An epigenomic approach to understanding the mechanism of nucleosome retention in mouse spermatozoa

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch–Naturwissenschaftlichen Fakultät

der Universität Basel

von

Serap Erkek

aus Turkei

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel auf Antrag von

Prof. Antoine Peters
(Fakultätsverantwortlicher und Referent)

Prof. Gernot Längst
(Koreferent)

Basel, der 26 Februar, 2013

Prof. Jörg Schibler
(Dekan)

Table of Contents

Summary

In mammals fusion of sperm and oocyte gives rise to a totipotent embryo. Origin of totipotency of the early embryo is highly debated: whether it is achieved by inheritance of the epigenetic states of the gametes or by reprogramming of such parental epigenetic marks in the embryo.

Oocyte and sperm differ in their potential to transmit epigenetic information. The oocyte is full of maternal transcripts, proteins and its DNA is packed into nucleosomes while a spermatozoon is in highly compact structure and the majority of its histones are exchanged by protamines. It has been determined that in mature sperm histone-to-protamine exchange is not complete, still around 10% and 1% of histones are retained in human and mouse sperm, respectively. During the initial period of my PhD, we and others showed that in human sperm retained histones are not randomly distributed in the genome but to some extent are enriched at loci important for developmental and signaling pathways. We obtained similar findings in mouse sperm at single loci. Nevertheless, genome-wide localization of nucleosomes in mouse sperm and the main principles defining specific nucleosome retention were not known. In my project, the major aim was to determine the logic of nucleosome retention by using mouse sperm as the model system. In addition, I investigated transcript dynamics during late spermatogenesis to identify characteristics of the transcriptomes in maturing germ cells and spermatozoa.

By taking a genome-wide approach we have identified that combinatorial effects of sequence composition, histone variants, histone modifications and gene expression uniquely package sperm DNA. Importantly, H3.3 constitutes the main histone H3 variant retained in mature sperm and localizes to CpG islands. The majority of the genomic regions containing H3.3 are marked by H3K4me3. H3.3 retention in sperm reflects high nucleosome turnover in round spermatids. Canonical histone H3 variants H3.1 and H3.2 are present in low amounts in mature sperm and their retention pattern mostly shows the history from non-replicating round spermatids. GC-rich genomic regions marked by H3K27me3 retain H3.1/H3.2, likely related to low nucleosome turnover in round spermatids. Investigating transcript dynamics during later stages of spermatogenesis showed that overall transcript levels towards sperm development are static. Nevertheless, our data relating changes in gene expression to changes in chromatin states highly suggest for ongoing transcriptional activity during differentiation of spermatids into sperm.

Overall, we identified that histone modification states of retained nucleosomes and spermatozoal RNA pool highly relate to early embryonic gene expression, which argues that sperm carries critical information to the early embryo.

Chapter 1: General introduction

1.1 Epigenetic states and inheritance

1.1.1 What is epigenetics?

Epigenetics is a term first raised during 1940s by Conrad Waddington. Waddington was studying embryonic development and he wanted to make connections between genetics and development. He proposed that embryonic development takes place as a result of combinatorial actions of different gene activities via formation of gene networks. Waddington combined the words epigenesis- processes which result in development of an organism and genetics to define epigenetics (Van Speybroeck, 2002). In his terms epigenetics is defined as "the study of processes by which the genotype gives rise to the phenotype" (Morange, 2002).

In 1958 Nanney used the term "epigenetic control systems" which are involved in existence of different cellular phenotypes by the regulated expression of the same "genetic potentialities" and he emphasized the stability of epigenetic control systems nevertheless still emphasizing the point that cellular memory could be changed depending on the environmental conditions (Nanney, 1958).

During 1970s, DNA methylation was proposed to be involved in gene regulation and development (Griffith and Mahler, 1969; Holliday and Pugh, 1975; Riggs, 1975). Soon after these proposals, discoveries which showed that DNA methylation could be involved in gene silencing (especially to explain X chromosome inactivation) without a need of genetic mutation started to change the definition of epigenetics. By 1990s, epigenetics started to be referred as "inheritance which is not based on DNA sequence changes" (Holliday, 2006).

Nowadays, epigenetics is defined as "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" (Wu and Morris, 2001). It has been also suggested that to be called "epigenetic", a process should have three main properties: it should be heritable, there should be a mechanism for the propagation of the state and finally it should be reversible (Bonasio et al., 2010).

To conclude, epigenetics is a term with changing definitions over time and depending on new scientific discoveries. Personally, I find Waddington`s definition of epigenetics "the study of processes by which the genotype gives rise to the phenotype" more logical. In my terms, I would define epigenetics as "systems in a cell nucleus which direct execution of cellular regulation programs with interaction with the genome itself and other molecules/complexes". I avoid using heritability in my definition, because although epigenetic inheritance has been suggested to

function in many systems as it will be discussed below, actually, it is not known whether the molecule initially involved in specification of a certain regulation program is inherited during DNA replication and whether inheritance of this molecule is indeed required.

1.1.2 Inheritance of epigenetic states

Multicellular organisms consist of a multitude of distinct cell types and each cell type has a unique gene expression program. Epigenetic mechanisms play an essential role in specification as well as maintenance of a particular cellular identity. Various processes such as DNA methylation, Polycomb group (PcG) and Trithorax group (TrxG) of proteins have been described to be involved in epigenetic memory (Ng and Gurdon, 2008). Detailed functioning of these systems will be discussed later in section 1.2.

Examples for transcriptional memory have been demonstrated both in yeast, *Drosophila*, and mammalian systems. In yeast treatment of cells with TSA, which is a histone deacetylase inhibitor, results in expression of centromeric markers with increased histone acetylation levels. The established expression state is maintained over successive generations without TSA treatment (Ekwall et al., 1997). In *Drosophila*, embryonically induced activated or repressed state of a regulatory element dependent on action of PCG and TrxG proteins is heritable (Cavalli and Paro, 1998). In mammalian somatic cells, CpG methylation is involved in inheritance of silenced state (Feng et al., 2006).

Studies which focus on position effect variegation (PEV) also provided some insights about epigenetic inheritance. PEV refers to a phenomenon where genes change their expression status depending on whether they will be close to transcriptionally permissive or impermissive environment. In *Drosophila* by using a gene whose expression could be easily identified such as eye color, it was shown that gene expression states established depending on the chromatin environment the gene is present are inherited over successive cell divisions (Margueron and Reinberg, 2010).

All the examples above illustrate the maintenance of an established state, nevertheless, the molecular details of how the cellular memory works and actually which molecules are inherited are not known.

1.1.3 Transgenerational epigenetic inheritance

It is widely believed that after fertilization epigenetic states carried by oocytes and sperm to the early embryo are reprogrammed to ensure totipotent state of the

zygote. Although currently, it is unclear what exactly happens to the germline set epigenetic states during pre-implantation development, there are certainly some examples which show that epigenetic states of the germ cells at some loci are retained after fertilization. Below, several these kinds of examples will be introduced.

- Genomic imprinting

Genomic imprinting represents a mechanism where certain genes are expressed in a parent-specific origin manner. For this class of genes, paternal and maternal alleles have different epigenetic configuration which affect their transcriptional outcome. Imprinted genes resist reprogramming after fertilization. Nevertheless, in the germline, these genes are subject to reprogramming again to enable the parent-specific origin of expression.

- Epialleles

Studies performed in plants represent one example of transgenerational epigenetic inheritance. One of the most common examples concerns the flower symmetry in *Linaria vulgarus*. Wild type flowers of this plant are bilateral symmetrical. It has been identified that when the *Lcyc* gene, controlling dorsoventral asymmetry, is methylated and silenced, flowers become radially symmetrical. This naturally present epimutation has been also shown to be heritable across the genarations (Cubas et al., 1999).

One of the most well known examples in mammals is the agouti viable yellow allele (A^{vy}) of the *agouti* locus. *Agouti* locus encodes for a protein that is involved in switching of pigment color from brown/black to yellow in melanocytes (Duhl et al., 1994). This locus contains an intracisternal A particle (IAP) retrotransposon element inserted which is close the promoter region of the gene. Expression of the allele inversely correlates with methylation status of IAP element. Coat color is transmitted only maternally, which argues for the clearance of the epigenetic marks in paternal germline at this locus (Daxinger and Whitelaw, 2012). However, DNA methylation status inherited from A^{vy} maternal allele is lost in blastocysts, which suggests that DNA methylation is not the mark implicated in the transmission of the phenotype (Blewitt et al., 2006). In addition, transgenerational epigenetic inheritance happens at *axin-fused* allele ($Axin^{Fu}$) via methylation differences of associated IAP element. Different from *agouti* locus, epigenetic state of $Axin^{Fu}$ could me both maternally and paternally inherited (Daxinger and Whitelaw, 2012).

- Paramutation

Paramutation refers to allelic interaction which results in meiotically heritable gene expression. Most known examples of paramutation come from the studies in plants for loci encoding pigment colors. In maize, *b1* gene is involved in purple color pigmentation. *B-I* allele is responsible for dark pigmentation, whereas *B`* is involved in light pigmentation. Upon crossing of *B-I* with *B`*, *B-I* is changed into *B`* (*B`**). Presence of tandem repeats upstream of the gene have been shown to involved in the paramutation. Bidirectional transcription from the repeat elements results in production of siRNAs which is thought to induce to change the epigenetic state of *B-I* into *B`*. Nevertheless, exact molecular mechanism of this process remains to be determined (Arteaga-Vazquez and Chandler, 2010).

An example of paramutation in mouse was shown at *Kit* locus, which encodes for a tyrosine receptor kinase and involved in germ cell differentiation and melanogenesis. Heterozygous mutant of *Kit* where there is an insertion of *lacZ* sequence downstream of the ATG start site show white tail phenotypes. It was identified that wild type mice with heterozygous parents still have the white tail phenotype. It has been shown that actually the unusual amount of *Kit* mRNA produced during late spermatogenesis and its accumulation in mature sperm could be responsible for the inheritance of the phenotype. It was proposed that unusual *Kit* mRNA levels result in a degradation product which causes white tail phenotype as injecting miRNA against *Kit* mRNA in early embryos results in the same phenotype (Rassoulzadegan et al., 2006).

- Environmentally induced epigenetic changes

Epigenetic states could be influenced by environment that the organism is exposed to and could be transmitted to the subsequent generations. Nevertheless, in the context of epigenetic effects induced by environment it is important to differentiate between transgenerational epigenetic effects and transgenerational epigenetic inheritance. To be called as inherited, epigenetic states should be carried by gametes to the early embryo. On the other hand, epigenetic effects could simply result from the environment to which the embryo is exposed to in-utero (Daxinger and Whitelaw, 2012).

A study in humans investigated the individuals who were periconceptual exposed to famine during Dutch Hunger Winter for epigenetic differences of the imprinted *IGF2* gene. It was identified that *IGF2* gene had lower levels of DNA methylation as compared to control non-exposed individuals. This study was one of the first showing that environmental changes during early gestation in humans could result in epigenetic changes (Heijmans et al., 2008). Another study performed in mice

showed that pregnant females exposed to ethanol consumption during 0.5-8.5 days of gestation gave offspring with epigenetic changes at agouti viable yellow (A^{vy}) locus. Ethanol consumption has been found be associated with hypermethylation at A^{vy} locus and transcriptional silencing. In addition, ethanol consumption resulted in postnatal growth restriction phenotype and craniofacial dysmorphism (Kaminen-Ahola et al., 2010).

Transient embryonic exposure of pregnant rats during E8 to E14 day of gestation to the anti-androgen vinclozolin results in offspring with spermatogenic defects. Born males have significant apoptosis of the germ cells in testicular tubules, reduced sperm count and motility. Importantly, reduced spermatogenic capacity is transmitted to the subsequent generations (Anway et al., 2006). A few years after this study, it was shown that sperm from F3 generation (with F0 mothers treated with vinclozolin) had DNA methylation differences at some promoter regions as compared to control sperm (non-treated) (Guerrero-Bosagna et al., 2010). Although this finding could argue for a germ-line transmitted epigenetic change, analysis of the sperm from F1 and F2 generations should be also included to really claim that identified DNA methylation changes are the transmitted mark.

Another study in mice investigated the effect of paternal diet on gene expression of the offspring. Offspring of the males who had been fed with low-protein diet showed upregulation of the genes involved in lipid and cholesterol synthesis. Analysis of the liver from offspring showed that offspring from the males fed with low-protein diet had DNA methylation changes in promoter regions of some genes, especially at the promoter region of *Ppara* gene, which is involved in hepatic gene expression response. Nevertheless, analysis of the promoter region of *Ppara* gene from the sperm cells of the respective fathers did not show any DNA methylation changes. In addition, authors performed RNA and chromatin analysis on the sperm, and detected very subtle changes on some genes. Therefore, the exact mechanism how the offspring is affected from the paternal diet change and whether this goes through epigenetic inheritance via the gametes remains to be determined (Carone et al., 2010).

A recent example in rats studied the adaption of wound healing in rats. Males were subjected to repeated injury in F0 and F1 generations, and F2 generation was analyzed for hepatocellular damage and wound healing. It was identified that injury did not cause any heritable damage in liver but hepatic wound-healing was suppressed. As a result of this adaptation, the expression of antifibrogenic factor peroxisome proliferator-activated receptor γ (*Ppar-γ*) was increased and the expression of profibrogenic factor transforming growth factor β1 (*Tgf-β1*) was

determined to be decreased. *Ppar-γ* gene in liver from the injured animals was hypomethylated. Although analysis of sperm from the injured animals did not show any DNA methylation changes at *Ppar-γ* gene, *Ppar-γ* was enriched for H2A.Z and H3K27me3 in sperm. It has been suggested that a soluble factor from the liver of injured animals could change chromatin of sperm (Zeybel et al., 2012). Nevertheless, the ultimate connection between chromatin changes in sperm and its effect on expression of *Ppar-γ* gene remains to be determined.

1.2 Epigenetic mechanisms in gene regulation and maintenance of genome stability

I have discussed the epigenetic states and examples of epigenetic inheritance functioning in keeping certain cellular identity also potentially transmitting information across the generations via germline. But, how do mechanistically epigenetic systems work? In below sections, I will introduce the mechanism of action of the best known and characterized epigenetic mechanisms.

1.2.1 State of genomic DNA in a eukaryotic cell

DNA in a eukaryotic cell is not naked in the nucleus but packed into a structure called chromatin. Around 147 bp of DNA is wrapped around an octamer of histone proteins with 2 copies of histone H2A, H2B, H3 and H4 making the fundamental unit of chromatin, which are nucleosomes (Luger et al., 1997) Nucleosomes are connected by linker DNA, making an array known as “beads on a string” or primary structure. Presence of linker histone H1 or another architectural protein and short range nucleosome-nucleosome interactions further compact the nucleosome structure into a 30nm fiber forming a secondary structure. In addition long range fiber-fiber interactions could form tertiary chromatin structure which could be also stabilized by presence of other chromatin proteins (Luger et al., 2012) (Figure 1).

Packing of DNA into nucleosome structure itself is actually inhibitory to gene expression and other processes such as binding of transcription factors. Additional level of complexity comes from the fact that DNA itself is methylated, N-terminal tails of histones are chemically modified and there might be compositional differences within a nucleosome caused by presence of different histone variants (Probst et al., 2009). Each of these possibilities, or combination of them and further interactions with other proteins and RNA might have different roles in gene regulation and genome stability.

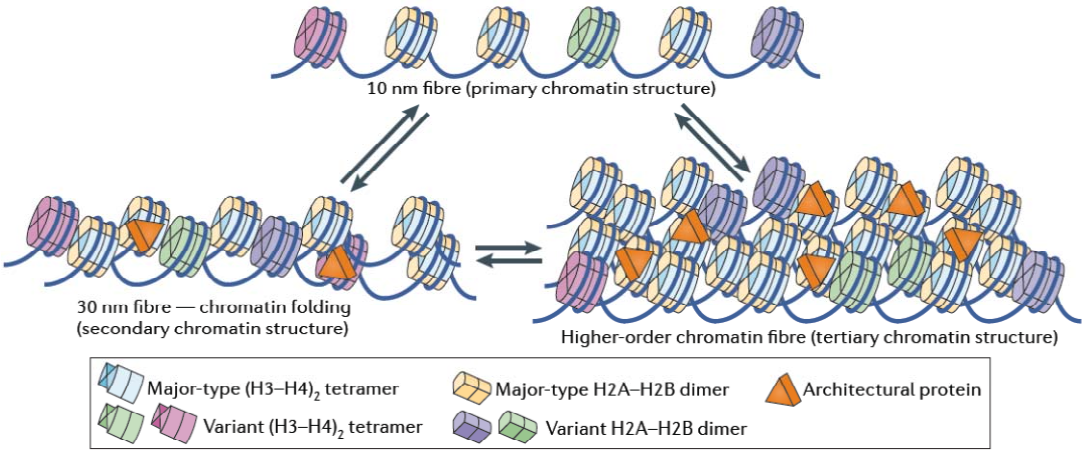


Figure 1: Packaging of DNA into chromatin. Figure illustrates the packing of DNA into nucleosomes making the primary, secondary and tertiary structures of the chromatin (adapted from (Luger et al., 2012), permission to use: license number: 3124281314453).

1.2.2 Mechanism and function of DNA methylation

DNA methylation was described many years ago. Initially identified as restriction modification mechanism in bacteria, analysis of the nucleotides from animals and plants showed that these organisms had also quite some 5-methylcytosine and almost all of this methylation happens within CpG dinucleotide (Cedar and Bergman, 2012).

CpG dinucleotide states, frequency and organization in the genome

CpG dinucleotide might be in many different states: unmetyhlated, methylated (5mC) and hydroxymethylated (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Methylation occurs on the 5 position of pyrimidine ring of cytosine base pair (5mC). Oxidation of 5mC could create 5hmC, 5fC and 5caC (Ito et al., 2011). Both methylated and unmethylated forms of cytosine are prone to deamination which could give rise to thymidine and urasil, respectively. Urasil, not normally present in DNA could be efficiently removed and replaced by cytosine via DNA repair system. However, thymine as regular component of DNA can not be efficiently repaired and this causes mutations. In vertebrates, approximately 80% of CpGs are methylated and this creates a mutational pressure that decreases frequency of CpGs genome-wide. Nevertheless, in vertebrates short unmethylated regions in the genome exists, so called CpG islands (CGI). CGIs have 10 times more CG frequency compared to bulk of the genome and constitutes about the 1% of the genome. In invertebrates such as *Drosophila melanogester* and *Caenorhabditis*

elegans, there is almost no DNA methylation, therefore CpGs exist at the expected frequency and there are no CpG islands (Bird, 2011). About half of CGI intersect with promoter regions, the other half of CGI lie within intragenic or intergenic sequences and they might be associated with novel promoter activity for instance by recruiting RNA polymerase II and called as “orphan CGI” (Deaton and Bird, 2011).

One of the main questions arising immediately is that how CGI were initially generated. Adrian Bird and colleagues suggest possible explanations for this process. Initially, it was thought CGI cannot be targeted by DNA methyltransferases because of sequence or organization characteristics of these elements. Since some CGI get DNA methylation during development, this hypothesis is likely not true. Alternatively, CGI could be actively demethylated to keep them in an unmethylated state. In this context, Tet1 protein which is involved in conversion of 5mC to 5hmC and having a CpG binding-CXXC domain might be the one of the best candidates functioning in such a mechanism. It has been argued that actually most favorable explanation comes from the fact that CGIs are associated with promoter regions. Binding of transcription factors at CGI, presence of RNA polymerase II at CGI and the fact that the great majority of genes with CGI are expressed in early embryo or in testis could all support this explanation (Deaton and Bird, 2011; Illingworth and Bird, 2009). A supportive finding to this idea also comes from a recent study. It has been shown that transcription factor binding in the genome shapes DNA methylation levels genome-wide in mouse, creating unmethylated regions (UMRs), lowly methylated regions (LMRs), and fully methylated regions (FMRs) in the genome (Stadler et al., 2011).

Mammalian DNA methyltransferases

In mammals, there are three main type of DNA methyltransferases (Dnmts) exist, Dnmt1, Dnmt3a, and Dnmt3b, and one Dnmt3 related protein called Dnmt3L with no catalytic activity. Dnmt1 is ubiquitously expressed in proliferating cells and it acts as maintenance Dnmt to keep DNA methylation patterns across cellular divisions (Jurkowska et al., 2011). Dnmt1 is required for early development as its disruption causes embryonic lethality after gastrulation and various defects with genomic imprinting and X chromosome inactivation. Nevertheless, absence of Dnmt1 in embryonic stem cells (ESCs) shows an effect only upon differentiation (Li et al., 1992).

Dnmt3 family, consisting of Dnmt3a, Dnmt3b and Dnmt3L (catalytically inactive and functioning in germ cells) has been determined to be responsible from de novo DNA methylation in mammals (Jurkowska et al., 2011). In mice, *Dnmt3a* and

Dnmt3b are highly expressed in (ESCs), but their expression is down-regulated after differentiation (Okano et al., 1998). *De novo* methylation does not take place in ESCs upon disruption of both *Dnmt3a* and *Dnmt3b*. Heterozygous deletion of *Dnmt3a* and *Dnmt3b* in mice does not cause any major problems. Homozygous deletion of *Dnmt3a* gives rise to viable offspring with an early death at 1 month age. Nevertheless, homozygous deletion of *Dnmt3b* causes some developmental problems starting at E9.5 and no viable offspring is produced. It was identified that *de novo* methylation patterns were lost both in single homozygous or double homozygous mutants of Dnmt3 enzymes suggesting overlapping functions although some specific functions for the methylation of minor satellite repeats were observed for *Dnmt3b* (Okano et al., 1999).

How are DNA methylation patterns established de novo?

What targets Dnmt3 enzymes to certain regions of the genome specifically? At this point, different mechanisms might be proposed. One mechanism might be that Dnmt3 enzymes recognize DNA via specific protein domains. PWWP domain present in human DNMT3b has been shown to be required for the catalytic activity of the protein. Furthermore, this domain could directly interact with DNA (Klose and Bird, 2006). Another possibility for the targeting of Dnmt3s might be that there could be DNA methylation centers for instance containing repetitive elements directing DNA methylation. In ESCs, such a region was identified. APRT methylation center, containing B1 repetitive elements upstream of *Aprt* gene is DNA methylated in *Dnmt1* deficient ESCs arguing for an intrinsic affinity towards de novo DNA methylation (Bird, 2002). In addition, *de novo* DNA methyltransferases Dnmt3a and Dnmt3b have flanking sequence preferences around target CpG sites. However, presence of Dnmt3L decreases the effect of sequence preferences of these enzymes and leads to formation of more uniform DNA methylation patterns (Wienholz et al., 2010). It has been also shown that RNA interference (RNAi) could guide de novo methylation. In plants, it was identified that RNAi could target DNA methylation to the homologous DNA sequence. Nevertheless, it remains to be determined whether this actually happens also in mammals because of the contradictory results found (Klose and Bird, 2006). As DNA is packaged into nucleosomes, chromatin states might have an effect on DNA methylation status, too. It has been identified that methylation of lysine 4 on histone H3 is inhibitory to binding of Dnmt3 enzymes (Ooi et al., 2007; Zhang et al., 2010). Finally, accessibility of DNA sequences to Dnmt3 enzymes might have a role in targeting since mutations in chromatin remodeling factors have been shown to affect DNA methylation patterns (Bird, 2002).

Reprogramming DNA methylation patterns in early embryogenesis and germ line

Fusion of two highly differentiated gametes, sperm and oocyte gives rise to a totipotent embryo which could give rise to all cells of an organism including the germ cells. It is thought that DNA methylation marks are globally reprogrammed after fertilization in early embryo and during germ cell specification to ensure correct embryonic cell identities and guarantee the potential of germ cells to form a totipotent embryo. Nevertheless, it is already known that there are genomic regions which could escape reprogramming (Gill et al., 2012; Smallwood and Kelsey, 2012).

Primordial germ cells (PGCs) before entry into gonads are highly methylated with similar patterns observed for somatic cells. It was identified that during primordial germ cell migration to gonads, starting from E12.5 to E13.5, DNA methylation levels are significantly reduced (Hajkova et al., 2002). Nevertheless, erasure of DNA methylation does not occur in the same way along the entire genome. For instance, intracisternal A-particle (IAP) class repeat elements remain partially methylated. However, DNA metylation patterns at imprinting control regions (ICRs) are erased. Actually, a recent genome-wide study showed that demethylation already starts from E9.5 and continues until E13.5 happening with different dynamics depending on the characteristic of the genomic region (Seisenberger et al., 2012). There are different mechanisms suggested to explain demethylation events occurring in PGCs. Presence of DNA single strand breaks and higher levels of DNA methylation levels observed with *Aid* (a cytidine deaminase) deficiency in primordial germ cells argued for the involvement of DNA repair and Tet1-mediated conversion of 5mC to 5hmC in this process (Hajkova et al., 2010; Popp et al., 2010). Nevertheless, a recent study showed that DNA demethylation events happen in a passive manner as revealed by the comparison of DNA demethylation kinetics to doubling times of PGCs (Kagiwada et al., 2012). In addition, data from another recent study supported the passive DNA demethylation mechanism based on the fact that during E9.5 to E13.5, PGCs undergo cellular divisions, and the data showing strand specific hemimethylated CpGs and exclusion of Dnmt1 from early PGCs (Seisenberger et al., 2012).

Following erasure of DNA methylation marks and sex determination of the early embryo, de novo methylation pattens start to be established in female and male germlines. In females, *de novo* methylation takes place in growing oocytes arrested in meiotic prophase I. In males, it initiates before the onset of meiosis in mitotically arrested spermatogonia and is completed before birth (Smallwood and Kelsey, 2012).

In germ line, certain genomic regions become differentially methylated. Among those are ICRs and repetitive elements (Smallwood and Kelsey, 2012). Recently, several genome-wide studies identified differentially methylated CGI or gametic differentially methylated regions (gDMRs). Number of identified gDMRs differed slightly depending on the method used for DNA methylation mapping (Kobayashi et al., 2012; Smallwood et al., 2011; Smith et al., 2012).

Dnmt3 enzymes play an essential role in establishment of germ cell specific DNA methylation patterns and they are all expressed both in female and male germ lines (Smallwood and Kelsey, 2012). Offspring from *Dnmt3a* conditionally mutant (deletion in primordial germ cells by using TNAP-Cre) females die in utero around E10.5 and analysis of embryos from E10.5 showed that maternal imprints were lost. Conditional deletion of *Dnmt3a* in males results in defective spermatogenesis, azoospermia, and loss of paternal imprints. On the other hand, conditional deletion of *Dnmt3b* did not give any apparent phenotype (Kaneda et al., 2010). Later, another study from the same group actually showed that Dnmt3b function is important to keep paternal imprint at *Rasgrf1* ICR (Kato et al., 2007). Male mice homozygous for disrupted *Dnmt3L* are sterile and offspring from females homozygous for disrupted *Dnmt3L* die before midgestation. Analysis of the mutant females showed that maternal imprints were abolished (Bourc'his et al., 2001). Following this, another study showed that Dnmt3L is also required for establishment of paternal imprints, acquisition of DNA metylation at certain repeat elements, and lack of Dnmt3L results in meiotic arrest of spermatogenesis caused by structural changes in chromatin (Webster et al., 2005). A genome-wide study showed that Dnmt3a and Dnmt3L are required for oocyte specific DNA methylation patterns at CGI, extending the function of Dnmt3 enzymes beyond setting correct imprinting patterns (Smallwood et al., 2011). Recently, it was also shown that although Dnmt3L is required for setting of maternal imprints, to a certain extent DNA methylation of some retrotransposons are not dependent on Dnmt3L in oocytes (Kobayashi et al., 2012).

After fertilization another wave of DNA methylation reprogramming takes place. Paternal genome undergoes a rapid and active demethylation process, on the other hand, for maternal genome this takes place slowly and in a passive manner (dependent on DNA replication) (Gill et al., 2012; Smallwood and Kelsey, 2012). Mechanism involved in active demethylation of paternal genome has been proposed to be linked to conversion of 5mC to 5hmC on the paternal genome. Maternal deletion of dioxygenase *Tet3* blocks the conversion of 5mC to 5hmC and 5mC stays constant on the paternal genome. Embryos lacking maternal Tet3 show a significant developmental failure (Gu et al., 2011). Starting with blastocyst implantation and the

determination of cell lineages, cell type specific DNA methylation patterns start to be reset (Smallwood and Kelsey, 2012).

There has been in a significant interest in identifying whether demethylation events occurring after fertilization happens everywhere in the genome or at certain genomic regions. In this context, ICRs are one of the best examples known to escape this reprogramming and represent a one generational cycle of epigenetic inheritance (Gill et al., 2012). In addition, recently there have been several studies analyzing the fate of DNA methylation states in oocyte and sperm upon fertilization at a genome-wide scale. Smallwood and colleagues showed that methylation levels of most CGIs in blastocyst, which were originally methylated in oocytes, were significantly higher than that would be expected from a passive demethylation process (Kobayashi et al., 2012; Smallwood et al., 2011). In addition, it was identified that about half of the gDMRs partially resist to DNA demethylation (Kobayashi et al., 2012).

It has been shown that there are two main transitions of DNA methylation patterns during early embryonic development, one happening from sperm to zygote with a substantial decrease in DNA methylation (Smith et al., 2012) and one happening from early inner cell mass (ICM) to post-implantation development with a significant increase in DNA methylation (Borgel et al., 2010; Smith et al., 2012). After fertilization, demethylation of genomic regions in sperm mainly occurs at intergenic sites. Certain LINE elements which are fully methylated in sperm shows the most dramatic demethylation (Smith et al., 2012).

Mechanistics of DNA methylation: What reads DNA methylation mark?

DNA methylation is inhibitory to transcriptional activity, but how is this mechanistically achieved? DNA methylation sensitive CpG binding proteins have been described to be involved in this process (Bird, 2002). Two families of proteins could bind methylated DNA: Methyl CpG binding proteins (MBDs) and Kaiso protein family. MBDs mainly consist of MBD1, MBD2, MBD3, MBD4 and MeCP2. MBDs share a homologous MBD domain. Importantly, mutation in MeCP2 causes Rett syndrome, a neurological disorder. Kaiso protein family, consisting of Kaiso, Zbtb4 and Zbtb38, have zinc finger domains. These proteins form histone deacetylase complexes and establish silent chromatin (Bogdanovic and Veenstra, 2009).

In addition to methyl CpG binding proteins, there are proteins which are repelled by DNA methylation. For instance, CXXC-domain containing proteins bind to DNA only if it is unmethylated (Allen et al., 2006).

Function of DNA methylation in gene and repetitive element silencing

The role of DNA methylation has been described in many cellular contexts. Despite the clear role of DNA methylation in silencing, it is not thought as a mechanism in the initial establishment of silenced states. There are several studies which support this hypothesis. Silencing of exogenous retroviral elements occurs without the need of DNA methylation initially per se in embryonic stem cells (ESCs) (Bird, 2002). In addition, it has been shown that *de novo* methylation requires nucleosome for the functioning of Dnmt3 enzymes and methylation of lysine 4 on histone H3 is inhibitory to DNA methylation (Ooi et al., 2007).

Studies which used knock-out mouse models of Dnmt3 enzymes clearly showed that DNA methylation is involved in establishment of parental specific imprinting patterns. Nevertheless, maintenance of parental specific imprinting patterns does not only involve DNA methylation, but in addition additional action of other proteins. For instance, recently it has been shown that Kruppel-associated box-containing zinc finger protein (Zfp57) regulates maintenance of DNA methylation at imprinted loci in early embryos and embryonic stem cells via recognizing a methylated hexanucleotide sequence (Li et al., 2008; Quenneville et al., 2011; Zuo et al., 2011).

DNA methylation is also involved in long term silencing of inactive X chromosome in mammals. Genes on the inactive X chromosome has been found to hypermethylated but not on active X chromosome. DNA methylation plays a role in maintenance of silencing rather than initiation on inactive X chromosome (Chang et al., 2006).

DNA methylation is also required for efficient silencing of the transposable elements in the genome. Transposable elements are highly methylated and silenced in somatic cells. In the context that transposition could create DNA damage, it has been suggested that DNA methylation might be thought as host-defense system initially evolved to counteract parasitic sequence threads (Yoder et al., 1997).

1.2.3 Chromatin based epigenetic mechanisms

1.2.3.1 Nucleosomal organization of genome

Genomic DNA is packed around histone proteins to form the fundamental unit of chromatin which are nucleosomes. There have been a significant number of studies asking the question what kind of criteria determines the packaging of DNA into nucleosomes. At this point, there are two main distinctions: nucleosome occupancy and nucleosome positioning. Nucleosome positioning is defined as “the probability that a nucleosome starts at a given base pair within the genome”. On the

other hand, nucleosome occupancy refers to “presence or absence of nucleosomes over specific DNA sequences in the genome and differs from nucleosome positioning in that the former does not care where the nucleosome starts as long as the base pair is covered by it” (Arya et al., 2010). If we think that nucleosomes cover 147 bp DNA, there might be 147 different nucleosomes covering a given base pair. In this way, nucleosome occupancy could also be defined as “147-bp-moving average of the nucleosome positioning probabilities” (Segal and Widom, 2009).

Several studies mapped genome-wide nucleosome occupancy in many eukaryotes. It has been shown that transcriptional activity creates nucleosome free regions around the transcriptional start sites (TSS) of genes (Li et al., 2011; Schones et al., 2008). This effect is more apparent for the genes which have low GC-content promoters since high GC-content promoters are mostly devoid of nucleosomes without being necessarily associated with transcriptional activity (Fenouil et al., 2012; Li et al., 2011; Ramirez-Carrozzi et al., 2009). Exonic sequences were identified to have high nucleosome occupancy, which is suggested to regulate RNA splicing (Andersson et al., 2009; Choi, 2010; Schwartz et al., 2009). A recent study investigating genome-wide nucleosome occupancy in embryonic stem cells (ESCs) showed that the majority of the transcription factor binding sites were depleted from nucleosomes in ESCs. In differentiated cells most probably because of the lack of expression of the factors, those sites were occupied by nucleosomes (Teif et al., 2012).

Nucleosome occupancy and positioning might be influenced by several factors. It has been shown that DNA sequence itself could predict nucleosome occupancy (Kaplan et al., 2009). Analysis of 5-base-pair sequences for average nucleosome occupancy both *in vivo* and *in vitro* identified that 5-mers enriched in A and T mononucleotides had the lowest nucleosome occupancy whereas 5-mers composed of C and G mononucleotides had relatively higher nucleosome occupancies. Later, another study determined that G+C content and frequency of AAAA are the most important sequence features explaining nucleosome occupancy *in vitro* such that G+C content explains 50% of variation in nucleosome occupancy (Tillo and Hughes, 2009).

DNA sequences with power to position the nucleosomes had been a great interest for long time. A study identified the most powerful DNA sequences to position nucleosomes by measuring their affinity for histone octamer via *in vitro* nucleosome reconstitution experiments (Lowary and Widom, 1998). Actually, one such sequence from this experiment called “601” is widely used in the studies of chromatin structure. It has been also shown that 10 bp periodic presence of AA/TT/AT and GC

dinucleotides in a given DNA sequence favors nucleosome positioning. It has been suggested that this kind of patterning allows the proper bending of DNA over nucleosomes (Arya et al., 2010).

Given the role of DNA methylation in genome regulation, it affects the nucleosomal organization of the genome. A study in *Arabidopsis thaliana* showed that genomic regions associated with nucleosomes had higher levels of DNA methylation compared to flanking DNA (Chodavarapu et al., 2010). In contrast, by using both *in vitro* and *in vivo* assays, it was also determined that packaging of DNA into nucleosomes was inhibitory to DNA methylation (Felle et al., 2011). In addition, another study identified that CGI-promoters which are DNA methylated are occupied by nucleosomes whereas DNA methylation and nucleosome occupancy were anti-correlated at CTCF binding sites (Kelly et al., 2012). Ultimate relationship between DNA methylation and nucleosome occupancy could be context specific and depend on which comes first.

Action of physical forces plays important roles in nucleosome positioning. Among those, we could think of binding of transcription factors, RNA polymerase II occupancy and enzymatic action of chromatin remodelers. It has been suggested that physical presence of a factor will have an effect on the positioning of neighboring nucleosome, which will further affect the positioning of the next nucleosomes. This type of nucleosome phasing is often referred as “statistical positioning”. It is expected that nucleosome positioning strength decreases as the distance from the physical barrier increases. In this context, RNA polymerase II binding at TSS is one of the best examples, positioning -1 and +1 and a few more downstream nucleosomes at the TSS (Arya et al., 2010). Having a defined sequence binding motif, transcription factor CTCF has also a substantial power for positioning nucleosomes (Ohlsson et al., 2010). In addition, transcription factors could affect nucleosome positioning by competing with nucleosomes itself. At the end, nucleosome positioning might depend on the relative affinities of nucleosomes and transcription factors for a given sequence and relative concentration of the factors (Segal and Widom, 2009). In addition, chromatin remodeling factors which can move nucleosomes to different positions or remove nucleosomes completely by using ATP play important roles in nucleosome positioning. In this respect, SWI/SNF family of remodelers could eject and slide nucleosomes, and ISWI family was found to be involved in positioning of nucleosomes (Arya et al., 2010; Segal and Widom, 2009). Importantly, a study showed that nucleosome positioning around TSS requires chromatin remodeling factors, especially the use of ATP in yeast cells (Zhang et al., 2011). Another study in *Drosophila* mapped genome-wide distribution of main chromatin remodeler families

genome-wide. It was identified that different remodelers might be targeted to distinct DNA sequences and could counteract DNA-sequence driven positioning differentially (Moshkin et al., 2012).

To conclude, it might be argued that intrinsic DNA sequence preferences, transcription factors, chromatin remodelers and other chromatin binding factors have unique and essential roles in nucleosomal organization of the genome. At the end, combinatorial action of all these features might be necessary to create proper chromatin configuration in different cell types and distinct cellular processes.

1.2.3.2 Post-translational histone modifications

N-terminal tails of the histone tails protrude from the nucleosome core and they are post-translationally modified. These modifications include methylation, acetylation, phosphorylation, ubiquitination, sumoylation and some others. Initially suggested by Jenuwein and Allis, histone modifications could create a “histone code” which would be involved in regulation of diverse cellular processes (Jenuwein and Allis, 2001). There are writer and reader proteins which deposit the modification on histones and recognize the deposited mark, respectively (Justin et al., 2010). Methylation of lysine 4 and 27 on histone H3 are the best studied examples of histone modification mediated gene regulation and the details of function of these modifications will be discussed in more detail in section 1.2.3.3. Below, I will introduce function of several histone modifications and molecular machineries involved in deposition and recognition of the mark, with a special focus on histone methylation.

- Histone acetylation

Histones could be acetylated and deacetylated by histone acetyltransferases (HATs) and histone deacetyltransferases (HDACs), respectively. One of the most conserved functions of histone acetylation is di-acetylation of lysines 5 and 12 on newly synthesized H4 which will be deposited into newly replicating DNA. Acetylation marks are recognized by histone chaperones and an ordered chromatin structure is ensured. Immediately after deposition onto DNA, histones become deacetylated to allow new patterns of histone acetylation (Shahbazian and Grunstein, 2007; Turner, 2000).

Histone acetylation is also involved in chromatin compaction, heterochromatin spreading and transcription. Acetylation neutralizes the positive charge on histone tails which would lead to disruption of electrostatic interactions between histones and DNA, thus favoring a looser chromatin structure. In yeast, deacetylation of histone

tails has been shown to be involved in heterochromatin formation spreading via recognition of deacetylated forms by Sir proteins. As acetylation creates a more open chromatin environment, it promotes transcription by facilitating the binding of transcriptional machinery at promoter regions. In addition, acetylation could be recognized by bromodomain-containing proteins, which could activate transcription. On the contrary, as expected histone deacetylation plays a role in gene repression (Shahbazian and Grunstein, 2007).

- Methylation of histone H3 on lysine 4

H3K4me1 is often associated with enhancer function. H3K4me2 and H3K4me3 have been associated with transcriptional activity and they could also be found at promoter regions of genes poised for activation. Actually, it was also shown that H3K4me3 is found almost at all of high CpG-content promoters of the genes regardless of their expression (Greer and Shi, 2012; Zhou et al., 2011). In this respect, it was identified that Cfp1 which has an unmethylated CpG binding CXXC domain establishes H3K4me3 by recruiting H3K4 methyltransferase Setd1 (Thomson et al., 2010). Plant-homeo-domain (PHD) containing proteins plays and essential role in recognition of H3K4me3 and directing further downstream events (Justin et al., 2010).

The first enzyme catalyzing H3K4me is Set1 identified in yeast (Miller et al., 2001). Mammalians have around 10 known H3K4methyltransferases. Six of them, Set1a, Set1b, Mixed-lineage leukemia (Mll) /Trithorax family members Mll 1-4 are close homologs of Set1 identified in yeast and function in complexes called COMPASS (Hublitz et al., 2009). In addition, there are enzymes which demethylate H3K4me. It was identified that LSD1 demethylates H3K4 and is involved in gene repression (Shi et al., 2004). Nevertheless, LSD1 does not only demethylate H3K4, as it was shown later that it could function as a H3K9 demethylase as well (Metzger et al., 2005).

- Methylation of histone H3 on lysine 36

H3K36me is best known for its role in transcription. It has been found that transcribed genes had a progressive shift from H3K36me1 to H3K36me3, starting from promoter regions of the genes and going through 3’ of genes. H3K36 methyltransferase Set2 is involved in H3K36me during transcriptional elongation by associating with hyperphosphorylated form of RNAP II. H3K36me3 mark prevents abnormal transcription along the coding sequences by recruiting a histone

deacetylase complex. Protein complexes with members having a PWWP domain recognize H3K36me (Wagner and Carpenter, 2012).

In addition, H3K36me has been implicated in transcriptional repression in certain cellular contexts and several organisms (Wagner and Carpenter, 2012). Nevertheless, the exact role of H3K36me in silencing and whether actually it is required to establish or maintain a silent state need to be identified.

- Methylation of histone H3 on lysine 27

H3K27 methylation is processive. H3K27me1 is involved in constitutive heterochromatin formation. In plants, H3K27me1 is deposited by ATRX5 and ATRX6, which are not conserved in mammals. Although in mammals, H3K27me1 is detectable, it is not clear how this mark is initially created; it could be deposited by an unknown enzyme or could be created via demethylation of H3K27me2/3. H3K27me2/3 is catalyzed by enzymatic subunits, EZH1 and EZH2, of the Polycomb repressive complex 2 (PRC2). H3K27me3 is the stable mark and H3K27me2 has been seen as an intermediate and does not have a significant role in gene regulation (Margueron and Reinberg, 2011).

H3K27me3 is associated with gene repression and silenced chromatin formation. Chromodomain-containing CBX proteins, which are the subunits of Polycomb repressive complex 1 (PRC1), bind to H3K27me3 (Hublitz et al., 2009). Importantly, it has been also shown that EED which is another PRC2 member binds to H3K27me3 mark and this binding is required for the propagation of H3K27me3 (Margueron et al., 2009). H3K27 could also be demethylated. Histone demethylases UTX and JMJD3 specifically demethylates H3K27me. These demethylases are important in differentiation and development and their absence could create different defects (Hublitz et al., 2009).

- Methylation of histone H3 on lysine 9

H3K9me is involved in transcriptional silencing, DNA methylation and heterochromatin formation. H3K9me is mainly catalyzed by G9A, SETDB1, PRMD2 and SUV39H1-2. Lysine demethylases, JHDM2 (KDM3), JMJD2 (KDM4) and PHF8 are involved in demethylation of H3K9me (Krishnan et al., 2011).

