **236**, 962-4 (1987).
21. Gardiner-Garden, M., Ballesteros, M., Gordon, M. & Tam, P.P. Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. *Molecular and cellular biology* **18**, 3350-6 (1998).
22. Wykes, S.M. & Krawetz, S.A. The structural organization of sperm chromatin. *The Journal of biological chemistry* **278**, 29471-7 (2003).
23. Pittoggi, C. et al. A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. *Journal of cell science* **112 (Pt 20)**, 3537-48 (1999).
24. Arpanahi, A. et al. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* **19**, 1338-49 (2009).
25. Mohn, F. et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-66 (2008).
26. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-6 (2009).
27. Tillo, D. & Hughes, T.R. G+C content dominates intrinsic nucleosome occupancy. *BMC bioinformatics* **10**, 442 (2009).
28. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-22 (2011).
29. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-98 (2008).

30. Ramirez-Carrozzi, V.R. et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-28 (2009).
31. Fenouil, R. et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome research* (2012).
32. Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K.H. The nucleosome map of the mammalian liver. *Nature structural & molecular biology* **18**, 742-6 (2011).
33. Kelly, T.K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research* (2012).
34. Stadler, M.B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-5 (2011).
35. Kacem, S. & Feil, R. Chromatin mechanisms in genomic imprinting. *Mammalian genome : official journal of the International Mammalian Genome Society* **20**, 544-56 (2009).
36. Smallwood, S.A. & Kelsey, G. De novo DNA methylation: a germ cell perspective. *Trends in genetics : TIG* **28**, 33-42 (2012).
37. Vavouri, T. & Lehner, B. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS genetics* **7**, e1002036 (2011).
38. Molaro, A. et al. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029-41 (2011).
39. Elsaesser, S.J., Goldberg, A.D. & Allis, C.D. New functions for an old variant: no substitute for histone H3.3. *Current opinion in genetics & development* **20**, 110-7 (2010).
40. Orsi, G.A., Couble, P. & Loppin, B. Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *The International journal of developmental biology* **53**, 231-43 (2009).
41. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell research* **21**, 421-34 (2011).
42. van der Heijden, G.W. et al. Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. *Nat Genet* **39**, 251-8 (2007).
43. Pina, B. & Suau, P. Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. *Developmental biology* **123**, 51-8 (1987).
44. Goldberg, A.D. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-91 (2010).
45. Zeng, F. & Schultz, R.M. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Developmental biology* **283**, 40-57 (2005).
46. Delaval, K. et al. Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *The EMBO journal* **26**, 720-9 (2007).
47. Posfai, E. et al. Polycomb function during oogenesis is required for mouse embryonic development. *Genes & development* **26**, 920-32 (2012).
48. Hackenberg, M. et al. CpGcluster: a distance-based algorithm for CpG-island detection. *BMC bioinformatics* **7**, 446 (2006).

Figure legends

Figure 1: Nucleosome occupancy in sperm is highly dependent on CpG composition.

(**a**) Nucleosome occupancy and GC percentage at representative CpG and non-CpG island loci in mouse sperm. (**b**) Density plot showing the distribution of nucleosome enrichment ± 1kb around transcriptional start sites (TSS) of genes classified according to GC composition of their promoters: high, intermediate, and low GC content (HCP, ICP, LCP). (**c**) Nucleosome occupancy and GC percentage at an intergenic region in sperm. (**d**, **e**) Correlation of single nucleotide frequencies (left) and dinucleotide frequencies normalized for single nucleotide composition (right) with nucleosome enrichment in sperm (**d**) and in mouse liver[32] (**e**) in 1kb regions tiling the mouse genome. (**f**) Average profiles for nucleosome occupancy in mouse sperm and liver[32] ± 3kb around TSS of genes.

Figure 2: Nucleosome occupancy correlates negatively with DNA methylation in sperm.

(**a**) Box-plot showing distributions of nucleosome enrichments in 1 kb regions of different DNA methylation states (genome-wide) (with the central bar marking the median, lower and upper limits of the box marking the 25th and 75th percentiles, and the whiskers extending 1.5 times the interquartile range from the 25th and 75th percentiles). (**b**) Scatter plot showing the correlation of nucleosome occupancy with average DNA methylation according to Kobayashi and coworkers[6] in 1kb windows genome-wide. (**c**) Panels show the relationship between number of CpGs in CGI and width of CGI as a function of nucleosome enrichment in sperm. CGIs[48] were grouped into 4 classes according to their DNA methylation status in sperm[6]. (**d**) Correlation of observed to predicted nucleosome enrichment that was calculated by a linear model integrating CpG dinucleotide frequency and DNA methylation status in 1 kb windows (R=0.789).

Figure 3: Histone variant specific packaging of sperm DNA.

(**a**) Western blots showing relative levels of chromatin bound H3.1/ H3.2, H3.3, and total H3 in embryonic stem cells (ESC), round spermatids (RS) and sperm. (**b**) Occupancy of nucleosomes, H3.3 and H3.1/H3.2 histones and GC percentage at the *Fgf9* locus in sperm. (**c**) Scatter plots showing the correlation between observed and predicted nucleosome occupancies (1kb windows) in relation to relative enrichment of H3.3 (left) and H3.1/H3.2 (right) in sperm.

Figure 4: CpG density and gene expression associate with nucleosome eviction in round spermatids.