H3K9me positively correlates with DNA methylation, and for several processes, it has been demonstrated that loss of H3K9me could lead to loss of DNA methylation as well (Cheng and Blumenthal, 2010). H3K9me2 is mainly associated with facultative heterochromatin and has a role in processes such as X chromosome inactivation (Trojer and Reinberg, 2007). Furthermore, it is enriched at lamina-

associated domains (LADs) in the nucleus (Guelen et al., 2008). Recently, it was also identified that globally H3K9me2 is found at transcriptionally silent genomic regions both in ESCs and differentiated cells with similar levels (Lienert et al., 2011). On the other hand, H3K9me3 is more involved constitutive heterochromatin formation, and represented within centromeric, pericentromeric and telomeric domains. At this point, recognition of H3K9me3 by chromodomain-containing HP1 plays an important role in chromatin compaction and higher order organization (Campos and Reinberg, 2009). In addition, it was also determined that H3K9me3 could mark olfactory receptor genes in the olfactory epithelium (Magklara et al., 2011).

1.2.3.3 Polycomb – Trithorax system and epigenetic memory

During initial hours of *Drosophila* embryonic development, expression levels of homeotic genes depends on the local concentrations of activators and repressors present in the embryo in a position specific manner. Once the homeotic gene expression patterns are established, expression states of the genes are maintained throughout the development without further need of activators or repressors. It has been shown that Polycomb group (PcG) and and Trithorax group (TrxG) of proteins are involved in the maintenance of silent or repressed states via associating with Polycomb/Trithorax group response elements (PRE/TREs), respectively (Ringrose and Paro, 2007).

PcG proteins are highly conserved proteins from *Drosophila to* mammals. To date, 5 different PcG complexes were discovered in various organisms. PRC1, PRC2 are the major characterized PcG complexes (Lanzuolo and Orlando, 2012).

PRC1 consists of CBX proteins, BMI1, MEL18, PHC proteins, RING1A, RING1B, and some others. PRC1 catalyses lysine 119 monoubiquitination on histone H2A (H2AK119Ub1) via E3 ligase activity of RING1B, and to some extent via RING1A (Lanzuolo and Orlando, 2012). In vitro, PRC1 complex has been shown to repress transcription and compact nucleosomal arrays (Simon and Kingston, 2009). In addition, Ring1A, Ring1B and H2AK119Ub1 mark have been shown to be involved in poising of RNAPII at developmental promoters in mouse embryonic stem cells (Lanzuolo and Orlando, 2012).

PRC2 complex mainly consists of EZH1, EZH2, SUZ12, EED, RbAp46/48 and JARID2. As already mentioned, EZH1/2 catalyzes H3K27me2/3. JARID2 via its binding ability to GC rich DNA could facilitate loading of PRC2 onto chromatin. Nevertheless, findings which show context dependent roles of JARID2 on PRC2 regulation make the exact role of JARID2 unclear (Lanzuolo and Orlando, 2012; Margueron and Reinberg, 2011).

Although TrxG proteins are implicated in many cellular processes, still it is a bit unclear whether main function of TrxG proteins is to antagonize PcG or activate gene expression in general. TrxG proteins are evolutionarily conserved, and consist of three main classes. First class is SET domain-containing enzymes which catalyze H3K4me3. These include MLL group of proteins as already discussed under the section describing function of H3K4me3. The second class consists of ATP-dependent chromatin remodeling complexes. Among those, there are SWI/SNF family, ISWI family which recognize H3K4me3 via their PHD domain, and chromodomain helicase DNA-binding (CHD) family which recognizes H3K4me3 via their chromodomains. Finally, there is a third class which recognizes DNA directly. These proteins are targeted to TRE elements in *Drosophila*. However, no TRE elements have been discovered in mammals. In mammals, recognition of DNA elements could be achieved by recognition of CpG islands via CXXC-domain containing proteins, which could further target H3K4 methyltransferases (Schuettengruber et al., 2011).

How are PcG and TrxG complexes targeted?

Exact mechanism of targeting of PcG and TrxG proteins to the chromatin is not known. There are several lines of evidence which hints about targeting, nevertheless since in general one mechanism does not explain the targeting of whole complexes, it is believed that combinatorial and context dependent mechanisms recruit PcG and TrxG complexes.

In *Drosophila*, PcG proteins bind to PRE, however in mammals, up to date to specific and global DNA element for targeting has not been identified yet except for PRE-like sequences identified only for a few loci (Sing et al., 2009; Woo et al., 2010). Genome-wide studies showed that PcG localizes to CGI. PRC2 component JARID2 binds GC-rich sequences; nevertheless it may not be fully responsible from targeting of PcG to CGI. In addition, PRC1 component RYBP could interact with YY1, transcription factor component of PhoRC complex, and is involved in recruitment of PRC1 and PRC2. Actually, recently long non-coding RNA (ncRNAs) has been shown to be play important roles in PcG targeting. Long nc-RNA could mediate PcG spreading and promote targeting. A genome-wide study identified over 9000 RNAs which were associated with PRC2 component EZH2, which shows how complicated PcG targeting could be (Lanzuolo and Orlando, 2012; Margueron and Reinberg, 2011).

Similar to PcG, in *Drosophila* TrxG proteins are targeted to TRE elements. However, in mammals no TRE element has been identified yet. As already

described, TrxG proteins could be targeted to CGI via CpG binding protein CXXC1 (Schuettengruber et al., 2011). Furthermore, already existing H3K4me3 might be involved in targeting of MLL via binding of PHD domain of MLL to H3K4me3 (Milne et al., 2010). Role of ncRNAs has been also shown for TrxG recruitment and gene activation at single loci basis, nevertheless it remains to be determined whether this is a general mechanism (Schuettengruber et al., 2011).

Role of PcG during development, pluripotency and differentiation

There have been many studies investigating the role of PcG during early embryonic development by using mouse models. *Ezh2* null mutant mice die early post-implantation and show gastrulation defects (O'Carroll et al., 2001). Similarly disruption of other PRC2 components Eed and Suz12 in mice results in early post-implantation lethality (Faust et al., 1995; Pasini et al., 2004). In addition, zygotic depletion of PRC1 component, *Ring1b* results in lethality during gastrulation (Valk-Lingbeek et al., 2004), and embryos maternally deficient for *Ring1b* and *Ring1a* can not develop beyond 2-cell stage (Posfai et al., 2012).

Role of PcG proteins has been also implicated in maintenance of pluripotency and differentiation. Many studies by using ESCs where components of PcG complexes are absent showed that actually PcG function is not required for the maintenance of pluripotency. However, these cells experience problems when they are induced to differentiate. It has been proposed that PcG function is required to prevent aberrant expression of developmental regulators and impairment of this system could interfere with execution of correct developmental programs and lineage commitment (Margueron and Reinberg, 2011; Vastenhouw and Schier, 2012).

Bivalent chromatin states

Genome-wide studies in ESCs showed that promoter regions of developmental regulators are marked by both H3K4me3 and H3K27me3 and they are referred as “bivalent” domains. The function of such domains has been a debate for long time. It has been proposed that presence of H3K27me3 ensures repression of lineage-specific genes and marking by H3K4me3 poises lineage specific genes for activation as shown in Figure 2 (Vastenhouw and Schier, 2012).

Since H3K4me3 and H3K27me3 mappings come from a population of cells, co-association of H3K4me3 and H3K27me3 may simply result from cellular heterogeneity. Nevertheless, several studies showed the existence of bivalent domains in zebrafish blastomeres and mouse epiblast cells, although no bivalent domains were found in *Xenopus* blastomeres and *Drosophila* embryos (Vastenhouw

and Schier, 2012). It was also shown that PRC2 activity is inhibited by H3K4me3 only if they are on the same tail (Schmitges et al., 2011). Another recent study showed the existence of H3K4me3 and H3K37me3 within the same nucleosome (Voigt et al., 2012), further supporting the occurrence of bivalency.

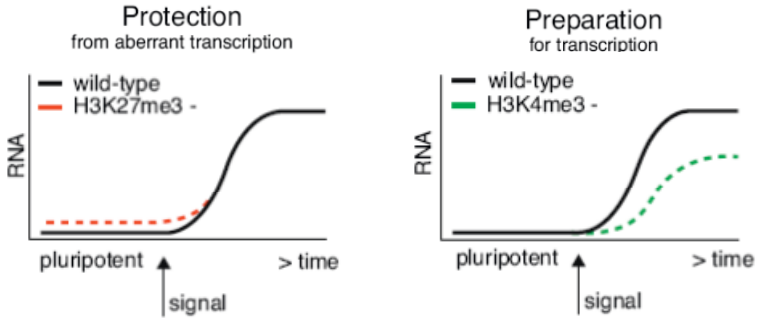


Figure 2: Model for the role of H3K27me3 and H3K4me3 in pluripotency and differentiation. H3K27me3 protects aberrant transcription of lineage-specific genes before differentiation signal. Marking by H3K4me3 ensures the efficient induction of gene expression during differentiation (modified from (Vastenhouw and Schier, 2012), permission to use: license number: 3124290173575).

One apparent question is that how bivalent states are initially established? Studies performed in sperm of human, mouse and zebrafish (Brykczynska et al., 2010; Hammoud et al., 2009; Wu et al., 2011) showed that developmental regulators were already bivalent and transmission of these marks might be responsible from bivalent domains in ESCs. Nevertheless, it has been also argued that these marks might be erased after fertilization and re-established during early embryonic development. Currently, it is still a technical challenge to perform high-throughput analysis with early embryos and oocytes because of limited sample sizes. Future technologies could allow the identification of similarities / differences in the chromatin states of sperm, oocyte and early embryo.

1.2.3.4 Histone variants

Canonical histone genes are found in clusters in the genome and their synthesis is dependent on DNA synthesis coupled processes such as DNA replication and repair. There are other histone genes, which are in general found in the genome as single copy, and could be constitutively expressed, these are called as “non-canonical” histone variants (Talbert and Henikoff, 2010). Below, I will introduce the function and genomic localization of main histone variants, with a special focus on histone H3 variants.

H3 variants

In mammals, 5 H3 variant have been described until now: Canonical H3 variants H3.1 and H3.2, and non-canonical variants, H3.3, centromere-specific H3 variant CENP-A, and testis specific H3 variant H3t. There are also two recently identified primate specific H3 variants, H3.X and H3.Y. Alignment of these variants is shown in Figure 3 (Szenker et al., 2011).

Centromere-specific H3 variant CENP-A is involved in the assembly of kinetochore structure and it shows only 50-60% identity with canonical H3s (Talbert and Henikoff, 2010). H3.X and H3.Y have been identified in a study by re-annotation of two genes similar to H3.1 initially annotated as pseudogenes. These variants are expressed in some human cell at low levels and stress conditions could increase their expression (Wiedemann et al., 2010). H3t is a testis specific histone variant with a specific expression in spermatocytes (Witt et al., 1996). Nevertheless, the exact role of this variant is not known yet.

H3.3 is the most studied H3 variant and it functions in distinct cellular processes. H3.3 protein is produced from two different mRNAs H3.3A and H3.3B with polyA tails. H3.3A and H3.3B genes are different in their nucleotide sequence but give rise to same protein (Szenker et al., 2011).

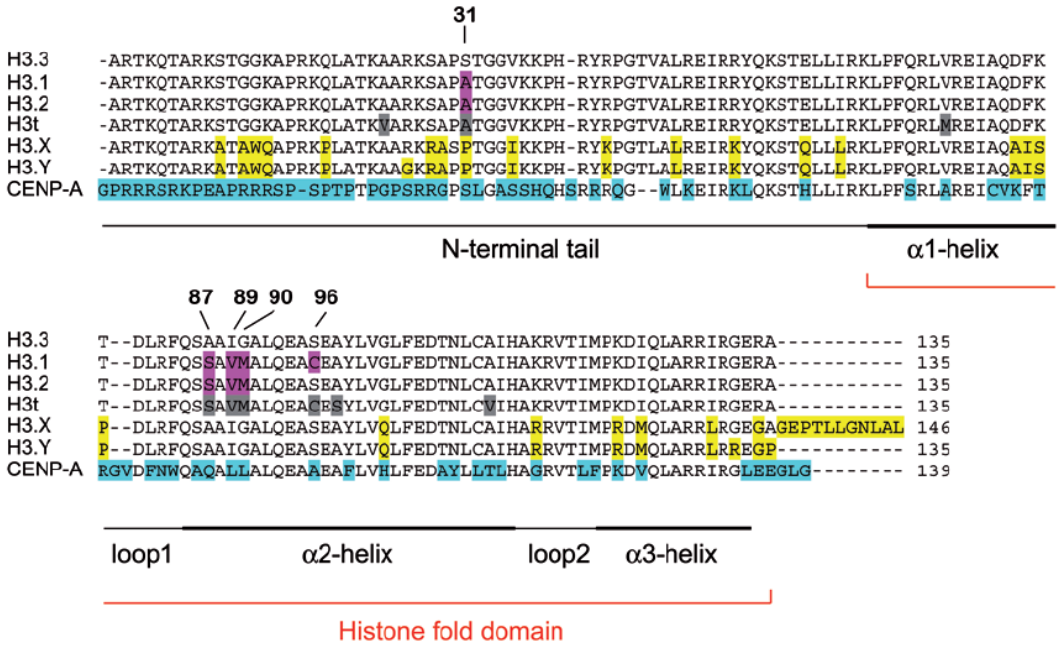


Figure 3: Alignment of histone H3 variants in human. Major amino acid changes and histone domains are indicated (adapted from (Szenker et al., 2011), permission to use: license number: 3124290407528)

One of the most prominent roles of H3.3 is in transcription. A genome-wide study in ESCs by using tagged H3.3 showed that H3.3 is enriched around TSS of high-CpG content promoters independent of transcription with slight variations depending on the method of chromatin preparations as native or fixed. Nevertheless, expressed genes are marked by increasing levels of H3.3 along the gene bodies and transcription end sites. H3.3 chaperone Hira has been shown to be required for enrichment of H3.3 at genic regions (Goldberg et al., 2010).

Several studies suggested that H3.3 containing nucleosomes especially when they are together with H2A variant H2A.Z are sensitive to salt disruption and form unstable nucleosomes and enriched at nucleosome free regions of active genes (Jin and Felsenfeld, 2007; Jin et al., 2009). Nevertheless, another study using sedimentation techniques at different salt concentrations showed that neither H3.3 nor H2A.Z or their combination affected nucleosome stability significantly (Thakar et al., 2009). In addition, since the regions occupied by H3.3 are more accessible and GC rich, the reason why H3.3 containing nucleosomes would be more susceptible to disruption *in vivo* might be caused of accessibility issues.

In a study investigating interaction partners of H3.3, two proteins Atrx and Daxx have been identified as chaperones of H3.3 and both of these proteins localize to heterochromatin. Atrx-Daxx complex is required for H3.3 deposition at telomeric regions. The exact role of H3.3 deposition at heterochromatic loci is not entirely clear. Nevertheless, S31 phosphorylation is unique to H3 variant H3.3 and S31 phosphorylation on H3.3 is detected on several heterochromatic regions (Goldberg et al., 2010; Lewis et al., 2010).

H3.3 is also required for fertility. In *Drosophila*, null mutations of two H3.3 results in reduced viability and complete sterility both in females and males. Importantly, male meiosis is impaired and chromatin reorganization can not happen properly (Sakai et al., 2009). In mice, 50% animals with homozygous mutation in H3.3A die immediately after fertilization, and surviving animals especially males have reduced fertility (Couldrey et al., 1999).

The majority of sperm DNA comes into embryo packed in protamines. Upon fertilization, protamines are rapidly exchanged by maternally provided H3.3 (Elsaesser et al., 2010). Furthermore H3.3 play a role in the remodeling of chromatin during meioic sex chromosome inactivation, and this will be discussed in more detail in section 2.2, Box1.

As already pointed out, canonical histone H3s, H3.1 and H3.2 are incorporated in chromatin in DNA replication/repair coupled way (Szenker et al., 2011). Analysis of chromatin from post-replicative cells such as differentiating

neurons, activated lymphocytes and myotubes (Bosch and Suau, 1995; Wu et al., 1983; Wunsch and Lough, 1987) showed that differentiated non-proliferating cells accumulate H3.3 and lose canonical H3 variants H3.1 and H3.2 gradually.

H3.3 has been shown to play a role in cancer also. Sequencing of the exomes from glioblastoma patients showed that nearly 30% of the patients had mutations on H3.3 especially on histone tails critical for histone modifications, implicating the role of chromatin organization in cancer (Schwartzentruber et al., 2012).

H2A variants

H2A has two main variants: H2A.Z and H2A.X. H2A.Z has been shown to function in a variety of processes, from transcription to heterochromatin formation. In yeast, it was identified that H2A.Z could positively regulate transcription by interacting with transcriptional machinery, and H2A.Z is positioned directly upstream and downstream of TSS of all genes in euchromatin (Adam et al., 2001; Raisner et al., 2005). Similar to this finding, it was determined that H2A.Z is positioned in the -3, -2, +1, +2, and +3 nucleosomes surrounding TSS of active genes (Schones et al., 2008). Furthermore, the role of H2A.Z in recruitment of RNAPII has been also shown in human cells (Hardy et al., 2009). It has been suggested that H2A.Z randomly accumulates in the genome and transcription depletes H2A.Z from the body of transcribed genes, and conversely H2A.Z accumulates at the genes found in heterochromatin because of the lack of transcription (Hardy et al., 2009). Enrichment of H2A.Z at pericentric heterochromatin during early mammalian development and its potential function in centromere formation have been also described (Greaves et al., 2007; Rangasamy et al., 2003). It is apparent that there might be contradictions for the function of H2A.Z. Differential functioning of this variant might depend on post-translational histone modifications, interaction with chromatin remodeling complexes (Talbert and Henikoff, 2010).

H2A variant H2A.X mainly functions in double-strand breaks and DNA repair. Upon double-strand breaks, H2A.X gets phosphorylated and this phoshorylated form could recruit DNA repair proteins and chromatin remodeling complexes (Talbert and Henikoff, 2010). The roles of H2A variants MacroH2A and H2A.Bbd in spermatogenesis are discussed under section 2.2.

1.2.3.5 Inheritance of chromatin states over cell cycle

Although packaging of DNA into nucleosomes, post-translational histone modifications and histone variants regulate genome function at multiple leves, the

exact mechanism how histones are inherited during cell division is not known. There are two main hypotheses. One hypothesis is that histones are distributed onto DNA strands in a semi-conservative manner. With this hypothesis, it is expected that histone modifications should exist symmetrically in nucleosomes and there should be a machinery which could orderly distribute histones on leading and lagging strands. These requirements make semi-conservative model a bit unlikely. Second hypothesis claims that histones are randomly distributed to DNA strands during replication, and mechanistically this seems more acceptable. Nevertheless, with this model for a histone modification to be transmitted effectively, it should be enriched on several adjacent nucleosomes (Margueron and Reinberg, 2010).

It has been shown that some chromatin modifiers attach to replication machinery to via PCNA and this could enable replication coupled deposition of histone marks (Zhu and Reinberg, 2011). A recent study showed that PRC1 can stay on chromatin and bridge nucleosomes *in vitro* (Lo et al., 2012). In the context of replication coupled deposition of histone marks, one of the main questions is that whether all histone marks are copied directly upon replication or it takes gradually in an replication uncoupled way. Future technologies could allow identification of which histone/histone marks needs to be deposited in a replication coupled way (Zhu and Reinberg, 2011).

1.2.4 RNA based epigenetic mechanisms

Recent high-throughput data showing that almost 65% of the mammalian genome is transcribed and the fact that only 2% of mammalian genome is coding for protein-coding genes has significantly broadened our view about role of noncoding RNA in genome regulation (Wery et al., 2011).

Small noncoding RNAs

Small noncoding RNAs include transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and 19-30 nt long RNAs involved in RNA interference. Small interfering RNAs could be classified in 3 main classes: short-interfering RNAs (siRNAs), micro RNAs (miRNAs) and PIWI-interacting RNAs (piRNAs) (Wery et al., 2011).

miRNAs were first discovered in *C.elegans* during 90s as small RNAs regulating development of this organism. Later, miRNAs were also identified in other animals and plants and has been shown to be involved in differentiation, apoptosis and proliferation. miRNAs are initially transcribed into primary transcripts (pri-miRNA), which is processed in the nucleus forming the pre-miRNA and transported

into cytoplasm where it is cleaved by Dicer to generate 22 nt length miRNAs. miRNAs, which are complementary to their target mRNA 3` UTRs, function in gene regulation by interfering with translational initiation machinery, affecting mRNA stability or degradation of mRNAs (Ul Hussain, 2012).

siRNAs can be classified in two groups: exogenous siRNAs and endogenous siRNAs. Exogenous siRNAs are produced from exogenous dsRNA coming from viruses, transgenes, and are 21-25 nt long. Cleavage of exogenous dsRNA by Dicer produces siRNAs functioning in gene silencing, heterochromatin formation or antiviral silencing. Endogenous siRNAs has been described in plants, mammals and some other eukaryotes and could arise from dsRNA originated from different genomic regions. Although the general function of endogenous siRNAs is not known, they have been shown to function in silencing of transposons in flies and direction of DNA methylation and histone methylation in plants (Li and Liu, 2011).

piRNAs are longer than siRNAs and miRNAs, generally between 24-32 nt length. piRNAs are highly enriched in germ cells (Li and Liu, 2011). piRNAs expressed during spermatogenesis interact with PIWI-family of proteins, MIWI, MIWI2 and MILI. piRNAs are expressed in spermatocytes and spermatids, and mainly function in suppression of retrotransposon expression (Bak et al., 2011).

Long noncoding RNAs

Long noncoding RNAs (lncRNAs) are generally defined as RNAs longer than 200 bp and do not have any protein coding potential (Rinn and Chang, 2012). Initially lncRNAs gained attention in connection to their role in X chromosome inactivation and genomic imprinting (lncRNAs, Xist, Air, H19). Nevertheless, recent high-throughput techniques revealed that function of lncRNAs go beyond these functions. As already described, lncRNAs have been shown to interact with PcG complexes and they function in gene repression and gene activation by interacting with TrxG proteins (Lee, 2012). One of the main questions in lncRNA field is that how lncRNAs exert their function. In this context, several hypotheses have been put forward: lncRNAs bind to DNA binding factors, they act as scaffolds to bring protein complexes in close proximity, they recruit chromatin modifying enzymes through RNA:DNA interactions and lncRNA could be targeted via chromosomal looping (Rinn and Chang, 2012).

1.3 Male germline

1.3.1 Primordial germ cells

PGCs arise from proximal epiblast cells during embryonic day (E) 7.25 as approximately 40 cells. Bone morphogenetic protein (Bmp) signals generated by extraembryonic ectoderm have been shown to be required for specification of PGCs. Importantly, B-lymphocyte-induced maturation factor 1 (Blimp1) was determined to be expressed specifically in founder PGCs around E6.25 (Saitou, 2009). Blimp1 is a transcriptional repressor and represses somatic cell genes particularly Hox gene family while enhances expression of PGC specific genes (Richardson and Lehmann, 2010). PR-domain containing protein (Prmd14) is another regulator functioning in the specification of PGCs. It starts to be expressed in Blimp1-positive PGCs during E6.25 and continues to be expressed until E13.5-14.5 (Saitou, 2009).

Starting from E7.5, PGCs start to move from primitive steak to endoderm, reach to hindgut at E8-9.5, and finally migrate to genital ridges at E10.5-11.5 (Richardson and Lehmann, 2010) and their migration is completed by E13.5 (Jan et al., 2012).

At the beginning of post-implantation development and before PGCs are specified, epiblast cells undergo changes in DNA methylation and chromatin compaction. Nevertheless, once the PGCs are specified from epiblast cells, these changes need to be reprogrammed to ensure acquisition of totipotency (Magnusdottir et al., 2012). There are epigenetic reprogramming events taking place during migration of PGCs. Genome-wide there is reduction of H3K9me2 while there is global gain of H3K27me3. There is also genome-wide DNA demethylation as discussed in section 1.2.2. In addition, X chromosome is reactivated in female PGCs. (Magnusdottir et al., 2012; Saitou, 2009).

Once PGCs reach at genital ridges, they are called as “gonocytes” (Jan et al., 2012). Around E10, there are no differences in female and male genital ridges and they have bi-potential to form female or male gonads. Sex-specific gonad development starts with the expression of sex-determining region on the chromosome Y (*Sry*) in somatic cells of XY genital ridges during E10-10.5. Sry upregulates Sox9, which is involved in differentiation of Sertoli cells (supporting germ cells in testis) and expression of male specific genes. Testis cords are formed by E12.5. In XX genital ridges, as Sry is not present, female-specific genes starts to be expressed at E11.5-12.5 and ovary development is initiated (Kashimada and Koopman, 2010).

Around E13.5, female germ cells stop proliferation in the developing ovary, enter meiosis and they are arrested at prophase-I around birth. At puberty, arrested oocytes restart and complete meiosis I and they arrest at meiosis II. Meiosis II is completed after fertilization. On the other hand, male germ cells arrest at G1 phase

of mitotic cycle at E13.5. Immediately after birth, they restart proliferation and after a few days, they enter in meiosis to produce spermatids which will differentiate into sperm (Bowles and Koopman, 2007).

1.3.2 Spermatogenesis

1.3.2.1 Proliferative phase

Spermatogenesis is a highly coordinated process that continuously generates mature spermatozoa. It is characterized by three different phases. The first phase is the proliferative phase in which spermatogonia undergo a series of mitotic divisions and differentiate into primary spermatocytes. In the second meiotic phase, exchange of genetic information by recombination takes place and haploid spermatids are produced. The third, post-meiotic phase termed spermiogenesis involves morphogenetic events, which are required for the production of mobile mature sperm. All stages of spermatogenesis are completed in seminiferous tubules and interaction of germ cells with supporting Sertoli cells creates seminiferous epithelium (Oatley and Brinster, 2008).

In mice, spermatogonial stem cells arise from gonocytes during day 0-6 after birth (Oatley and Brinster, 2008). Spermatagonial stem cells present at the basement of seminiferous tubules are called A single spermatogonia (A_s). A_s could either divide to generate new spermatogonial stem cells or enter in differentiation to produce spermatogonia (Jan et al., 2012).

Differentiation of A_s results in A paired spermatogonia (A_{pr}). Apr undergoes additional mitotic divisions to form 4, 8, and 16 A aligned spermatogonia (A_{al}) which could differentiate into A_1 spermatogonia (Oatley and Brinster, 2008). Importantly, starting from differentiation of A_s and onwards, cytokinesis is not complete and produced spermatogonia become connected by an intercellular bridge (Jan et al., 2012).

Further divisions of A_1 spermatogonia give rise to A_2, A_3 and A_4 spermatogonia. A_4 spermatogonia finally differentiate into type B spermatogonia (Oatley and Brinster, 2008). Mitotic division of type B spermatogina gives rise to pre-leptotene spermatocytes which enters in meiosis (Jan et al., 2012).

1.3.2.2 Meiotic phase

Spermatocytes after going through S and G2 phases enter in meiotic prophase I. Initiation of meiosis depends on retinoic acid. RNA-binding protein Dazl is involved in response of spermatocytes to retinoic acid, which triggers the expression of *Stra8*. Stra8 is required for initiation of meiosis as Stra8 deficient mice can not

produce later stage spermatocytes and spermatids (Jan et al., 2012). Because meiotic prophase I consists of many critical steps required for the formation of functional germ cells, it is divided into 4 stages. Without going into details of what is happening exactly in each stage, formation of synapsis, meiotic recombination and meiotic sex chromosome inactivation (MSCI) are the major events happening during meiotic prophase I. The details of meiotic recombination and MSCI are discussed in section 2.2, Box1 and Box2, respectively. Completion of meiotic divisions I and II gives rise to haploid round spermatids which would enter into spermiogenesis to produce mature spermatozoa (Ahmed and de Rooij, 2009).

Proper organization of the chromatin during meiotic prophase I is also critical for progression into later stages of meiosis. Mice lacking H3K9 methyltransferases *Suv39h1* and *Suv39h2* can not go beyond meiotic prophase I stage because of defects associated with lack of H3K9me3 at pericentric heterochromatin (Peters et al., 2001). In addition, deficiency in another H3K9 methyltransferase *G9a* impairs progression into meiosis (Tachibana et al., 2007). Meiosis is also affected by changes in DNA methylation. Deficiency in Dnmt3L results in transcriptional activation of retrotransposons and nonhomologous synapsis. Furthermore, absence of MILI, a member of PIWI-like proteins, causes decreased levels of DNA methylation and activation of LINE1 retrotransposons and impairing meiosis (Kota and Feil, 2010).

1.3.2.3 Spermiogenesis

Spermiogenesis is subdivided into 16 substeps. Because of the overlapping appearance of the first 8 and last 4 stages, cycle of seminiferous epithelium is divided into 12 stages. Duration of the cycle of seminiferous epithelium has been determined to take 8.6 days with different length of time spent in each stage (Ahmed and de Rooij, 2009).

Morphological events

There are a number of morphological changes happening during differentiation of spermatids into sperm. Development of sperm tail, required for motility and composed of a microtubular structure called axenome, starts in round spermatids and completed towards the end of spermiogenesis. During nuclear elongation, a ring like microtubular structure called manchette surround the base of nucleus and is involved in the development of sperm head. Cytoplasm is removed to ensure the compact structure of sperm (Jan et al., 2012).

Chromatin remodeling events

During development of round spermatids into mature sperm, there are extensive chromatin remodeling events which happening at multiple steps (Gaucher et al., 2010) (Figure 4). One of the major events is genome-wide exchange of histones with protamines to ensure the compact packaging of sperm DNA. In addition, post-translational histone modifications and incorporation of histone variants have been shown to play essential roles during genome-wide remodeling during spermiogenesis, and their role is discussed in more detail in section 2.2 under the subtitle “Global chromatin remodeling during spermiogenesis”. Although, there are multiple pieces of data addressing the chromatin organization in spermiogenesis, actually the details and the order of events taking place are not known. Below, I will introduce the function of transition proteins, protamines, DNA strand breaks occurring during chromatin remodeling and genomic regions surviving genome-wide exchange of histones with protamines and still keeping nucleosomes.

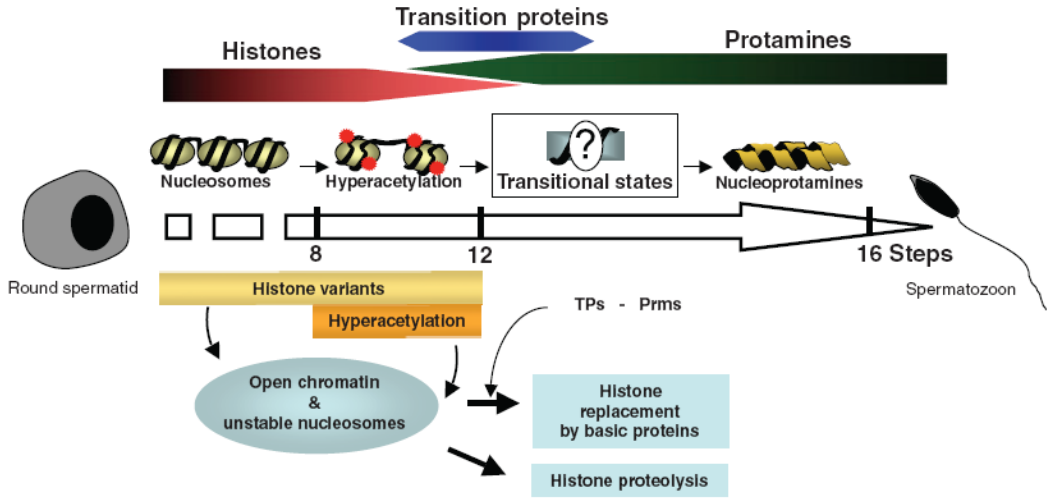


Figure 4: Chromatin remodeling during spermiogenesis. Genome-wide histones are initially exchanged by transition proteins which will be finally replaced by protamines. Hyperacetylation of histones creates a more accessible chromatin environment which facilitates histone-to-protamine exchange (Adapted from (Gaucher et al., 2010), permission to use: license number: 3124300488355)

Role of transition proteins

In mammals, histones are not directly replaced by protamines, but initially replaced by transition proteins (TPs) (Gaucher et al., 2010). There are two TPs, TP1

and TP2. TP mRNAs could be first detected in early round spermatids (Heidaran and Kistler, 1987; Heidaran et al., 1988). After their synthesis, TPs are not translated immediately. TP1 was determined to be translated 3-4 days after its synthesis in elongating spermatids (Heidaran and Kistler, 1987). At the protein level, they are first detected in step 10-11 spermatids (Alfonso and Kistler, 1993; Heidaran et al., 1988) and they reach maximal level of expression during step 12-13 spermatids constituting the 90% of chromatin basic proteins. TP1 and TP2 are structurally different from each other and TP1 is approximately 2.5 more times more abundant than TP2. TP1 is around 6.5 kDa and TP2 is 13 kDa (Yu et al., 2000). Although TP1 is a DNA melting protein (Singh and Rao, 1987), TP2 is zincmetalloprotein and it has the ability to bind to CpG-rich sequences in a specific manner (Kundu and Rao, 1996). In addition, TP2 is rapidly phosphorylated soon after its synthesis but TP1 is not (Green et al., 1994). It has been proposed that phosphorylation of TP2 soon after its synthesis initially inhibits its ability to condense chromatin. TP2 first binds to CpG-rich sequences with its zinc fingers and subsequently with dephosphorylation of TP2, chromatin condensation is initiated (Meetei et al., 2002). In addition, it was shown that TP1 and TP2 condense DNA in vitro and binding of TP2 to DNA was 6 times more stronger (Brewer et al., 2002). In the context of CpG-rich specific binding of TP2 to DNA, it was identified that GC versus AT-rich DNA show mutually exclusive localization pattern in elongating spermatids (Kolthur-Seetharam et al., 2009). HSPA2 has been determined to be a chaperone of TP1 and TP2, but the functional role of this chaperone still needs to be addressed (Govin et al., 2006). In addition, it was shown that TP2 could interact with histone chaperone nucleoplasmin 3 (Npm3) and acetylation of TP2 could negatively affect its interaction with Npm3 (Pradeepa et al., 2009). In the context of nuclear transport of TPs, Importin-4 is involved in nuclear transport of TP2 via interaction with its nuclear localization signal, while TP1 is transported to nucleus with passive diffusion (Pradeepa et al., 2008).

TPs are not fully required for the development of functional sperm. Mice either lacking TP1 or TP2 do not have significant abnormalities although there seems to be reduced fertility (Adham et al., 2001; Yu et al., 2000; Zhao et al., 2001). In mice, lacking both TPs, nuclear shaping and histone-to-protamine exchange takes place normally nevertheless, there are defects associated with chromatin condensation, protamine 2 remains as uncleaved precursor and mice are sterile (Zhao et al., 2004). Although overall effects associated with TPs seemed to be a gene dosage effect, there could be some functional differences.

In fact, it is not known what exactly transition proteins are doing in chromatin remodeling events during spermiogenesis. For instance, there is not solid data which

shows that they are first replacing histones and finally they are replaced by protamines.

Role of protamines

Protamines are a group highly basic, arginine rich family of proteins and they are involved in compact packaging of sperm DNA in many species. In mammals, there two main protamines, protamine 1 (P1) and protamine 2 (P2). P2 is only found in primates and many rodents (Balhorn, 2007). Protamines are transcribed in round spermatids (step 7-9) and their mRNAs are stored in cytoplasm for about 1 week until they are translated in elongating spermatid stage (Stern et al., 1983). 3`UTR of protamine mRNA is important for their timed and efficient translation (Braun et al., 1989). In mice carrying a transgene for protamine 1 lacking its regular 3`UTR, protamine 1 mRNA is prematurely translated in round spermatids (step 7-8), whereas normally it is first detected in step 12 spermatids. Furthermore, differentiation of spermatids is blocked and mice are infertile (Lee et al., 1995). P1 is around 49-50 amino acids. P2 is first synthesized as a precursor and upon binding on to DNA N-ter part of the protein is cleaved and final form is around 63 amino acids (Balhorn, 2007). Protamines are phosphoryated soon after their synthesis and after binding to DNA they are dephosphorylated. It was suggested that initial interaction of phosphorylated protamines with DNA could facilitate optimum nucleoprotamine structure formation (Meetei et al., 2002).

P1 binds around 10-11 bp of DNA and P2 15 bp of DNA. Protamines could form bisulfide bridges which contribute to a more compact structure. Protamines bind the major groove on DNA and neutralize the negative charge forming toroidal structures. These toroids constitute the basic unit of nucleoprotamines and they have about 50 kb DNA. Protamines could package DNA into 1/20 volume of a somatic nucleus (Balhorn, 2007).

In human sperm P1/P2 ratio nearly equals 1 (Oliva, 2006), but this ratio could change significantly among the species (Corzett et al., 2002). Both P1 and P2 are required for fertility and haploinsufficiency in either of them creates infertility (Cho et al., 2001). Later, it was determined that protamine deficiency leads to damaged DNA. Although protamine 2 deficient sperm could activate oocyte by intracytoplasmic sperm injection (ICSI), the majority of the embryos could not develop up to blastocyst stage (Cho et al., 2003).

Transient DNA strand breaks during chromatin remodeling

In elongating spermatids, DNA strand breaks were detected coincident with chromatin remodeling (Marcon and Boissonneault, 2004). It is proposed that normally DNA supercoils are constrained by nucleosomes and removal of nucleosomes will results in a great number of supercoils. Nevertheless, appearance of DNA breaks could eliminate this supercoiling. Another study showed that topoisomerase II beta (Top2b) was involved in the introduction of DNA breaks happening in elongating spermatids (Leduc et al., 2008). In the same study, it was shown that phosphorylated form of H2A.X and tyrosyl-DNA phosphodiesterase 1 (TDP-1), an enzyme known to resolve topoisomerase-mediated DNA damage, localizes to strand breaks to initiate DNA repair process. It has been suggested that transition proteins and protamines could also be involved in repair process due to their ability to neutralize negative charge on DNA and bring DNA ends together (Boissonneault, 2002).

During spermiogenesis, the presence of DNA strand breaks triggers poly(ADP-ribose) (PAR) formation. Mice deficient for poly(ADP-ribose) polymerase Parp1 have problems with nuclear condensation and are subfetile (Meyer-Ficca et al., 2009). It was also shown that Parp1 and Parp2 could modulate Top2b activity *in vitro* such that Top2b first creates DNA strand breaks and Parp1 and Parp2 are activated, in turn Parp proteins could inhibit Top2b (Meyer-Ficca et al., 2011b). In addition, impairment of PAR metabolism with targeted disruption and chemical inhibition results in abnormal retention of nucleosomes in mature sperm (Meyer-Ficca et al., 2011a).

Genomic regions packed into nucleosomes in mature spermatozoa

In human and mouse, it was determined that histone-to-protamine exchange is not complete, still around 10% and 1% of histones are retained in human and mouse sperm, respectively (Brykczynska et al., 2010).

Initially, a study demonstrated the existence of sequence specific presence of DNA packed into nucleosomes or protamines (Gatewood et al., 1987). Another study determined that genes encoding for epsilon and gamma globulin expressed in embryonic yolk sac were associated with histones in human sperm whereas beta and delta globulin genes which are not expressed in embryonic yolk sac were not associated with histones (Gardiner-Garden et al., 1998). By treating human sperm chromatin with micrococcal nuclease, it was shown that DNA is organized into nucleosomes in 148 bp periodicities and telomeric DNA is associated with nucleosomes (Zalenskaya et al., 2000).

In addition to single gene-based analysis, genome-wide analysis of retained nucleosomes in human and mouse sperm provided significant evidence that retained

nucleosomes in mature sperm are not randomly distributed in the genome. Two studies in human sperm showed that nucleosomes are retained at loci which are important for developmental and signaling pathways. Importantly, retained histones are also post-translationally modified and different histone marks show differential association with the regulatory elements of genes (Brykczynska et al., 2010; Hammoud et al., 2009). Retained histones are differentially post-translationally modified at single loci-basis also for mouse sperm, suggesting for the evolutionary conservation of specific nucleosome retention (Brykczynska et al., 2010). In zebrafish, histone-to-protamine exchange does not take place. Instead, zebrafish sperm contains relatively higher levels of linker histone H1. Nevertheless, conceptually, similar pattern of histone modifications at developmental regulators were identified in zebrafish sperm in comparison to human and mouse sperm (Wu et al., 2011).

1.4 Rationale and scope of the thesis

In mammals, fusion of two highly specialized gametes, sperm and oocyte gives rise to a totipotent embryo and this totipotent embryo could generate all cells of a multicellular organism. One of the critical questions is how the totipotent state of the early embryo is achieved. At this point, we propose two models (Gill et al., 2012). In the first scenario, which is the epigenetic reprogramming model, it is assumed that epigenetic marks carried by sperm and oocyte to the early embryo are reprogrammed to ensure the totipotent state. Initially, because of the reprogramming of DNA methylation upon fertilization it was assumed that similarly other chromatin marks should be reprogrammed. Nevertheless, currently even for DNA methylation there is genome-wide data showing that there are many loci resisting genome-wide DNA demethylation (Kobayashi et al., 2012; Smallwood et al., 2011). The alternative scenario to reprogramming model is the inheritance model. This model proposes that chromatin states brought by sperm and oocyte are required for the developmental potential of the early embryo.

In our lab, the ultimate question to which we would like to find an answer to is to what extent epigenetic states of sperm and oocyte are involved in the establishment of the proper gene expression program of early embryonic development. Importantly, sperm and oocyte differ in their potential to transmit epigenetic information. Whereas oocyte is full of maternal transcripts and proteins and its DNA is packed in nucleosomes, sperm is in a highly condensed structure and the majority of its histones are exchanged by protamines. Given the little of amount of histones retained in mature sperm, initially it was thought that sperm did not have any

significant potential to transmit epigenetic information. Nevertheless, as already discussed, retained nucleosomes associate with specific genomic regions in sperm. Furthermore, it was identified that genes whose promoter regions are marked by active H3K4me3 in sperm are mostly associated with early embryonic gene activity. On the other hand, the majority of the genes whose promoter regions marked by repressive H3K27me3 are never expressed during early embryonic development (Brykczynska et al., 2010).

Although genome-wide localization of nucleosomes in human sperm has been determined (Brykczynska et al., 2010; Hammoud et al., 2009), genome-wide mapping of nucleosomes in mouse sperm is not known. Furthermore, as already discussed the details actually how chromatin remodeling happens during spermiogenesis and how nucleosomes are specifically retained are not known. In my project, the major aim was to determine the logic of nucleosome retention in spermatozoa and for this aim I used mouse sperm as the model system. Mainly, I determined what special features 1% of mouse genome had to keep nucleosomes. By taking a genome-wide approach, I tried to find out the relative contributions of sequence composition, DNA methylation, histone variants, gene expression and histone modifications in selective nucleosome retention. In addition to this major aim, I worked on transcript dynamics during differentiation of round spermatids into sperm by performing RNA-seq analysis at specific points of spermiogenesis. Ultimate goal of this transcript dynamics analysis is to correlate changes in chromatin states to gene expression and to investigate the potential of sperm in transmission of RNA to the early embryo.

Cited literature

Adam, M., Robert, F., Larochelle, M., and Gaudreau, L. (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol *21*, 6270-6279.

Adham, I.M., Nayernia, K., Burkhardt-Gottges, E., Topaloglu, O., Dixkens, C., Holstein, A.F., and Engel, W. (2001). Teratozoospermia in mice lacking the transition protein 2 (Tnp2). Mol Hum Reprod *7*, 513-520.

Ahmed, E.A., and de Rooij, D.G. (2009). Staging of mouse seminiferous tubule cross-sections. Methods Mol Biol *558*, 263-277.

Alfonso, P.J., and Kistler, W.S. (1993). Immunohistochemical localization of spermatid nuclear transition protein 2 in the testes of rats and mice. Biol Reprod *48*, 522-529.

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. Embo J *25*, 4503-4512.

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res *19*, 1732-1741.

Anway, M.D., Memon, M.A., Uzumcu, M., and Skinner, M.K. (2006). Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. J Androl *27*, 868-879.

Arteaga-Vazquez, M.A., and Chandler, V.L. (2010). Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev *20*, 156-163.

Arya, G., Maitra, A., and Grigoryev, S.A. (2010). A structural perspective on the where, how, why, and what of nucleosome positioning. J Biomol Struct Dyn *27*, 803-820.

Bak, C.W., Yoon, T.K., and Choi, Y. (2011). Functions of PIWI proteins in spermatogenesis. Clin Exp Reprod Med *38*, 61-67.

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. Genome Biol *8*, 227.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.
Bird, A. (2011). The dinucleotide CG as a genomic signalling module. J Mol Biol *409*, 47-53.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet *2*, e49.

Bogdanovic, O., and Veenstra, G.J. (2009). DNA methylation and methyl-CpG binding proteins: developmental requirements and function. Chromosoma *118*, 549-565.

Boissonneault, G. (2002). Chromatin remodeling during spermiogenesis: a possible role for the transition proteins in DNA strand break repair. FEBS Lett *514*, 111-114.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612-616.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Bosch, A., and Suau, P. (1995). Changes in core histone variant composition in differentiating neurons: the roles of differential turnover and synthesis rates. Eur J Cell Biol *68*, 220-225.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Bowles, J., and Koopman, P. (2007). Retinoic acid, meiosis and germ cell fate in mammals. Development *134*, 3401-3411.

Braun, R.E., Peschon, J.J., Behringer, R.R., Brinster, R.L., and Palmiter, R.D. (1989). Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. Genes Dev *3*, 793-802.

Brewer, L., Corzett, M., and Balhorn, R. (2002). Condensation of DNA by spermatid basic nuclear proteins. J Biol Chem *277*, 38895-38900.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schubeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol *17*, 679-687.

Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. Annu Rev Genet *43*, 559-599.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D.*, et al.* (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell *143*, 1084-1096.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505-518.

Cedar, H., and Bergman, Y. (2012). Programming of DNA methylation patterns. Annu Rev Biochem *81*, 97-117.

Chang, S.C., Tucker, T., Thorogood, N.P., and Brown, C.J. (2006). Mechanisms of X-chromosome inactivation. Front Biosci *11*, 852-866.

Cheng, X., and Blumenthal, R.M. (2010). Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. Biochemistry *49*, 2999-3008.

Cho, C., Jung-Ha, H., Willis, W.D., Goulding, E.H., Stein, P., Xu, Z., Schultz, R.M., Hecht, N.B., and Eddy, E.M. (2003). Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice. Biol Reprod *69*, 211-217.

Cho, C., Willis, W.D., Goulding, E.H., Jung-Ha, H., Choi, Y.C., Hecht, N.B., and Eddy, E.M. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. Nat Genet *28*, 82-86.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J.*, et al.* (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388-392.

Choi, J.K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol *11*, R70.

Corzett, M., Mazrimas, J., and Balhorn, R. (2002). Protamine 1: protamine 2 stoichiometry in the sperm of eutherian mammals. Mol Reprod Dev *61*, 519-527.

Couldrey, C., Carlton, M.B., Nolan, P.M., Colledge, W.H., and Evans, M.J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. Hum Mol Genet *8*, 2489-2495.

Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature *401*, 157-161.

Daxinger, L., and Whitelaw, E. (2012). Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nat Rev Genet *13*, 153-162.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L., and Barsh, G.S. (1994). Neomorphic agouti mutations in obese yellow mice. Nat Genet *8*, 59-65.

Ekwall, K., Olsson, T., Turner, B.M., Cranston, G., and Allshire, R.C. (1997). Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell *91*, 1021-1032.

Elsaesser, S.J., Goldberg, A.D., and Allis, C.D. (2010). New functions for an old variant: no substitute for histone H3.3. Curr Opin Genet Dev *20*, 110-117.

Faust, C., Schumacher, A., Holdener, B., and Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. Development *121*, 273-285.

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., and Langst, G. (2011). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res *39*, 6956-6969.

Feng, Y.Q., Desprat, R., Fu, H., Olivier, E., Lin, C.M., Lobell, A., Gowda, S.N., Aladjem, M.I., and Bouhassira, E.E. (2006). DNA methylation supports intrinsic epigenetic memory in mammalian cells. PLoS Genet *2*, e65.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., and Andrau, J.C. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res.

Gardiner-Garden, M., Ballesteros, M., Gordon, M., and Tam, P.P. (1998). Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. Mol Cell Biol *18*, 3350-3356.

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.

Gaucher, J., Reynoird, N., Montellier, E., Boussouar, F., Rousseaux, S., and Khochbin, S. (2010). From meiosis to postmeiotic events: the secrets of histone disappearance. Febs J *277*, 599-604.

Gill, M.E., Erkek, S., and Peters, A.H. (2012). Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? Curr Opin Cell Biol *24*, 387-396.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Govin, J., Caron, C., Escoffier, E., Ferro, M., Kuhn, L., Rousseaux, S., Eddy, E.M., Garin, J., and Khochbin, S. (2006). Post-meiotic shifts in HSPA2/HSP70.2 chaperone activity during mouse spermatogenesis. J Biol Chem *281*, 37888-37892.

Greaves, I.K., Rangasamy, D., Ridgway, P., and Tremethick, D.J. (2007). H2A.Z contributes to the unique 3D structure of the centromere. Proc Natl Acad Sci U S A *104*, 525-530.

Green, G.R., Balhorn, R., Poccia, D.L., and Hecht, N.B. (1994). Synthesis and processing of mammalian protamines and transition proteins. Mol Reprod Dev *37*, 255-263.
Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet *13*, 343-357.

Griffith, J.S., and Mahler, H.R. (1969). DNA ticketing theory of memory. Nature *223*, 580-582.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Guerrero-Bosagna, C., Settles, M., Lucker, B., and Skinner, M.K. (2010). Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS One *5*.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev *117*, 15-23.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78-82.

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.

Hardy, S., Jacques, P.E., Gevry, N., Forest, A., Fortin, M.E., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet *5*, e1000687.

Heidaran, M.A., and Kistler, W.S. (1987). Transcriptional and translational control of the message for transition protein 1, a major chromosomal protein of mammalian spermatids. J Biol Chem *262*, 13309-13315.

Heidaran, M.A., Showman, R.M., and Kistler, W.S. (1988). A cytochemical study of the transcriptional and translational regulation of nuclear transition protein 1 (TP1), a major chromosomal protein of mammalian spermatids. J Cell Biol *106*, 1427-1433.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A *105*, 17046-17049.

Holliday, R. (2006). Epigenetics: a historical overview. Epigenetics *1*, 76-80.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

Hublitz, P., Albert, M., and Peters, A.H. (2009). Mechanisms of transcriptional repression by histone lysine methylation. Int J Dev Biol *53*, 335-354.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett *583*, 1713-1720.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Jan, S.Z., Hamer, G., Repping, S., de Rooij, D.G., van Pelt, A.M., and Vormer, T.L. (2012). Molecular control of rodent spermatogenesis. Biochim Biophys Acta *1822*, 1838-1850.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. Chembiochem *12*, 206-222.

Justin, N., De Marco, V., Aasland, R., and Gamblin, S.J. (2010). Reading, writing and editing methylated lysines on histone tails: new insights from recent structural studies. Curr Opin Struct Biol *20*, 730-738.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2012). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. Embo J.

Kaminen-Ahola, N., Ahola, A., Maga, M., Mallitt, K.A., Fahey, P., Cox, T.C., Whitelaw, E., and Chong, S. (2010). Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model. PLoS Genet *6*, e1000811.

Kaneda, M., Hirasawa, R., Chiba, H., Okano, M., Li, E., and Sasaki, H. (2010). Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes Cells.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kashimada, K., and Koopman, P. (2010). Sry: the master switch in mammalian sex determination. Development *137*, 3921-3930.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet *16*, 2272-2280.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y.*, et al.* (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet *8*, e1002440.

Kolthur-Seetharam, U., Pradeepa, M.M., Gupta, N., Narayanaswamy, R., and Rao, M.R. (2009). Spatiotemporal organization of AT- and GC-rich DNA and their association with transition proteins TP1 and TP2 in rat condensing spermatids. J Histochem Cytochem *57*, 951-962.

Kota, S.K., and Feil, R. (2010). Epigenetic transitions in germ cell development and meiosis. Dev Cell *19*, 675-686.

Krishnan, S., Horowitz, S., and Trievel, R.C. (2011). Structure and function of histone H3 lysine 9 methyltransferases and demethylases. Chembiochem *12*, 254-263.

Kundu, T.K., and Rao, M.R. (1996). Zinc dependent recognition of a human CpG island sequence by the mammalian spermatidal protein TP2. Biochemistry *35*, 15626-15632.

Lanzuolo, C., and Orlando, V. (2012). Memories from the polycomb group proteins. Annu Rev Genet *46*, 561-589.

Leduc, F., Maquennehan, V., Nkoma, G.B., and Boissonneault, G. (2008). DNA damage response during chromatin remodeling in elongating spermatids of mice. Biol Reprod *78*, 324-332.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science *338*, 1435-1439.

Lee, K., Haugen, H.S., Clegg, C.H., and Braun, R.E. (1995). Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. Proc Natl Acad Sci U S A *92*, 12451-12455.

Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A *107*, 14075-14080.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, L., and Liu, Y. (2011). Diverse small non-coding RNAs in RNA interference pathways. Methods Mol Biol *764*, 169-182.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell *15*, 547-557.

Li, Z., Schug, J., Tuteja, G., White, P., and Kaestner, K.H. (2011). The nucleosome map of the mammalian liver. Nat Struct Mol Biol *18*, 742-746.

Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B., and Schubeler, D. (2011). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS Genet *7*, e1002090.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012). A bridging model for persistence of a polycomb group protein complex through DNA replication in vitro. Mol Cell *46*, 784-796.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol *13*, 436-447.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C.*, et al.* (2011). An epigenetic signature for monoallelic olfactory receptor expression. Cell *145*, 555-570.

Magnusdottir, E., Gillich, A., Grabole, N., and Surani, M.A. (2012). Combinatorial control of cell fate and reprogramming in the mammalian germline. Curr Opin Genet Dev *22*, 466-474.

Marcon, L., and Boissonneault, G. (2004). Transient DNA strand breaks during mouse and human spermiogenesis new insights in stage specificity and link to chromatin remodeling. Biol Reprod *70*, 910-918.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet *11*, 285-296.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Meetei, A.R., Ullas, K.S., Vasupradha, V., and Rao, M.R. (2002). Involvement of protein kinase A in the phosphorylation of spermatidal protein TP2 and its effect on DNA condensation. Biochemistry *41*, 185-195.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Meyer-Ficca, M.L., Ihara, M., Lonchar, J.D., Meistrich, M.L., Austin, C.A., Min, W., Wang, Z.Q., and Meyer, R.G. (2011a). Poly(ADP-ribose) metabolism is essential for proper nucleoprotein exchange during mouse spermiogenesis. Biol Reprod *84*, 218-228.

Meyer-Ficca, M.L., Lonchar, J., Credidio, C., Ihara, M., Li, Y., Wang, Z.Q., and Meyer, R.G. (2009). Disruption of poly(ADP-ribose) homeostasis affects spermiogenesis and sperm chromatin integrity in mice. Biol Reprod *81*, 46-55.

Meyer-Ficca, M.L., Lonchar, J.D., Ihara, M., Meistrich, M.L., Austin, C.A., and Meyer, R.G. (2011b). Poly(ADP-ribose) polymerases PARP1 and PARP2 modulate topoisomerase II beta (TOP2B) function during chromatin condensation in mouse spermiogenesis. Biol Reprod *84*, 900-909.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A *98*, 12902-12907.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G., and Allis, C.D. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol Cell *38*, 853-863.

Morange, M. (2002). The relations between genetics and epigenetics: a historical point of view. Ann N Y Acad Sci *981*, 50-60.

Moshkin, Y.M., Chalkley, G.E., Kan, T.W., Reddy, B.A., Ozgur, Z., van Ijcken, W.F., Dekkers, D.H., Demmers, J.A., Travers, A.A., and Verrijzer, C.P. (2012). Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. Mol Cell Biol *32*, 675-688.

Nanney, D.L. (1958). Epigenetic Control Systems. Proc Natl Acad Sci U S A *44*, 712-717.

Ng, R.K., and Gurdon, J.B. (2008). Epigenetic inheritance of cell differentiation status. Cell Cycle *7*, 1173-1177.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol *21*, 4330-4336.

Oatley, J.M., and Brinster, R.L. (2008). Regulation of spermatogonial stem cell self-renewal in mammals. Annu Rev Cell Dev Biol *24*, 263-286.

Ohlsson, R., Bartkuhn, M., and Renkawitz, R. (2010). CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. Chromosoma *119*, 351-360.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.
Oliva, R. (2006). Protamines and male infertility. Hum Reprod Update *12*, 417-435.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. Embo J *23*, 4061-4071.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A.*, et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T.C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M.*, et al.* (2012). Polycomb function during oogenesis is required for mouse embryonic development. Genes Dev *26*, 920-932.

Pradeepa, M.M., Manjunatha, S., Sathish, V., Agrawal, S., and Rao, M.R. (2008). Involvement of importin-4 in the transport of transition protein 2 into the spermatid nucleus. Mol Cell Biol *28*, 4331-4341.

Pradeepa, M.M., Nikhil, G., Hari Kishore, A., Bharath, G.N., Kundu, T.K., and Rao, M.R. (2009). Acetylation of transition protein 2 (TP2) by KAT3B (p300) alters its DNA condensation property and interaction with putative histone chaperone NPM3. J Biol Chem *284*, 29956-29967.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol *10*, 192-206.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., and Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233-248.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114-128.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D.J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. Embo J *22*, 1599-1607.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat Rev Mol Cell Biol *11*, 37-49.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet *14*, 9-25.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem *81*, 145-166.

Saitou, M. (2009). Germ cell specification in mice. Curr Opin Genet Dev *19*, 386-395.

Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Curr Biol *19*, 1816-1820.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol *12*, 799-814.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet *25*, 335-343.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Mol Cell *48*, 849-862.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem *76*, 75-100.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697-708.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell *138*, 885-897.

Singh, J., and Rao, M.R. (1987). Interaction of rat testis protein, TP, with nucleic acids in vitro. Fluorescence quenching, UV absorption, and thermal denaturation studies. J Biol Chem *262*, 734-740.

Smallwood, S.A., and Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. Trends Genet *28*, 33-42.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet *43*, 811-814.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Stern, L., Kleene, K.C., Gold, B., and Hecht, N.B. (1983). Gene expression during mammalian spermatogenesis. III. Changes in populations of mRNA during spermiogenesis. Exp Cell Res *143*, 247-255.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell Res *21*, 421-434.

Tachibana, M., Nozaki, M., Takeda, N., and Shinkai, Y. (2007). Functional dynamics of H3K9 methylation during meiotic prophase progression. Embo J *26*, 3346-3359.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol *11*, 264-275.

Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Hofer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat Struct Mol Biol *19*, 1185-1192.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausio, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852-10857.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D.*, et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC bioinformatics *10*, 442.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell *28*, 1-13.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Ul Hussain, M. (2012). Micro-RNAs (miRNAs): genomic organisation, biogenesis and mode of action. Cell Tissue Res *349*, 405-413.

Valk-Lingbeek, M.E., Bruggeman, S.W., and van Lohuizen, M. (2004). Stem cells and cancer; the polycomb connection. Cell *118*, 409-418.

Van Speybroeck, L. (2002). From epigenesis to epigenetics: the case of C. H. Waddington. Ann N Y Acad Sci *981*, 61-81.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Curr Opin Cell Biol *24*, 374-386.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol *13*, 115-126.

Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R.*, et al.* (2005). Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A *102*, 4068-4073.

Wery, M., Kwapisz, M., and Morillon, A. (2011). Noncoding RNAs in gene regulation. Wiley Interdiscip Rev Syst Biol Med *3*, 728-738.

Wiedemann, S.M., Mildner, S.N., Bonisch, C., Israel, L., Maiser, A., Matheisl, S., Straub, T., Merkl, R., Leonhardt, H., Kremmer, E.*, et al.* (2010). Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. J Cell Biol *190*, 777-791.

Wienholz, B.L., Kareta, M.S., Moarefi, A.H., Gordon, C.A., Ginno, P.A., and Chedin, F. (2010). DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. PLoS Genet *6*.

Witt, O., Albig, W., and Doenecke, D. (1996). Testis-specific expression of a novel human H3 histone gene. Exp Cell Res *229*, 301-306.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell *140*, 99-110.

Wu, C., and Morris, J.R. (2001). Genes, genetics, and epigenetics: a correspondence. Science *293*, 1103-1105.

Wu, R.S., Tsai, S., and Bonner, W.M. (1983). Changes in histone H3 composition and synthesis pattern during lymphocyte activation. Biochemistry *22*, 3868-3873.

Wu, S.F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res *21*, 578-589.

Wunsch, A.M., and Lough, J. (1987). Modulation of histone H3 variant synthesis during the myoblast-myotube transition of chicken myogenesis. Developmental biology *119*, 94-99.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Yu, Y.E., Zhang, Y., Unni, E., Shirley, C.R., Deng, J.M., Russell, L.D., Weil, M.M., Behringer, R.R., and Meistrich, M.L. (2000). Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc Natl Acad Sci U S A *97*, 4683-4688.

Zalenskaya, I.A., Bradbury, E.M., and Zalensky, A.O. (2000). Chromatin structure of telomere domain in human sperm. Biochem Biophys Res Commun *279*, 213-218.

Zeybel, M., Hardy, T., Wong, Y.K., Mathers, J.C., Fox, C.R., Gackowska, A., Oakley, F., Burt, A.D., Wilson, C.L., Anstee, Q.M.*, et al.* (2012). Multigenerational epigenetic adaptation of the hepatic wound-healing response. Nat Med *18*, 1369-1377.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., and Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res *38*, 4246-4253.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977-980.

Zhao, M., Shirley, C.R., Mounsey, S., and Meistrich, M.L. (2004). Nucleoprotein transitions during spermiogenesis in mice with transition nuclear protein Tnp1 and Tnp2 mutations. Biol Reprod *71*, 1016-1025.

Zhao, M., Shirley, C.R., Yu, Y.E., Mohapatra, B., Zhang, Y., Unni, E., Deng, J.M., Arango, N.A., Terry, N.H., Weil, M.M.*, et al.* (2001). Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. Mol Cell Biol *21*, 7243-7255.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7-18.
Zhu, B., and Reinberg, D. (2011). Epigenetic inheritance: uncontested? Cell Res *21*, 435-441.

Zuo, X., Sheng, J., Lau, H.T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., and Li, X. (2011). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem *287*, 2107-2118.

Chapter 2: Results

2.1 Published manuscript:

Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa

This article was first published in Nature Structural and Molecular Biology on May 16, 2010, volume 17, pages 679-687, doi:10.1038/nsmb.1821.

nature structural & molecular biology

Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa

Urszula Brykczynska[1,4], Mizue Hisano[1], Serap Erkek[1], Liliana Ramos[2], Edward J Oakeley[1,4], Tim C Roloff[1], Christian Beisel[3], Dirk Schübeler[1], Michael B Stadler[1] & Antoine H F M Peters[1]

In higher eukaryotes, histone methylation is involved in maintaining cellular identity during somatic development. As most nucleosomes are replaced by protamines during spermatogenesis, it is unclear whether histone modifications function in paternal transmission of epigenetic information. Here we show that two modifications important for Trithorax- and Polycomb-mediated gene regulation have methylation-specific distributions at regulatory regions in human spermatozoa. Histone H3 Lys4 dimethylation (H3K4me2) marks genes that are relevant in spermatogenesis and cellular homeostasis. In contrast, histone H3 Lys27 trimethylation (H3K27me3) marks developmental regulators in sperm, as in somatic cells. However, nucleosomes are only moderately retained at regulatory regions in human sperm. Nonetheless, genes with extensive H3K27me3 coverage around transcriptional start sites in particular tend not to be expressed during male and female gametogenesis or in preimplantation embryos. Promoters of orthologous genes are similarly modified in mouse spermatozoa. These data are compatible with a role for Polycomb in repressing somatic determinants across generations, potentially in a variegating manner.

By classical Mendelian inheritance, genetic information is transmitted through the generations, underlying phenotypic diversity in sexually reproducing organisms. Nevertheless, non-Mendelian inheritance of traits across generations has been reported in various higher eukaryotes[1]. Furthermore, the low reproductive efficacy of nuclear transfer versus natural reproduction[2] suggests that resetting and maturation of chromatin states during gametogenesis is crucial for early embryogenesis, arguing for a transgenerational epigenetic contribution at conception.

In mammals, the dimorphic gametes differ greatly in their potential to transmit epigenetic information encoded in histones and the associated post-translational modifications[3–5]. Whereas in oocytes chromatin retains a nucleosomal conformation, marked by histone methylations[4], most histones are replaced by protamines at the end of spermatogenesis[6,7] (**Fig. 1a**). Following gamete fusion, maternally provided histones replace protamines that subsequently become post-translationally modified by oocyte-derived factors[4,5]. Despite such major remodeling, histones have been reported to reside at specific sequences in human and mouse spermatozoa[7–11] and to remain associated with the paternal genome during *de novo* nucleosome formation upon fertilization[12].

During somatic development, Polycomb (PcG) and Trithorax (TrxG) group proteins serve conserved chromatin-based repressive and antirepressive roles in epigenetic memory of cell identity, for example, by controlling the expression of developmental regulators that drive differentiation[13,14]. In mammals, PcG proteins function in at least two distinct Polycomb repressive complexes (PRCs). PRC2, which consists of Eed, Suz12 and Ezh2 or Ezh1, catalyzes trimethylation on histone H3 Lys27 (H3K27me3)[15–17], a modification associated with gene repression in development[18,19]. Eed binds through its WD40 domains to trimethylated lysine histone residues associated with repressive chromatin. This interaction was shown to stimulate the enzymatic activity of PRC2 toward H3K27 and to be essential for early *Drosophila melanogaster* development[20]. Furthermore, the catalytic SET domain of Ezh2 is required for long-term repression[21]. These data suggest that PRC2 has an intrinsic ability to propagate H3K27me3 across cellular generations and that this mark functions in transcriptional memory of the repressed state. Mammalian TrxG proteins of the mixed lineage leukemia (MLL) protein family counteract the repressive function of PcG proteins and mediate H3K4 methylation, a mark associated with transcriptional activity[14].

It was recently reported that human spermatozoa show strong enrichments of nucleosomes, H3K27me3, H3K4me2 and H3K4me3 at regulatory regions of genes with developmental and other specific functions[11]. The molecular basis for the selective retention of nucleosomes at regulatory regions is unclear, as is whether the presence of specific histone modifications could serve regulatory functions in the following generation. Here we report that H3K27me3 and H3K4me2 are retained at regulatory sequences in mature human spermatozoa, as shown previously[11]. Our computational analyses show that they mark promoters of distinct gene classes with defined expression programs during spermatogenesis and early embryonic development. Notably, we show that promoters of orthologous genes are similarly marked in

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [2]Fertility Laboratory, Department of Obstetrics and Gynaecology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. [3]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [4]Present addresses: Department of Zoology and Animal Biology, University of Geneva, Sciences III, Geneva, Switzerland (U.B.) and Novartis Institutes for Biomedical Research, Basel, Switzerland (E.J.O.). Correspondence should be addressed to A.H.F.M.P. (antoine.peters@fmi.ch).

Received 3 September 2009; accepted 24 March 2010; published online 16 May 2010; doi:10.1038/nsmb.1821

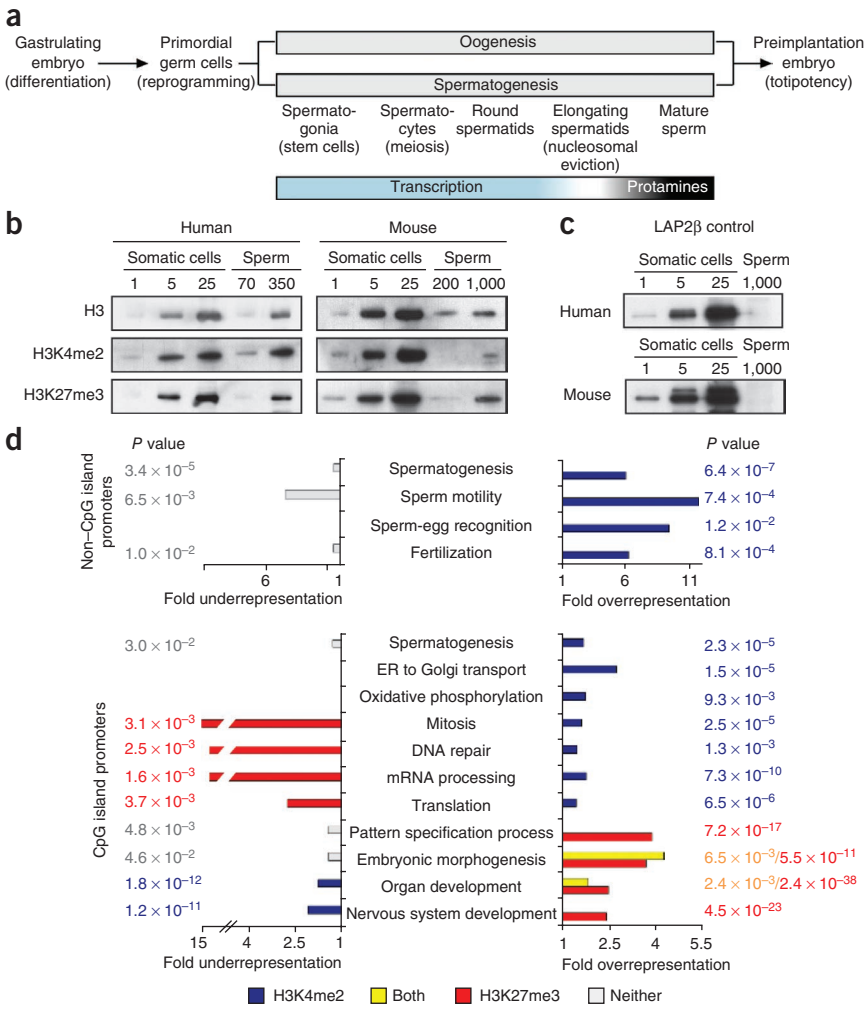


Figure 1 Methylated histones are present in human sperm and localize to distinct promoter sets. (**a**) The mammalian germ cell and embryonic development. Primordial germ cells, specified in the proximal epiblast, undergo epigenetic reprogramming including global DNA demethylation. During spermatogenesis, male germ cells first proliferate (spermatogonia), then undergo meiosis (spermatocytes) and convert into spermatozoa after transcriptional arrest and global exchange of histones by protamines (elongating spermatids). Fertilization of the oocyte results in the totipotent early embryo. (**b**) Presence of histone H3, H3K27me3 and H3K4me2 in human and mouse spermatozoa as measured by western blotting analysis. Cell numbers are given in thousands. (**c**) Absence of signal for LAP2β, a marker for somatic and immature germ cells[22], shows the purity of sperm samples used. (**d**) Selection of gene ontology–based gene functions significantly over- and underrepresented among modified promoters in human sperm (in comparison to all annotated promoters on the array within a given CpG-density class). H3K4me2 and H3K27me3 occupy sets of genes with mutually exclusive functions, with spermatogenic and housekeeping functions for H3K4me2-marked genes and developmental functions for H3K27me3-marked genes. All significantly over- and underrepresented gene ontology terms are listed in **Supplementary Table 2**.

mouse spermatozoa. In human sperm, however, we observe genome-wide distribution of nucleosomes with only a minor enrichment at regulatory sequences. In light of the evolutionary conservation and the fact that histone levels are one order of magnitude lower in human spermatozoa versus somatic cells, we propose a role for histone methylation in transgenerational chromatin inheritance as a function of histone-modification occupancy across regulatory regions of loci in sperm.

RESULTS

H3K4me2 and H3K27me3 mark functionally distinct gene sets

To evaluate the presence of histones and associated modifications in mature human and mouse spermatozoa, we performed western blotting analyses on highly purified human and mouse spermatozoa (**Fig. 1b**). As described previously, we observed ~10% of histone H3 in human sperm compared to histone H3 levels in human somatic cells[7]; in mouse sperm, we detected ~1% of histone H3 (**Fig. 1b**). We further detected H3K27me3 and H3K4me2 in spermatozoa of both species (**Fig. 1b**). The absence of a signal for lamina-associated polypeptide 2β (LAP2β) (**Fig. 1c**), a marker for somatic and immature germ cells[22], showed the purity of the sperm samples used. To define the chromosomal localization of modified histones, we developed a chromatin immunoprecipitation (ChIP) approach for H3K4me2 and H3K27me3 that is compatible with the highly condensed chromatin state present in human spermatozoa. Following ChIP on cross-linked chromatin isolated from a pool of human spermatozoa obtained from nine fertile donors, we amplified and hybridized precipitated genomic DNA to an oligonucleotide array representing 18,152 human promoters, each spanning 2.7 kb around the transcriptional start site (TSS) (further referred to as ChIP-chip experiments). We applied a hidden Markov model–based peak-finding algorithm (**Supplementary Figs. 1a** and **2a** and **Supplementary Methods**) and identified more than 1,609 and 4,555 promoters that are marked by H3K27me3 and H3K4me2, respectively, in three independent ChIP-chip experiments. For 458 promoters, we detected the presence of both modifications (**Supplementary Table 1**). Thus, more than 30% of all tested human promoters are positive for these histone modifications. In independent sperm samples, single-gene analyses of 41 selected promoters confirmed that promoters are uniquely modified by either one or both modifications (**Supplementary Fig. 2b**,**c**). As a separate validation, we performed ChIP experiments under native conditions and obtained similar results, arguing for an overall conservation of promoter distributions of H3K27me3 and H3K4me2 in mature human spermatozoa (**Supplementary Fig. 2d**).

Next, we addressed whether promoters bound by H3K27me3 and/or H3K4me2 share sequence features or characteristic functions of the associated genes. We grouped promoters according to their CpG density. We observed that H3K27me3 in sperm was restricted to CpG island–containing promoters (**Supplementary Fig. 3a**), as in somatic cells[18,23]. Gene ontology analysis showed that many developmental regulatory genes (for example, *SOX2*, *CDX2*, *GATA6* and *BMP4*, *BRACHYURY (T)*) and *HOX* genes are strongly overrepresented among H3K27me3-marked genes, some of which were also marked by H3K4me2 (**Fig. 1d**, **Supplementary Fig. 2** and **Supplementary Table 2**). H3K4me2 was overrepresented among promoters of genes that regulate various spermatogenic processes (for example, *PRM1*, *PGK2*, *BRDT* and *TSH2B*) (**Fig. 1d**, **Supplementary Fig. 2** and

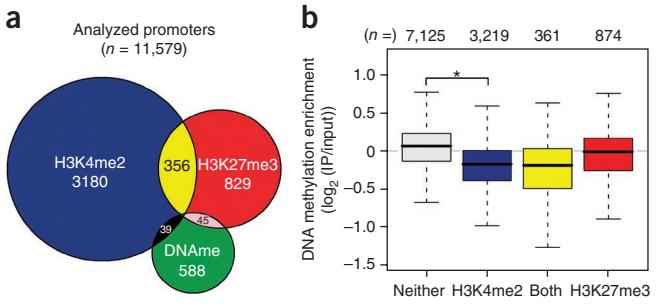


Figure 2 DNA methylation of CpG islands is mutually exclusive with H3K4 methylation in sperm. (**a**) Venn diagram showing low frequency of co-occupancy of histone modifications and DNA hypermethylation[24] on CpG island promoters (see also **Supplementary Fig. 3**). DNA-methylated promoters are significantly underrepresented among promoters marked by H3K4me2 or H3K27me3 (one-sided hypergeometric test; $P = 1.85 \times 10^{-58}$ and $P = 1.24 \times 10^{-3}$, respectively). (**b**) Boxplot showing relative enrichment for DNA methylation in sperm[24] at genes positive or negative for the tested chromatin marks (with the central bar marking the median, lower and upper limits of the box marking the 25th and 75th percentiles, and the whiskers extending the 1.5 interquartile range from the box). H3K4me2-marked genes show significantly lower levels of DNA methylation than genes with neither histone mark (Wilcoxon test; $^*P < 2.2 \times 10^{-16}$). IP, immunoprecipitation.

Supplementary Table 2). Promoters of genes functioning in cellular homeostasis and gene expression (for example, *RPS3*, *SFRS6*, *DICER1* and *PRMT5*) were particularly overrepresented among H3K4me2-marked genes and underrepresented among H3K27me3-marked genes (**Fig. 1d**). These data show that functionally distinct gene sets are marked by the two modifications in sperm.

Mutually exclusive histone and DNA methylation at promoters

In mammals, paternal transmission of DNA methylation is required for imprinted gene regulation in the subsequent generation. To determine a possible interplay between histone and DNA methylation pathways during gametogenesis, we evaluated the DNA methylation status at CpG island promoters in human spermatozoa because DNA hypermethylation confers transcriptional repression at such promoters[24]. We compared promoters that were previously classified as either DNA methylated or unmethylated[24] and observed that both histone modifications were largely mutually exclusive with DNA methylation (**Fig. 2a**). Furthermore, because developmental genes were not overrepresented among sperm targets of DNA methylation (data not shown), Polycomb and DNA methylation mark distinct gene targets in the germ line, as in soma. When analyzing DNA methylation levels without applying a defined cutoff for the methylated state, we observed that DNA methylation levels were significantly lower at promoters marked by H3K4me2 than at those harboring neither mark (**Fig. 2b**). This suggests that H3K4 and DNA methylation are largely antagonistic during spermatogenesis, a notion that is consistent with data observed in somatic cells[18,24,25].

Histone methylation status in sperm versus somatic cells

To relate genomic localization to chromatin regulation during development, we compared ChIP patterns in sperm to those generated in human embryonic stem cells (hESCs) and primary fibroblasts[24,26–28]. Most H3K4me2 targets in sperm were equally marked in somatic cells (**Fig. 3a** and **Supplementary Fig. 3b**). These targets are associated with gene ontology functions in cellular homeostasis and gene regulation (data not shown). However, a substantial number of H3K4me2 targets in sperm lack this modification in hESCs and

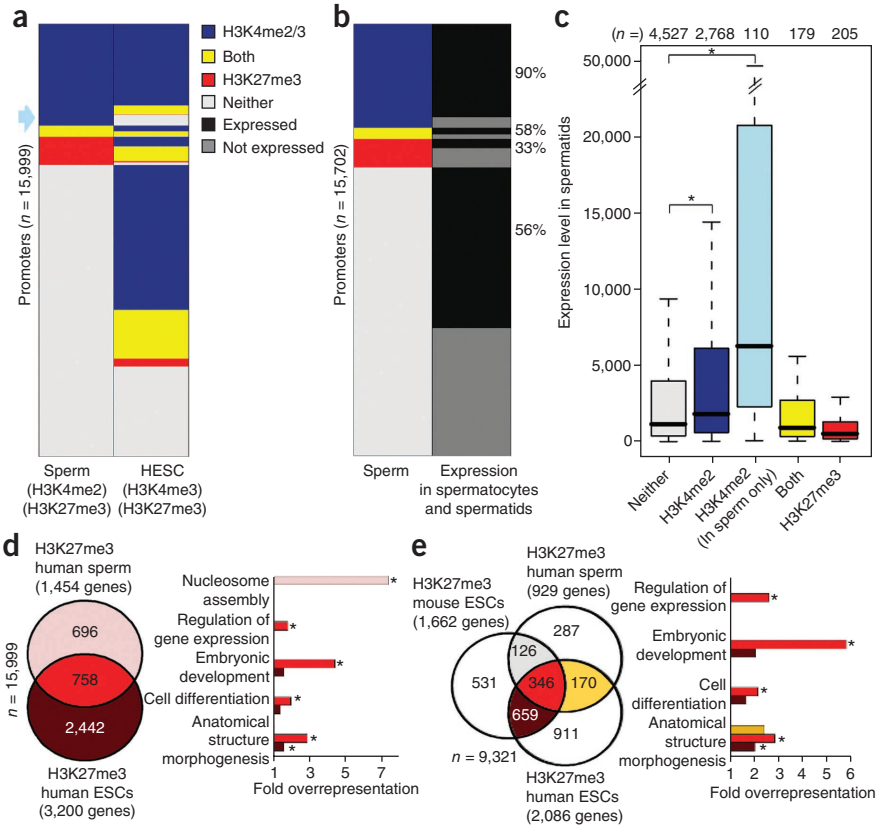


Figure 3 Spermatogenic and highly expressed genes are marked by H3K4me2 in sperm. (**a**) State map showing clustering of 15,999 genes according to the chromatin status of their promoters in human sperm and hESCs[26]. Arrow indicates genes marked by H3K4me2 in sperm only (see also **Supplementary Fig. 3**). (**b**) State map showing comparison of modification status at 15,702 promoters in sperm with gene expression status in human spermatocytes and spermatids[29]. Of the promoters marked by H3K4me2, 90% control genes that are actively transcribed in spermatogenesis. Percentages represent fractions of genes expressed. (**c**) Box plot showing expression levels in human spermatids for genes that are differentially modified in sperm. Genes under the control of promoters that are H3K4me2 targets in sperm but not in soma (light blue), also indicated by the blue arrow in **a**, are the most highly expressed. Other H3K4me2 genes (dark blue) show significantly higher levels of expression than do genes with neither mark (Wilcoxon test; $^*P$ value $< 2.2 \times 10^{-16}$). (**d**,**e**) Venn diagram and gene ontology term graphs showing overrepresentation of developmental gene functions among H3K27me3-marked genes shared between sperm and ESCs[26] of human (**d**) and ESCs[18] of mouse (**e**) (* indicates gene ontology terms with $P < 1.0 \times 10^{-10}$). **Supplementary Table 2** lists enrichments at all relevant gene ontology terms.

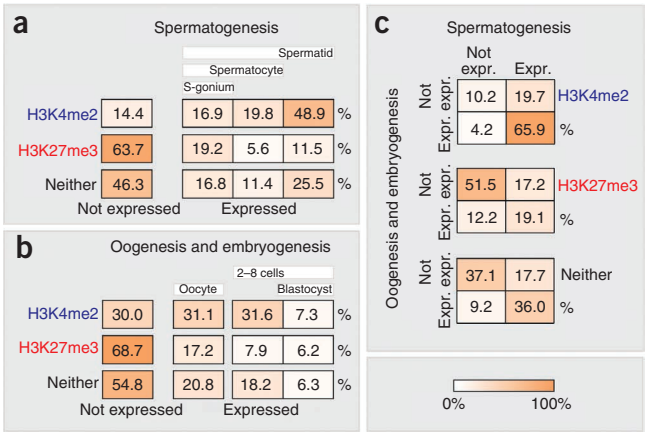


Figure 4 H3K27me3 and H3K4me2 in sperm reflect differential history and potential for expression during development. (**a**,**b**) Classification of mouse genes (n = 9,859) according to their expression status during spermatogenesis[31] (**a**), oogenesis and embryogenesis[32] (**b**) (indicated by percentages and colors) in relation to the histone-modification status at orthologous genes in human sperm. For the H3K4me2, H3K27me3 and 'neither' modification states, 2,293, 738 and 6,561 genes, respectively, represents 100%. Genes were classified as 'expressed' or 'not expressed' according to the last stage of spermatogenesis (in **a**) or to the first stage of embryogenesis (in **b**) in which their mRNAs were detected. We used the first stage of embryonic expression as the criterion for those genes expressed in oocytes and embryos. (**c**) Genes were classified according to their expression state during spermatogenesis versus oogenesis and early embryogenesis. Both H3K27me3- and H3K4me2-marked genes show significantly different distribution from genes with neither mark (Chi-square test; $P < 2.6 \times 10^{-14}$ for all comparisons). Expression classification of genes marked by both modifications is not shown. The intensity of coloring correlates with the percentage of genes expressed. Expr., expressed; S-gonium, spermatogonium.

fibroblasts (blue arrow in **Fig. 3a** and in **Supplementary Fig. 3b**), suggesting testis-specific regulation. Consistently, those genes are highly expressed in human spermatocytes and round spermatids[29] (**Fig. 3b**,**c** and **Supplementary Fig. 4a**), two cell populations that represent the meiotic and haploid stages of spermatogenesis preceding the final stage of global transcriptional repression in elongating spermatids (**Fig. 1a**). Along similar lines, gene ontology analysis revealed a strong overrepresentation of spermatogenetic functions among this group of H3K4me2-marked genes (data not shown).

Figure 3a further shows that only subsets of genes containing either H3K4me2 or H3K27me3 in hESCs are also marked in sperm. Genes retaining H3K4me2 in sperm were more likely to be expressed and at substantially higher levels during spermatogenesis than were genes without H3K4me2 or with H3K27me3 (**Fig. 3b**,**c** and **Supplementary Fig. 4a**). These data argue that H3K4me2 in sperm largely reflects robust transcription during the final stages of spermatogenesis[30], whereas H3K27me3 is likely to represent PcG-mediated transcriptional repression at preceding developmental stages.

In sperm, only 28% of H3K27me3 promoters also contained H3K4me2 (**Fig. 3a**). Compared to hESCs, this represents a three-fold underrepresentation of doubly marked or 'bivalent' promoters, particularly among CpG island promoters in human sperm (**Supplementary Fig. 3a**). These data point toward a specific regulation of H3K4 methylation at CpG island promoters during human spermatogenesis, as distinct from soma.

To obtain a closer insight into genes marked by H3K27me3 in sperm and/or hESCs, we performed in-depth gene ontology analyses. Among the targets uniquely marked in sperm, only histone genes were overrepresented (**Fig. 3d**). Closer analyses revealed that more than 70% of the 66 canonical histone genes localized in the large histone gene clusters on chromosomes 1 and 6 were marked by H3K27me3 (as well as H3K4me2) in sperm, whereas in hESCs, these genes were marked by only H3K4me2 (data not shown)[26]. In contrast, histone variant genes operating beyond DNA replication were not uniformly marked by H3K27me3 in human sperm, but they did harbor H3K4me2 in hESCs (data not shown). These data argue for a cluster-wide marking of canonical histone genes by repressive H3K27me3 that may result from entry into meiosis and/or the cell-cycle exit that is associated with terminal differentiation of male germ cells during spermiogenesis.

Notably, developmental gene ontology terms were more overrepresented among targets shared by sperm and hESCs, as compared to targets unique to hESCs (**Fig. 3d** and **Supplementary Table 2**). Furthermore, we compared targets in human and mouse ESCs and in human sperm. We observed that genes that were modified in all three cell types were more significantly overrepresented for developmental gene functions than were those genes that were modified in only two or one cell types ($P < 1.0 \times 10^{-10}$) (**Fig. 3e** and **Supplementary Table 2**). We conclude that many PRC2 targets are evolutionarily conserved between germline and ESCs of humans and mice.

Transcriptional history and potential of marked genes

To understand the origin and possible future function of modifications present in sperm, we investigated how the observed chromatin patterns relate to expression at multiple developmental time points during gametogenesis and after fertilization. Owing to the absence of comprehensive expression data sets for the human germ line and embryo, we inferred expression states from data in mice[31,32]. The validity of such a cross-species approach was supported by direct comparative expression analyses, which revealed high expression levels for those orthologs expressed in human and mouse spermatocytes and/or spermatids and low expression levels for those expressed only in germ cells of one species (**Supplementary Fig. 4**). We considered only genes with one ortholog, and we classified them as inactive or active at each developmental time point. **Figure 4** shows the percentages of genes that are active or never expressed at various developmental stages. Similar to the results in human cells (**Fig. 3b**), most H3K4me2-marked genes were expressed in mouse spermatocytes and/or spermatids (**Fig. 4a**). Moreover, many H3K4me2 targets were also expressed in oocytes or became activated in two- or eight-cell embryos (**Fig. 4b**). Comparison of all developmental stages confirmed that more than 65% of H3K4me2-marked genes expressed in oocytes and/or embryos were indeed also expressed during spermatogenesis (**Fig. 4c**). These data suggest that, in sperm, H3K4me2 preferentially marks genes with housekeeping functions, commonly expressed in the germ line and during embryogenesis.