(**a**) Average profiles of H3.3 and H3.1/H3.2 enrichments ± 3kb around transcriptional start sites (TSS) and transcriptional end sites (TES) in sperm and round spermatids. (**b**) Average profiles of H3.1/H3.2 and H3.3 enrichments around TSS and TES in round spermatids. Genes were classified according to expression status in round spermatids and the percentage of CpGs within ± 1kb windows around TSS of genes (left: CpG % < 3, right: CpG % ≥ 3).

Figure 5: Extent of nucleosome turnover in round spermatids relates to histone variant specific retention in sperm.

(**a**) Scatter plots showing the correlation between H3.3 (left) and H3.1/H3.2 (right) enrichments in sperm versus round spermatids (RS) at genomic regions enriched for nucleosomes in sperm. Enriched regions are classified as "weak", "intermediate" and "strong" according to their relative occupancy by nucleosomes in sperm. (**b**) Scatter plots showing the correlation between percentage of CpGs at TSS (± 1kb) and nucleosome turnover (H3.3 over H3.1/H3.2 ratio) in RS in relation to relative enrichment of H3.3 (left), and H3.1/H3.2 (right) in sperm.

Figure 6: Combinatorial effects of CpG density, histone variants and histone modifications uniquely package sperm DNA.

(**a**, **b**) Scatter plots showing the correlation of the percentage of CpGs with enrichment of variant and canonical H3 histones (TSS; ± 1kb) in sperm in comparison to the enrichment of H3K4me3 (**a**) and H3K27me3 (**b**). (**c**) Heatmap of genes illustrating expression status in RS, CpG density, nucleosome coverage in sperm, histone variant and modification coverage around TSS (± 3kb) in round spermatids (RS) and sperm. Feature density shows the scaled read densities from ChIP-seq experiments. Genes (n=19180) were grouped using k-means into five clusters (1 to 5) containing 1346, 5358, 4468, 2902 and 5106 genes, respectively. 1000 genes were randomly selected for visualization. **d**, Variance partitioning analysis (see Online Methods for details) assessing the unique contribution of different variables (most in RS) to the relative enrichments of H3.3 (left) and H3.1/H3.2 (right) around TSS (± 1 kb) in sperm. Combinatorial effects refer to variation which is common to different combination of variables included.

Figure 7: Model of nucleosome retention during spermiogenesis.

(**a**, **b**) Expression states of genes belonging to different clusters (Fig. 6c) during oogenesis and early embryogenesis. We classified genes as "not expressed", "oocyte", "2-8 cell" and "blastocyst" as described before[11]. Embryonic expression was

classified according to the first expression stage during development. Genes transcribed in oocytes and 2-8 cell embryos or in oocytes and blastocyst embryos were classified as “2-8 cell” or “blastocyst” (**a**) or as “oocyte” (**b**). We matched 14032 of 19180 Refseq genes for expression during oogenesis and embryogenesis[45]. Numbers of genes in each cluster are 1097, 4419, 3431, 2417, 2668, respectively. Statistical significances: *P < 1.0e-06 (Fisher`s exact test). (**c**) Model of nucleosome turnover and retention during spermiogenesis. H3.3 nucleosomes, marked by H3K4me3, become stably incorporated at unmethylated CGIs in response to cessation of global histone turnover and transcription in late round spermatids. Reduced turnover of H3K27me3-marked H3.1/H3.2 nucleosomes in round spermatids promotes retention of such nucleosomes in spermatozoa. Nucleosome retention at unmethylated CGIs would be mediated by unknown CGI-binding factors suppressing nucleosome eviction or alternatively could result from a reduced affinity of protamines for CG-rich DNA. In the presence of DNA methylation, protection against eviction is lost due to the inability of the CGI-binding factor(s) to bind to methylated DNA.

Online Methods

Biological Sample Collection

Mouse sperm were collected from C57BL/6J mice by using swim-up procedure as described[11]. To isolate round spermatids, testicular cells were prepared from 28 day old C57BL/6J mice. Isolated cells were subjected to Hoechst (Invitrogen, cat. num. 33342) staining for 30 min at 37°C and round spermatids were collected via Fluorescent Activated Cell Sorter (FACS) with 90 % purity. All experiments were performed in accordance with the Swiss animal protection laws (license 51, Kantonales Veterinäramt, Basel, Switzerland) and institutional guidelines.

Mononucleosomal DNA preparation and native ChIP

We performed chromatin isolation from mature sperm under native conditions as described[11]. MNase treatment for sperm was performed with 15 U (Roche Nuclease S7, cat. num. 10107921001) at 37°C for 5 min per 2 million spermatozoa. Round spermatid chromatin was isolated in a similar way, except for omission of DTT treatment used for sperm. MNase treatment for round spermatids was performed with 5 U at 37°C for 30 minutes per 1 million cells. Chromatin immunoprecipitation (ChIP) was carried out with antibodies against H3.3 (Millipore 17-10245-ChIP grade (1st replicate), Millipore 09-838 (2nd replicate)), H3.1 and H3.2 (H3.1/H3.2)[49,42],

H3K4me3 (Millipore 17-614) and H3K27me3 (Millipore 07-449) by using approximately 15-20 million sperm or 5 million round spermatids and ~5 µg antibody per ChIP. Both mononucleosomal and immunoprecipitated DNA were resolved by 5% polyacrylamide electrophoresis and 150 bp DNA was gel-purified. Input genomic DNA control was prepared by treating sperm with DTT and detergents as in mononucleosomal preparation, followed by isolation of genomic DNA and subsequent sonication. The reproducibility of nucleosome isolations and ChIPs experiments was demonstrated by the use of biological replicates (Supplementary Fig. 7).