Targets of H3K27me3 in sperm show opposing behavior, as two out of three targets were never expressed during spermatogenesis (**Fig. 4a**). Almost 20% of H3K27me3 targets marked in sperm were expressed in spermatogonia. Expression of several of these genes, such as *c-Kit*, *Stra8* and *Dnmt3a*, has been shown to be required for that developmental stage[33–35]. Thus, although not directly investigated in this study, PcG-mediated repression may dynamically regulate target genes at specific stages of germ cell development, as observed in other differentiation systems[18,19]. In oocytes, most H3K27me3 targets were not expressed, nor did they become activated in early embryos, reminiscent of the situation in spermatogenesis (**Fig. 4b**). Exceptions to this were several key regulatory genes of embryonic and extra-embryonic differentiation, such as *Cdx2*, *Elf5* and *Bmp4*, that

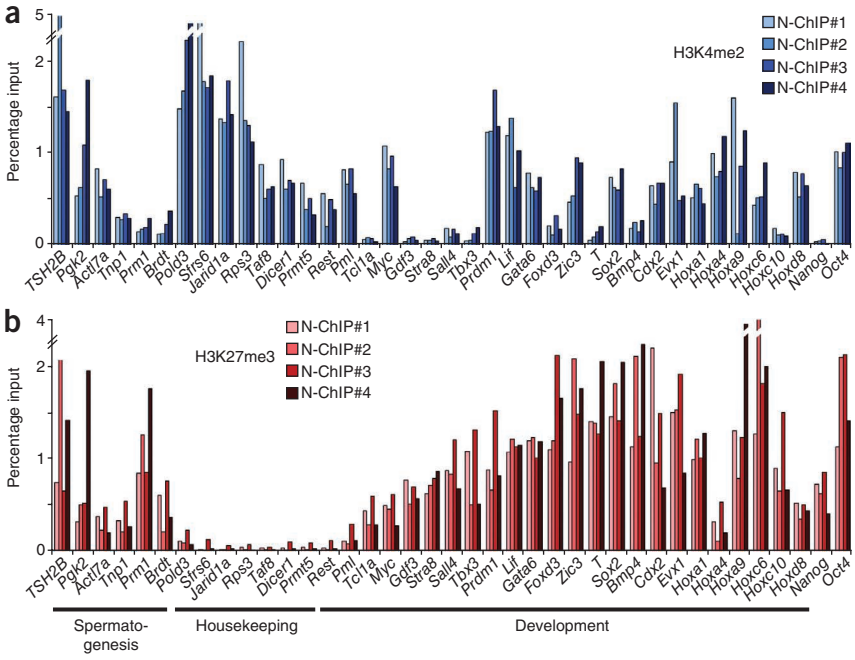


Figure 5 Evolutionary conservation of H3K27me3- and H3K4me2-marked promoters in mouse spermatozoa. (**a**,**b**) H3K4me2 (**a**) and H3K27me3 (**b**) status at the promoters of 39 mouse genes, orthologous to the human genes analyzed in **Supplementary Figure 2**. Results are represented as a percentage of the material immunoprecipitated from input chromatin under native conditions, as determined by real-time qPCR analyses. Linear scaling was applied for the replicas presented (for N-ChIP#1 to N-ChIP#4 of H3K4me2: 1×, 0.5×, 3.5×, 1.6×; for N-ChIP#1 to N-ChIP#4 of H3K27me3: 1×, 0.25×, 1×, 0.15×). Genes were selected on the basis of gene function and modification status at promoters of human orthologous genes.

were repressed during spermatogenesis and oogenesis but became activated in the early embryo[36,37]. More than 50% of H3K27me3 targets, however, were never expressed during spermatogenesis[31], oogenesis[38] and early embryogenesis[32] (**Fig. 4c** and **Supplementary Fig. 5**). Even at earlier stages of gametogenesis such as developing primordial germ cells (PGC), more than 90% of the genes in this group were not transcribed (**Supplementary Fig. 5**). As H3K27me3 target genes in sperm are highly enriched for key regulators of lineage specification and differentiation in soma (**Fig. 1d**), their repressed state throughout germ-cell development and in totipotent early embryos suggests that this modification may serve transgenerational gene-regulatory functions.

Evolutionary conservation between human and mouse sperm

If H3K4 and H3K27 methylation does indeed perform transcriptional regulatory functions across generations, we would expect them to have evolutionarily conserved targets in the sperm of humans and mice. A high level of conservation would imply selection for maintenance of histone modifications at promoters of specific target genes during the extensive chromatin remodeling taking place in elongating spermatids. As global transcription is shut down in elongating spermatids and mature spermatozoa, the presence of histone methylation at selected loci could exert its gene-regulatory function only after fertilization, in line with a role in transgenerational epigenetic inheritance.

To address conservation, we profiled H3K4me2 and H3K27me3 at promoters of 39 mouse genes, which are orthologous to the human genes analyzed before (**Supplementary Fig. 2**). We developed a ChIP procedure with an increased immunoprecipitation efficiency to accommodate the lower abundance of histones and the higher level of chromatin compaction in mouse spermatozoa as compared to human sperm. We performed ChIP under native conditions followed by quantitative PCR (qPCR) detection. For many testis-specific and housekeeping genes, we observed, as in human cells, strong enrichments for H3K4me2 (**Fig. 5a**), consistent with their expression in spermatids. Notably, some testis-specific genes, but none of the housekeeping genes, were also marked by H3K27me3 (**Fig. 5b**). The observed double marking is specific to mouse sperm and suggests that these testis-specific genes acquire H3K27me3 following their expression in spermatocytes and round spermatids, possibly to safeguard their repression after fertilization. For developmental regulatory genes, most promoters tested were marked by H3K27me3, as observed in humans. Several genes also harbored H3K4me2. In summary, the H3K4me2 status at testis-specific and housekeeping genes, as well as the H3K27me3 status at developmental genes, is highly conserved between mouse and human spermatozoa.

Distribution of nucleosomes in the human sperm genome

In human spermatozoa, as compared with soma, approximately 10% of histones are retained[7] (**Fig. 1b**). To evaluate the nucleosomal occupancy at regulatory regions marked by histone modifications, we developed a micrococcal nuclease (MNase) assay that enables the isolation of mononucleosomal DNA from the entire genome of human spermatozoa (**Supplementary Fig. 6**). Using this assay, we isolated duplicate samples of nucleosomal sperm DNA from three individuals and subjected the samples to high-throughput sequencing (Illumina GAII). We normalized the data to that for genomic DNA that was sonicated after being pretreated under conditions identical to those of the MNase-treated samples.

We observed a regular distribution of nucleosomes along the entire genome of all sperm samples with only a slight enrichment around the TSSs of genes (**Fig. 6**). For genome-wide nucleosomal enrichments, we measured high correlations between biological replicates and between individuals (the Pearson correlation coefficients for the 500-bp window were 0.84–0.88 between biological replicates and 0.73–0.84 between individuals) (**Supplementary Fig. 7a**). These data demonstrate the high technical reproducibility of our methodology and a rather low interindividual variability in the nucleosomal distribution in human sperm genomes.

To study in more detail the nucleosomal distribution in human sperm, we pooled the six data sets and quantified the read counts in different parts of the genome relative to the corresponding genomic size. We measured a 2.2-fold overrepresentation of nucleosomes in 1-kb regions upstream and downstream of the TSSs of Refseq genes. At exons, we observed a 1.6-fold overrepresentation, whereas at intronic and intergenic regions nucleosomal read counts conformed the corresponding genomic size. Comparing the nucleosomal occupancy with the H3K4me2 and H3K7me3 status of promoter regions, as identified by the ChIP-chip metholodogy, showed that positively marked promoters are enriched in nucleosomes (**Supplementary Fig. 7b**).

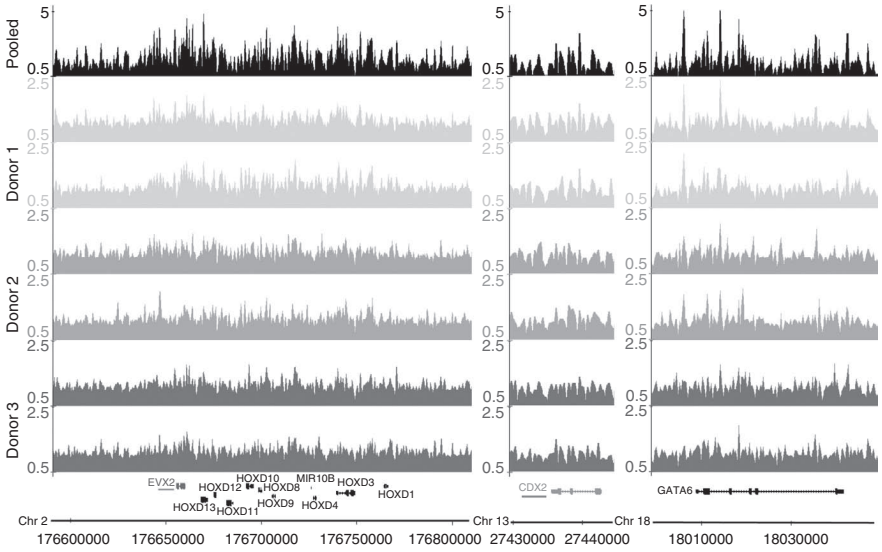


Figure 6 Even distribution of nucleosomes along the human sperm genome with modest enrichments around TSSs. Distribution of nucleosomes at the *HOXD* cluster, *CDX2* and *GATA6* loci in sperm of three human individuals. Panels present enrichments of mononucleosomal DNA generated by MNase treatment and normalized to sonicated control DNA. Biological replicates of individuals are indicated in different shades of gray. Black panels present enrichments of pooled high-throughput sequence reads. Black and gray colors of gene diagrams reflect plus and minus orientation of transcription. Chr, chromosome.

These data indicate that nucleosomes, possibly modified, are retained with some preference at regulatory sequences in human sperm.

Several reports suggested selective retention of nucleosomes at repetitive sequences in human and mouse spermatozoa[9,10]. To address this question in a genome-wide manner, we determined the nucleosomal enrichment values for all genomic coordinates corresponding to the major classes of long terminal repeat (LTR) and non-LTR repetitive sequences and to telomeric and simple repeats. For each class we observed that a fraction of sequences are enriched in nucleosomes (data not shown). However, we failed to observe a correlation between enhanced nucleosomal retention and the expression status of individual repeats in human testis or embryonic tissues (data not shown)[39]. Thus, the role of histone occupancy at selective repeats in human sperm for gene regulation following fertilization remains unclear.

Nucleosomes and histone modifications in human sperm

Recently, the genome-wide distribution of nucleosomes and histone modifications in human sperm was determined by high-throughput sequencing of nucleosomal DNA isolated after MNase treatment and immunoprecipitation under native chromatin conditions[11]. In contrast to our data, this study reported higher enrichments of nucleosomes at sequences around TSSs in spermatozoa of different donors (7.6-fold versus 2.2-fold)[11] (**Supplementary Fig. 8a**). The differential enrichments around TSSs may be due to technical differences in the preparation of mononucleosomal DNA from human sperm by micrococcal digestion. In contrast to what occurs in the protocol used in ref. 11, under our experimental conditions the entire genome is digested (**Supplementary Fig. 6**), thereby enabling detection of nucleosomes from open and more closed regions of the human sperm genome[7–9,40,41].

Similar to the nucleosomal distributions presented in **Figure 6**, the previous study reported that H3K4me3- or H3K27me3-modified nucleosomes are broadly distributed around the TSSs of many genes[11] (**Supplementary Fig. 8b**). We further observe strong correlations between the average immunoprecipitation enrichment obtained for each promoter region in the current ChIP-chip experiments and for the normalized number of ChIP-seq reads aligned to the corresponding promoter region[11] (**Supplementary Fig. 8c**). Of the promoter regions that we found to be enriched in H3K4 and H3K27 methylation, 99% and 96%, respectively, were also enriched in the ChIP-seq experiments. The strong overlap between the two data sets, despite differences in detection sensitivity due to the ChIP procedures used, argues for a widespread marking of promoters by histone methylation in human spermatozoa.

Histone modification coverage and paternal transmission

As approximately 10% of histones are maintained in human sperm, the ~2.2-fold nucleosomal enrichment at sequences near the TSSs of genes argues that only about 20% of nucleosomes are retained at regulatory sequences in individual spermatozoa. The likelihood of paternal transmission of epigenetic information by histone modifications at any locus is therefore expected to depend on the size of the region modified. If a region is widely marked, several modified nucleosomes could be transmitted even if only 20% of nucleosomes are retained during spermiogenesis. To assess the likelihood of such a mode of transmission for individual loci we determined the extent to which sequences flanking the TSS contain modified nucleosomes, a variable we termed 'modification coverage' (**Fig. 7a**). Because, in general, genes are modified on both sides of their TSS (**Supplementary Fig. 8b**), we calculated the modification coverage in a 6-kb window around the TSS (±3 kb) using the data from ref. 11. As for somatic cells, we found that H3K27me3-marked TSS proximal regions are, overall, more widely enriched in modified nucleosomes than in H3K4me3-marked regions (**Fig. 7a**, above). For example, 46% and 32% of H3K27me3- and H3K4me3-marked loci, respectively, have a modification coverage of ≥0.5 (between 3,000 and 6,000 bp). Thus, H3K27me3-marked loci may be more likely to transmit their modification status than are H3K4me3-marked loci.

Notably, gene ontology analyses show that developmental gene functions are more overrepresented among genes that are more broadly marked by H3K27me3 in spermatozoa (**Supplementary Fig. 8d**). The observed correlation probably reflects an intrinsic characteristic of PcG targets with developmental gene functions[42]. The data suggest that modification coverage may be relevant for paternal transmission of epigenetic information. To further evaluate this notion, we related the modification coverage to the modification status as determined in our ChIP-chip experiments and to the expression states of genes during gametogenesis and embryogenesis[31,32]. Compared to all enriched TSS proximal regions (**Fig. 7a**, above), the modification coverage was significantly higher for those promoters that we had identified as enriched in H3K4me2 or H3K27me3 in the ChIP-chip experiments (Wilcoxon test: $P < 2.2 \times 10^{-16}$ for either mark) (**Fig. 7a**, middle and below). Furthermore, genes with a higher H3K27me3 coverage at their TSS were more likely to be repressed during gametogenesis and early embryogenesis than were genes with a low H3K27me3 coverage

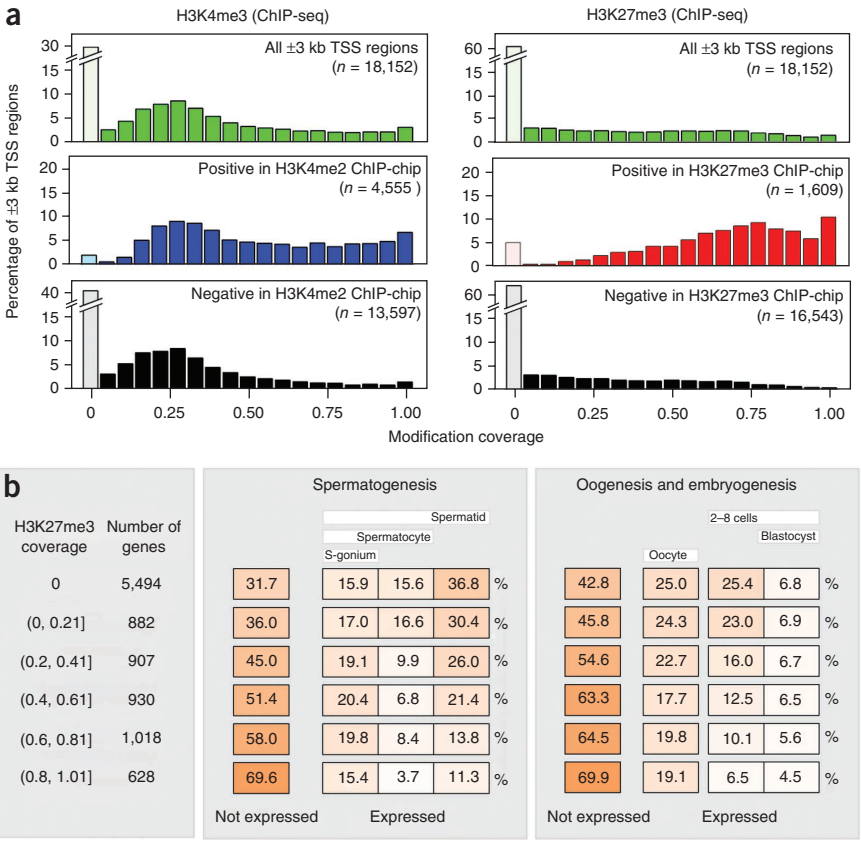


Figure 7 Histone modification coverage in human sperm. (**a**) Above, modification coverage around TSSs in human sperm. Genes, present on the HG18 NimbleGen array, were classified according to the extent to which sequences around TSSs (±3 kb) are modified by H3K4me3 or H3K27me3 (as determined by the MACS peak-finding program) (n = 18,152; data from ref. 11). Middle and bottom, modification coverage of genes identified as positive or negative for H3K4me2 or H3K27me3 in ChIP-chip experiments in the present study. ChIP-chip–positive genes (middle) have a significantly higher modification coverage compared to all genes (above) (Wilcoxon test; P value $< 2.2 \times 10^{-16}$ for H3K4me2 and for H3K27me3). (**b**) Classification of mouse genes (n = 9,859) according to their expression status during spermatogenesis[31], and during oogenesis and embryogenesis[32] (indicated by percentages and colors) in relation to the H3K27me3 coverage at TSS-proximal regions of orthologous genes in human sperm (data from ref. 11). For classification of expression states, see the legend of **Figure 4**. S-gonium, spermatogonium.

(**Fig. 7b**). Together, these data are compatible with a role for H3K27me3 in paternal transmission of the repressed state, which is likely to be a function of the modification coverage at TSS-proximal sequences.

DISCUSSION

During sperm development in animals, histones are replaced by sperm-specific histones, protamine-like proteins or protamines. In mammals and other organisms, however, a certain fraction of histones remain present in mature spermatozoa, providing a possibility for epigenetic inheritance[7–11,43–45]. Here we show that promoters with distinct gene functions and developmental expression patterns are selectively marked by active and/or repressive histone methylation in human and mouse spermatozoa. We further demonstrate comparable genome-wide retention of nucleosomes in purified spermatozoa of three fertile human individuals with modest enrichments around the TSSs of genes. Given the ten-fold lower levels of histones in human sperm versus somatic cells[6,7] (**Fig. 1b**), the low nucleosomal occupancy at regulatory sequences probably reduces the potential for transgenerational inheritance of chromatin-encoded epigenetic information by individual spermatozoa. However, we observe a strong positive correlation between the extent to which H3K27me3-modified nucleosomes are present around the TSSs of genes in spermatozoa (modification coverage) and gene repression in early embryos. We therefore propose a role for histone methylation in paternal chromatin inheritance as a function of the number of modified nucleosomes that are retained across regulatory regions of loci in sperm. Depending on the modification coverage at loci, paternal transmission of histone modification–encoded epigenetic information may be subject to variegation.

The proposed model predicts that, in embryos, sperm-inherited modified nucleosomes remain retained in the paternal genome during its remodeling by maternally provided histones in the course of pronucleus formation. Studies on the pronuclear localization of replication-dependent versus replication-independent H3 variants in early zygotes suggest that sperm-inherited histones are indeed retained within the paternal human and mouse genomes during pronucleus formation[12,46,47] (R. Kunzmann and A.H.F.M.P., unpublished data).

The model further implies that transgenerationally inherited marks need to be maintained during subsequent preimplantation development. For H3K27me3, the modification becomes microscopically detectable at the paternal genome in the one-cell embryo, concurrent with replication and before global genome activation[4,5,48]. This staining may reflect *de novo* targeting of PRC2 to unmodified sites through sequence-specific mechanisms[14,49]. Alternatively, given the intrinsic ability of PRC2 to propagate H3K27me3 (ref. 20), establishment of broad H3K27me3 domains in one-cell embryos before genome activation may be seeded by H3K27me3-modified nucleosomes inherited from sperm. The positive correlation observed between the H3K27me3 coverage around the TSSs of genes in sperm, and gene repression in gametes and early embryos supports the second hypothesis. Furthermore, the persistent presence of H3K27me3 at the originally inactive X chromosome in cloned preimplantation embryos[50], or at the maternal genome in one-cell embryos that are maternally and zygotically deficient for *Ezh2* (ref. 4), strongly argues for the lack of substantial zygotic and/or maternal H3K27me3 demethylase activity in early embryos. Together, these studies support a model in which paternally inherited H3K27me3 is propagated through subsequent preimplantation development, contributing to the propagation of totipotency across generations.

We speculate that Polycomb serves a similar regulatory function at maternal alleles during oogenesis and early embryogenesis. If maternal alleles of the H3K27me3 targets identified in sperm would be differentially regulated, the situation would resemble classical imprinting. The undisturbed embryonic patterning observed in gynogenones and parthenogenones argues against such a scenario.

Furthermore, live-born offspring with two maternal genomes are obtained at a respectable frequency by genetic manipulation of only two imprinted loci[51]. Therefore, there is no strong argument for a restriction of the transgenerational contribution by Polycomb to the regulation of developmental genes on the paternal genome only.

For H3K4me2, it is unknown whether the mark functions only in the process of transcription or whether it also serves a role in epigenetic memory of the active state in proliferating cells. Nuclear-transfer experiments performed in *Xenopus laevis* oocytes provided evidence for a role of H3K4 in transcriptional memory[52]. In *Caenorhabditis elegans*, deficiency for the H3K4me2 demethylase Lsd1 (KDM1) caused progressive sterility over many generations that correlated with transgenerational accumulation of H3K4me2 in the germ line and increased expression of spermatogenic genes in the soma[53]. These data argue that programmed H3K4 demethylation, possibly of testis-expressed genes, is required for germline immortality in *C. elegans*. In mouse and human embryos, H3K4 methylation is established along the paternal genome within the first cell cycle[46,54], providing a potential means for somatic transmission. Nevertheless, the fate of germline-inherited H3K4me2 at, for example, housekeeping versus testis-specific genes remains to be tested.

Molecular genetic experiments will be required to elucidate the extent and functional significance of methylation at distinct histone residues and loci for transgenerational inheritance. Beyond the intrinsic variation in modification coverage between loci, there may be variability in establishment and retention of modified histones between spermatozoa and/or individuals in response to environmental influences and/or innate cues, such as incomplete chromatin remodeling during spermatid elongation[55]. Hence, transmission of histone-encoded epigenetic information may prove to constitute a previously unknown transgenerational mechanism for phenotypic variation[56].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. NCBI Gene Expression Omnibus: Data have been deposited with accession code GSE19892.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank T. Jenuwein (Max-Planck Institute of Immunobiology, Freiburg, Germany) and R. Foisner (Max F. Perutz Laboratories, Medical University of Vienna, Vienna, Austria) for providing antisera. We are grateful to F. Zilbermann and the Friedrich Miescher Institute (FMI) animal facility for excellent technical assistance. We thank P. de Boer and R. Terranova for critical reading of the manuscript and all members of the Peters and Schübeler laboratories for discussions. U.B. and S.E. acknowledge the Boehringer Ingelheim Fonds for their PhD fellowships. M.H. was supported by the Swiss National Science Foundation and the Japan Society for the Promotion of Science. T.C.R. was supported by the Swiss Cancer Leagues. Research in the Peters laboratory is supported by the Novartis Research Foundation, the Swiss National Science Foundation, the European Network of Excellence "The Epigenome" and the EMBO YIP program.

AUTHOR CONTRIBUTIONS

U.B., M.H., S.E. and A.H.F.M.P. conceived and designed the experiments; U.B., M.H. and S.E. performed the experiments; L.R. provided purified samples of human spermatozoa; U.B., M.H., S.E., E.J.O., M.B.S. and A.H.F.M.P. analyzed the data; C.B. contributed to deep-sequencing; T.C.R. and D.S. provided advice on data analyses and the manuscript; U.B., S.E. and A.H.F.M.P. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Chong, S. & Whitelaw, E. Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* **14**, 692–696 (2004).
2. Hochedlinger, K. & Jaenisch, R. Nuclear transplantation, embryonic stem cells, and the potential for cell therapy. *N. Engl. J. Med.* **349**, 275–286 (2003).
3. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
4. Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat. Genet.* **40**, 411–420 (2008).
5. Albert, M. & Peters, A.H. Genetic and epigenetic control of early mouse development. *Curr. Opin. Genet. Dev.* **19**, 113–121 (2009).
6. Balhorn, R., Gledhill, B.L. & Wyrobek, A.J. Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16**, 4074–4080 (1977).
7. 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**, 962–964 (1987).
8. Gardiner-Garden, M., Ballesteros, M., Gordon, M. & Tam, P.P. Histone- and protamine-DNA association: conservation of different patterns within the β-globin domain in human sperm. *Mol. Cell. Biol.* **18**, 3350–3356 (1998).
9. Wykes, S.M. & Krawetz, S.A. The structural organization of sperm chromatin. *J. Biol. Chem.* **278**, 29471–29477 (2003).
10. Pittoggi, C. *et al.* A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. *J. Cell Sci.* **112**, 3537–3548 (1999).
11. Hammoud, S.S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478 (2009).
12. van der Heijden, G.W. *et al.* Sperm-derived histones contribute to zygotic chromatin in humans. *BMC Dev. Biol.* **8**, 34 (2008).
13. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* **6**, 846–856 (2006).
14. Hublitz, P., Albert, M. & Peters, A.H. Mechanisms of transcriptional repression by histone lysine methylation. *Int. J. Dev. Biol.* **53**, 335–354 (2009).
15. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
16. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
17. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
18. Mohn, F. *et al.* Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
19. Ezhkova, E. *et al.* Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136**, 1122–1135 (2009).
20. Margueron, R. *et al.* Role of the Polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
21. Hansen, K.H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).
22. Alsheimer, M., Fecher, E. & Benavente, R. Nuclear envelope remodelling during rat spermiogenesis: distribution and expression pattern of LAP2/thymopoietins. *J. Cell Sci.* **111**, 2227–2234 (1998).
23. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
24. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
25. Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6**, 34 (2006).
26. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299–312 (2007).
27. Zhao, X.D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298 (2007).
28. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20**, 1123–1136 (2006).
29. Chalmel, F. *et al.* The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. USA* **104**, 8346–8351 (2007).
30. Geremia, R., Boitani, C., Conti, M. & Monesi, V. RNA synthesis in spermatocytes and spermatids and preservation of meiotic RNA during spermiogenesis in the mouse. *Cell Differ.* **5**, 343–355 (1977).
31. Namekawa, S.H. *et al.* Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667 (2006).
32. Zeng, F. & Schultz, R.M. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Dev. Biol.* **283**, 40–57 (2005).
33. Ohta, H., Tohda, A. & Nishimune, Y. Proliferation and differentiation of spermatogonial stem cells in the *w/wv* mutant mouse testis. *Biol. Reprod.* **69**, 1815–1821 (2003).

34. Anderson, E.L. *et al.* Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc. Natl. Acad. Sci. USA* **105**, 14976–14980 (2008).
35. Kaneda, M. *et al.* Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).
36. Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102 (2005).
37. Ng, R.K. *et al.* Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* **10**, 1280–1290 (2008).
38. Pan, H., O'Brien, M.J., Wigglesworth, K., Eppig, J.J. & Schultz, R.M. Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development *in vitro*. *Dev. Biol.* **286**, 493–506 (2005).
39. Faulkner, G.J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
40. Arpanahi, A. *et al.* Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res.* **19**, 1338–1349 (2009).
41. Gineitis, A.A., Zalenskaya, I.A., Yau, P.M., Bradbury, E.M. & Zalensky, A.O. Human sperm telomere-binding complex involves histone H2B and secures telomere membrane attachment. *J. Cell Biol.* **151**, 1591–1598 (2000).
42. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
43. Chu, D.S. *et al.* Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. *Nature* **443**, 101–105 (2006).
44. Dorus, S. *et al.* Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat. Genet.* **38**, 1440–1445 (2006).
45. Ooi, S.L. & Henikoff, S. Germline histone dynamics and epigenetics. *Curr. Opin. Cell Biol.* **19**, 257–265 (2007).
46. 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. *Mech. Dev.* **122**, 1008–1022 (2005).
47. Torres-Padilla, M.E., Bannister, A.J., Hurd, P.J., Kouzarides, T. & Zernicka-Goetz, M. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int. J. Dev. Biol.* **50**, 455–461 (2006).
48. Minami, N., Suzuki, T. & Tsukamoto, S. Zygotic gene activation and maternal factors in mammals. *J. Reprod. Dev.* **53**, 707–715 (2007).
49. Schuettengruber, B. & Cavalli, G. Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531–3542 (2009).
50. Bao, S. *et al.* Initiation of epigenetic reprogramming of the X chromosome in somatic nuclei transplanted to a mouse oocyte. *EMBO Rep.* **6**, 748–754 (2005).
51. Kawahara, M. *et al.* High-frequency generation of viable mice from engineered bi-maternal embryos. *Nat. Biotechnol.* **25**, 1045–1050 (2007).
52. Ng, R.K. & Gurdon, J.B. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat. Cell Biol.* **10**, 102–109 (2008).
53. Katz, D.J., Edwards, T.M., Reinke, V. & Kelly, W.G.A. *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308–320 (2009).
54. van der Heijden, G.W. *et al.* Parental origin of chromatin in human monopronuclear zygotes revealed by asymmetric histone methylation patterns, differs between IVF and ICSI. *Mol. Reprod. Dev.* **76**, 101–108 (2009).
55. Ramos, L. *et al.* Incomplete nuclear transformation of human spermatozoa in oligo-astheno-teratospermia: characterization by indirect immunofluorescence of chromatin and thiol status. *Hum. Reprod.* **23**, 259–270 (2008).
56. Chong, S. *et al.* Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat. Genet.* **39**, 614–622 (2007).

ONLINE METHODS

Sperm sample collection and purification. Human spermatozoa were obtained from normospermic men and purified by density gradient centrifugation using Pure Sperm solution (Nidacom). Mature mouse spermatozoa were obtained from caudal epididymis of CD1 and C57BL/6J mice (for western blotting and ChIP, respectively) using the swim-up assay. For details on sperm sample preparation, see **Supplementary Methods**.

Western blotting. Proteins from mouse and human sperm were isolated as described[57] with minor modifications. Additional steps of sonication (two times 30 s at 40% amplitude using a Branson sonicator) and extraction with 1.6 M urea, 1 M NaCl and 0.28 M β-mercaptoethanol for 30 min at 37 °C were included. After precipitation with 20% (w/v) trichloroacetic acid, protein pellets were boiled for 20 min in SDS sample buffer and separated by SDS-PAGE. Western blot analyses were performed using the following antibodies and dilutions: polyclonal H3K4me2 (Upstate, catalog no. 07030) (1:1,000), polyclonal H3K27me3 (Upstate, catalog no. 07449) (1:1,000), polyclonal H3 (Abcam, catalog no. 17921) (1:10,000), monoclonal LAP2β[58] (1:5). Protein extracts from WI38 human primary lung fibroblasts and CCE mouse ESCs were used as controls and were prepared as described above.

Cross-linked chromatin immunoprecipitation. ChIP-chip experiments were performed on a pool of nine donor samples to average possible variability between individuals. H3K4me2 and H3K27me3 ChIPs were carried out in parallel on identical sets of samples. For each ChIP assay, 2×10^7 spermatozoa were used. After thawing, pooled samples were washed with PBS to remove the cryopreservation medium (5 min at 800*g*). ChIP experiments were performed as described[24] with several modifications. Fixation was performed with 0.5% (w/v) paraformaldehyde for 10 min at room temperature (22–24 °C). Lysis was performed in the presence of 0.5% (w/v) SDS and 10 mM DTT for 1 h at room temperature. *N*-Ethylmaleimide was added to quench the DTT, and the samples were diluted 2.5 times before sonication. Sonication was performed six times for 20 s (Branson sonicator, amplitude 70%) to obtain chromatin with fragment sizes of 300–700 bp.

Sperm chromatin was then used for immunoprecipitation at 4 °C overnight with 5 µg of antibody: H3K4me2 (Upstate, catalog no. 07030) or H3K27me3 (Upstate, catalog no. 07449). The following steps included incubation with protein A–Sepharose beads and washing as described in ref. 24. Cross-link reversal, DNA isolation and amplification with the WGA2 amplification kit (Sigma) was performed according to ref. 59. For amplification, 50 ng of input DNA and entire ChIP DNA were used. For each H3K27me3 array experiment, three simultaneously prepared ChIP samples were pooled and used for the amplification. For each H3K4me2 array experiment, one ChIP sample was used for amplification. A set of four genes was tested for each sample by qPCR, and all showed similar bound-to-input ratios before and after amplification. Amplified DNA was hybridized to a tiling microarray (NimbleGen Systems) representing 18,029 promoter regions (2,200 bp upstream to 500 bp downstream of TSSs) of all RefSeq annotated human genes. Sample labeling, hybridization and array scanning were performed by NimbleGen Systems according to standard procedures.

Validation of microarray results was performed by ChIP-qPCR analyses using the SYBR Green PCR Master Mix (Applied Biosystem) and the ABI Prism 7500 real-time PCR machine and is presented in **Supplementary Figure 2** (for a list of primers, see **Supplementary Table 3**). ChIP was performed on pools of sperm obtained from donors different from those used for ChIP-chip analyses.

Analyses of ChIP-chip data. For each probe, the $\log_2$ ratio of precipitated over input DNA was calculated. Loess normalization and a signal intensity cut off were applied to correct for labeling dye artifacts and to remove noise. A hidden Markov model peak-finding algorithm was applied to identify regions enriched in analyzed modifications. Peaks were mapped to the closest Ensembl annotated TSS (release 48, genome build hg18). Single mouse orthologous genes were identified using Ensembl criteria (http://www.ensembl.org/info/docs/compara/homology_method.html). Gene ontology analyses were performed using GO Stat (http://gostat.wehi.edu.au). To relate histone-modification data to expression status of linked genes, we processed publicly available Affymetrix CEL files using Genedata's Expressionist pro 5.0 (Genedata AG). For details on data processing and analyses, see **Supplementary Methods**.

Nucleosomal DNA preparation and native chromatin immunoprecipitation. Native ChIP on human and mouse sperm was performed according to the protocol by Umlauf and colleagues (http://www.epigenome-noe.net/researchtools/protocol.php?protid=22) with modifications[60]. For one ChIP on human sperm, three donor samples were pooled. For both mouse and human samples, 1×10^7 spermatozoa were used per ChIP assay. Before ChIP, mouse spermatozoa were treated with 50 mM DTT in PBS at room temperature for 2 h, followed by *N*-ethylmaleimide treatment and washing with PBS. Subsequently, human spermatozoa were lysed on ice for 30 min in a buffer containing 0.3 M sucrose, 15 mM Tris (pH 7.5), 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5% deoxycholate, 0.25% nonidet P-40 and 10 mM DTT. Mouse spermatozoa were lysed for 10 min under identical conditions except for DTT (0.5 mM). Chromatin digestion was performed by MNase as described[60]. Immunoprecipitation was then carried out with H3K4me2 (Upstate, catalog no. 07030) or H3K27me3 (Upstate, catalog no. 07449) antibodies following the published protocol[60]. Real-time PCR was performed using the SYBR Green PCR Master Mix (Applied Biosystems) and an ABI Prism 7500 Real-time PCR machine (for list of primers, see **Supplementary Table 3**).

Mononucleosomal DNA was prepared using the same protocol, subjected to DNA electrophoresis and isolated from polyacrylamide gels. For one mononucleosomal DNA preparation, one donor sample was used. DNA was prepared for deep sequencing using the ChIP-Seq sample preparation kit from Illumina and sequenced on the Genome Analyzer 2 following manufacturer's protocols. For details on data processing and analyses, see **Supplementary Methods**.

Statistical analyses. Statistical tests were performed using the R software (http://www.r-project.org). In **Figures 2b** and **3c**, a two-sided Wilcoxon rank-sum test was used, as a nonparametric test of location for non-normal data. In **Figure 2a**, the one-sided hypergeometric test (R "Phyper" function) was used to measure the probability of observing an overlap equal to or smaller than that obtained from the real data. Associations in **Figure 4** were tested using Pearson's Chi-square test on the raw count data. The *P* values reported for enriched gene ontology terms (**Figs. 1d** and **3d**,**e**) were obtained using GO Stat.

57. Lee, K., Haugen, H.S., Clegg, C.H. & Braun, R.E. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc. Natl. Acad. Sci. USA* **92**, 12451–12455 (1995).
58. Dechat, T. *et al.* Detergent-salt resistance of LAP2α in interphase nuclei and phosphorylation-dependent association with chromosomes early in nuclear assembly implies functions in nuclear structure dynamics. *EMBO J.* **17**, 4887–4902 (1998).
59. O'Geen, H., Nicolet, C.M., Blahnik, K., Green, R. & Farnham, P.J. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* **41**, 577–580 (2006).
60. Umlauf, D., Goto, Y. & Feil, R. Site-specific analysis of histone methylation and acetylation. *Methods Mol. Biol.* **287**, 99–120 (2004).

doi:10.1038/nsmb.1821

Chapter 2: Results

2.2 Published review:

Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming?

This article was first published in Current Opinion in Cell Biology in June 24, 2012, volume 3, pages 387-96, doi: 10.1016/j.ceb.2012.03.002.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Cell Biology

Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming?

Mark E Gill[1], Serap Erkek[1,2] and Antoine HFM Peters[1]

At fertilization, fusion of two differentiated gametes forms the zygote that is capable of forming all of the varied cell lineages of an organism. It is widely thought that the acquisition of totipotency involves extensive epigenetic reprogramming of the germline state into an embryonic state. However, recent data argue that this reprogramming is incomplete and that substantial epigenetic information passes from one generation to the next. In this review we summarize the changes in chromatin states that take place during mammalian gametogenesis and examine the evidence that early mammalian embryogenesis may be affected by inheritance of epigenetic information from the parental generation.

Addresses
[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland
[2] Faculty of Science, University of Basel, CH-4003 Basel, Switzerland

Corresponding author: Peters, Antoine HFM (antoine.peters@fmi.ch)

Current Opinion in Cell Biology 2012, **24**:387–396

This review comes from a themed issue on
Nucleus and gene expression
Edited by Asifa Akhtar and Karla Neugebauer

Available online 23rd March 2012

0955-0674/$ – see front matter

DOI 10.1016/j.ceb.2012.03.002

Introduction

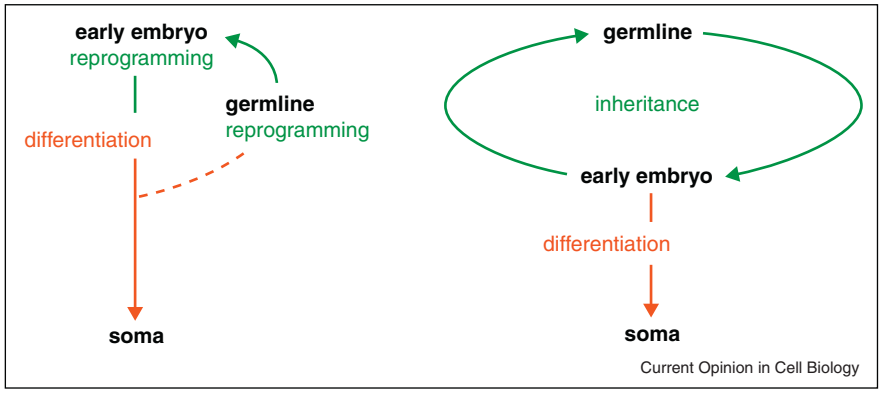
Embryos develop from the fusion of two highly specialized haploid germ cells, oocytes and sperm, yet possess during early development the property of totipotency, the ability to give rise to every cell type in an organism. The acquisition of developmental potency is thought to arise from reprogramming of the parental germline epigenetic state to a new epigenetic state in the early embryo. The full extent of this reprogramming is, however, unknown, leaving open the question of how much epigenetic information present in mature gametes is retained in the embryo.

Epigenetic information is generally considered to represent heritable information in genome function that is not encoded by the DNA sequence. On the basis of studies of embryonic development and cellular differentiation, several mechanisms have been proposed for the transmission of epigenetic information. These include chemical alteration of DNA itself (most often by methylation or related chemical groups), post-translational modifications of histones, and transmission of RNA. Recent genome-wide chromatin analyses in a variety of cell types revealed that regulatory sequences of genes as well as of repetitive sequences are generally associated with DNA and/or distinct histone modifications and chromatin associated proteins [1]. The presence of a defined 'epigenome' in embryonic stem (ES) cells raises the question about the ontogeny of such a pluripotency-related chromatin program: whether it is newly defined during pre-implantation development or inherited from gametes.

To establish a conceptual framework for future experiments, we propose two opposite models for epigenetic control of mammalian pre-implantation development (Figure 1): The *reprogramming model* proposes that chromatin states in gametes become reset upon fertilization to enable the acquisition of totipotency. This classical model is in part based on global changes in DNA methylation along the paternal and maternal genomes of pre-implantation embryos, as visualized by immuno-fluorescence microscopy [2] and the ability of the cytoplasm of oocytes and zygotes to partially reprogram the epigenome of differentiated nuclei [3]. Likewise, during primordial germ cell (PGC) specification, somatic epigenetic programs acquired during early embryogenesis are erased and subsequently replaced by female-specific and male-specific germ cell programs. In the case of DNA methylation in the context of genomic imprinting, however, part of the germline program escapes reprogramming in early embryos, safe guarding parental-specific expression during somatic development [4].

In contrast, the *inheritance model* proposes that chromatin states in mature germ cells are inherited by embryos to direct transcriptional activation or silencing upon zygotic genome activation and subsequent development. Chromatin programs in mature germ cells may either be specified during gametogenesis, or originate from pre-implantation embryos or even from parental gametes, thereby enabling intergenerational or transgenerational inheritance of epigenetic information. Transmission of such pre-patterned chromatin states, constituting a default '*intrinsic intergenerational/transgenerational inheritance program*', could contribute to the high level of developmental potential of early embryos. We further postulate that '*acquired intergenerational/transgenerational inheritance*' results from a temporal exposure to environmental cues altering the cellular state of the germline in

Figure 1



The *reprogramming* model (left) and *inheritance* model (right) of chromatin-based epigenetic control of early embryonic development. See text for explanation.

one generation and changing transitorily, for one or multiple generations, the execution of the default intrinsic inheritance program, and thereby affecting the phenotype of offspring [5]. Acquired inheritance would depend on the interaction between germ cells and somatic cells in the organism, either directly between gonadal somatic cells and germ cells, or indirectly by hormonal signaling. In general, acquired inheritance (and not the intrinsic inheritance program) is referred to as transgenerational inheritance [6].

Studies in non-mammalian model organisms have provided evidence for various mechanisms playing a role in intergenerational/transgenerational epigenetic inheritance. For instance, studies in plants demonstrated that altered patterns of DNA methylation (so-called epialleles) can be inherited over successive generations [7•,8]. Transmission of the chromatin state of a chromosomal regulatory element through both mitosis and meiosis has been shown in *Drosophila* [9•,10]. In *zebrafish* embryos, homeostatic and developmentally regulated genes are marked by permissive and repressive histone modifications before zygotic genome activation. This prepatterning reflects in part chromatin marking observed in *zebrafish* sperm, providing means for intergenerational inheritance [11,12]. In the context of RNA-based inheritance, transmission of silencing induced by siRNAs (for instance in *Caenorhabditis elegans*) [13] or maternal piRNAs (in *Drosophila*) [14•]) has also been established. Below, we review the current knowledge on chromatin dynamics during germ cell development and intergenerational epigenetic inheritance in mouse.

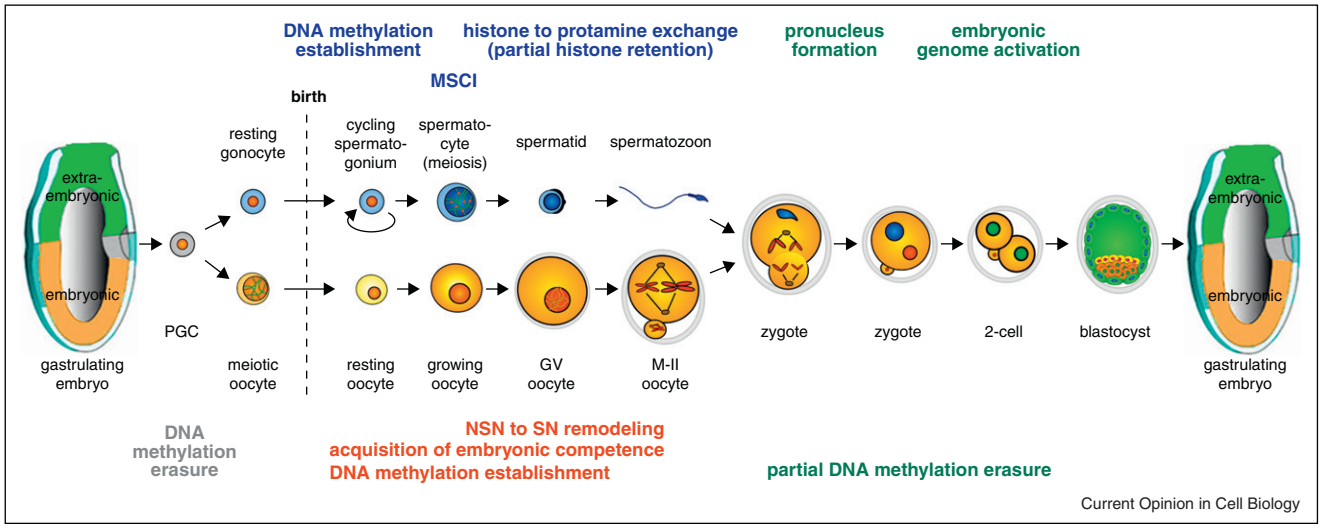
Germ cell development in mammals

In several well-studied model organisms germ cells arise via asymmetric distribution of germline inducing factors, known as germplasm, providing an obvious mechanism for intergenerational inheritance [15]. This is not the case in mammals, where PGCs are specified from a small population of posterior proximal epiblast cells (Figure 2) [16,17]. Before undergoing overt cellular differentiation, PGCs undergo changes in gene expression state and DNA and histone methylation levels that are believed to represent re-programming of these cells [18•,19]. The balance between epigenetic reprogramming versus inheritance in PGCs is, however, not understood, in part due to the technical challenges of working with these rare and relatively inaccessible cells. Following the early events of germline reprogramming, germ cells in male and female embryos embark on drastically different differentiation programs, which may suggest different capacities for intergenerational inheritance between the sexes.

DNA methylation

Methylation of the fifth position of cytosine on DNA (5mC) represents one well-studied mechanism of epigenetic transmission in the mouse. This mechanism provides evidence for both germline reprogramming and germline inheritance. During PGC migration to the developing gonad, germ cell chromatin undergoes substantial changes [19]. During the same time period levels of 5mC in germ cells also begin to decrease [18•]. PGC DNA demethylation continues after their arrival in the gonad and affects around 80–90% of the genome [20•]. The mechanism behind this demethylation has been the subject of many years of research. It was hypothesized that, following a paradigm established in plants [21], deamination and subsequent DNA repair might be involved in DNA demethylation. Deletion of the cytidine deaminase *Aid* in the mouse provided some support for this model as PGCs in these animals display modest DNA hypermethylation [20•]. Further support was provided by the observation that PGC demethylation is also associated with the presence of DNA single stranded breaks [22]. A similar mechanism for demethylation of the paternal

Figure 2



Life cycle of mammalian gametogenesis and embryogenesis. Primordial germ cells (PGCs) arise from proximal epiblast cells and undergo extensive erasure of DNA methylation and chromatin changes during migration to and upon entry into the gonad. Directed by the somatic gonadal environment, germ cells enter the male or female fate. Male germ cells, initially called gonocytes, are cell cycle arrested and start to establish male-specific DNA methylation patterns. During subsequent meiotic prophase, the X and Y chromosomes undergo meiotic sex chromosome inactivation (MSCI) characterized by major chromatin remodeling events. Following meiotic divisions, haploid spermatids undergo extensive nuclear and morphological changes including an almost genome-wide exchange of histones by protamines. Regulatory sequences, however, retain nucleosomes providing means for epigenetic inheritance. Female germ cells enter meiotic prophase in the embryo and complete the meiotic divisions upon hormonal induction in the adult ovary and fertilization by sperm. During the growing phase, oocytes establish DNA methylation at genes and imprinting control regions, undergo chromatin remodeling and acquire competence to direct embryogenesis. Upon fertilization, parental genomes form two pronuclei that are epigenetically distinct, reflecting the history of parental germline specific chromatin remodeling events. Paternal and maternal genomes undergo active and passive erasure of DNA methylation. The asymmetry in chromatin states at paternal and maternal chromosomes may potentially regulate activation and repression of *de novo* gene expression in pre-implantation embryos thereby directing embryogenesis.

genome may exist in the early embryo as well [23]. Another recent advance was the discovery that the hydroxylated form of methylated DNA (5hmC) is present in substantial quantities in mammalian genomes and that this modified base could be formed from 5mC by proteins of the Tet family [24,25•]. Three different Tet proteins exist in mice. While deletion of *Tet1* leads to no obvious mutant phenotypes [26], *Tet2* seems to function as a tumor suppressor during hematopoiesis [27,28]. Maternal deletion of *Tet3* leads to defects in early development with 50% of maternally *Tet3*-deficient embryos failing to develop to term [29]. What differences exist between these two classes of *Tet3*-deficient embryos remains unknown. Likewise, it is unclear whether this observed embryonic lethality is due to impaired 5mC to 5hmC conversion of the paternal genome in the zygote versus changes to the maternal epigenome, inherited from the mutant oocyte.

While removal of DNA methylation across the majority of the genome in PGCs would seem to preclude the inheritance of this mark, certain DNA sequences, such as the Intracisternal A Particle (IAP), appear rather resistant DNA demethylation in PGCs [30]. A well-known example of maternal epigenetic inheritance in the mouse is the agouti viable yellow (A^{vy}) allele of the agouti locus which contains an IAP retrotransposon sequence inserted near the promoter of the gene [31]. Gene expression from this allele correlates positively with the degree of DNA methylation of the IAP sequence in soma of the individual itself as well as of the mother. Furthermore, the methylation state is susceptible to environmental cues such as maternal diet [32]. Nonetheless, DNA methylation at the A^{vy} allele inherited from the oocyte is lost in pre-implantation embryos arguing that DNA methylation is not the primary mediator of intergenerational epigenetic inheritance [33]. Comparable observations were made for the *Axin-fused* allele [34,35].

Unlike IAP retrotransposons, DNA methylation at imprinted control regions (ICRs) is erased in PGCs [30]. ICRs become subsequently methylated in either the male or female germline and maintain their methylation state following fertilization to direct parental-specific expression during embryonic and postnatal development. Thus, developmental changes in DNA methylation at ICRs represent a one generational cycle of epigenetic reprogramming (in PGCs) and inheritance

(in early embryos). Intriguingly, while DNA methylation is clearly involved in epigenetic inheritance of imprinted DNA methylation, recent work indicates that feedback to DNA sequence via the Krüppel-associated box-containing zinger finger protein Zfp57 is required for efficient maintenance of imprinted DNA methylation in early embryos [36] and in ES cells [37,38].

While imprinted loci have been extensively studied, the role of inherited DNA methylation throughout the genome is less well understood. Recently, two genome-wide studies of DNA methylation on developing mouse oocytes revealed that DNA methylation established in the female germline by the *de novo* DNA methyltransferase Dnmt3a and its non-catalytic paralog Dnmt3l correlates with DNA methylation profiles measured in blastocyst embryos [39••,40•]. Surprisingly, the DNA methylation level in blastocysts was several-fold higher than expected on the basis of the widely cited model of active and passive demethylation that are thought to reprogram the paternal and maternal genomes during pre-implantation development. Kobayashi and colleagues also demonstrated that ICRs versus retrotransposons have a differential requirement for *Dnmt3l* for DNA methylation establishment [40•]. Nonetheless, embryos derived from oocytes deficient for *Dnmt3a* or *Dnmt3l* are capable of developing to midgestation in the mouse [41,42] suggesting that maternally inherited DNA methylation does not play a critical role in pre-implantation development.

Oogenesis and chromatin

Female germ cells initiate meiosis during fetal life, arresting their cell cycle at the diplotene stage of meiotic prophase (Figure 2). Beginning a few days after birth (and continuing periodically throughout the reproductive lifespan of the organism) oocytes are recruited into a growing phase, which associates with increased transcriptional activity. At the end of this growing phase, oocytes are induced to resume meiosis by a surge of luteinizing hormone (LH), re-arresting at metaphase of meiosis II (M-II). It has long been known that morphological changes in chromatin are associated with growing oocyte development [43]. During the early growing phase, chromatin of mouse oocytes exists in a de-condensed configuration known as the Non-Surrounded Nucleolus (NSN) state [44–46]. As oocytes reach the final stage of oocyte growth they undergo a change in chromatin state, forming condensed rings of chromatin (containing pericentric heterochromatin [47] around the prenucleolar body), forming the Surrounded Nucleolus (SN) state [44–46]. The transcriptional activity of oocytes correlates with their chromatin configuration: NSN oocytes display transcriptional activity while SN oocytes are transcriptionally repressed. However, the transcriptional repression found in mature oocytes is not dependent on their SN chromatin state. Despite failing to achieve the SN chromatin configuration [47], oocytes deficient for the histone chaperone Nucleoplasmin 2 (*Npm2*) undergo transcriptional repression [48]. Importantly, oocyte chromatin configuration does, however, correlate strongly with embryonic competence [49,50].

The linker histone H1 variant *H1foo* is specifically expressed in growing oocytes [51,52]. Knockdown of this variant in growing oocytes by morpholino antisense oligonucleotides leads to a reduced capacity of these oocytes to resume meiosis [53]. Oocyte chromatin also contains the H2A variant macroH2A throughout the growing phase, as well as during meiotic resumption [54]. This variant also remains associated with maternal chromatin following fertilization [54], providing a possible mediator for inheritance. As expected for non-replicating cells, growing oocytes do not incorporate the replication-dependent H3 variants, H3.1 and H3.2 [55••]. They do, however, robustly incorporate the replication-independent variant H3.3, suggesting that ongoing changes in chromatin composition occur during oocyte growth [55••]. H3.3 incorporation was observed in the nuclei of oocytes where transcription had ceased, indicating the existence of continuous, transcription-independent nucleosome turnover in oocytes [55••].

Changes in histone modifications during oogenesis have been extensively cataloged in many publications (reviewed in [56]); however, only a limited number of studies have identified functional roles for these modifications. Growing oocyte development is associated with increasing levels of histone acetylation, followed by abrupt de-acetylation during meiotic resumption [57]. Modulation of HDAC (histone de-acetylase) activity in oocytes can alter the condensation of chromatin in these cells, with increased HDAC activity leading to premature chromatin condensation [58], and HDAC inhibition leading to chromatin de-condensation [48]. Genetic ablation of *Hdac1* and *Hdac2* severely impairs oocyte growth and transcription leading to female sterility [59]. Histone de-acetylation is also important for the development of oocytes following meiotic resumption, with failure to de-acetylate histones impairing chromosome elongation and alignment during M-II. These defects are thought to be caused in part by failure centromeric heterochromatin binding by the chromatin remodeler Atrx in these cells [60•].

Histone methylation appears to be inherited by embryos through the female germline. In early embryos, heterochromatic sequences are differentially marked in pronuclei of maternal and paternal origins [61••]. Maternal heterochromatin carries *Suv39h2*-dependent H3K9 trimethylation, which is inherited from the oocyte and is required for maintenance of the canonical heterochromatic state in embryos. In paternal heterochromatin, which lacks this mark, proteins of the Polycomb

Repressive Complex 1 (a major repressive complex implicated in epigenetic repression of, e.g. developmental regulatory genes during development [62]) mediate transcriptional repression of heterochromatin associated satellites sequences [61••]. In addition to roles for histone methylation associated with repressive chromatin, trimethylation of H3K4, an active mark, has also been shown to be important in the female germline. Oocyte-specific deletion of *Mll2*, an H3K4 methyltransferase, results in decreased levels of H3K4 methylation in peri-ovulatory oocytes and impairs either ovulation or pre-implantation development, depending on the timing of conditional deletion during oogenesis [63•].

Global chromatin remodeling during spermiogenesis

In contrast to female germline development, meiosis begins in male germ cells only after birth (Figure 2). With respect to chromatin, changes are clearly detectable from the onset of meiosis. Chromatin changes may be responsible for localization of meiotic recombination (Box 1) and the specialized meiotic behavior of the sex chromosomes (Box 2). Following meiosis, haploid round spermatids undergo dramatic and extensive chromatin remodeling. This process results in the genome-wide exchange of histones by spermatogenesis-specific basic DNA packaging proteins (initially transition proteins and ultimately protamines) [64,65]. Nevertheless, in human and mouse approximately 10 and 1% of histones are retained in spermatozoa, respectively [66••]. Since these retained histones harbor post-translational modifications, as in somatic cells, they may function as mediators of epigenetic inheritance between generations.

Some histone variants such as testis-specific histone H2B variant, TH2B and testis-specific histone H1, H1t are present from early spermatogenic cells to the round spermatid stage [67–69]. However, variants of H2A such as H2AL1/2, H2A.Bbd, or of H1, such as H1t2 and Hils1, are specifically incorporated during the histone-to-protamine exchange in round/elongating spermatids [70–73]. While H2AL1/2 mark pericentric heterochromatin during

Box 1 Prdm9 and chromatin in meiotic recombination.

Chromatin plays a role in one of the fundamental processes of gametogenesis: meiosis. The products of meiosis contain a genetic complement that differs from that of their parental cells because of the process of meiotic recombination. Recombination does not occur at equal rates throughout the genome, but instead preferentially occurs at specific sites, known as 'hotspots,' where the incidence of DNA double strand breaks (DSBs) is increased [91]. Little interspecific conservation of hotspot usage has been observed, and in mice various genetic strain backgrounds also display different recombination patterns [92]. In species from yeast to mammals, recombination hotspots correlate with enrichment for H3K4me3 [93,94]. In mammals, these sites are bound by Prdm9 (also known as Meisetz), a protein capable of methylating H3K4 [95] and which is required for the completion of both male and female meiosis in mice [96]. It was shown that variations in hotspot usage correlate with variation in the DNA binding domain of Prdm9 [95,97]. It is currently unknown whether Prdm9's role in promoting DSB formation and completion of meiosis depends on its methyltransferase activity towards H3K4. Whether additional chromatin factors (beyond DNA sequence) serve to recruit Prdm9 to hotspots also remains unknown. Interestingly, hotspots strongly affect the inheritance of specific sequences, as their tendency to undergo gene conversion drives their removal from the genome [98]. This suggests that a chromatin-based system utilized during spermatogenesis contributes to paternal inheritance of DNA sequences.

Box 2 Meiotic sex chromosome inactivation.

The heteromorphic nature of mammalian sex chromosomes means that these chromosomes cannot fully homologously pair during male meiosis. These chromosomes are held together during meiotic prophase by the formation of a single crossover at X–Y homologous sequences known as the pseudoautosomal region [99]. The unpaired regions of the chromosomes form a subdomain on the nuclear periphery termed the sex body where they undergo transcriptional silencing in a process known as MSCI (meiotic sex chromosome inactivation) [100]. MSCI represents a specific instance of a common response to the existence of unpaired DNA regions during meiosis, first identified in the fungus *Neurospora*, known as MSUC (meiotic silencing of unpaired chromatin) [101]. The establishment of MSCI requires phosphorylation of the histone variant H2AX by the kinase ATR to form γH2AX [101]. While ATR is initially targeted to chromatin by the tumor suppressor BRCA1 [101], spreading of γH2AX along the sex chromosomes occurs in an MDC1-dependent manner [102]. Several additional histone variants and modifications have been proposed to regulate MSCI. For instance *Rnf8*, an E3 ubiquitin ligase is responsible for accumulation of ubiquitinated H2A on meiotic sex chromosomes. However, although deletion of *Rnf8* leads to a loss of this signal, MSCI is not impaired suggesting that H2A ubiquitination is not required for MSCI [82]. Additionally macro-H2A, a variant of H2A possessing a large non-histone domain, is strongly associated with the X and Y chromosomes during the early stages of MSCI, becomes lost (and, it is thought to be replaced with H2AZ) during late meiotic prophase [103,104]. Intriguingly, canonical replication-loaded H3 (known in mammals as H3.1/H3.2) is depleted from sex chromosomes and replaced with the replication-independent variant H3.3 during the process of MSCI [105]. What functional role this global nucleosome replacement plays in MSCI and how this process is regulated by the known regulators of MSCI (such as ATR or MDC1) remain unknown. Subsequent to the initiation of MSCI and nucleosomal remodeling, meiotic sex chromosomes are dynamically methylated at various lysine residues of histone H3 and H4 [91,105], as well as de-acetylated on the same histones.

Following meiosis, most single copy genes on the X and Y chromosomes remain in a transcriptionally silent state known as Post-meiotic Sex Chromatin (PMSC) [106,101,104], while genes present in multicopy arrays show expression in postmeiotic cells [107]. The role that PMSC plays in paternal inheritance in the early embryo is unknown. It has been proposed that imprinted inactivation of the paternal X chromosome in the pre-implantation embryo is established via a continuation of the PMSC state [108,106]. However, this view remains controversial, as several studies indicated that imprinted X inactivation is initiated in the early embryo either in a *Xist*-dependent or *Xist*-independent manner [109–111]. In addition, the maternal X chromosome harbors a strong imprint that prevents its inactivation in early mouse embryos [112,113].

spermatogenesis, they are quickly displaced from paternal heterochromatin after fertilization [74], potentially restricting a functional role in paternal inheritance. H2A.Bbd was found to destabilize nucleosomes, thus its presence in nucleosomes was suggested to facilitate replacement of histones by protamines [71]. Recently, another H2A variant, H2A.Lap1, has been shown to be loaded onto the X chromosome and autosomes in round spermatids and suggested to have a role in transcription of repressed genes [75]. Among testis-specific H1 variants, it was shown that Hils1 localizes to the same sites as Tnp2 and Prm1 [73], while deficiency for H1t2 was found to result in reduced sperm mobility and condensation defects [72].

In addition to incorporation of histone variants, post-translational modifications of histones and their read-out play a critical role for genome-wide chromatin remodeling during late spermatogenesis. Histones are hyperacetylated beginning at the round spermatid stage with increasing levels detectable in the elongating spermatid stage [76,77]. This hyperacetylation is thought to facilitate global chromatin remodeling through creation of a more accessible chromatin environment. Hyperacetylated histones are recognized by the bromodomain containing protein Brdt [78,79]. Deletion of the first bromodomain of Brdt in mice results in abnormal chromatin remodeling and male infertility [80•]. Brdt-mutant sperm is, however, capable of supporting normal embryonic development when used for intracytoplasmic sperm injection (ICSI) [80•]. Histones are also ubiquitinated in elongating spermatids [81]. In particular, the E3 ubiquitin ligase Rnf8 is required for normal histone eviction during spermatogenesis [82]. *Rnf8*-deficient spermatids are also deficient for spermatid histone acetylation, suggesting interplay between these two pathways [82]. Recently, a global survey of histone modifications identified crotonylation as a new histone modification that was found at high levels in elongating spermatids. Histone crotonylation correlates with gene expression with a specific enrichment on sex chromosomes in spermatogenic cells. The exact role that histone crotonylation plays in spermatogenesis (and elsewhere) remains to be determined [83••].

The role of several other chromatin regulators in spermatogenesis has been studied by knockout mouse models. Deficiency for *Jmjd1a*, a demethylase for H3K9me1/2, was found to block spermatid elongation and cause infertility [84]. In contrast, the absence of *Kdm4d*, a demethylase for H3K9me3, did not influence the progression of spermatogenesis, perhaps because of functional redundancy with other histone demethylases [85]. Other recent studies showed that the histone methyltransferase Mll5 is required for proper spermatid maturation and fertility [86,87].

Although there have been many studies investigating the role of histone variants and chromatin regulators in reorganization of the paternal genome, in general the mechanisms of their actions remain unknown. Several recent studies have shown that histones (and their modifications) retained in human sperm are not randomly distributed, but are instead enriched at regulatory elements of genes [88••,89,66••]. Intriguingly, differential histone modifications associate with functionally distinct set of genes suggesting that transmission of retained histones might guide transcription during early embryonic development [88••,66••].

Conclusions

Embryonic development requires epigenetic reprogramming to regenerate totipotency from a germline state in each generation. This reprogramming is likely counterbalanced by the inheritance of epigenetic information present in mature gametes. Many questions remain as to the relative contributions of these two forces in early embryogenesis and what functional roles inherited epigenetic information plays in mammalian embryogenesis.

Establishing that a system actually utilizes, for example, chromatin-based mechanisms for epigenetic inheritance is, however, difficult. For example, the recent finding that the maintenance of imprinted DNA methylation in embryos and ES cells requires feedback to the underlying genome via the Zfp57 protein, binding to specific methylated DNA sequences within ICRs, sheds a new light on 'epigenetic' inheritance of DNA methylation during early development [36–38]. Furthermore, a number of chromatin modifying enzymes have been shown to possess catalytic activity towards non-histone proteins, potentially complicating the interpretation of genetic deficiency studies [90].

Nonetheless, to provide evidence for the 'intrinsic intergenerational/transgenerational inheritance program' model, more in depth molecular genetic studies are needed to elucidate the germline function of major epigenetic regulators shown to be involved in inheritance during somatic development. To understand the function of paternally transmitted (modified) histones for embryogenesis, it is necessary to dissect the mechanisms of histone eviction versus retention during spermiogenesis. Ultimately, it will be required to study the effect of germline expression of histones with residue specific mutations on parent-of-origin transcription during embryogenesis.

Acknowledgements

M. Gill greatly acknowledges his EMBO long-term fellowship and S. Erkek her Boehringer Ingelheim Fonds PhD fellowship. Research in the Peters lab is supported by the Novartis Research Foundation, the Swiss National Science Foundation (31003A_125386 and NRP 63 – Stem Cells and Regenerative Medicine), SystemsX.ch (Cell plasticity), the Japanese Swiss Science and Technology Cooperation Program, and the EMBO Young Investigator Program.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ram O, Goren A, Amit I, Shoresh N, Yosef N, Ernst J, Kellis M, Gymrek M, Issner R, Coyne M *et al.*: **Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells**. *Cell* 2011, **147**:1628-1639.

2. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T: **Demethylation of the zygotic paternal genome**. *Nature* 2000, **403**:501-502.

3. Egli D, Rosains J, Birkhoff G, Eggan K: **Developmental reprogramming after chromosome transfer into mitotic mouse zygotes**. *Nature* 2007, **447**:679-685.

4. Bartolomei MS, Ferguson-Smith AC: **Mammalian genomic imprinting**. *Cold Spring Harb Perspect Biol* 2011, **3**.

5. Daxinger L, Whitelaw E: **Transgenerational epigenetic inheritance: more questions than answers**. *Genome Res* 2010, **20**:1623-1628.

6. Daxinger L, Whitelaw E: **Understanding transgenerational epigenetic inheritance via the gametes in mammals**. *Nat Rev Genet* 2012, **13**:153-162.

7. • Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P *et al.*: **Assessing the impact of transgenerational epigenetic variation on complex traits**. *PLoS Genet* 2009, **5**:e1000530.

This study was one of the first studies showing that epialleles can be stably inherited over many successive generations in plants.

8. Reinders J, Wulff BB, Mirouze M, Mari-Ordonez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J: **Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes**. *Genes Dev* 2009, **23**:939-950.

9. • Cavalli G, Paro R: **The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis**. *Cell* 1998, **93**:505-518.

With this study, for the first time in *Drosophila* it was shown that the chromatin state of a chromosomal element, established by Polycomb and Trithorax group of proteins, could be transmitted via both mitosis and meiosis.

10. Cavalli G, Paro R: **Epigenetic inheritance of active chromatin after removal of the main transactivator**. *Science* 1999, **286**:955-958.

11. Lindeman LC, Andersen IS, Reiner AH, Li N, Aanes H, Ostrup O, Winata C, Mathavan S, Muller F, Alestrom P *et al.*: **Prepatterning of developmental gene expression by modified histones before zygotic genome activation**. *Dev Cell* 2011, **21**:993-1004.

12. Wu SF, Zhang H, Cairns BR: **Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm**. *Genome Res* 2011, **21**:578-589.

13. Alcazar RM, Lin R, Fire AZ: **Transmission dynamics of heritable silencing induced by double-stranded RNA in *Caenorhabditis elegans***. *Genetics* 2008, **180**:1275-1288.

14. • Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ: **An epigenetic role for maternally inherited piRNAs in transposon silencing**. *Science* 2008, **322**:1387-1392.

piRNAs play an important role in silencing of transposable elements in germ cells. This study showed that piRNAs transmitted by maternal germline are critical for embryonic silencing of transposable elements inherited from the male and female germlines.

15. Johnson L, Llewellyn CH, van Jaarsveld CH, Cole TJ, Wardle J: **Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort**. *PLoS One* 2011, **6**:e19918.

16. Saitou M, Payer B, Lange UC, Erhardt S, Barton SC, Surani MA: **Specification of germ cell fate in mice**. *Philos Trans R Soc Lond B: Biol Sci* 2003, **358**:1363-1370.

17. Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M: **A signaling principle for the specification of the germ cell lineage in mice**. *Cell* 2009, **137**:571-584.

18. • Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y: **Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice**. *Dev Biol* 2005, **278**:440-458.

This study is one of the initial examples describing the major changes in chromatin states occurring during development of primordial germ cells. During migration of germ cells to the gonad, H3K9 dimethylation and DNA methylation levels are greatly reduced, whereas H3K27 trimethylation levels are significantly increased.

19. Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M: **Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice**. *Development* 2007, **134**:2627-2638.

20. • Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W: **Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency**. *Nature* 2010, **463**:1101-1105.

This study measured DNA methylation levels in PGCs and showed that *Aid*-deficiency leads to globally higher levels of DNA methylation in primordial germ cells.

21. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL: **DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation**. *Cell* 2006, **124**:495-506.

22. Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA: **Genome-wide reprogramming in the mouse germline entails the base excision repair pathway**. *Science* 2010, **329**:78-82.

23. Wossidlo M, Arand J, Sebastiano V, Lepikhov K, Boiani M, Reinhardt R, Scholer H, Walter J: **Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes**. *EMBO J* 2010, **29**:1877-1888.

24. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain**. *Science* 2009, **324**:929-930.

25. • Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L *et al.*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1**. *Science* 2009, **324**:930-935.

5-hmC is quite widespread in the mouse genome, as determined by this study. It was also shown that Tet proteins can convert 5-mC to 5-hmC.

26. Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi SW, Page DC *et al.*: **Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development**. *Cell Stem Cell* 2011, **9**:166-175.

27. Li Z, Cai X, Cai CL, Wang J, Zhang W, Petersen BE, Yang FC, Xu M: **Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies**. *Blood* 2011, **118**:4509-4518.

28. Kunimoto H, Fukuchi Y, Sakurai M, Sadahira K, Ikeda Y, Okamoto S, Nakajima H: **Tet2 disruption leads to enhanced self-renewal and altered differentiation of fetal liver hematopoietic stem cells**. *Sci Rep* 2012, **2**:273.

29. Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG *et al.*: **The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes**. *Nature* 2011, **477**:606-610.

30. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA: **Epigenetic reprogramming in mouse primordial germ cells**. *Mech Dev* 2002, **117**:15-23.

31. Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP: **Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage**. *Genes Dev* 1994, **8**:1463-1472.

32. Cropley JE, Suter CM, Beckman KB, Martin DI: **Germ-line epigenetic modification of the murine A vy allele by nutritional**

supplementation. *Proc Natl Acad Sci U S A* 2006, **103**:17308-17312.

33. Blewitt ME, Vickaryous NK, Paldi A, Koseki H, Whitelaw E: **Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice**. *PLoS Genet* 2006, **2**:e49.

34. Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E: **Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission**. *Proc Natl Acad Sci U S A* 2003, **100**:2538-2543.

35. Fernandez-Gonzalez R, Ramirez MA, Pericuesta E, Calle A, Gutierrez-Adan A: **Histone modifications at the blastocyst Axin1(Fu) locus mark the heritability of in vitro culture-induced epigenetic alterations in mice**. *Biol Reprod* 2010, **83**:720-727.

36. Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC: **A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints**. *Dev Cell* 2008, **15**:547-557.

37. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A *et al.*: **In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions**. *Mol Cell* 2011, **44**:361-372.

38. Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G *et al.*: **Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain**. *J Biol Chem* 2011, **287**:2107-2118.

39. •• Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G: **Dynamic CpG island methylation landscape in oocytes and preimplantation embryos**. *Nat Genet* 2011, **43**:811-814.
By RRBS the authors determined genome-wide DNA methylation levels in mouse oocytes of various stages of development. The paper also shows that oocyte established methylation patterns are partially maintained in the embryo.

40. • Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y *et al.*: **Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks**. *PLoS Genet* 2012, **8**:e1002440.
This is the second study to examine global patterns of DNA methylation in mammalian oocytes and early embryos. The authors establish that methylation of different classes of sequences display differential requirements for Dnmt3l.

41. Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, Sasaki H: **Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation**. *Genes Cells* 2010, **15**:169-179.

42. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH: **Dnmt3L and the establishment of maternal genomic imprints**. *Science* 2001, **294**:2536-2539.

43. Chouinard LA: **A light- and electron-microscope study of the oocyte nucleus during development of the antral follicle in the prepubertal mouse**. *J Cell Sci* 1975, **17**:589-615.

44. Mattson BA, Albertini DF: **Oogenesis: chromatin and microtubule dynamics during meiotic prophase**. *Mol Reprod Dev* 1990, **25**:374-383.

45. Debey P, Szollosi MS, Szollosi D, Vautier D, Girousse A, Besombes D: **Competent mouse oocytes isolated from antral follicles exhibit different chromatin organization and follow different maturation dynamics**. *Mol Reprod Dev* 1993, **36**:59-74.

46. Zuccotti M, Piccinelli A, Giorgi Rossi P, Garagna S, Redi CA: **Chromatin organization during mouse oocyte growth**. *Mol Reprod Dev* 1995, **41**:479-485.

47. Burns KH, Viveiros MM, Ren Y, Wang P, DeMayo FJ, Frail DE, Eppig JJ, Matzuk MM: **Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos**. *Science* 2003, **300**:633-636.

48. De La Fuente R, Viveiros MM, Burns KH, Adashi EY, Matzuk MM, Eppig JJ: **Major chromatin remodeling in the germinal vesicle (GV) of mammalian oocytes is dispensable for global transcriptional silencing but required for centromeric heterochromatin function**. *Dev Biol* 2004, **275**:447-458.

49. Zuccotti M, Giorgi Rossi P, Martinez A, Garagna S, Forabosco A, Redi CA: **Meiotic and developmental competence of mouse antral oocytes**. *Biol Reprod* 1998, **58**:700-704.

50. Inoue A, Nakajima R, Nagata M, Aoki F: **Contribution of the oocyte nucleus and cytoplasm to the determination of meiotic and developmental competence in mice**. *Hum Reprod* 2008, **23**:1377-1384.

51. Tanaka M, Hennebold JD, Macfarlane J, Adashi EY: **A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog**. *Development* 2001, **128**:655-664.

52. Tanaka M, Kihara M, Hennebold JD, Eppig JJ, Viveiros MM, Emery BR, Carrell DT, Kirkman NJ, Meczekalski B, Zhou J *et al.*: **H1FOO is coupled to the initiation of oocytic growth**. *Biol Reprod* 2005, **72**:135-142.

53. Furuya M, Tanaka M, Teranishi T, Matsumoto K, Hosoi Y, Saeki K, Ishimoto H, Minegishi K, Iritani A, Yoshimura Y: **H1foo is indispensable for meiotic maturation of the mouse oocyte**. *J Reprod Dev* 2007, **53**:895-902.

54. Chang CC, Ma Y, Jacobs S, Tian XC, Yang X, Rasmussen TP: **A maternal store of macroH2A is removed from pronuclei prior to onset of somatic macroH2A expression in preimplantation embryos**. *Dev Biol* 2005, **278**:367-380.

55. •• Akiyama T, Suzuki O, Matsuda J, Aoki F: **Dynamic replacement of histone H3 variants reprograms epigenetic marks in early mouse embryos**. *PLoS Genet* 2011, **7**:e1002279.
During early embryogenesis, paternal DNA becomes remodeled as the sperm chromatin is decondensed. This work reveals that maternal chromatin also undergoes remodeling during early development and reveals the timing of these changes.

56. De La Fuente R: **Chromatin modifications in the germinal vesicle (GV) of mammalian oocytes**. *Dev Biol* 2006, **292**:1-12.

57. Kim JM, Liu H, Tazaki M, Nagata M, Aoki F: **Changes in histone acetylation during mouse oocyte meiosis**. *J Cell Biol* 2003, **162**:37-46.

58. Verdel A, Seigneurin-Berny D, Faure AK, Eddahbi M, Khochbin S, Nonchev S: **HDAC6-induced premature chromatin compaction in mouse oocytes and fertilised eggs**. *Zygote* 2003, **11**:323-328.

59. Ma P, Pan H, Montgomery RL, Olson EN, Schultz RM: **Compensatory functions of histone deacetylase 1 (HDAC1) and HDAC2 regulate transcription and apoptosis during mouse oocyte development**. *Proc Natl Acad Sci U S A* 2012, **109**:E481-E489.

60. • De La Fuente R, Viveiros MM, Wigglesworth K, Eppig JJ: **ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes**. *Dev Biol* 2004, **272**:1-14.
As oocytes differentiate, chromatin structure becomes more condensed. This study showed that HDAC activity is required for this condensation and that this activity is important for progression through meiosis.

61. •• Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K, Brykczynska U, Kolb C, Otte AP, Koseki H, Orkin SH *et al.*: **PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos**. *Nat Genet* 2008, **40**:411-420.
In early embryos male and female chromatin is marked by different repressive chromatin marks. This paper demonstrates that H3K9me3 established by the *Suv39h2* enzyme in oocytes serve as a default intergenerational signal for constitutive heterochromatin formation in the next generation. In its absence, Polycomb Repressive Complex 1 proteins form an alternative repressive state at major satellites underlying constitutive heterochromatin.

62. Morey L, Helin K: **Polycomb group protein-mediated repression of transcription**. *Trends Biochem Sci* 2010, **35**:323-332.

63. • Andreu-Vieyra CV, Chen R, Agno JE, Glaser S, Anastassiadis K, Stewart AF, Matzuk MM: **MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing**. *PLoS Biol* 2010:8.
MLL2 functions in oocytes to provide normal levels of H3K4me3 and deficiency for this gene during oogenesis results in developmental failure.

64. Balhorn R, Gledhill BL, Wyrobek AJ: **Mouse sperm chromatin proteins: quantitative isolation and partial characterization**. *Biochemistry* 1977, **16**:4074-4080.

65. Gatewood JM, Cook GR, Balhorn R, Bradbury EM, Schmid CW: **Sequence-specific packaging of DNA in human sperm chromatin**. *Science* 1987, **236**:962-964.

66. •• Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schubeler D, Stadler MB, Peters AH: **Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa**. *Nat Struct Mol Biol* 2010, **17**:679-687.
This study shows that different histone modifications associate with distinct promoter sets in human sperm. H3K4me2 marks genes involved in spermatogenesis and cellular homeostasis, while H3K27me3 marks developmental promoters. Importantly, the study also shows that histone modification states in human and mouse sperm are highly conserved.

67. Unni E, Zhang Y, Kangasniemi M, Saperstein W, Moss SB, Meistrich ML: **Stage-specific distribution of the spermatid-specific histone 2B in the rat testis**. *Biol Reprod* 1995, **53**:820-826.

68. Drabent B, Bode C, Bramlage B, Doenecke D: **Expression of the mouse testicular histone gene H1t during spermatogenesis**. *Histochem Cell Biol* 1996, **106**:247-251.

69. Steger K, Klonisch T, Gavenis K, Drabent B, Doenecke D, Bergmann M: **Expression of mRNA and protein of nucleoproteins during human spermiogenesis**. *Mol Hum Reprod* 1998, **4**:939-945.

70. Govin J, Escoffier E, Rousseaux S, Kuhn L, Ferro M, Thevenon J, Catena R, Davidson I, Garin J, Khochbin S *et al.*: **Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis**. *J Cell Biol* 2007, **176**:283-294.

71. Ishibashi T, Li A, Eirin-Lopez JM, Zhao M, Missiaen K, Abbott DW, Meistrich M, Hendzel MJ, Ausio F H2A: **Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis**. *Nucleic Acids Res* 2010, **38**:1780-1789.

72. Martianov I, Brancorsini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, Sassone-Corsi P, Davidson I: **Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis**. *Proc Natl Acad Sci U S A* 2005, **102**:2808-2813.

73. Yan W, Ma L, Burns KH, Matzuk MM: **HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis**. *Proc Natl Acad Sci U S A* 2003, **100**:10546-10551.

74. Wu F, Caron C, De Robertis C, Khochbin S, Rousseaux S: **Testis-specific histone variants H2AL1/2 rapidly disappear from paternal heterochromatin after fertilization**. *J Reprod Dev* 2008, **54**:413-417.

75. Soboleva TA, Nekrasov M, Pahwa A, Williams R, Huttley GA, Tremethick DJ: **A unique H2A histone variant occupies the transcriptional start site of active genes**. *Nat Struct Mol Biol* 2011, **19**:25-30.

76. Hazzouri M, Pivot-Pajot C, Faure AK, Usson Y, Pelletier R, Sele B, Khochbin S, Rousseaux S: **Regulated hyperacetylation of core histones during mouse spermatogenesis: involvement of histone deacetylases**. *Eur J Cell Biol* 2000, **79**:950-960.

77. Faure AK, Pivot-Pajot C, Kerjean A, Hazzouri M, Pelletier R, Peoc'h M, Sele B, Khochbin S, Rousseaux S: **Misregulation of histone acetylation in Sertoli cell-only syndrome and testicular cancer**. *Mol Hum Reprod* 2003, **9**:757-763.

78. Pivot-Pajot C, Caron C, Govin J, Vion A, Rousseaux S, Khochbin S: **Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein**. *Mol Cell Biol* 2003, **23**:5354-5365.

79. Moriniere J, Rousseaux S, Steuerwald U, Soler-Lopez M, Curtet S, Vitte AL, Govin J, Gaucher J, Sadoul K, Hart DJ *et al.*: **Cooperative binding of two acetylation marks on a histone tail by a single bromodomain**. *Nature* 2009, **461**:664-668.

80. • Shang E, Nickerson HD, Wen D, Wang X, Wolgemuth DJ: **The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation**. *Development* 2007, **134**:3507-3515.
Brdt is a testis specific bromodomain-containing protein which could bind to acetylated lysines. Deletion of first bromodomain of Brdt in mice resulted in sterile males although intracytoplasmic sperm injection showed that mutant sperm could support embryonic development.

81. Baarends WM, Hoogerbrugge JW, Roest HP, Ooms M, Vreeburg J, Hoeijmakers JH, Grootegoed JA: **Histone ubiquitination and chromatin remodeling in mouse spermatogenesis**. *Dev Biol* 1999, **207**:322-333.

82. Lu LY, Wu J, Ye L, Gavrilina GB, Saunders TL, Yu X: **RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis**. *Dev Cell* 2010, **18**:371-384.

83. •• Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N *et al.*: **Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification**. *Cell* 2011, **146**:1016-1028.
Using a proteomics approach, the authors identified many novel post-translational modification sites on histones. From those new modifications, histone lysine crotonylation was found to mark active genes and potentially regulate male germ cell differentiation.

84. Liu Z, Zhou S, Liao L, Chen X, Meistrich M, Xu J: **Jmjd1a demethylase-regulated histone modification is essential for cAMP-response element modulator-regulated gene expression and spermatogenesis**. *J Biol Chem* 2009, **285**:2758-2770.

85. Iwamori N, Zhao M, Meistrich ML, Matzuk MM: **The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility**. *Biol Reprod* 2011, **84**:1225-1234.

86. Madan V, Madan B, Brykczynska U, Zilbermann F, Hogeveen K, Dohner K, Dohner H, Weber O, Blum C, Rodewald HR *et al.*: **Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog MLL5**. *Blood* 2009, **113**:1444-1454.

87. Yap DB, Walker DC, Prentice LM, McKinney S, Turashvili G, Mooslehner-Allen K, de Algara TR, Fee J, de Tassigny X, Colledge WH *et al.*: **Mll5 is required for normal spermatogenesis**. *PLoS One* 2011, **6**:e27127.

88. •• Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR: **Distinctive chromatin in human sperm packages genes for embryo development**. *Nature* 2009, **460**:473-478.
This study was the first to show genome-wide localization of histones and histone modifications in human sperm. The paper demonstrates that histones localize to specific sequences in human sperm and that the histone modifications H3K4me3 and H3K27me3 mainly localize to developmental promoters.

89. Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D: **Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences**. *Genome Res* 2009, **19**:1338-1349.

90. Zhang K, Dent SY: **Histone modifying enzymes and cancer: going beyond histones**. *J Cell Biochem* 2005, **96**:1137-1148.

91. Petes TD: **Meiotic recombination hot spots and cold spots**. *Nat Rev Genet* 2001, **2**:360-369.

92. Baudat F, de Massy B: **Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot**. *PLoS Genet* 2007, **3**:e100.

93. Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A: **Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites**. *EMBO J* 2009, **28**:99-111.

94. Buard J, Barthes P, Grey C, de Massy B: **Distinct histone modifications define initiation and repair of meiotic recombination in the mouse**. *EMBO J* 2009, **28**:2616-2624.

95. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B: **PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice**. *Science* 2010, **327**:836-840.

96. Hayashi K, Yoshida K, Matsui Y: **A histone H3 methyltransferase controls epigenetic events required for meiotic prophase**. *Nature* 2005, **438**:374-378.

97. Grey C, Barthes P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B: **Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination**. *PLoS Biol* 2011, **9**:e1001176.

98. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P: **Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination**. *Science* 2010, **327**:876-879.

99. Kauppi L, Barchi M, Baudat F, Romanienko PJ, Keeney S, Jasin M: **Distinct properties of the XY pseudoautosomal region crucial for male meiosis**. *Science* 2011, **331**:916-920.

100. Turner JM: **Meiotic sex chromosome inactivation**. *Development* 2007, **134**:1823-1831.

101. Turner JM, Mahadevaiah SK, Ellis PJ, Mitchell MJ, Burgoyne PS: **Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids**. *Dev Cell* 2006, **10**:521-529.