RNA isolation

RNA from FACS sorted round spermatids was isolated by using the Qiagen RNeasy Mini kit. RNA integrity was confirmed by running RNA samples on Agilent 2100 Bioanalyzer mRNA pico arrays.

Library preparation and sequencing

Library preparation for ChIP-seq was done using the Illumina ChIP-seq DNA Sample Prep Kit (Cat# IP-102-1001). Before preparing RNA-seq libraries, rRNA from RNA was depleted by using the Ribo-Zero rRNA removal kit (Epicentre Biotechnologies). Strand specific RNA-seq libraries were prepared by following the Illumina directional mRNA-seq library preparation pre-release protocol. Quality of libraries was assessed by Agilent 2100 Bioanalyzer. Libraries were sequenced on Illumina GA II (36 bp reads) and Illumina Hiseq 2000 (51 bp reads).

Chromatin-bound (histone) fractionation and immunoblotting

Round spermatids were isolated from C57BL/6J mouse testes by centrifugal elutriation[50] and chromatin-bound fractionation was performed according to[51] with some modifications. Briefly, cells were resuspended in buffer A (10 mM HEPES pH7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 0.05% Nonidet P-40, 0.5 mM DTT with protease inhibitors) and incubated for 10 min on ice. After centrifugation, the nuclear pellet was collected and washed twice with buffer A. Nuclei were further lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, protease inhibitors). Then insoluble chromatin was collected by centrifugation, washed twice with buffer B and resuspended in 0.2 M HCl to extract histones. Sperm samples collected by swim-up procedure were initially treated with 50mM DTT at room temperature for 2 hours. Then, the chromatin bound fraction was isolated as described for round spermatids and was concentrated by trichloroacetic acid precipitation. Chromatin-bound extracts were analyzed by 15%

SDS-PAGE gels and transferred onto PVDF membranes that were incubated with antibodies against H3 (abcam ab1791), H3.3 (Millipore 17-10245) and H3.1/H3.2[49,42].

Processing and alignment of the reads

Filtering, alignment and processing of the reads for both ChIP-seq and RNA-seq were done as described[34]. Reads from native ChIP-seq experiments were shifted by 74 nucleotides, corresponding the ½ length of a nucleosome, towards their 3` end to account for the fragment length.

Genomic coordinates

All coordinate regions used in analyzing mouse ChIP-seq and RNA-seq data were based on mouse mm9 assembly (July 2007 Build 37 assembly by NCBI and Mouse Genome Sequencing Consortium). To obtain 1kb windows used in genome-wide analysis, the mouse genome was divided into non-overlapping 1kb windows. From these, the subset of mapable windows (as defined in[34]) was used in the subsequent analysis. Refseq coordinates were downloaded from UCSC[52] (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/refGene.txt.gz from August 16, 2009). For each gene, coordinates corresponding to the longest known transcript was selected.

Genomic regions were classified as promoter, exon, repeat, intron or intergenic as follows: Promoter is defined as the bases covering ±1 kb surrounding Refseq transcripts. Exons are exonic sequences of Refseq transcripts which are not overlapping ±1 kb TSS. Repeats are repeat elements of repeat masker (obtained http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/chr*_rmsk.txt.gz from Jan 30, 2009), which are not overlapping promoter/exon regions. Introns are intronic sequences of Refseq transcripts which are not overlapping promoter/exon/repeat. The remaining part of the genome, which is not promoter/exon/repeat/intron was classified as intergenic. Genomic regions used in analysis of published human ChIP-seq data were based on human hg18 assembly (March 2006 Build 36.1 assembly by NCBI and International Human Genome Sequencing Consortium). 1kb windows for human genome were generated in a similar way as for the mouse genome.

Classification of genes according to their promoter GC content

CpG classifications of the genes as high CpG (HCP), intermediate CpG (ICP) and low CpG (LCP) was performed according to criteria defined in[25]. For the classifications, coordinates ± 1kb surrounding TSS were used (Fig. 1b).

Calculation of observed/expected ratios for dinucleotide frequencies

Dinucleotide and single nucleotide counts per 1kb window were obtained using the R package Biostrings[53]. Observed/expected ratio was calculated as follows: $XY_{cnt}/(X_{cnt}*Y_{cnt})*(W_{size}-1)$, where XY_{cnt} is the dinucleotide count of XY in one 1kb window, X_{cnt} and Y_{cnt} are single nucleotide counts, and W_{size} is the window size (1kb).

CGI definition and usage

CpG island definitions are based on a CpG cluster algorithm[48]. The algorithm was run with default parameters on mm9 to obtain genomic coordinates of CGI.

UCSC tracks

Wiggle files were generated for 100 bp windows and uploaded to the UCSC genome browser[52]. Data was visualized using smoothing over 3 pixels (Fig.1a and 1c, Fig. 3b).

Quantification of enrichment levels genome-wide, at promoter regions and at nucleosome peaks

Enrichment levels for ChIP-seq experiments were calculated for 1kb windows, promoter regions of the genes (± 1kb surrounding transcriptional start sites (TSS)), and nucleosome peaks identified by hidden semi-Markov model (see Supplementary Note for the identification of nucleosome peaks). To calculate enrichment, total read counts mapping to a coordinate region were calculated for ChIP and control (input genomic DNA) samples. Then, these counts were normalized to account for different library sizes between ChIP and control samples. Enrichment for each region was calculated as the ratio between library size normalized read counts for ChIP and control samples according to the following formula: $\log2(((Cnt_{smp}/LSize_{smp}*\min(LSize_{smp}, LSize_{cnt}))+pscnt)/ ((Cnt_{cnt}/LSize_{cnt}*\min(LSize_{smp}, LSize_{cnt}))+pscnt))$, where Cnt_{smp} is the total number of reads mapping to the coordinate in ChIP sample, $LSize_{smp}$ is the total library size for the ChIP sample, Cnt_{cnt} is the total number of reads mapping to the coordinate in the control sample, $LSize_{cnt}$ is the total library size for the control sample, and pscnt is a constant number (pscnt=8), which was used to stabilize enrichments based on low read counts.