102. Ichijima Y, Ichijima M, Lou Z, Nussenzweig A, Camerini-Otero RD, Chen J, Andreassen PR, Namekawa SH: **MDC1 directs chromosome-wide silencing of the sex chromosomes in male germ cells**. *Genes Dev* 2011, **25**:959-971.

103. Turner JM, Mahadevaiah SK, Elliott DJ, Garchon HJ, Pehrson JR, Jaenisch R, Burgoyne PS: **Meiotic sex chromosome inactivation in male mice with targeted disruptions of Xist**. *J Cell Sci* 2002, **115**:4097-4105.

104. Greaves IK, Rangasamy D, Devoy M, Marshall Graves JA, Tremethick DJ: **The X and Y chromosomes assemble into H2A.Z-containing [corrected] facultative heterochromatin [corrected] following meiosis**. *Mol Cell Biol* 2006, **26**:5394-5405.

105. van der Heijden GW, Derijck AA, Posfai E, Giele M, Pelczar P, Ramos L, Wansink DG, van der Vlag J, Peters AH, de Boer P: **Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation**. *Nat Genet* 2007, **39**:251-258.

106. Namekawa SH, Park PJ, Zhang LF, Shima JE, McCarrey JR, Griswold MD, Lee JT: **Postmeiotic sex chromatin in the male germline of mice**. *Curr Biol* 2006, **16**:660-667.

107. Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JM: **The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression**. *Nat Genet* 2008, **40**:794-799.

108. Huynh KD, Lee JT: **Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos**. *Nature* 2003, **426**:857-862.

109. Kalantry S, Purushothaman S, Bowen RB, Starmer J, Magnuson T: **Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation**. *Nature* 2009, **460**:647-651.

110. Patrat C, Okamoto I, Diabangouaya P, Vialon V, Le Baccon P, Chow J, Heard E: **Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice**. *Proc Natl Acad Sci U S A* 2009, **106**:5198-5203.

111. Namekawa SH, Payer B, Huynh KD, Jaenisch R, Lee JT: **Two-step imprinted X inactivation: repeat versus genic silencing in the mouse**. *Mol Cell Biol* 2010, **30**:3187-3205.

112. Goto Y, Takagi N: **Tetraploid embryos rescue embryonic lethality caused by an additional maternally inherited X chromosome in the mouse**. *Development* 1998, **125**:3353-3363.

113. Tada T, Obata Y, Tada M, Goto Y, Nakatsuji N, Tan S, Kono T, Takagi N: **Imprint switching for non-random X-chromosome inactivation during mouse oocyte growth**. *Development* 2000, **127**:3101-3105.

Chapter 2: Results

2.3 Submitted manuscript:

Density and methylation state of CpG dinucleotides define histone variant specific retention of nucleosomes in mouse spermatozoa

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

1. Friedrich Miescher Institute for Biomedical Research (FMI), Maulbeerstrasse 66, 4058 Basel, Switzerland

2. Faculty of Sciences, University of Basel, 4056 Basel, Switzerland

3. Nephrology Research Laboratory, Department of Nephrology, Radboud University Nijmegen Medical Centre, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands

4. Swiss Institute of Bioinformatics, 4058 Basel, Switzerland.

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

Keywords:
Nucleosome, spermatozoa, histone variants, epigenetic inheritance

Nucleosomes are the principal packaging units of chromatin and critical for gene regulation and genome stability[1,2]. In mammals, a subset of nucleosomes fail to be replaced by protamines during spermatogenesis and are retained in mature spermatozoa providing opportunities for paternal epigenetic transmission[3]. In humans, the remaining 10% localize at regulatory elements of genes[4,5]. To assess evolutionary 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. In striking contrast to mammalian somatic cells[6,7,8,9] and haploid round spermatids, we observe high enrichment of nucleosomes at CpG-rich sequences throughout the genome, at conserved regulatory sequences as well as at intra- and intergenic regions and repetitive DNA. This preferred occupancy occurs mutually exclusive with DNA methylation both in mouse and human sperm. At unmethylated CpG-rich sequences, residing nucleosomes are largely composed of the H3.3 histone variant, and trimethylated at lysine 4 (H3K4me3). Both canonical H3.1/H3.2 and H3.3 variant histones are present at promoters marked by Polycomb-mediated H3K27me3, which is strongly predictive for gene repression in pre-implantation embryos. Our data indicate important roles of DNA sequence composition, DNA methylation, variant H3.3 and canonical H3.1/H3.2 histones and associated modifications in nucleosome retention versus eviction during the histone-to-protamine remodeling process in elongating spermatids and potentially in 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 concurs with extensive epigenetic reprogramming, affecting DNA methylation, histone modifications, replication timing and transcriptional activity in parental specific manners[10,11,12,13]. 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[14,15,16,17,18].

To assess the potential of paternal epigenetic inheritance by nucleosomes in mouse we first aimed at determining the position of nucleosomes in mature spermatozoa that are competent for fertilization and transmission of genetic and epigenetic information. We isolated motile spermatozoa from caudal epididymi and performed deep-sequencing of DNA associated with mono-nucleosomes that we had prepared by micrococcal nuclease (MNase) digestion of sperm chromatin. We

observed strong nucleosomal enrichment at promoter regions of many but not all genes (Fig. 1a; Supplementary Fig. 1). Classification of promoters according to their GC content, CpG ratio and length of CpG-rich region[19] showed that high-CpG (HCP) and intermediate-CpG (ICP) promoters are highly and moderately enriched in nucleosomes respectively (representing strong and weak CpG islands (CGIs)) while most promoters with low CpG content (LCP) lack nucleosomes (Fig. 1b). Nucleosomal enrichment is, however, 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 firstly determined single nucleotide frequencies in 1kb windows tiled throughout the genome. While guanine and cytosine strongly correlate with nucleosome occupancy genome-wide, adenine and thymine do not (Fig. 1d). We next assessed the contribution of different dinucleotides on nucleosome occupancy, independent of single nucleotide frequencies, by calculating the ratio of "observed over expected" frequencies for each dinucleotide. Remarkably, these genome-wide analysis revealed that CpG is the most important dinucleotide contributing to sequence related nucleosomal packaging of sperm DNA (Fig. 1d). In contrast, the GpC dinucleotide has almost no contribution while ApA and TpT dinucleotides contribute moderately. To determine whether CpG density constitutes an intrinsic DNA sequence preference for nucleosome formation, we reanalyzed data of an *in vitro* nucleosome reconstitution experiment with histone octamers assembled onto yeast genomic DNA[20]. Similar to[21], we observed a strong contribution of G+C to *in vitro* nucleosome formation, yet no specific contributions of either CpG nor GpC dinucleotides (Supplementary Fig. 2). Thus, the strong association of CpG density to nucleosome retention in mouse sperm does not reflect an intrinsic DNA preference of nucleosomes but represents a novel feature of CGIs in genome function and germ cell biology[22]. Consistently, motif analysis did not reveal any specific sequence compositions other than a strong correlation to GC composition (Supplementary Fig. 3).

The nucleosomal occupancy at CGIs in sperm strongly contrasts with the depletion of nucleosomes at CGI-promoters in somatic cells[6,7,9]. Indeed, we observed extensive nucleosomal depletion around TSS and a clear anti-correlation between nucleosome occupancy and CpG frequency in mouse liver[8] (Supplementary Fig. 4). However, in somatic cells, nucleosomes are not depleted at CGI-promoters repressed by Polycomb Group (PcG) proteins or by DNA methylation [23]. 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[24]. In contrast to expectation, our data shows that methylated genomic regions are devoid of nucleosomes in sperm (Fig. 1e). We observed a similar inverse relationship using genome-wide shotgun bisulfite sequencing data of mouse sperm (Supplementary Fig. 5a, 5b)[16]. The exclusive inverse relationship is nicely illustrated at imprinting control regions (ICRs) of the paternally imprinted genes *H19*, *Dlk1*/*Gtl2* and *Rasgrf1*, that are methylated and devoid of nucleosomes in mouse sperm. In contrast, ICRs that control maternally imprinted gene clusters and that are methylated in oocytes such as *Kcnq1ot1*, *Gnas*/*Nespas*, *Snrpn,* and *Peg10* contain nucleosomes in sperm (Supplementary Fig 6). Furthermore, LINE1 retroelements that are methylated in sperm and become demethylated after fertilization[10] lack nucleosomes in sperm (data not shown). Thus, DNA methylation established early during male germ cell development[25] prevents nucleosome retention during spermiogenesis, thereby precluding paternal transmission of chromatin states that were associated with DNA methylated sequences in immature male germ cells. This finding is consistent with differential reprogramming of DNA methylation in zygotes, generated by micro-insemination of round spermatids or mature spermatozoa[17]. Finally, we found a strong positive correlation between nucleosomal enrichment and density of CpG dinucleotides within CGIs that are devoid of DNA methylation (Supplementary Fig. 5c). Taken together, we can predict with high accuracy nucleosome occupancy in mouse sperm as a function of CpG dinucleotide frequency and DNA methylation level using a linear mathematical model (Fig. 1f).

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[4,5,26]. Our current sequence analysis demonstrates that nucleosome retention at CGIs is conserved in mouse and human spermatozoa (Supplementary Figure 7).

The unique nucleosomal organization in mature sperm, highly distinct from the chromatin state in somatic cells, emphasizes the unique nature of the chromatin remodeling processes taking place during the formation of mature sperm. Given the important roles of histone variants in chromatin dynamics during transcription, cellular differentiation, reproduction and development[27,28,29] we assessed whether canonical H3.1/H3.2 and variant H3.3 histones may serve specific functions in nucleosome eviction versus retention during spermiogenesis. We used antibodies highly specific for either H3.3 (Supplementary Fig.8) or H3.1/H3.2[30] to perform Western blot analysis and ChIP-sequencing experiments. Compared to proliferating embryonic stem cells

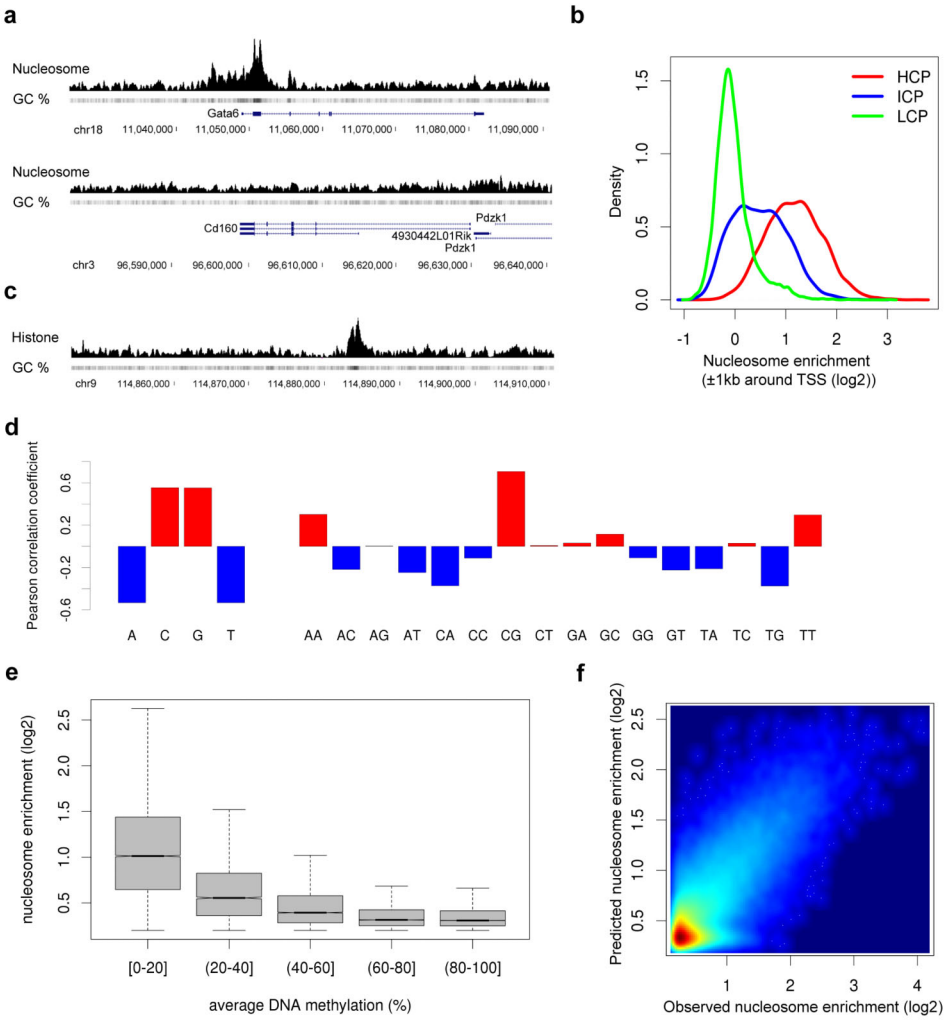


Figure 1: Nucleosome occupancy in sperm is highly dependent on CpG composition.
a, Nucleosome occupancy and GC percentage at representatitve CpG and non-CpG islands loci in mouse sperm. **b**, Density plot showing the distribution of nucleosome enrichment around TSS (±1kb) of genes classified according 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**, Correlation of single nucleotide frequencies (left) and single nucleotide composition normalized dinucleotide frequencies (right) with nucleosome enrichment in sperm in 1kb windows tiling the mouse genome. **e**, Distributions of nucleosome enrichments in regions of different DNA methylation status (1kb windows, genome-wide). **f**, 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).

(ESCs) and even to quiescent aging neurons[31], H3.3 is incorporated into chromatin of round spermatids and sperm to extremely high levels relative to H3.1/H3.2,

suggesting an extensive and rapid replacement of most canonical histones by the H3.3 variant, presumably upon entry into meiosis and/or during spermatid differentiation (Fig. 2a). In sperm, ChIP-sequencing profiles of H3.3 are highly similar to nucleosomal profiles whereas we fail to measure globally a correlation between H3.1/H3.2 and nucleosomal enrichments (Fig. 2b). Consistently, H3.3 enrichments are well predicted by our linear model that is largely based on CpG densities genome-wide. In contrast, H3.1/H3.2 enrichments are highly underestimated suggesting a CpG density-linked retention mechanism for H3.3 but not for H3.1/H3.2 containing nucleosomes (Fig. 2c).

To better understand the mechanisms of histone variant specific eviction and retention, 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 isolated round spermatids. We observed a widespread reduction in occupancy of H3.1/H3.2-nucleosomes as well as increased positioning of remaining nucleosomes around transcriptional start sites (TSS) of expressed genes that correlate well with mRNA levels of associated genes suggesting transcription coupled eviction of canonical histones (Fig. 2d; Supplementary Fig. 9a). Transcription-coupled depletion of H3.1/H3.2 around TSS is more pronounced in spermatids than that of H3.2-HA tagged nucleosomes in embryonic stem cells (Supplementary Fig. 9b)[32]. This is likely due to the progressive loss of H3.1/H3.2 replication-coupled histones during transcription in post-replicative germ cells like round spermatids that further extends the general replacement of canonical nucleosomes by H3.3 as indicated by western blot analysis (Fig. 2a). For H3.3 nucleosomes, we also measured some depletion around TSS that was more pronounced downstream of TSS at medium and highly expressed genes. Comparison of H3.3 over H3.1/H3.2 occupancy levels, however, suggests extensive transcription-coupled eviction of canonical histones and dynamic replacement by H3.3 nucleosomes in round spermatids, like in ESCs (Supplementary Fig. 9).

When comparing round spermatids to spermatozoa, we see 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 or *de novo* incorporation of H3.3 nucleosomes at CGIs during spermatid differentiation (Fig. 2d, 2e). On the other hand, 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 in spermatocytes and spermatids (Fig. 2e).

We next addressed whether the gain in nucleosomal occupancy at CGIs in sperm relates to transcription-coupled histone turnover seen in round spermatids (Supplementary Fig. 9) For nucleosomes in general as well as for H3.3-containing nucleosomes we observed that enrichments at promoter regions of genes (±1kb TSS) in sperm do not correlate with the expression level of genes in round spermatids but with the CpG percentage of the promoter regions (Fig. 3a). For H3.1/H3.2-nucleosomes in sperm, we observed enrichments at two classes of genes (Fig. 3a). First, CpG-poor promoters (<3% CpG) are relatively enriched for H3.1/H3.2, irrespective of their transcriptional status. Second, lowly and non-expressed CpG-rich promoters are slightly enriched. These later promoters contain also H3.3 nucleosomes. To explain this conundrum, we more closely examined the dynamics of H3.1/H3.2 in round spermatids. These analysis revealed a low level of eviction of canonical histones at non-expressed CpG-rich but not at non-expressed CpG-poor TSS regions (Supplementary Fig. 10), suggesting that the presence of H3.1/H3.2 and H3.3 histones at lowly and non-expressed CpG-rich promoters in sperm may be linked to a low level of default histone replacement at such CGIs in spermatids.

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. 11) indicating maintenance of the modification state during spermiogenesis. Interestingly, CGI promoters containing H3.3 nucleosomes are generally marked by H3K4me3 in sperm (Supplementary Fig. 12a). A fraction of these CGIs are also positive for H3K27me3, indicating the presence of bivalent promoters in sperm. Importantly, such bivalent CGI promoters also show some enrichment for H3.1/H3.2 histones in sperm (Supplementary Fig. 12b) suggesting that Polycomb-mediated H3K27me3 may suppress the default eviction of such histones at CGIs in round spermatids and consequently promote retention of pre-existing canonical histones (and similarly of newly incorporated H3.3 histones) during chromatin remodeling in elongating spermatids (Supplementary Fig. 12).

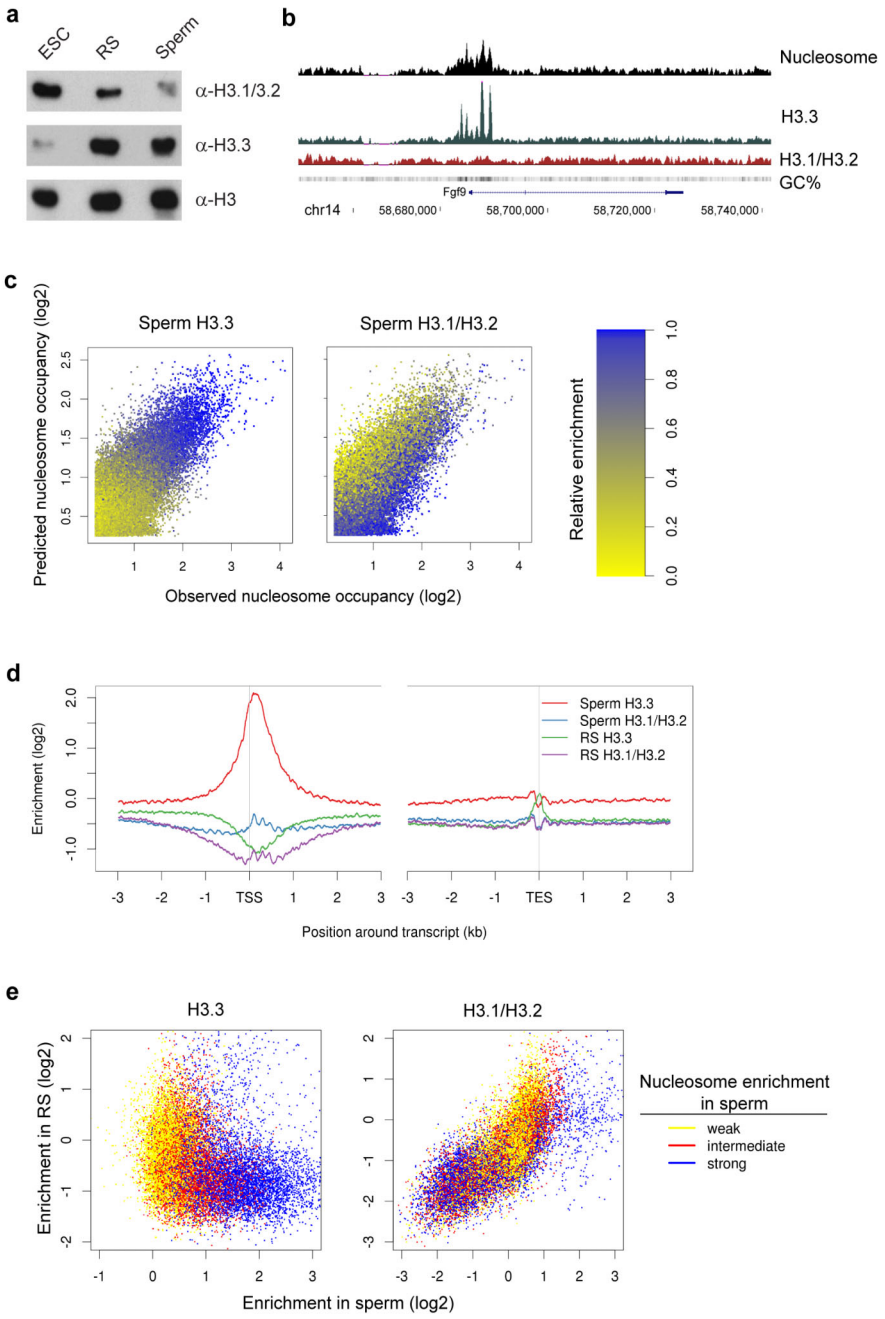


Figure 2: Histone variant specific packaging of sperm DNA.
a, Western blots showing relative levels of chromatin bound H3.1/H3.2, H3.3, and total H3 in embryonic stem cells (ESC), round spermatids (RS) and sperm. **b**, Occupancy of nucleosomes, H3.3 and H3.1/H3.2 histones and GC percentage at the *Fgf9* 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. **d**, Average profiles of H3.3 and H3.1/H3.2 enrichments around TSS (±3kb) and transcriptional end sites (TES; ±3kb) in sperm and RS. **e**, Scatter plots showing the correlation between histone H3.3 (left) and H3.1/H3.2 (right) enrichments in sperm versus 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.

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. 3b; Supplementary Fig. 13) 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 lineage specific differentiation (cluster 5) (Supplementary Table 1). To determine the extent by which different variables, measured in round spermatids, contribute to nucleosome occupancy in sperm, we performed a variance partitioning analysis for promoter regions (Fig. 3c). As expected, CpG percentage of promoters has the highest unique contribution to H3.3 occupancy in sperm (clusters 1-4 in Fig. 3b; Fig. 3c). In contrast, H3.1/H3.2 enrichments in sperm mostly relate to H3.1/H3.2 enrichments in round spermatids (cluster 5 in Fig. 3b; Fig. 3c). In addition, CGI promoters marked by H3K27me3 in round spermatids, constituting canonical Polycomb regulated genes, preferentially retain H3.1/H3.2 in sperm (cluster 4 in Fig. 3b; Fig. 3c). Beside unique contributions, we find that a significant portion of variation cannot be uniquely attributed to a single variable. We obtained similar results when performing the variance partitioning analysis in genome-wide 1kb windows (Supplementary Fig. 14), arguing that the principles we determined for retaining nucleosomes at promoters in sperm are generally applicable to the entire genome.

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. 3b in oocytes and in pre-implantation embryos[33,5]. Of germ line and housekeeping genes (clusters 1-3) that are moderately to highly expressed in spermatids and marked by H3K4me3 in spermatozoa, merely ~18% become *de novo* transcribed in pre-implantation embryos indicating a rather limited potential of H3K4 tri-methylated nucleosomes in sperm to predetermine transcription in early embryos (Supplementary Fig. 15). This may relate to the prevalent H3K4 methylation of CGIs observed in spermatids, ESCs and during somatic differentiation independent of their transcriptional status[22] (Supplementary Fig. 11a). In contrast, only ~10% of CGI promoters marked by H3K27me3 in sperm (cluster 4) are expressed in pre-implantation embryos (Supplementary Fig. 15). Consistently, the majority of Polycomb target genes in sperm are similarly modified by H3K27me3 in ESCs (Supplementary Fig. 11b), supporting a model of H3K27me3 mediating epigenetic inheritance of transcriptional repression between generations. Indeed, loss of Polycomb function in mature oocytes causes misexpression of canonical Polycomb target genes, marked by H3K27me3 in mouse sperm[34] and data not shown.

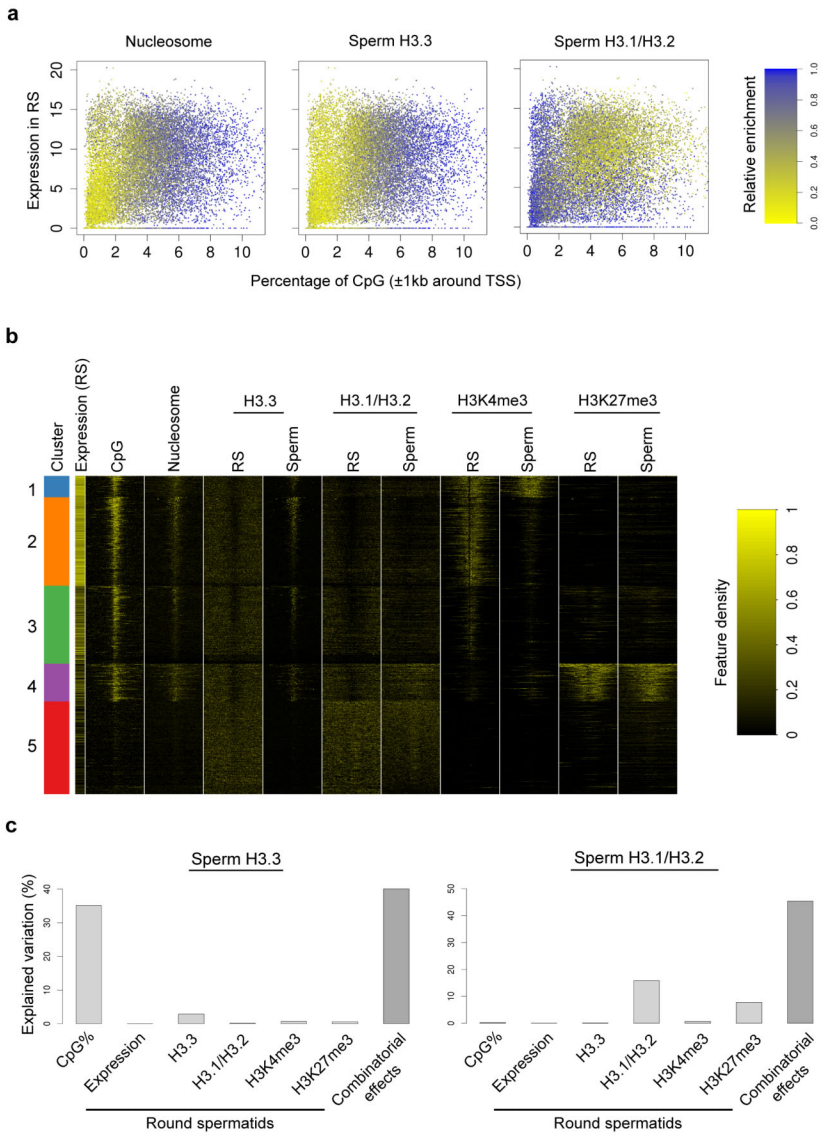


Figure 3: Combinatorial effects of CpG density, histone variants and histone modifications uniquely package sperm DNA.

a, Scatter plots showing the correlation between percentage of CpGs at TSS (±1kb) and expression of linked genes in RS in relation to relative enrichment of nucleosomes (left), H3.3 (middle), and H3.1/H3.2 (right) in sperm. **b**, Heatmap of genes illustrating expression status in RS, CpG density, nucleosome coverage in sperm, histone variant and modification coverage around TSS (±3kb) in 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. **c**, Variance partitioning analysis (see Methods for details) assessing the unique contribution of different variables to the relative enrichments of H3.3 (left) and H3.1/H3.2 (right) around TSS (±1kb) in sperm. Combinatorial effects refer to variation which is common to different combination of variables included.

Taken together, we demonstrate a remarkably strong occupancy of H3.3 containing nucleosomes at unmethylated CGIs in sperm that linearly correlates with CpG density. These data are in striking contrast to the default depletion of nucleosomes at CGIs in somatic cells[8,9] and in round spermatids that occurs independently of but is enhanced by transcription (Supplementary Fig. 10). Our findings support a model in which H3.3 nucleosomes present at CGIs in sperm are incorporated into chromatin and posttranslationally modified in late round spermatids in response to global ceasing of transcription and histone turnover in late round spermatids. This model implies that CpG-rich DNA would resist loading of transition proteins and protamines in elongating spermatids, thereby enabling nucleosome retention at CGIs in sperm (Fig. 4a). Alternatively, H3.3 nucleosomes may be *de novo* incorporated at CGIs during the histone-to-protamine exchange process in elongating spermatids after removal of an unknown intermediary place-holding factor (Fig. 4b). In round spermatids, Polycomb based repression may directly or indirectly reduce eviction of H3.1/H3.2 nucleosomes normally occurring at CGIs thereby promoting retention of such canonical, H3K27me3-labeled, nucleosomes in spermatozoa (Fig. 4c). Based on these data, genetic gain and loss-of function experiments can be designed to unambiguously determine the role of canonical and variant histones and their modifications in nucleosomal eviction during spermiogenesis and in epigenetic inheritance between generations.

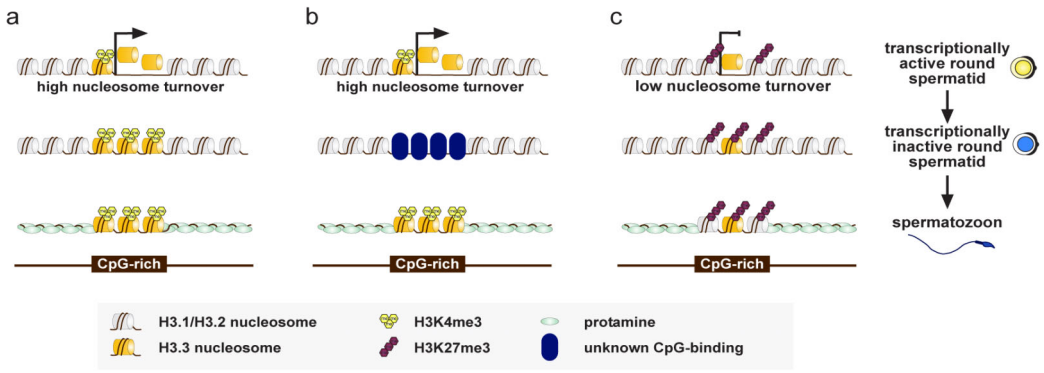


Figure 4: Models of nucleosome retention and eviction during spermiogenesis. **a**, Incorporation of H3.3 nucleosomes, marked by H3K4me3, at unmethylated CGIs in late round spermatids in response to global ceasing of transcription and histone turnover in late round spermatids. **b**, *De novo* incorporation of H3.3 nucleosomes at CGIs in elongating spermatids after removal of an unknown intermediary place-holding factor. **c**, Polycomb repression may reduce nucleosome turnover in round spermatids and promote retention of canonical H3.1/H3.2 nucleosomes, labeled with K27me3, in spermatozoa.

Methods Summary

Sample collection and sequencing of libraries for ChIP-seq and RNA-seq

Mouse spermatozoa were collected by swim-up procedure[5] from C57BL/6J mice. Round spermatids were isolated by FACS sorting using 28 day old BL6 mice. Mono-nucleosomal chromatin samples were prepared by MNase treatment. ChIP experiments for histone variants, H3.1/H3.2, H3.3, and histone modifications H3K4me3 and H3K27me3 were performed on native chromatin. ChIP-seq libraries were produced using the Illumina ChIP-seq DNA Sample Prep Kit (Cat# IP-102-1001). RNA isolated from round spermatids was subjected to Ribo-Zero rRNA removal kit (Epicentre) and subsequently strand specific libraries were prepared. ChIP-seq and RNA-seq libraries were sequenced either on Illumina GA II or Illumina Hiseq 2000.

Processing of the reads and analysis of the data

Reads were aligned to mm9 mouse genome assembly by using Bowtie allowing up to 2 mismatches. Genomic coordinates (all based on mouse mm9 assembly) were obtained from UCSC. Nucleosome and histone variant enrichments in 1kb windows or at promoter regions of the genes were calculated by taking the ratio of sample reads over genomic input reads in the respective windows. Nucleosome enrichment in 1kb windows was predicted by CpG dinucleotide frequencies and average DNA methylation by linear regression using lm() function of R (www.r-project.org). Nucleosome peak finding was performed by training a two state hidden semi-Markov model on nucleosome enrichments in genome-tiling 1kb windows and using the maximum likelihood path through the model, as implemented in the R-mhsmm package. Variance partitioning analysis to explain histone variant enrichment in sperm was carried out using the R package yhat. Expression in round spermatids was quantified as the sum of reads mapping to the respective transcripts, selecting one transcript with the most extreme coordinates as a representative for each gene.

References

1 Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**, 175-186, doi:10.1016/j.gde.2011.01.022 (2011).

2 Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* **13**, 436-447, doi:10.1038/nrm3382 (2012).

3 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**, 962-964 (1987).

4 Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478, doi:nature08162 [pii] 10.1038/nature08162 (2009).
5 Brykczynska, U. *et al.* Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* **17**, 679-687, doi:nsmb.1821 [pii] 10.1038/nsmb.1821 (2010).
6 Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898, doi:10.1016/j.cell.2008.02.022 (2008).
7 Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-128, doi:10.1016/j.cell.2009.04.020 (2009).
8 Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K. H. The nucleosome map of the mammalian liver. *Nature structural & molecular biology* **18**, 742-746, doi:10.1038/nsmb.2060 (2011).
9 Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome research*, doi:10.1101/gr.138776.112 (2012).
10 Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344, doi:10.1038/nature10960 (2012).
11 Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610, doi:nature10443 [pii] 10.1038/nature10443 (2011).
12 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, doi:10.1006/dbio.1996.8466 (1997).
13 Albert, M. & Peters, A. H. Genetic and epigenetic control of early mouse development. *Curr Opin Genet Dev* **19**, 113-121, doi:10.1016/j.gde.2009.03.004 (2009).
14 Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40**, 411-420, doi:ng.99 [pii] 10.1038/ng.99 (2008).
15 Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415-419, doi:10.1038/nature11093 (2012).
16 Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**, e1002440, doi:10.1371/journal.pgen.1002440 PGENETICS-D-11-01574 [pii] (2012).
17 Kishigami, S. *et al.* Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. *Developmental biology* **289**, 195-205, doi:10.1016/j.ydbio.2005.10.026 (2006).
18 Bui, H. T. *et al.* Essential role of paternal chromatin in the regulation of transcriptional activity during mouse preimplantation development. *Reproduction* **141**, 67-77, doi:10.1530/REP-10-0109 (2011).
19 Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).
20 Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-366, doi:10.1038/nature07667 (2009).
21 Tillo, D. & Hughes, T. R. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* **10**, 442, doi:10.1186/1471-2105-10-442 (2009).

22 Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-1022, doi:10.1101/gad.2037511 (2011).
23 Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research*, doi:10.1101/gr.143008.112 (2012).
24 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
25 Smallwood, S. A. & Kelsey, G. De novo DNA methylation: a germ cell perspective. *Trends Genet* **28**, 33-42, doi:10.1016/j.tig.2011.09.004 (2012).
26 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, doi:10.1371/journal.pgen.1002036 (2011).
27 Elsaesser, S. J., Goldberg, A. D. & Allis, C. D. New functions for an old variant: no substitute for histone H3.3. *Curr Opin Genet Dev* **20**, 110-117, doi:10.1016/j.gde.2010.01.003 (2010).
28 Orsi, G. A., Couble, P. & Loppin, B. Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *Int J Dev Biol* **53**, 231-243, doi:10.1387/ijdb.082653go (2009).
29 Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).
30 van der Heijden, G. W. *et al.* Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. *Nat Genet* **39**, 251-258, doi:ng1949 [pii] 10.1038/ng1949 (2007).
31 Pina, B. & Suau, P. Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. *Developmental biology* **123**, 51-58 (1987).
32 Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691, doi:10.1016/j.cell.2010.01.003 (2010).
33 Zeng, F. & Schultz, R. M. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Developmental biology* **283**, 40-57, doi:10.1016/j.ydbio.2005.03.038 (2005).
34 Posfai, E. *et al.* Polycomb function during oogenesis is required for mouse embryonic development. *Genes & development* **26**, 920-932, doi:10.1101/gad.188094.112 (2012).

Acknowledgements

We gratefully thank Sophie Dessus-Babus and Tim Roloff (FMI functional genomics group), Ina Nissen (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 a recipient of a Boehringer Ingelheim Fond fellowship. M.G. and R.M. are supported by an EMBO Long Term fellowship. Research in the Peters and Schübeler lab is supported by the Novartis Research Foundation and the Swiss initiative in Systems Biology, SystemsX.ch (Cell plasticity). The Peters lab further acknowledges support from the Swiss National Science Foundation (31003A_125386 and NRP 63 - Stem Cells and Regenerative Medicine), the Japanese Swiss Science and

Technology Cooperation Program, the FP7 Marie Curie Initial Training Network “Nucleosome4D” and the EMBO YIP 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.

Supplementary Information

Density and methylation state of CpG dinucleotides define histone variant specific retention of nucleosomes in mouse spermatozoa

Supplementary Figures and Legends

Supplementary Table 1: GO-term analysis for the gene clusters described in Fig. 3b.

Supplementary Methods

Biological Sample Collection

Mononucleosomal DNA preparation and native ChIP

RNA isolation

Library preparation and sequencing

Mononucleosome-BisSeq (MN-BisSeq) library preparation

Chromatin-bound (histone) fractionation and immunoblotting

Processing and alignment of the reads

Genomic coordinates

UCSC tracks

Peak identification

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

Classification of nucleosome peaks

Classification of genes according to their promoter GC content

Calculation of observed/expected ratios for dinucleotide frequencies

Analysis of bisulfite converted sequencing (BisSeq) data

CGI definition and usage

Genome-wide modeling of nucleosome occupancy

Plotting profiles around genomic regions

Quantifying expression in round spermatids

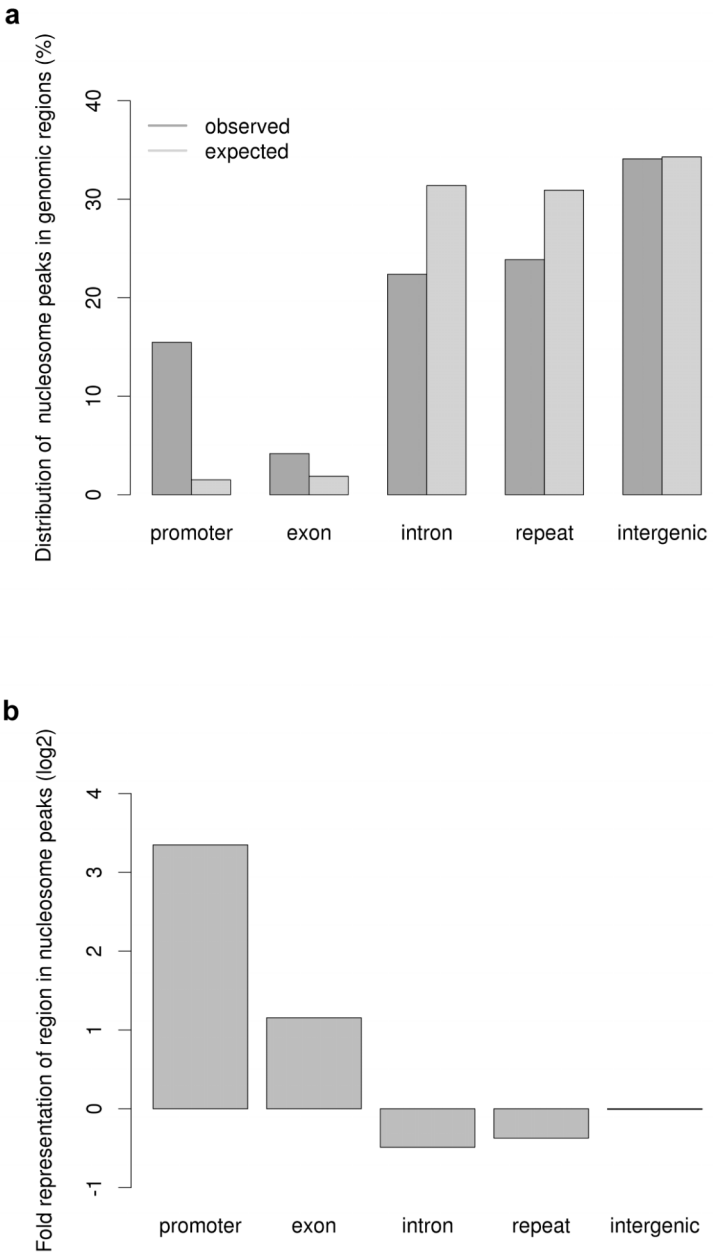
Heatmap plots

GO-term analysis

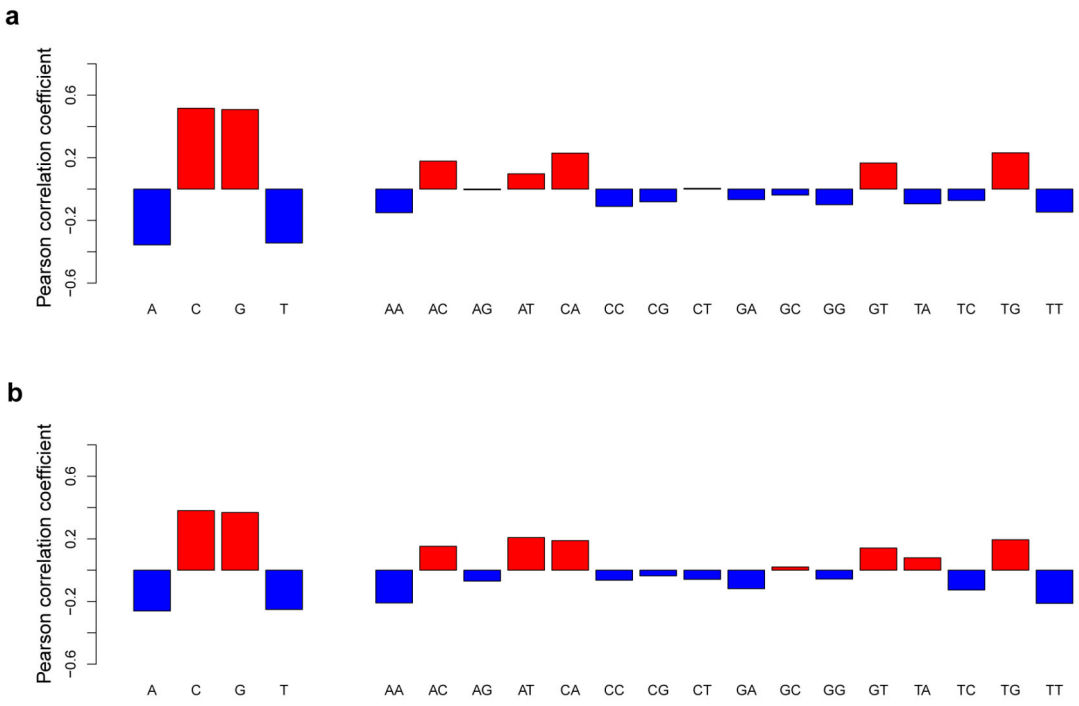
Variance partitioning analysis

Motif finding for histone peaks

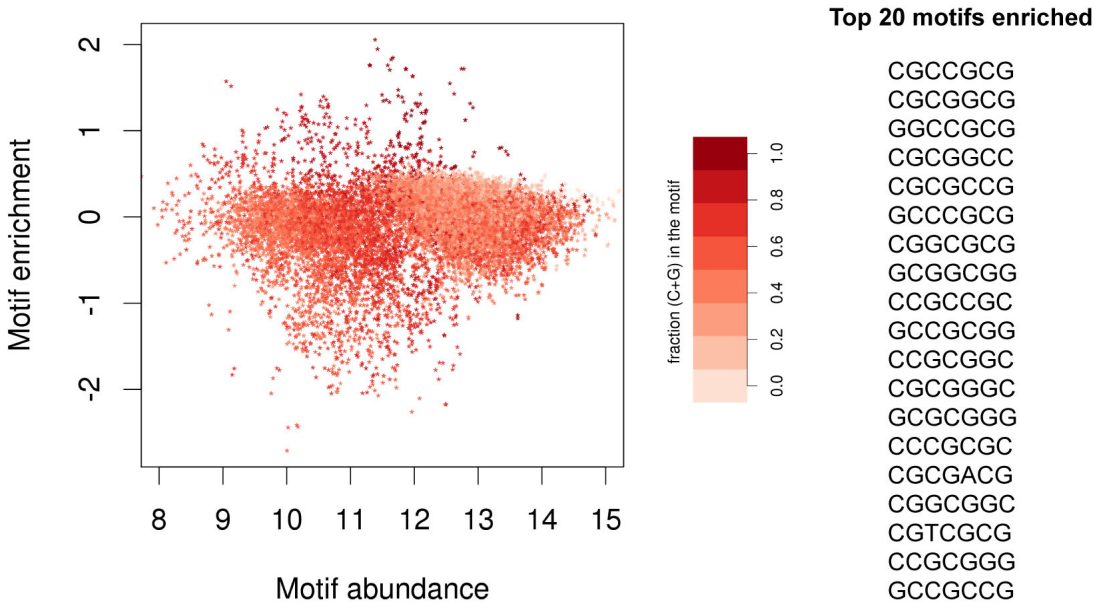
Defining expression for oogenesis or early embryogenesis



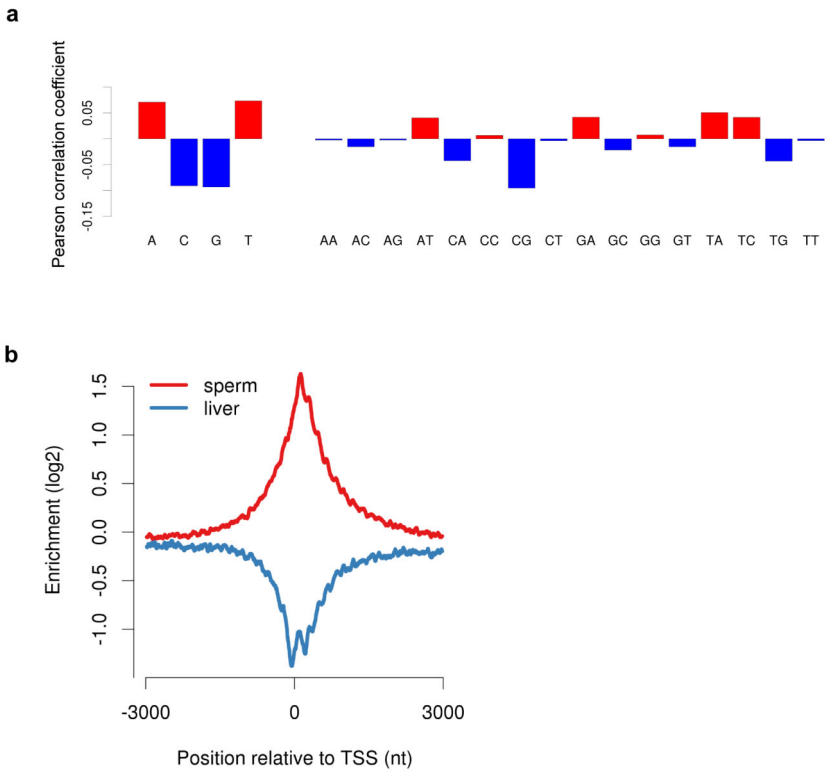
Supplementary Figure 1: Genome-wide distribution of nucleosome peaks.
The mouse genome is classified into promoter (± 1 kb around TSS), exon, intron, repeat and intergenic (non-repeat) regions. **a**, Observed and expected fraction of nucleosome peaks by genomic region. “Observed” refers to experimentally identified nucleosome peaks and “expected” is calculated assuming uniform distribution of the same number of peaks in the genome. **b**, Barplot showing the data from a as the log2 enrichments (observed/expected) of nucleosome peaks in different genomic regions.



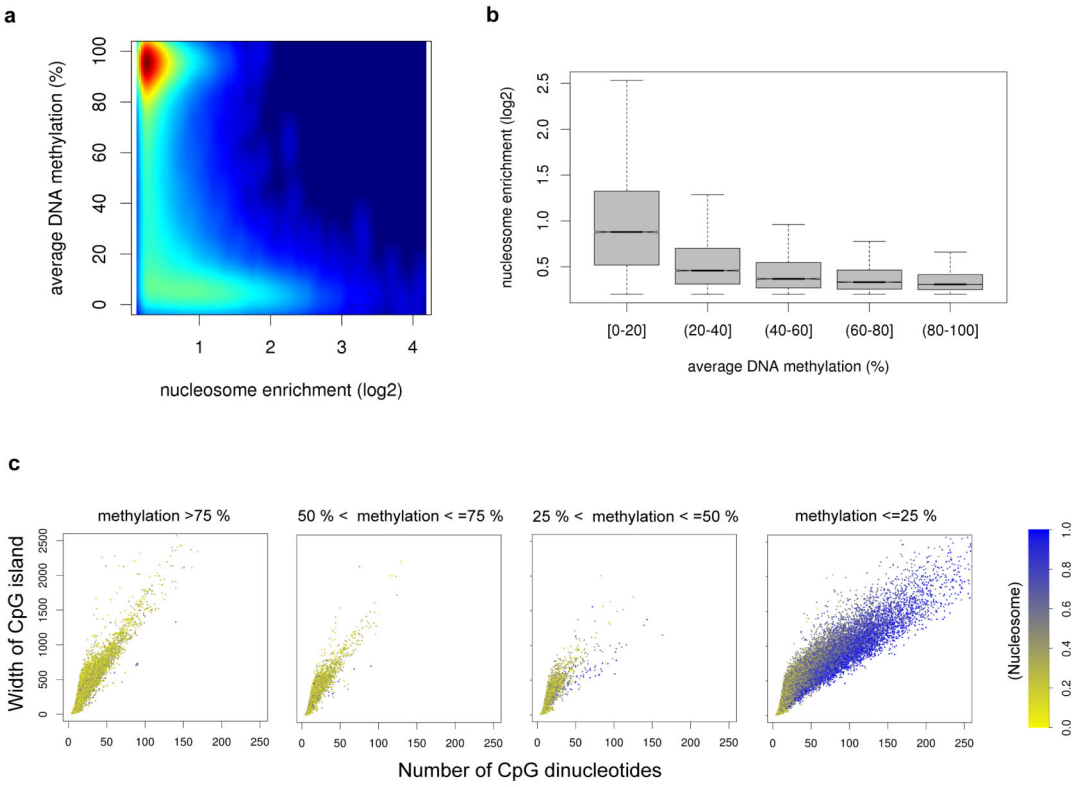
Supplementary Figure 2: Nucleosome occupancy in *Saccharomyces cerevisiae.*
Correlation of single nucleotide frequencies (left) and single nucleotide composition normalized dinucleotide frequencies (right) with nucleosome occupancy in *Saccharomyces cerevisiae*[20] in 1kb windows tiling the yeast genome. **a**, *in vitro* reconstituted nucleosomes. **b**, *in vivo* (YPD medium).



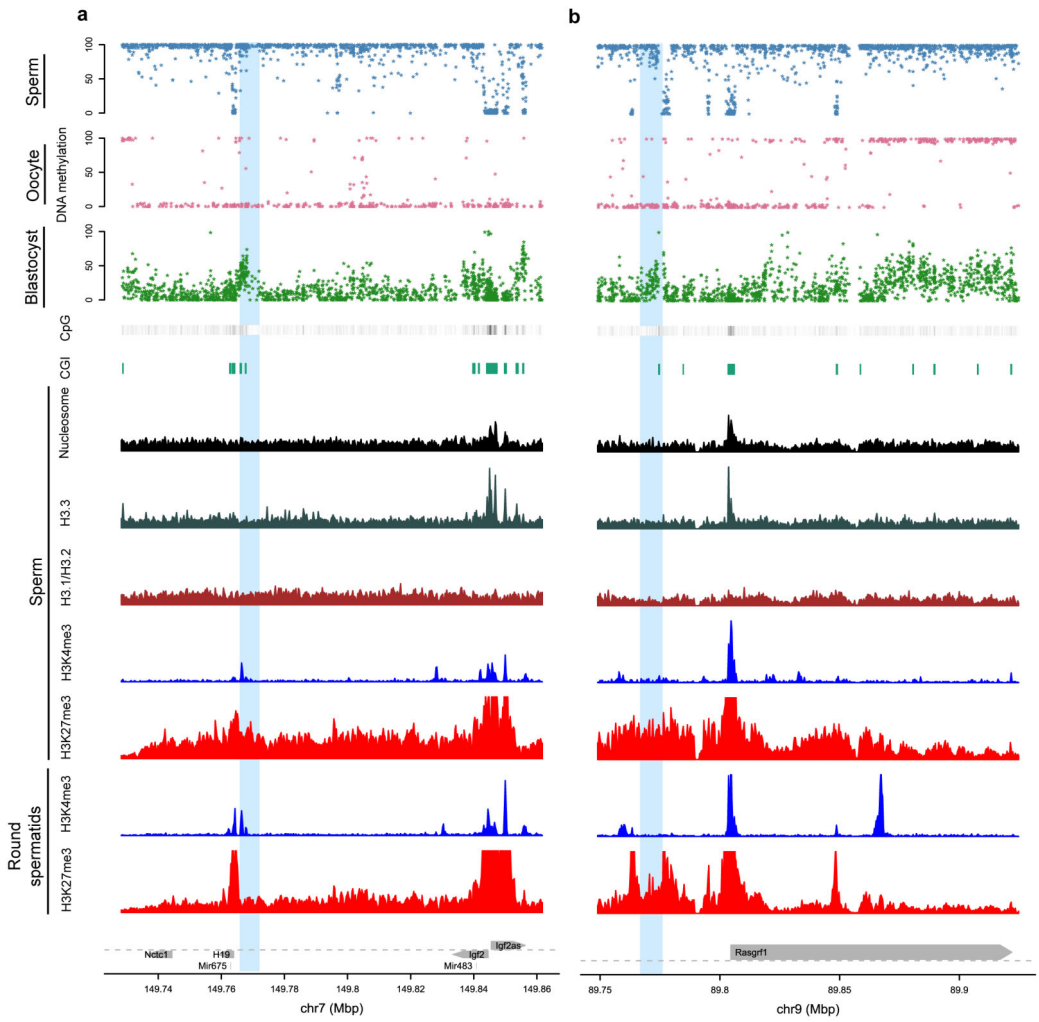
Supplementary Figure 3: Motif enrichment analysis for nucleosome peaks in sperm.
a, Motif abundance/enrichment plot showing the enrichment of 7-mer motifs in nucleosome peaks compared to background (see Methods for the description of background). **b**, Top 20 enriched motifs.



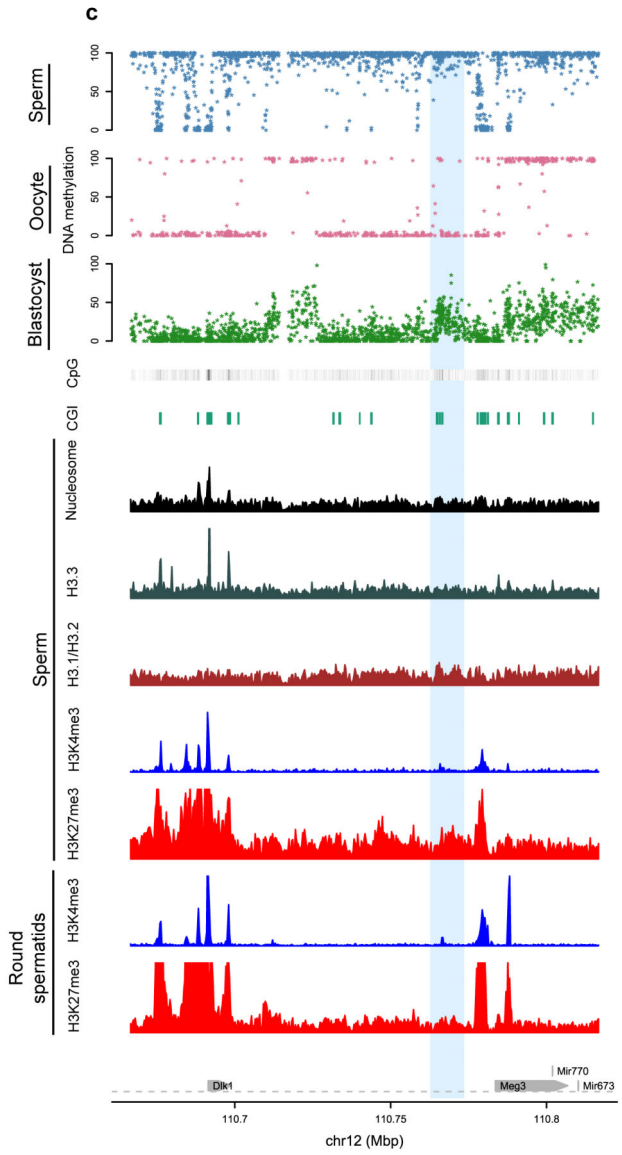
Supplementary Figure 4: Nucleosome occupancy in mouse liver.
a, Correlation of single nucleotide frequencies (left) and single nucleotide composition normalized dinucleotide frequencies (right) with nucleosome occupancy in mouse liver[8] in 1kb windows tiling the mouse genome. **b**, Average profiles for nucleosome occupancy in mouse sperm and liver around ± 3kb TSS of genes.



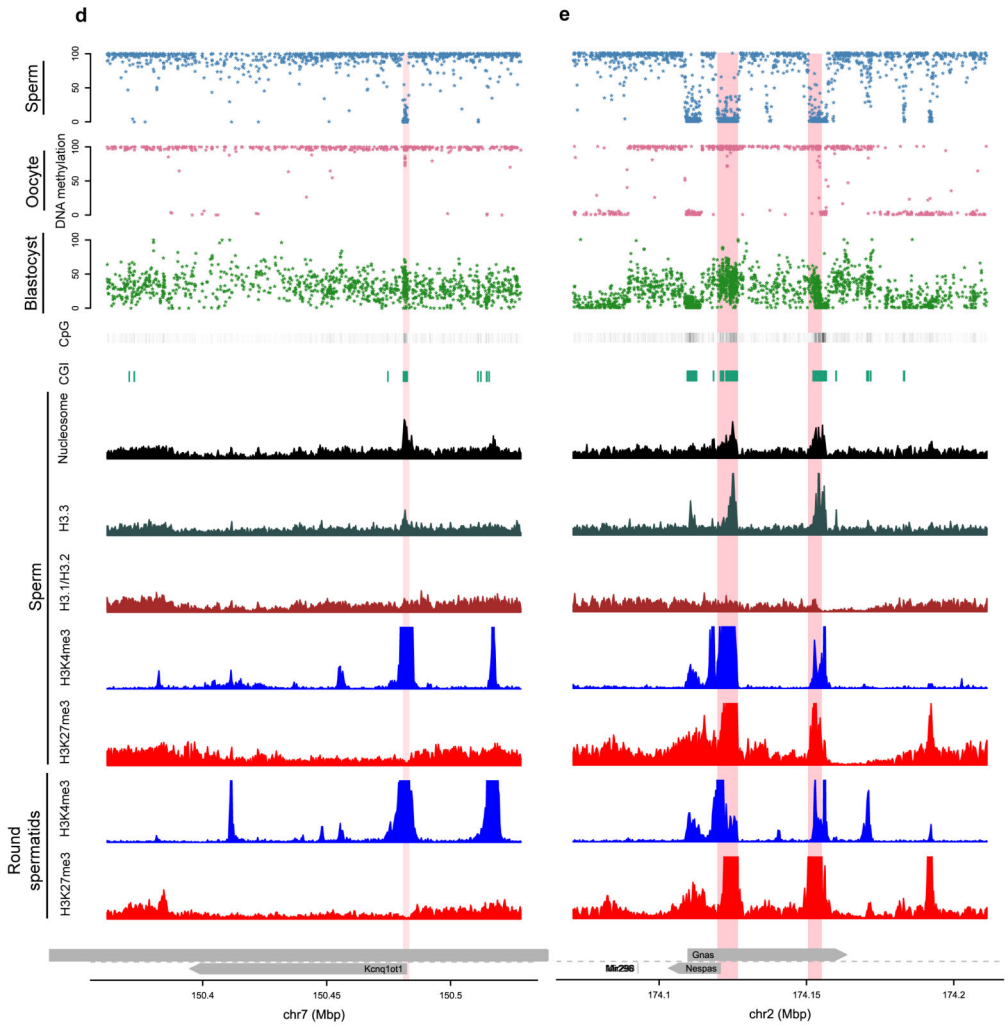
Supplementary Figure 5: Nucleosome occupancy correlates negatively with DNA methylation in sperm.
a, Scatter plot showing the correlation of nucleosome occupancy with average DNA methylation according to Kobayashi and coworkers[16] in 1kb windows genome-wide. **b**, Distributions of nucleosome enrichments in regions of different DNA methylation status (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[35] were grouped into 4 classes according to their DNA methylation status in sperm[16].



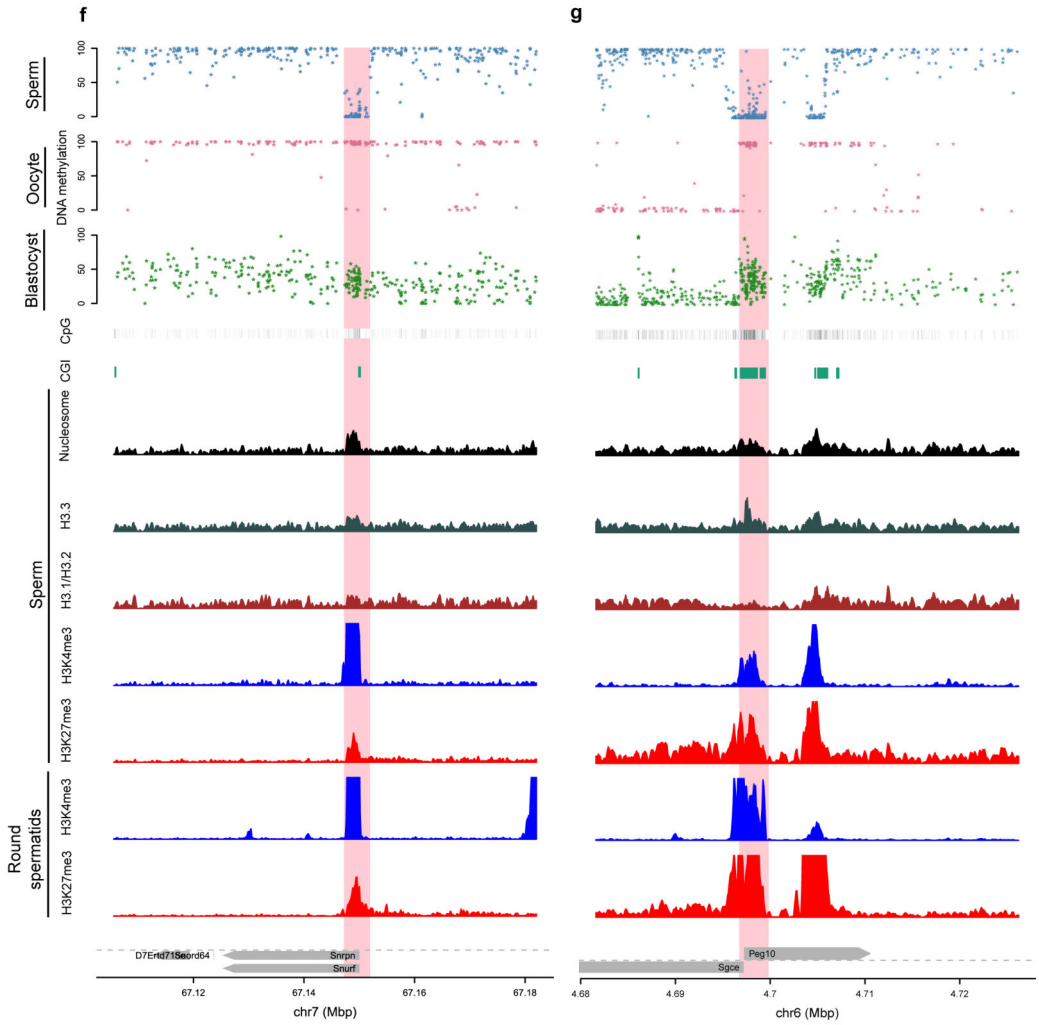
Supplementary Figure 6: Chromatin states of imprinting control regions (ICRs) in sperm and round spermatids.
From top to bottom, images show the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modification states in round spermatids. Imprinting control regions[16] for paternally imprinted genes are shown with light blue boxes and imprinting control regions for maternally imprinted genes are shown with light pink boxes. **a**, *H19*. **b**, *Rasgrf1*. **c**, *Dlk1-Meg3*. **d**, *Kcnq1ot1*. **e**, *Nespas-Gnas*. **f**, *Snrpn*. **g**, *Peg10*



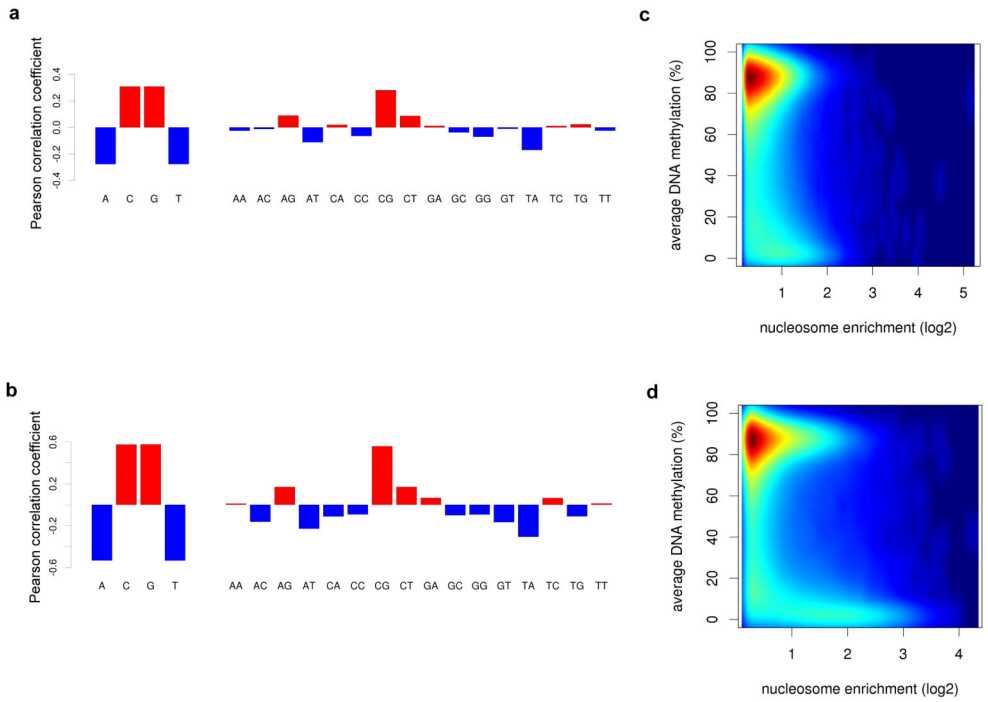
Supplementary Figure 6: Chromatin states of imprinting control regions (ICRs) in sperm and round spermatids.
From top to bottom, images show the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modification states in round spermatids. Imprinting control regions[16] for paternally imprinted genes are shown with light blue boxes and imprinting control regions for maternally imprinted genes are shown with light pink boxes. **a**, *H19*. **b**, *Rasgrf1*. **c**, *Dlk1-Meg3*. **d**, *Kcnq1ot1*. **e**, *Nespas-Gnas*. **f**, *Snrpn*. **g**, *Peg10*



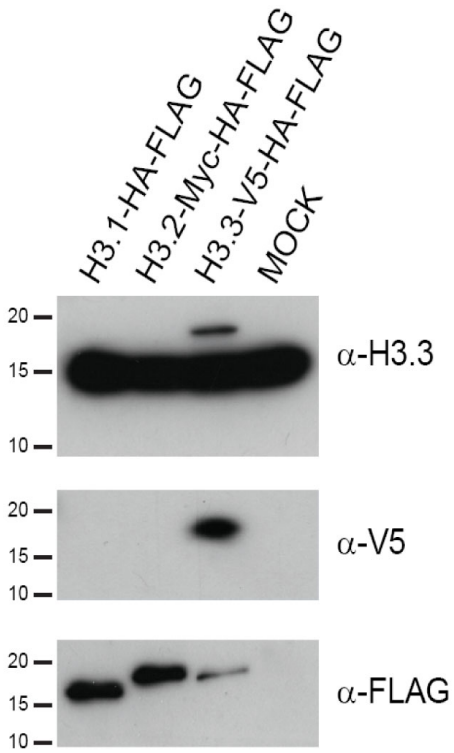
Supplementary Figure 6: Chromatin states of imprinting control regions (ICRs) in sperm and round spermatids.
From top to bottom, images show the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modification states in round spermatids. Imprinting control regions[16] for paternally imprinted genes are shown with light blue boxes and imprinting control regions for maternally imprinted genes are shown with light pink boxes. **a**, *H19*. **b**, *Rasgrf1*. **c**, *Dlk1-Meg3*. **d**, *Kcnq1ot1*. **e**, *Nespas-Gnas*. **f**, *Snrpn*. **g**, *Peg10*



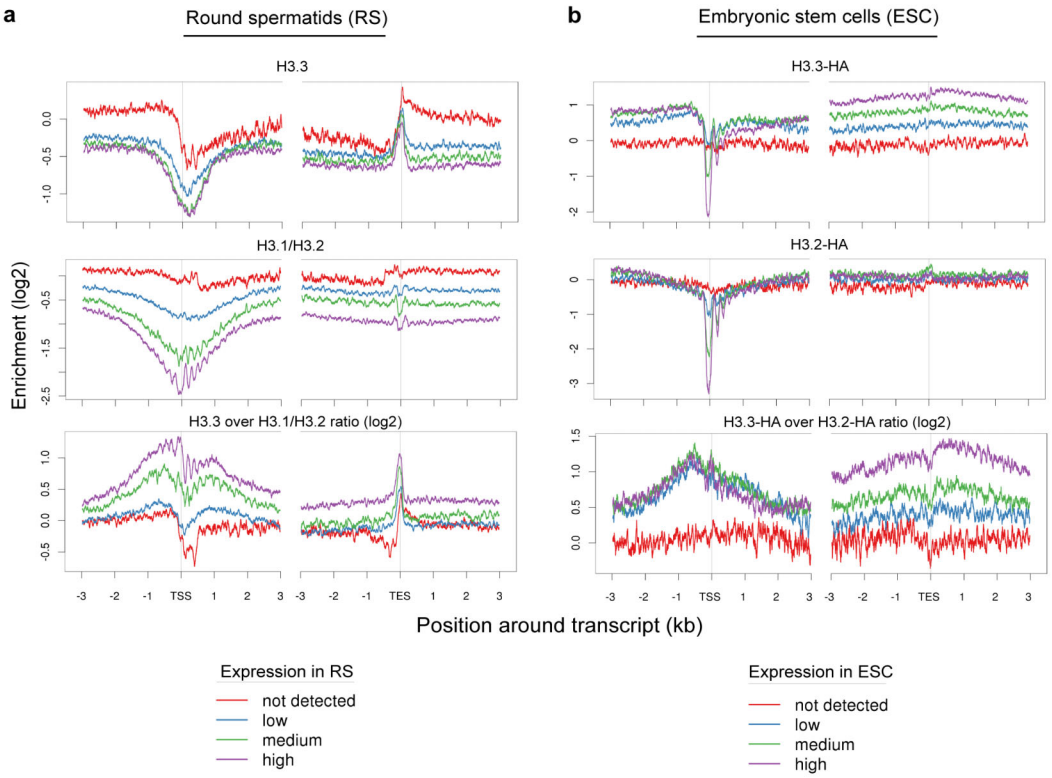
Supplementary Figure 6: Chromatin states of imprinting control regions (ICRs) in sperm and round spermatids.
From top to bottom, images show the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modification states in round spermatids. Imprinting control regions[16] for paternally imprinted genes are shown with light blue boxes and imprinting control regions for maternally imprinted genes are shown with light pink boxes. **a**, *H19*. **b**, *Rasgrf1*. **c**, *Dlk1-Meg3*. **d**, *Kcnq1ot1*. **e**, *Nespas-Gnas*. **f**, *Snrpn*. **g**, *Peg10*



Supplementary Figure 7: Principles defining nucleosome occupancy in mouse sperm is conserved for human sperm.
a, Correlation of single nucleotide frequencies (left) and single nucleotide composition normalized dinucleotide frequencies (right) with nucleosome occupancy in human sperm[5] in 1kb windows tiling the human genome. **b**, Same as in (a) except displaying human sperm nucleosome data from[4]. **c**, Correlation of human sperm DNA methylation[36] with nucleosome occupancy in human sperm[5] in 1kb windows tiling the human genome. **d**, Same as in (c) except displaying human sperm nucleosome data from[4].



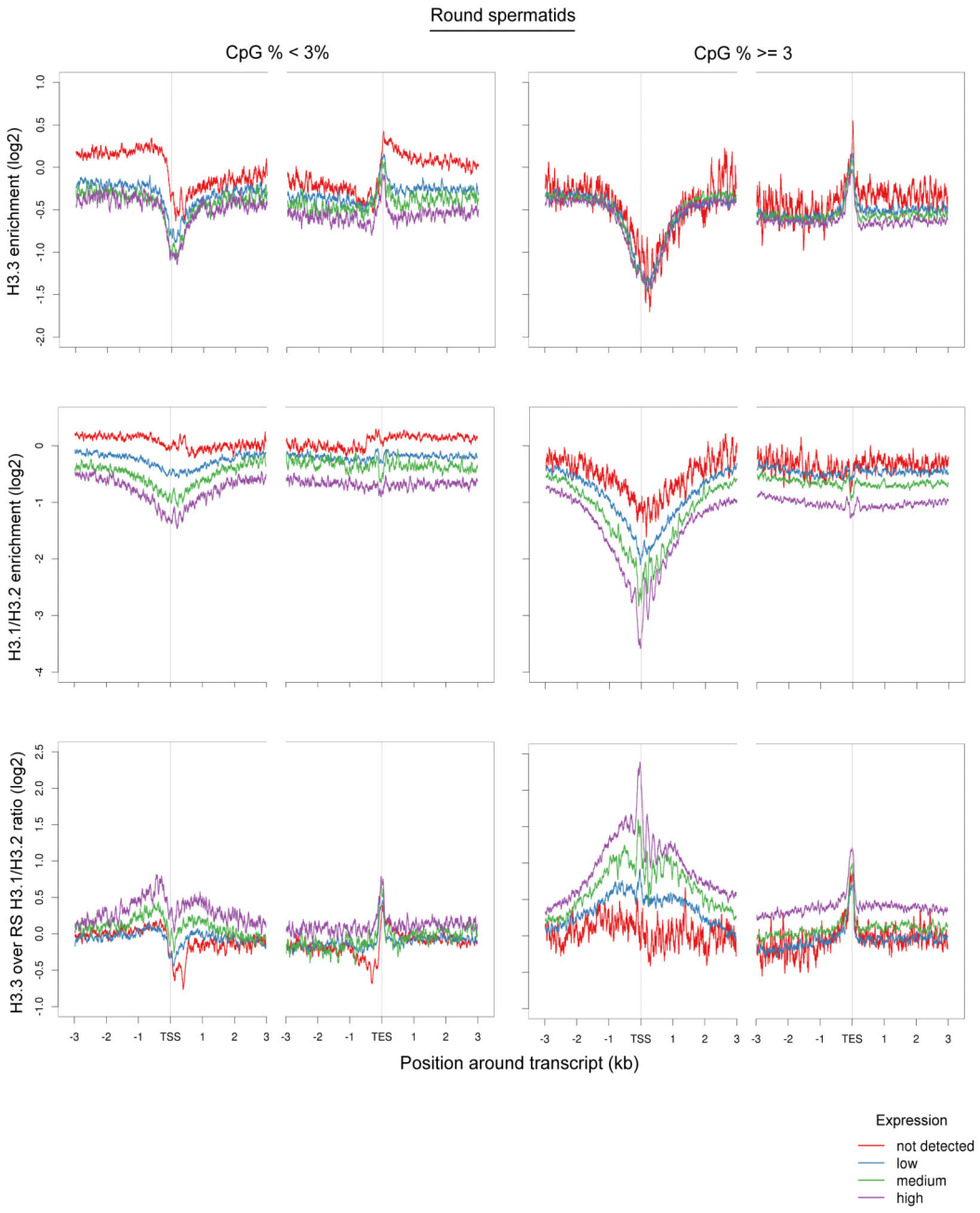
Supplementary Figure 8: Specificity of the H3.3 antibody.
Western blots showing endogenous and exogenously expressed histone variant H3.3 as detected by H3.3 and tag antibodies. 293 cells were transfected with constructs encoding tagged canonical H3.1, H3.2 and variant H3.3 histones.



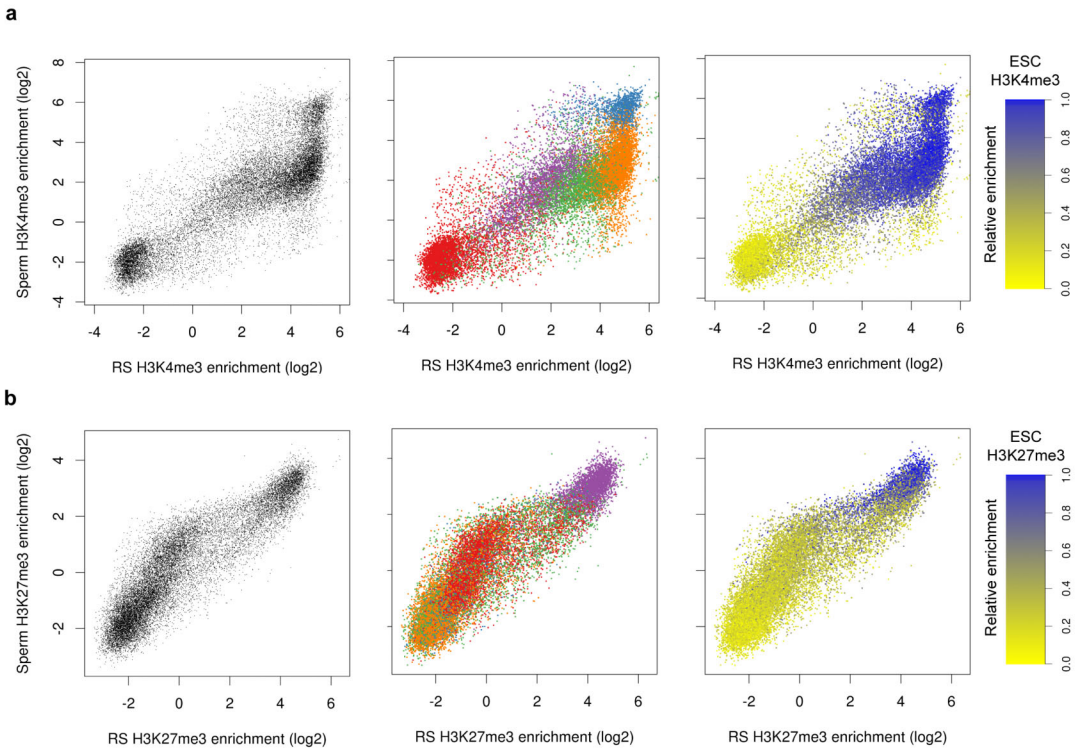
Supplementary Figure 9: Average profiles of canonical and variant histone occupancy along genes in relation to expression status in round spermatids and ESCs.

a, Average profiles of H3.3 and H3.1/H3.2 enrichments around transcriptional start sites (TSS; ±3kb) and transcriptional end sites (TES; ±3kb) in sperm and RS. Genes were classified according to expression status in RS. Transcripts without any aligned reads were classified as “not detected”. Remaining transcripts were classified on the basis of increasing expression values into three equally sized groups. In the bottom panel, the ratio between H3.3 over H3.1/H3.2 is shown.

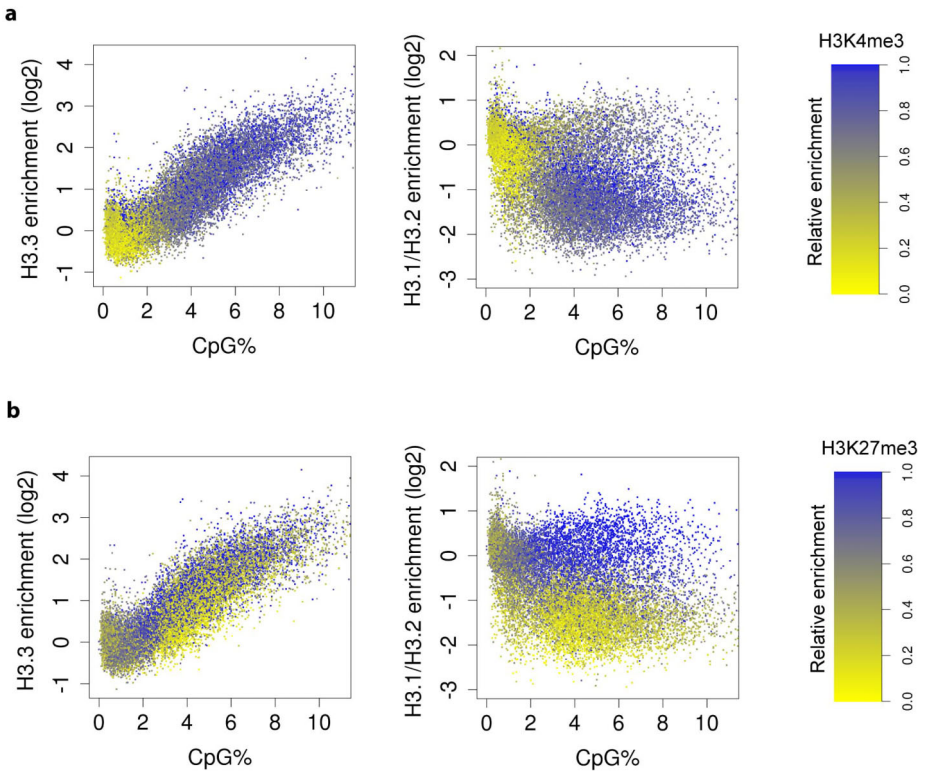
b, Average profiles of histone variants H3.3-HA and H3.2-HA in embryonic stem cells (ESC)[32]. Genes were classified as in (**a**) according to expression status in ESC[24]. In the bottom panel, the ratio between H3.3 over H3.1/H3.2 is shown.



Supplementary Figure 10: Average profiles of canonical and variant histone occupancy along genes with CpG-low and CpG-high promoters in relation to expression status in round spermatids.
Average profiles of H3.3 and H3.1/H3.2 enrichments around TSS (±3kb) and transcriptional end sites (TES; ±3kb) in sperm and RS. Genes were classified according to expression status in RS (as described in Supplementary Fig. 9a) and the percentage of CpGs within ±1kb windows around TSS of genes (left: CpG % < 3, right: CpG % ≥ 3).

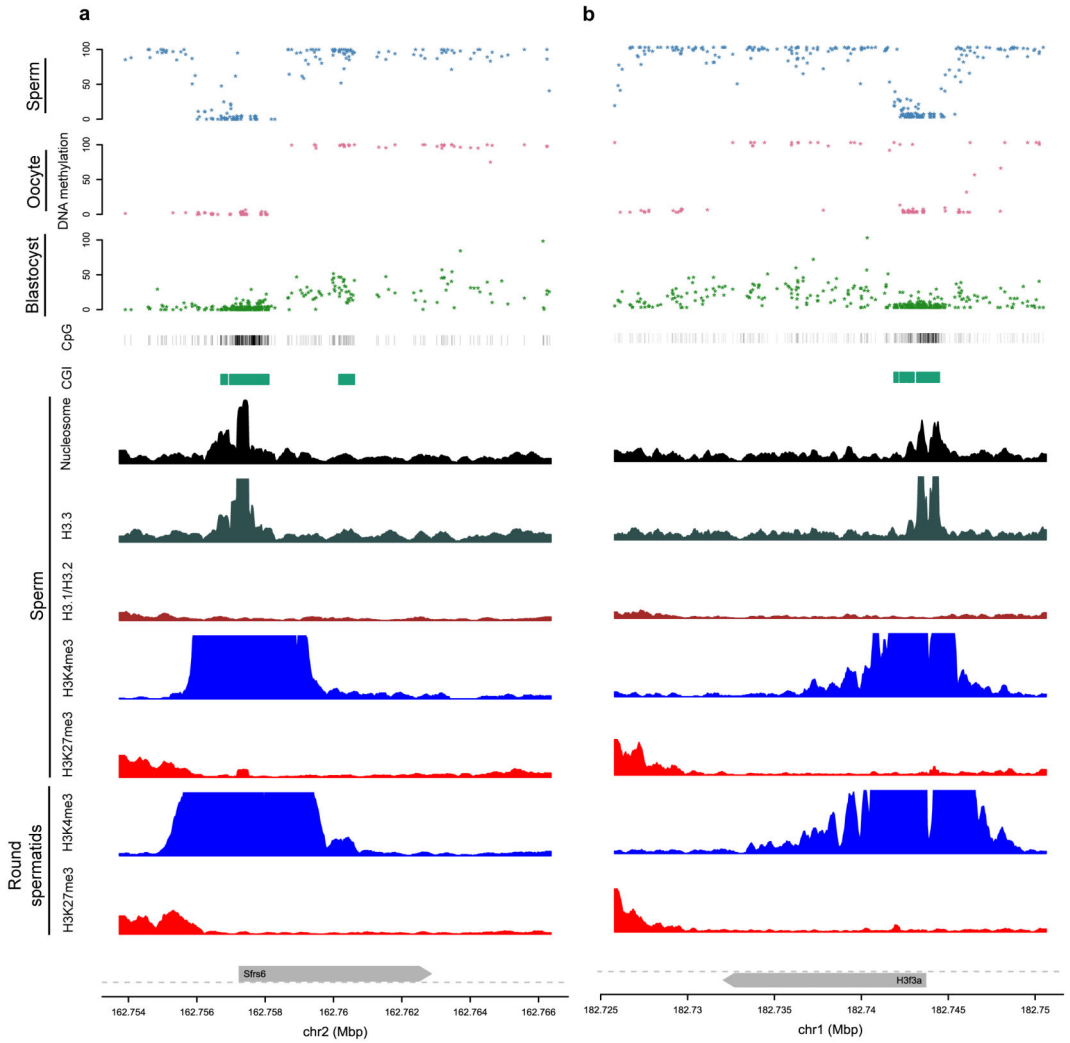


Supplementary Figure 11: Comparison of histone modification patterns between round spermatids and sperm.
a, Scatter plots showing the enrichments of H3K4me3 in round spermatids and sperm (left), in relation to the five gene clusters as described and color coded in Fig. 3b (middle), and in comparison to H3K4me3 levels in mouse ESCs[37]. **b**, H3K27me3 enrichments in RS, sperm and ESCs[24], displayed as in (**a**).