Plotting profiles around genomic regions

For each sample, reads mapping to the genomic regions of interest (Fig. 4, Supplementary Fig. 4b) were summed up for every base pair within the genomic

region analyzed. Average read counts per bp were calculated by dividing the total number of reads per bp to total number of genomic regions analyzed. To plot average enrichment values for multiple ChIP-seq samples on the same plot, counts were scaled by the library size and enrichment values were calculated as the ratio between scaled read counts of ChIP and control samples (sonicated sperm genomic DNA). Profiles were smoothed for plotting by taking the rolling mean over 40bp.

Heatmap plots

For ChIP-seq experiments, the number of reads covering each base pair in the region ±3 kb around TSS of genes was quantified. Read coverage was averaged in 50 bp windows along ±3kb TSS. Within each dataset, values were scaled to arrange between 0 - 1. CpG coverage around ±3 kb was obtained by Bioconductor package Biostrings and coverage intensities were scaled in a similar way like ChIP-seq features. Expression data for RS was calculated as log2 (read count per transcript). Clustering was performed by using k-means with k=5, empirically selected as the minimal value of k that resulted in distinct clusters consisting of homogenous members.

Variance partitioning analysis

Variance partitioning analysis was performed via using R package yhat[54]. Unique and combinatorial effects for each variable were obtained by using the function commonalityCoefficients().

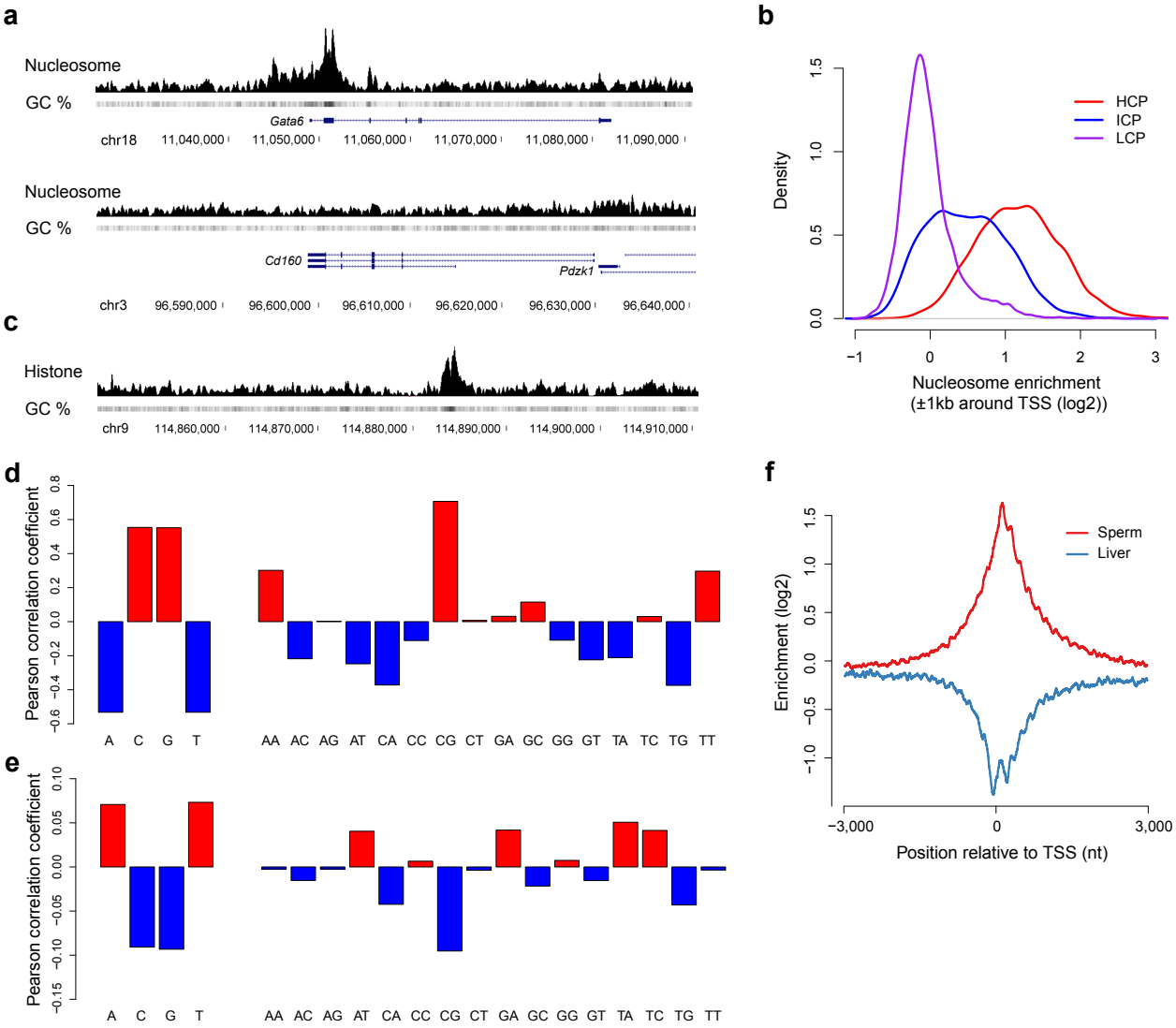
GO-term analysis

GO-term analysis was performed by using Bioconductor package topGO[55]. Enrichment tests were done by using Fisher`s exact test (Supplementary Table 1).

Methods-only references

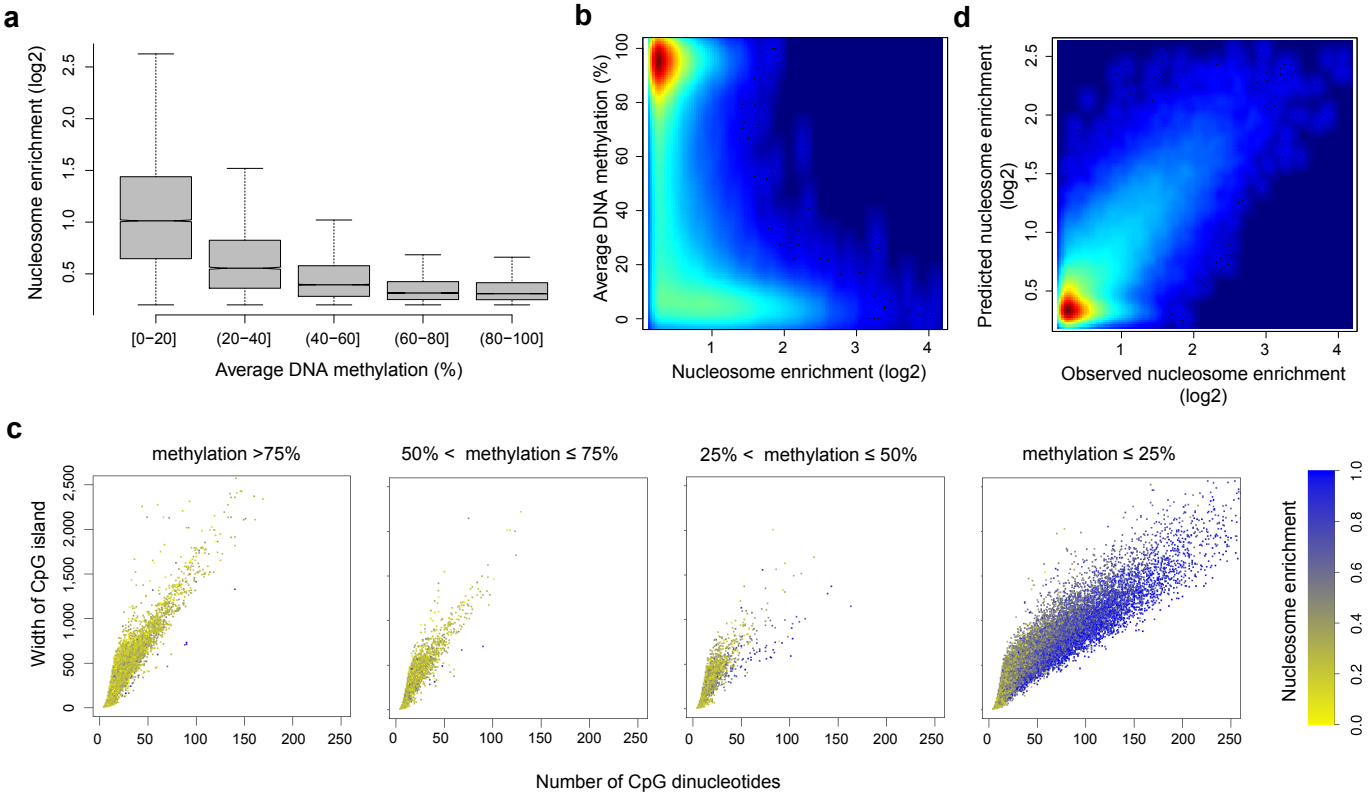
49. van der Heijden, G.W. et al. Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. *Mechanisms of development* **122**, 1008-22 (2005).
50. Barchi, M., Geremia, R., Magliozzi, R. & Bianchi, E. Isolation and analyses of enriched populations of male mouse germ cells by sedimentation velocity: the centrifugal elutriation. *Methods in molecular biology* **558**, 299-321 (2009).
51. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Molecular and cellular biology* **20**, 8602-12 (2000).

52. Kent, W.J. et al. The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).
53. H., P., P., A., R., G. & S., D. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.26.2.
54. K., N. & K., R.J. yhat: Interpreting Regression Effects. R package version 1.0-5. (2012).
55. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-7 (2006).



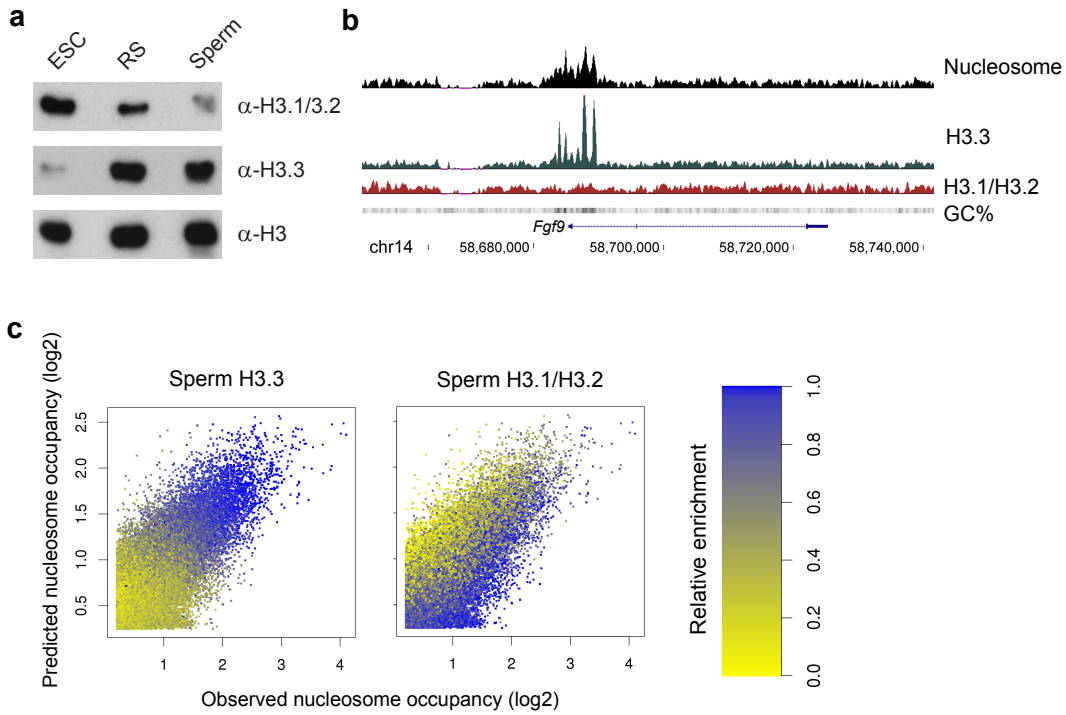
Erkek *et al.*

Figure 1



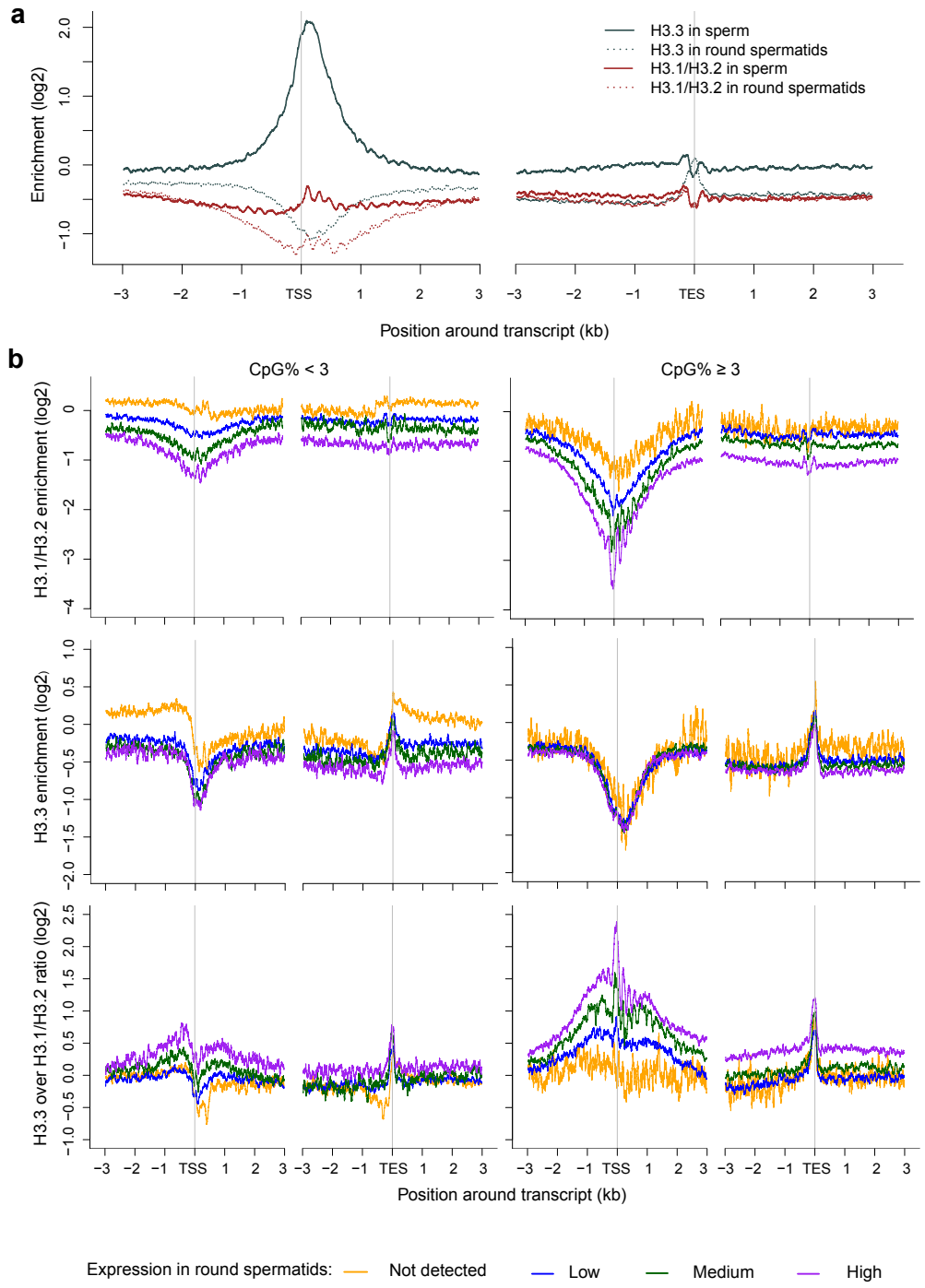
Erkek *et al.*

Figure 2



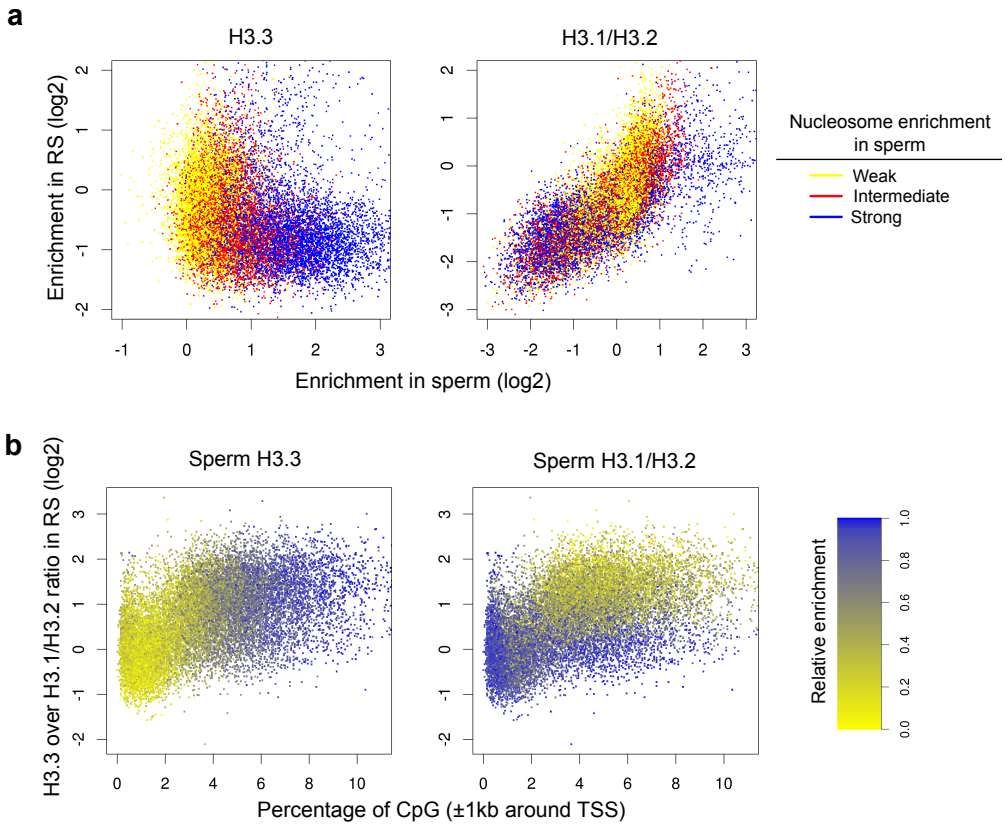
Erkek *et al.*

Figure 3



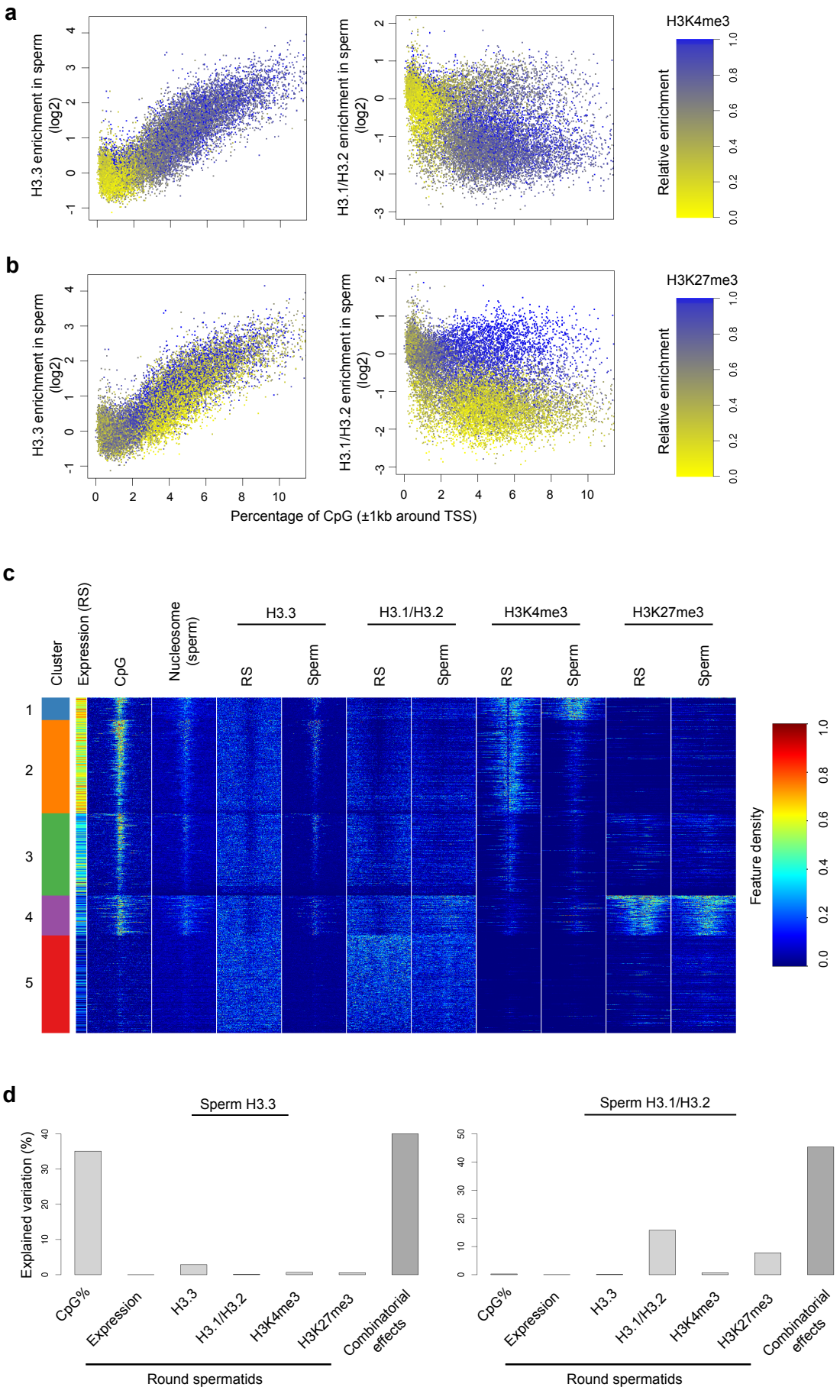
Erkek *et al.*

Figure 4



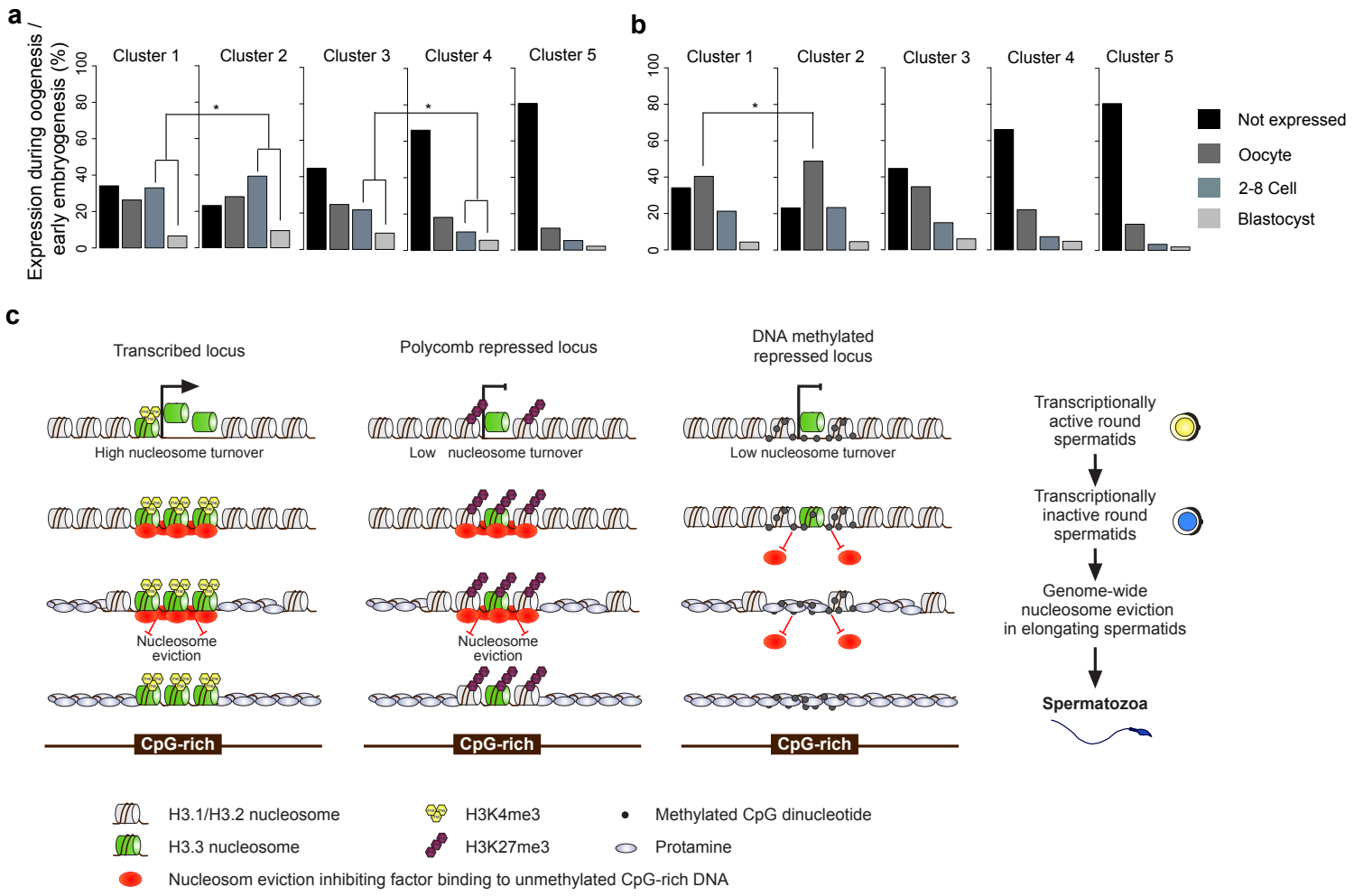
Erkek *et al.*

Figure 5



Erkek *et al.*

Figure 6



Erkek *et al.*

Figure 7