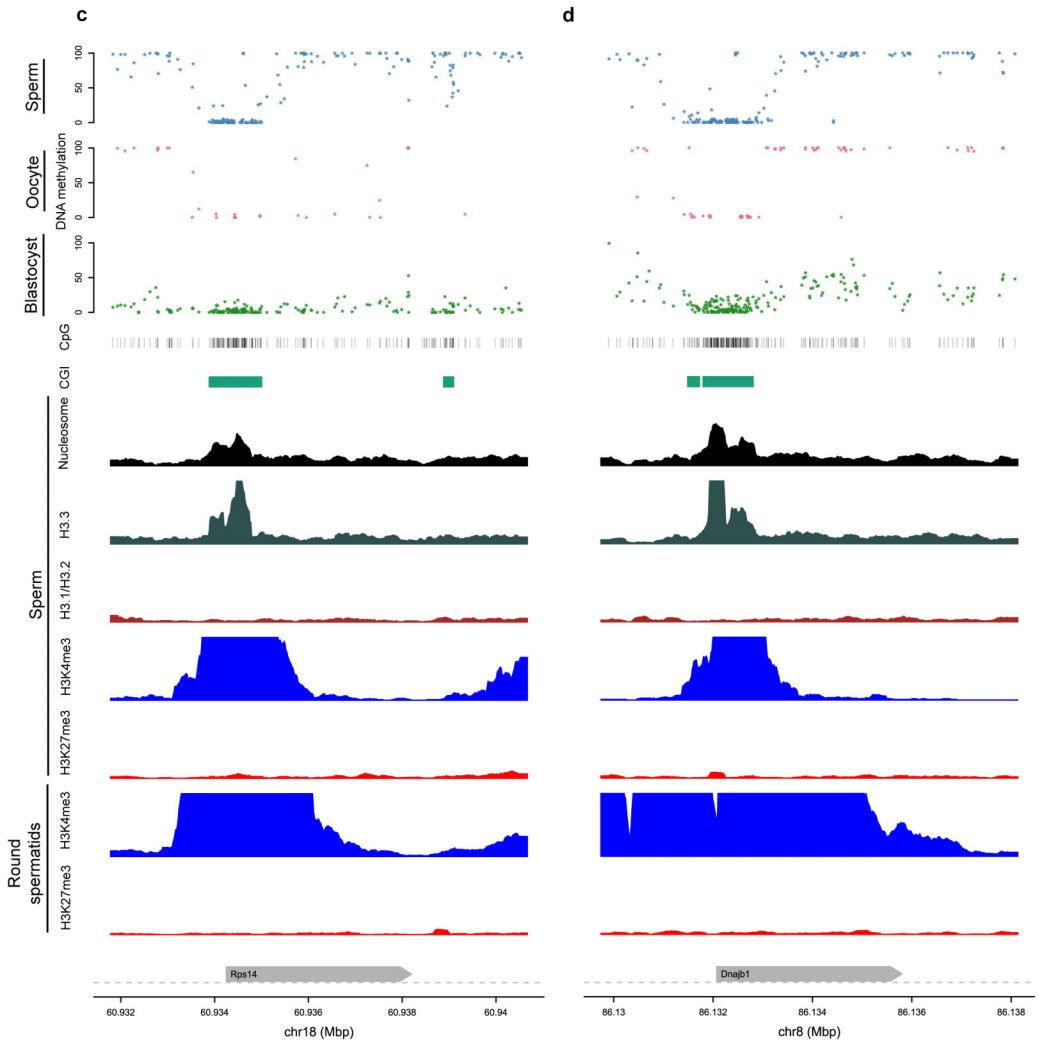


Supplementary Figure 12: Differential association of histone modifications with histone variants in sperm.

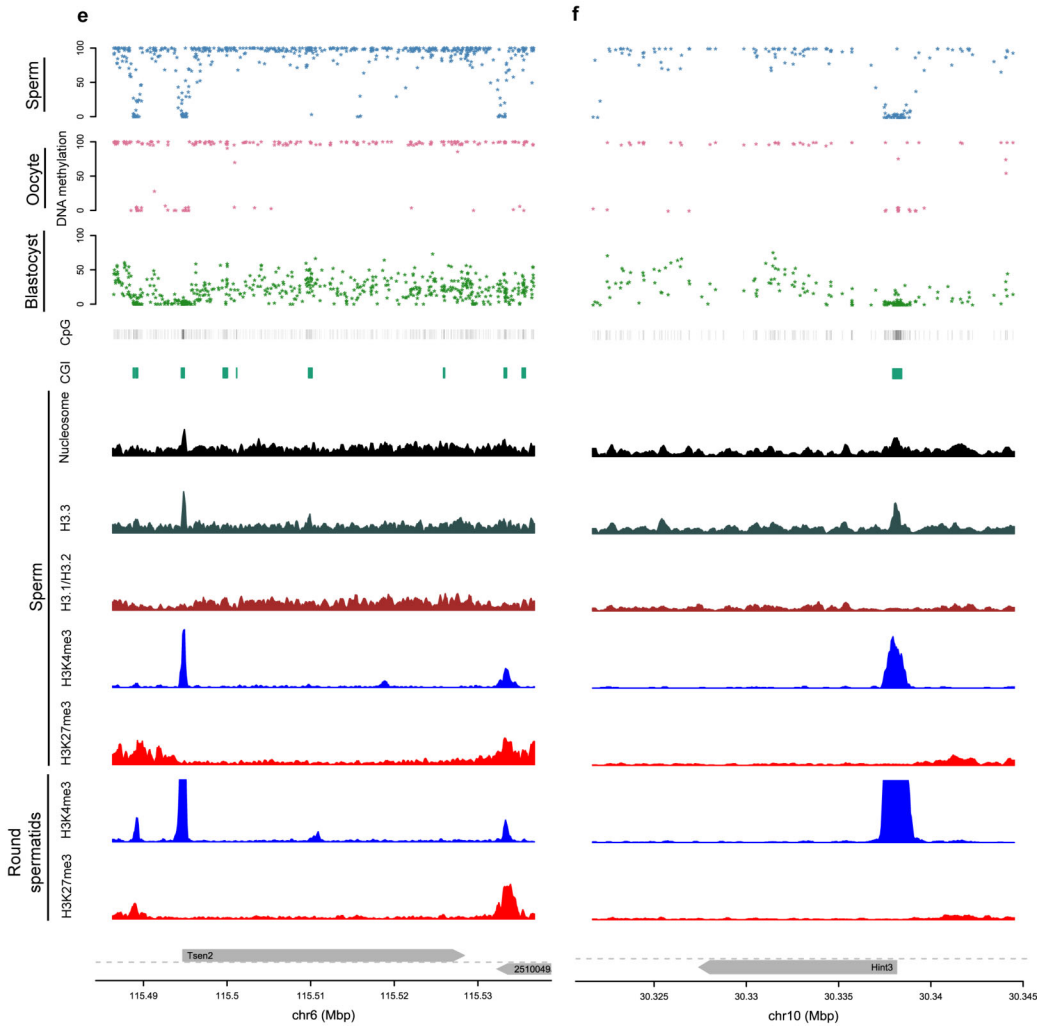
Scatter plots showing the correlation of the percentage of CpGs with enrichments of variant and canonical H3 histones (±1 kb around TSS) in sperm. The color of the points represents the enrichment of H3K4me3 (**a**) and H3K27me3 (**b)** in sperm.



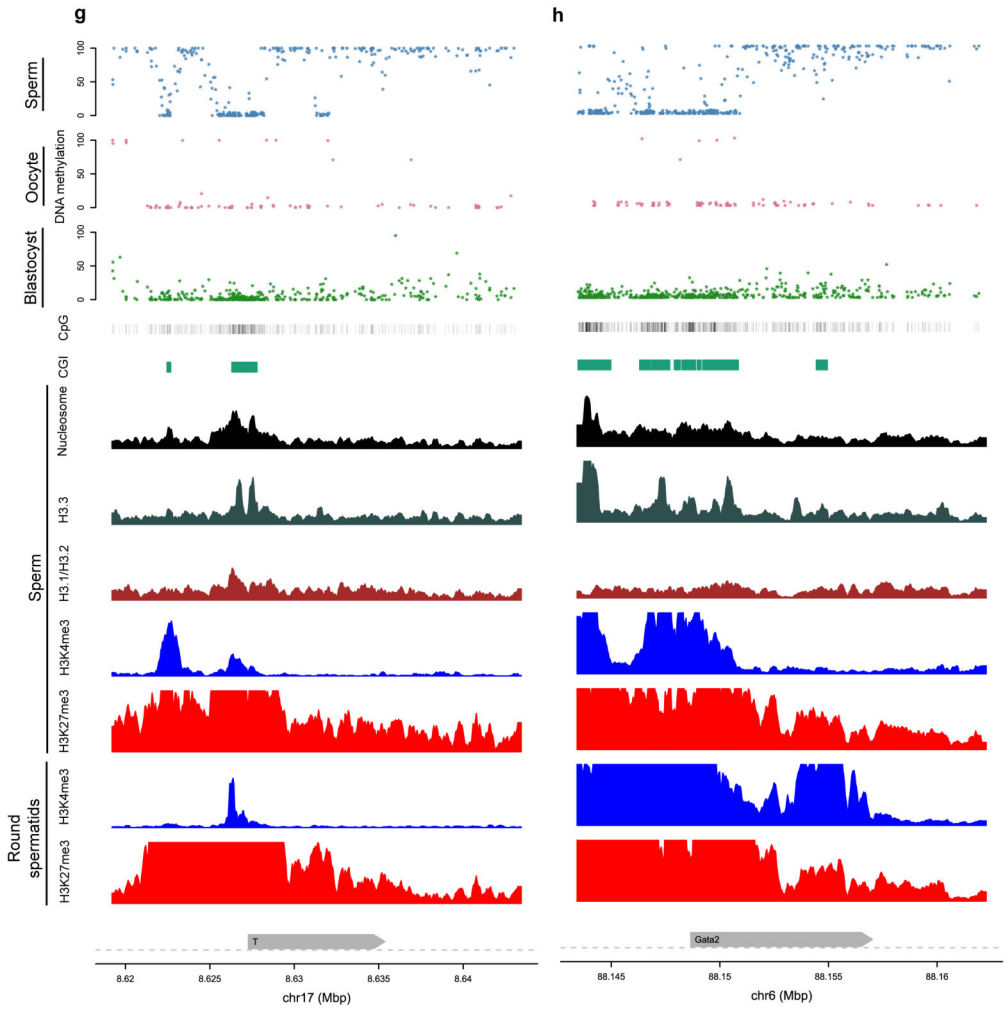
Supplementary Figure 13: Chromatin states of genes representative of the gene clusters described in Fig. 3b in sperm and round spermatids.
From top to bottom, images show at various loci the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modifications in round spermatids. **a,** and **b**, examples for cluster 1, *Sfrs6* and *H3f3a*. **c,** and **d**, examples for cluster 2, *Rps14* and *Dnajb1*. **e,** and **f**, examples for cluster 3, *Tsen2* and *Hint3*. **g,** and **h**, examples for cluster 4, *T* and *Gata2*. **i,** and **j**, examples for cluster 5, *Olfr* family and *Cts6*.



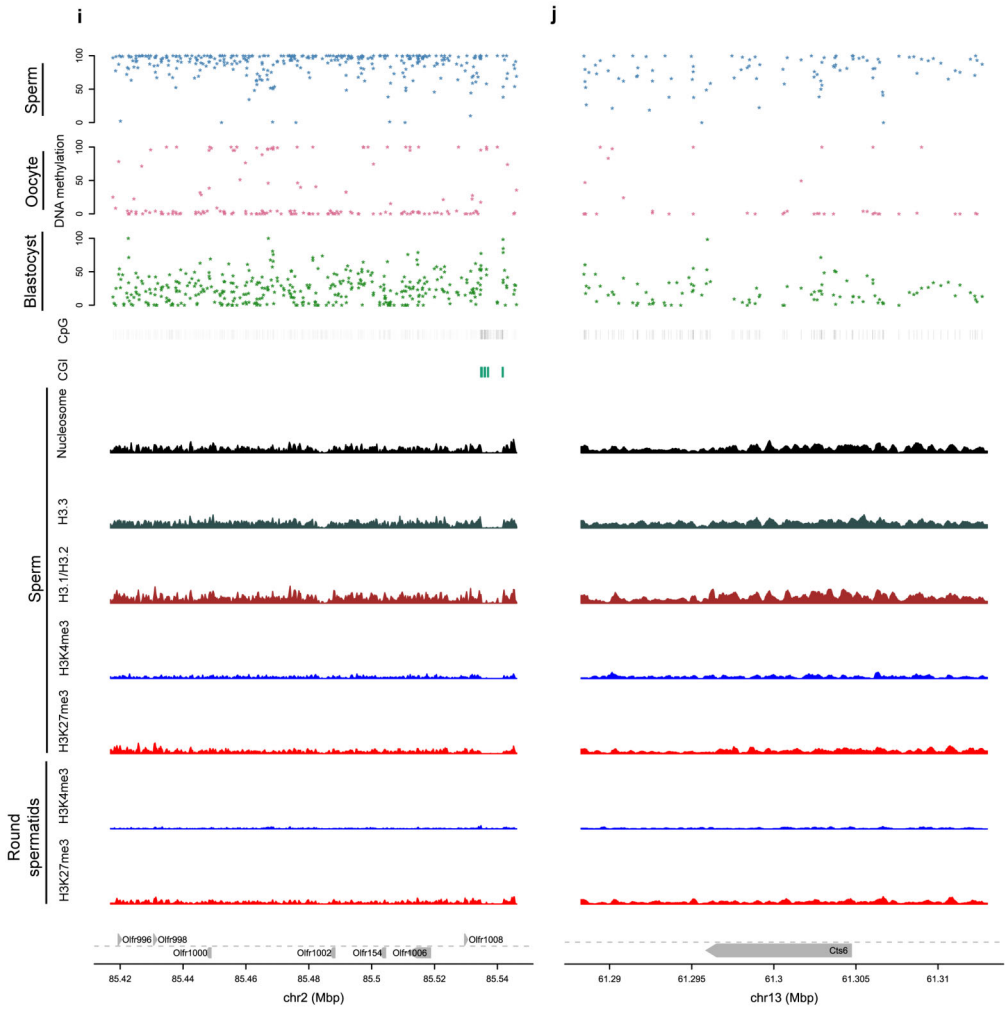
Supplementary Figure 13: Chromatin states of genes representative of the gene clusters described in Fig. 3b in sperm and round spermatids.
From top to bottom, images show at various loci the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modifications in round spermatids. **a,** and **b**, examples for cluster 1, *Sfrs6* and *H3f3a*. **c,** and **d**, examples for cluster 2, *Rps14* and *Dnajb1*. **e,** and **f**, examples for cluster 3, *Tsen2* and *Hint3*. **g,** and **h**, examples for cluster 4, *T* and *Gata2*. **i,** and **j**, examples for cluster 5, *Olfr* family and *Cts6*.



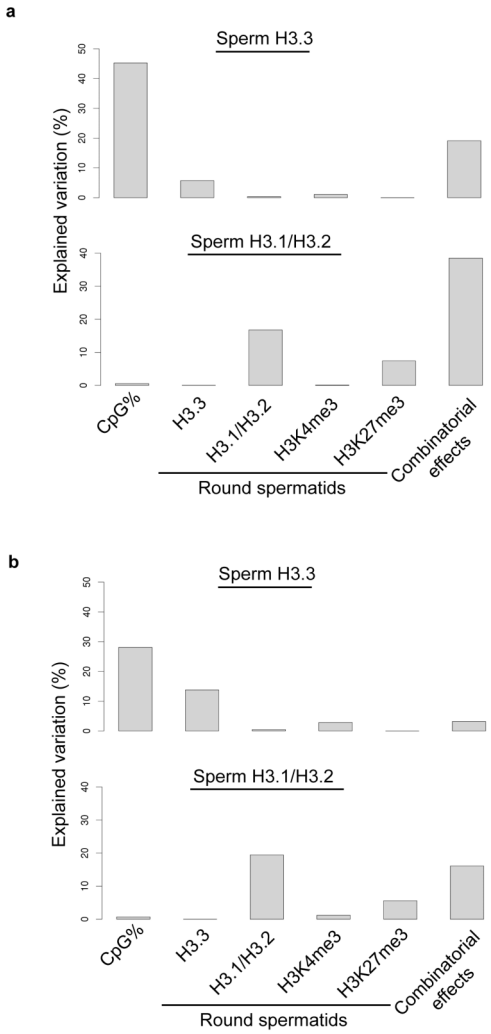
Supplementary Figure 13: Chromatin states of genes representative of the gene clusters described in Fig. 3b in sperm and round spermatids.
From top to bottom, images show at various loci the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modifications in round spermatids. **a,** and **b**, examples for cluster 1, *Sfrs6* and *H3f3a*. **c,** and **d**, examples for cluster 2, *Rps14* and *Dnajb1*. **e,** and **f**, examples for cluster 3, *Tsen2* and *Hint3*. **g,** and **h**, examples for cluster 4, *T* and *Gata2*. **i,** and **j**, examples for cluster 5, *Olfr* family and *Cts6*.



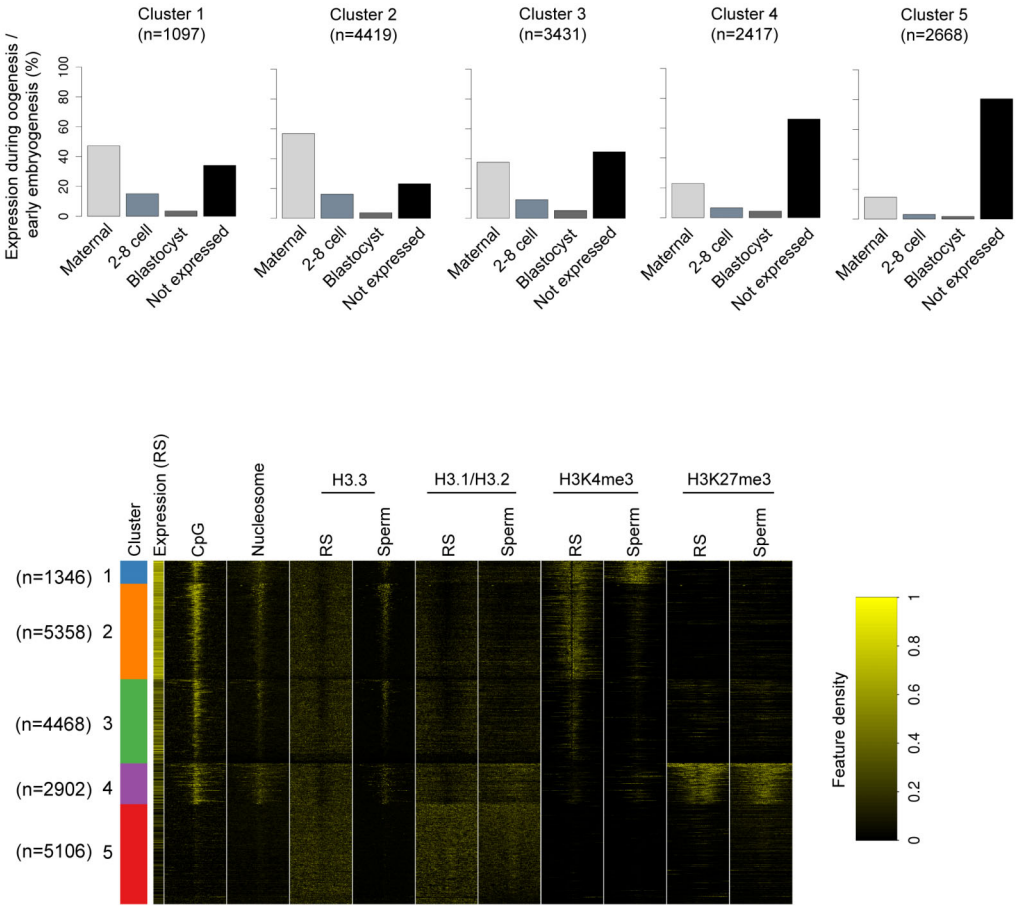
Supplementary Figure 13: Chromatin states of genes representative of the gene clusters described in Fig. 3b in sperm and round spermatids.
From top to bottom, images show at various loci the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modifications in round spermatids. **a,** and **b**, examples for cluster 1, *Sfrs6* and *H3f3a*. **c,** and **d**, examples for cluster 2, *Rps14* and *Dnajb1*. **e,** and **f**, examples for cluster 3, *Tsen2* and *Hint3*. **g,** and **h**, examples for cluster 4, *T* and *Gata2*. **i,** and **j**, examples for cluster 5, *Olfr* family and *Cts6*.



Supplementary Figure 13: Chromatin states of genes representative of the gene clusters described in Fig. 3b in sperm and round spermatids.
From top to bottom, images show at various loci the DNA methylation status for sperm, oocyte and blastocyst[16], CpG density, CGI localization, nucleosome, histone variant and histone modification states in sperm and histone modifications in round spermatids. **a,** and **b**, examples for cluster 1, *Sfrs6* and *H3f3a*. **c,** and **d**, examples for cluster 2, *Rps14* and *Dnajb1*. **e,** and **f**, examples for cluster 3, *Tsen2* and *Hint3*. **g,** and **h**, examples for cluster 4, *T* and *Gata2*. **i,** and **j**, examples for cluster 5, *Olfr* family and *Cts6*.



Supplementary Figure 14: Variance partitioning analysis of H3.3 and H3.1/H3.2 histone enrichments in sperm.
a, H3.3 enrichments (top panel) and H3.1/H3.2 enrichments (bottom panel) were measured genome-wide in 1kb windows and modeled as a linear combination of explanatory variables (CpG content and histone measurements in round spermatids). The unique contribution of each variable to observed sperm variation is shown in %. Combinatorial effects refer to variation which is explained by combinations of variables used in the analysis. **b**, analysis as in a but only using genomic windows not intersecting ± 1kb TSS.



Supplementary Figure 15: Expression status of genes belonging to different genes cluster during oogenesis / early embryogenesis.
Expression states during oogenesis and early embryogenesis are referred as "maternal", "2-8 cell", "blastocyst" and "not expressed" as described in[5]. Expression status of genes belonging to different clusters (Fig. 3b) is shown as % in the bar plots. From a total of 19180 genes, 14032 could be matched for expression status during oogenesis and/or early embryogenesis. Number of genes in each cluster is shown at the top of the bar plots. To facilitate comparison, data from Fig. 3b is presented in the bottom panel.

Supplementary Table 1: GO-term analysis for the gene clusters described in Fig. 3b.

Table provides the top 20 enriched GO-terms for each cluster.

Cluster 1-Go terms	Significant	Expected	classicFisher	Over-representation
protein modification by small protein conjugation or removal	69	29.44	2.40E-11	2.344
spermatogenesis	52	19.79	1.10E-10	2.628
male gamete generation	52	19.79	1.10E-10	2.628
protein ubiquitination	57	23	1.50E-10	2.478
cellular protein catabolic process	54	21.96	6.10E-10	2.459
ubiquitin-dependent protein catabolic process	48	18.67	1.20E-09	2.571
proteolysis involved in cellular protein catabolic process	52	21.26	1.50E-09	2.446
protein modification by small protein conjugation	58	25.03	1.50E-09	2.317
sexual reproduction	66	30.28	1.50E-09	2.180
cellular macromolecule metabolic process	458	364.86	2.10E-09	1.255
modification-dependent protein catabolic process	48	19.09	2.60E-09	2.514
cellular protein metabolic process	264	189.7	3.40E-09	1.392
modification-dependent macromolecule catabolic process	48	19.3	3.80E-09	2.487
protein catabolic process	60	28.6	3.50E-08	2.098
cellular macromolecule catabolic process	61	29.3	3.60E-08	2.082
proteasomal ubiquitin-dependent protein catabolic process	28	9.23	9.90E-08	3.034
gamete generation	55	26.08	1.10E-07	2.109
proteasomal protein catabolic process	29	9.93	1.40E-07	2.920
cellular metabolic process	582	499.04	4.20E-07	1.166
protein polyubiquitination	22	7.06	1.40E-06	3.116

Cluster 2-Go terms	Significant	Expected	classicFisher	Over-representation
RNA splicing	137	55.53	< 1e-30	2.467
mRNA processing	177	73.38	< 1e-30	2.412
mRNA metabolic process	206	87.26	< 1e-30	2.361
RNA processing	296	128.34	< 1e-30	2.306
nuclear division	177	79.61	< 1e-30	2.223
mitosis	177	79.61	< 1e-30	2.223
organelle fission	185	84.15	< 1e-30	2.198
M phase of mitotic cell cycle	180	81.88	< 1e-30	2.198
DNA repair	186	85	< 1e-30	2.188
M phase	250	117.86	< 1e-30	2.121
chromosome organization	310	160.64	< 1e-30	1.930
cell cycle phase	307	161.77	< 1e-30	1.898
cell cycle process	386	206.82	< 1e-30	1.866
mitotic cell cycle	266	143.36	< 1e-30	1.855
DNA metabolic process	330	185.29	< 1e-30	1.781
cell cycle	535	302.58	< 1e-30	1.768
organelle organization	800	491.84	< 1e-30	1.627

cellular protein metabolic process	1105	768.64	< 1e-30	1.438
cellular component organization or biogenesis at cellular level	1052	754.48	< 1e-30	1.394
nucleic acid metabolic process	1326	953.08	< 1e-30	1.391

Cluster 3-Go terms	**Significant**	**Expected**	**classicFisher**	**Over-representation**
metabolic process	2189	2023.32	2.10E-09	1.082
cellular metabolic process	1813	1669.8	1.30E-07	1.086
cofactor metabolic process	85	51	1.70E-07	1.667
small molecule metabolic process	513	430.73	1.50E-06	1.191
primary metabolic process	1837	1713.32	4.60E-06	1.072
cofactor biosynthetic process	41	21.76	8.40E-06	1.884
coenzyme metabolic process	68	42.35	1.20E-05	1.606
cellular lipid metabolic process	194	149.74	2.50E-05	1.296
type I interferon production	18	7.49	6.30E-05	2.403
cofactor transport	10	3.04	6.80E-05	3.289
cellular ketone metabolic process	201	159.1	9.40E-05	1.263
B cell homeostasis	15	5.85	9.50E-05	2.564
interferon-beta production	15	5.85	9.50E-05	2.564
B cell differentiation	30	15.91	0.00013	1.886
oxidation-reduction process	229	185.07	0.00013	1.237
biosynthetic process	1018	931.65	0.00014	1.093
lipid metabolic process	256	211.27	0.00023	1.212
regulation of interferon-beta production	14	5.62	0.00025	2.491
alcohol metabolic process	75	51.94	0.00027	1.444
glycerolipid biosynthetic process	41	24.8	0.0003	1.653

Cluster 4- Go terms	**Significant**	**Expected**	**classicFisher**	**Over-representation**
cell fate commitment	137	31.35	<1e-30	4.370
embryonic organ morphogenesis	142	33.23	<1e-30	4.273
axonogenesis	171	46.17	<1e-30	3.704
regionalization	168	45.39	<1e-30	3.701
pattern specification process	222	61.93	<1e-30	3.585
cell morphogenesis involved in neuron differentiation	188	54.44	<1e-30	3.453
cell morphogenesis involved in differentiation	246	71.91	<1e-30	3.421
neuron projection morphogenesis	192	56.31	<1e-30	3.410
brain development	201	59.28	<1e-30	3.391
central nervous system development	260	78.93	<1e-30	3.294
embryonic morphogenesis	243	75.35	<1e-30	3.225
regulation of nervous system development	227	70.67	<1e-30	3.212
regulation of cell development	240	76.75	<1e-30	3.127
organ morphogenesis	345	110.44	<1e-30	3.124
tissue morphogenesis	220	70.51	<1e-30	3.120
neuron projection development	244	79.25	<1e-30	3.079
neuron differentiation	369	120.27	<1e-30	3.068
neuron development	283	93.29	<1e-30	3.034
epithelium development	251	82.99	<1e-30	3.024

generation of neurons	400	133.06	<1e-30	3.006

Cluster 5- Go terms	Significant	Expected	classicFisher	Over-representation
detection of chemical stimulus involved in sensory perception of smell	1057	274.52	< 1e-30	3.850
detection of chemical stimulus involved in sensory perception	1085	282.99	< 1e-30	3.834
sensory perception of smell	1062	279.4	< 1e-30	3.801
detection of chemical stimulus	1092	288.64	< 1e-30	3.783
sensory perception of chemical stimulus	1107	295.06	< 1e-30	3.752
detection of stimulus involved in sensory perception	1091	292.75	< 1e-30	3.727
detection of stimulus	1118	310.21	< 1e-30	3.604
sensory perception	1204	371.59	< 1e-30	3.240
defense response to bacterium	110	36.21	< 1e-30	3.038
G-protein coupled receptor signaling pathway	1448	477.13	< 1e-30	3.035
neurological system process	1298	508.21	< 1e-30	2.554
system process	1411	594.75	< 1e-30	2.372
cell surface receptor signaling pathway	1659	781.7	< 1e-30	2.122
response to chemical stimulus	1531	739.07	< 1e-30	2.072
defense response	346	182.84	< 1e-30	1.892
signal transduction	1877	1142.24	< 1e-30	1.643
signaling	1941	1225.45	< 1e-30	1.584
cell communication	1959	1252.16	< 1e-30	1.564
cellular response to stimulus	2009	1318.67	< 1e-30	1.524
response to stimulus	2411	1606.8	< 1e-30	1.500

Supplementary Methods

Biological Sample Collection

Mouse sperm were collected from C57BL/6J mice by using swip-up procedure as described[5]. To isolate round spermatids, testicular cells were prepared from 28 day C57BL/6J mice. Isolated cells were subjected to Hoechst (Invitrogen, catalog number 33342) staining for 30 min at 37°C and round spermatids were collected via Fluorescent Activated Cell Sorter (FACS) with 90 % purity.

Mononucleosomal DNA preparation and native ChIP

Chromatin isolation from mature sperm was performed under native conditions as described[5]. MNase treatment for sperm was performed with 15 U (Roche Nuclease S7, catalog number 10107921001) at 37°C for 5 min per 2 million spermatozoa. Round spermatid chromatin was isolated in the same way, except for the DTT

treatment which was used for sperm. MNase treatment for round spermatids was performed with 5 U at 37°C for 30 minutes per 1 million cells. Immunoprecipitation was carried out with antibodies against H3.3 (Millipore 17-10245), H3.1/H3.2[38,30], H3K4me3 (Millipore 17-614) and H3K27me3 (Millipore 07-449) by using roughly 15-20 million sperm or 5 million round spermatids per immunoprecipitation. Both mononucleosomal DNA and immunoprecipitated DNA were resolved by 5% polyacrylamid electrophoresis and DNA was cut at the size of 150 bp. 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.

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 via 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).

Mononucleosome-BisSeq (MN-BisSeq) library preparation

The protocol was adapted from Illumina Genomic DNA Sample Preparation Guide. Briefly, 2 µg of mononucleosomal fraction DNA were end repaired by incubation at 20°C for 30 minutes with 200µM dNTP, 7.5 units of T4 DNA polymerase (NEB #M0203S), 5 units of DNA Polymerase I Large Fragment (Klenow) (NEB #M0210S), 25 units of T4 PNK (NEB #M0201S), 1x T4 DNA ligase buffer containing 10mM ATP (NEB). 3' ends of DNA fragments were adenylated by incubation at 37°C for 30 minutes with 100µM dATP, 1xNEB Buffer 2, 10 units Klenow Fragment (3´→5´ exo–) (NEB # M0212L). Single End adapter sequences were produced based on Illumina adapter sequences (Oligonucleotide sequences © 2006-2008 Illumina, Inc. All rights reserved). 5' P- GATXGGAAGAGXTXGTATGXXGTXTTXTGXTTG and 5' AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXT, where X is a methylated cytosine.

Adapters were ordered as single stranded oligos (Microsynth AG), resuspended in annealing buffer (10mM Tris pH7.5, 50mM NaCl, 1mM EDTA), annealed by heating at 95°C for 10 minutes and cooling down slowly. Annealed adapters were ligated to the DNA fragments as per manufacturer's instructions for genomic DNA library construction. Adapter-ligated DNA of ≈ 250 bp was isolated on 2% agarose gel electrophoresis. Gel purified DNA was then converted with sodium bisulfite using the Imprint® DNA Modification Kit (Sigma-Aldrich) following the manufacturer's instructions. Half of the bisulfite-converted, adapter-ligated DNA molecules wasenriched by 7 cycles of PCR with the following reaction composition: 2.5 U of uracil-insensitive *PfuTurboCx* Hotstart DNA polymerase (Stratagene), 5 µl 10X *PfuTurbo* reaction buffer, 25 µM dNTPs, 0.5µM of Illumina single end PCR primers. The thermocycling parameters were: 95°C 2 min, 98°C 30 sec, then 7 cycles of 98°C 15 sec, 65°C 30 sec and 72°C 3 min, ending with one 72°C 5 min elongation step. The reaction products were purified using the MinElute PCR purification kit (Qiagen, Valencia, CA), resolved by 2% agarose gel electrophoresis to separate the library from adapter-adapter ligation products, and purified from the gel using the MinElute gel purification kit (Qiagen, Valencia, CA). Quality of the libraries and template size distribution were assessed by running an aliquot of the library on an Agilent 2100 Bioanalyzer (Agilent Technologies).

Chromatin-bound (histone) fractionation and immunoblotting

Round spermatids were isolated from C57BL/6J mouse testes by centrifugal elutriation[39] and chromatin-bound fractionation was performed according to[40] 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[38,30].

Processing and alignment of the reads

Filtering, alignment and processing of the reads for both ChIP-seq and RNA-seq were done as described[24]. 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[24]) was used in the subsequent analysis. Refseq coordinates were downloaded from UCSC[41] (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/refGene.txt.gz from August 16, 2009). For each gene, only one transcript with the most extreme coordinates 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 rest 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.

UCSC tracks

Wiggle files were generated for 100 bp windows and uploaded to the UCSC genome browser[41]. Data was visualized using smoothing over 3 pixels (Fig.1a and 1c, Fig. 2b).

Peak identification

Peak identification for nucleosome data was performed by training a two state hidden semi-Markov model, R-mhsmm package[42] on nucleosome enrichments calculated for

1 kb windows. The two-state model (non-peak and peak states) was initialized using Gaussian emissions (means of 0, 1, and variances of 0.5, 0.5), a gamma sojourn distribution with shape=2 and scale=10, and initial state probabilities of 0.5, 0.5. Parameter estimation of the model was performed by using 1kb window enrichments on chr1 as a training data set and by selecting a maximum of 200 windows in one state. Model parameters were estimated using EM algorithm, and the fitted model with emission distribution means of -0.21, 0.57, variances of 0.071, 0.48, and sojourn distribution parameters shape of 1.32 , 0.38 , and scale of 52.84 , 7.86 was used to predict the maximum likelihood state path for the sequence of all 1kb windows in the genome. Adjacent 1kb windows with identical state labels were fused.

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

Enrichment levels for ChIP-seq experiments were calculated for the 1kb windows, promoter regions of the genes (± 1kb surrounding transcriptional start sites (TSS)), and nucleosome peaks identified by hidden semi-Markov model. To calculate enrichment, total read counts mapping to a coordinate region were calculated for IP sample and control (input genomic DNA). Then, these counts were normalized to account for different library sizes between IP sample and control. Enrichment for each region was calculated as the ratio between library size normalized read counts for IP sample and control 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 IP sample, $LSize_{smp}$ is the total library size for the IP 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 (8), which was used to stabilize enrichments based on low read counts.

Classification of nucleosome peaks

Nucleosome enrichments were quantified on the peaks identified by a hidden semi-Markov model. Peaks were classified into three equal sized groups according to enrichment levels, termed “weak”, “intermediate” and “strong” peak groups (Fig. 2e).

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[19]. 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[43]. 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).

Analysis of bisulfite converted sequencing (BisSeq) data

Read filtering and alignment of the BisSeq data from this study (bisulfite converted mononucleosome associated sperm DNA), published sperm whole methylomes (mouse sperm[16] and human sperm[36]) and published oocyte and blastocyst methylomes[16] were performed as described[24].

CGI definition and usage

CpG island definitions are based on a CpG cluster algorithm[35]. The algorithm was run with default parameters on mm9 to obtain genomic coordinates of CGI.

Genome-wide modeling of nucleosome occupancy

Modeling of genome-wide mouse sperm nucleosome occupancy (Fig. 1f) was performed by using enrichment values in 1kb windows. Nucleosome occupancy data was modeled by a linear model with CpG dinucleotide frequency and the average % DNA methylation in 1kb windows as regressors, including only windows with detectable levels of nucleosomes, defined as log2 nucleosome occupancy greater than 0.2. This filtering excludes the majority of 1kb windows without nucleosomes from the analysis which is required since the majority of the genome does not contain any nucleosomes in mouse sperm. In the model, DNA methylation data which was obtained by bisulfite converted sequencing of nucleosome associated DNA was used. Average DNA methylation was calculated by taking the ratio of total number of reads for methylated C over total number of reads for all C (methylated or umethylated) per window. Windows with less than 5 total reads for all Cs were excluded from the analysis. Finally, n=105571 (9% of all windows) were used in modeling of nucleosome occupancy in mouse sperm. The same windows were used to analyze the relationship between mouse sperm nucleosome occupancy and DNA methylation / sequence composition (Fig. 1d, Supplementary Fig. 5a and 5b).

To analyze the relationship between human sperm nucleosome occupancy[5,4] and DNA methylation / sequence composition (Supplementary Fig. 7), 1kb windows were processed in a similar way as for mouse sperm. Finally, n=894787 (29% of all windows) of[5] and n=515937 (17% of all windows) of[4] were used in the analysis.

Plotting profiles around genomic regions

For each sample, reads mapping to the genomic regions of interest (Fig. 2d, Supplementary Fig. 9 and 10) 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.

Quantifying expression in round spermatids

Round spermatid expression data was quantified by summing the total number of reads mapping to Refseq transcripts. Concerning the classification of the expression status, transcripts without any aligned reads were classified as “not detected”, and the remaining transcripts were classified on the basis of increasing expression values into three equally sized groups termed “low, “medium” and “high”.

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 arange 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.

GO-term analysis

GO-term analysis was performed by using Bioconductor package topGO[44]. Enrichment tests were done by using Fisher`s exact test (Supplementary Table 1).

Variance partitioning analysis

Variance partitioning analysis was performed via using R package yhat[45]. Unique and combinatorial effects for each variable were obtained by using the function commonalityCoefficients().

Motif finding for histone peaks

7-mer motif frequencies in the foreground and background sets were calculated by using Bioconductor package Biostrings. Foreground refers to histone peaks identified via the hidden semi-Markov model approach. As a background, we used CGI that do not intersect any of the peak regions in the foreground set. Motif enrichment (M) and abundance (A) values were calculated as follows: $M=\log2((fg_Z/fg_{Total}*\min(fg_{Total}, bg_{Total}))+pscnt)-\log2((bg_Z/bg_{Total}*\min(fg_{Total},bg_{Total}))+pscnt)$, and $A = (\log2((fg_Z/fg_{Total}*\min(fg_{Total},bg_{Total}))+pscnt)+\log2((bg_Z/bg_{Total}*\min(fg_{Total},bg_{Total)})+pscnt))*0.5$, where fg_Z is the number of counts of motif Z in the foreground, fg_{Total} is the total number of all motifs in foreground, bg_Z is the number of counts of motif Z in background, bg_{Total} is the total number of all motifs in the background, and pscnt is a constant number (8). Results were visualized in a motif enrichment abundance plot (MA plot) and the top 20 motifs enriched in the foreground set were displayed.

Defining expression for oogenesis or early embryogenesis

Expression data from[33] was processed as described in[5]. The first developmental stage in which expression was measured was used for the classification of expression as “maternal”, “2-8 cell” or “blastocytst”. The distinction between maternal and embryonic expression was made by comparing the expression levels in early embryos treated or untreated with α-amanitin.

Supplementary References

35 Hackenberg, M. *et al.* CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics* **7**, 446, doi:10.1186/1471-2105-7-446 (2006).

36 Molaro, A. *et al.* Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029-1041, doi:10.1016/j.cell.2011.08.016 (2011).

37 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).

38 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**, 1008-1022, doi:10.1016/j.mod.2005.04.009 (2005).

39 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 Mol Biol* **558**, 299-321, doi:10.1007/978-1-60761-103-5_18 (2009).
40 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**, 8602-8612 (2000).
41 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102. Article published online before print in May 2002 (2002).
42 J., O. C. & S., H. Hidden Semi Markov Models for Multiple Observation Sequences: The mhsmm Package for R. *Journal of Statistical Software* **39**, 1-22 (2011).
43 H., P., P., A., R., G. & S., D. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.26.2.
44 Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-1607, doi:10.1093/bioinformatics/btl140 (2006).
45 K., N. & K., R. J. yhat: Interpreting Regression Effects. R package version 1.0-5. (2012).

Chapter 2: Results

2.4 Manuscript in preparation:

Transcript and histone modification dynamics towards sperm development

Introduction

Germ cells carry critical information for the proper development of the embryo. In mammals, two highly differentiated gametes sperm and oocyte highly differ in their capacity to convey information to the next generation. The oocyte is seen as the main source of transcripts and proteins required for early embryonic development given the large pools of mRNAs that are generated during early oogenesis and their tightly regulated control (Kang and Han, 2011). On the other hand, sperm is thought to have a very little potential to transmit information as the majority of its histones are replaced by protamines during later stages of spermatogenesis and it is in a very compact structure. Nevertheless, recent genome-wide studies on human and mouse sperm showing the specific association of retained histones and their modifications with developmentally important loci pointed out the capacity of sperm to regulate early embryonic development (Brykczynska et al., 2010; Hammoud et al., 2009).

In addition to a potential function of the chromatin states of sperm, studies especially performed in human sperm showing the differential presence of certain RNAs between fertile and infertile man suggested that sperm born transcripts could have an effect on embryonic development or could be used as diagnostic markers (Hamatani, 2012).

It is currently believed that after round spermatid stage during differentiation into mature spermatozoa, transcription is shut-down and RNA transcribed in round spermatids is stored by post-transcriptional mechanisms to be translated in elongating spermatids. This argument mostly comes from the studies showing the absence of labeled RNA incorporation in late spermatids (Kierszenbaum and Tres, 1975), high accumulation of transcriptional machinery components in round spermatids compared to other later stages (Schmidt and Schibler, 1995), transcription of the sperm basic proteins such as transition proteins and protamines in round spermatids but their translation in elongating spermatids, and impairment of spermatogenesis with absence of certain RNA-binding proteins (Steger, 2001). Although all this data argues for a significant contribution of the post-transcriptional regulation in proper propagation of spermiogenesis, it does not directly exclude the occurrence of transcription after the round spermatid stage.

In mice, the first wave of spermiogenesis results in the appearance of early round spermatids around 23 days after birth, followed by the development of late round spermatids around the 28^{th} day, followed by nuclear elongation, and finally production of mature spermatozoa around the 35^{th} day. In this study, we investigated

the transcript dynamics during spermiogenesis by performing high-throughput sequencing of RNA in early round, late round, elongating spermatids and sperm. Our results show that overall gene expression levels among the four stages analyzed are static. Nevertheless, the changes in expression of genes and their association with chromatin states highly suggest for ongoing transcriptional activity during late spermiogenesis. In addition, analysis of the content of spermatozoal RNA pool showed a potential function of sperm RNA for early embryonic development.

Results

Genome-wide distribution of transcripts in spermatids and sperm

We analyzed the distribution of the reads in the genome for different cell types. In round and elongating spermatids almost 90 % of the reads map to already annotated Ensembl transcripts. Nevertheless, we determined that in sperm there is a significant amount of reads mapping to the repetitive part of the genome. We prepared RNA-seq libraries for all spermatid samples by removing rRNA from spermatid RNA. However, given the very small amount of RNA present in sperm, we could not perform rRNA depletion on sperm RNA and the majority of the reads for sperm sample map to rRNA (Supplementary Table 1). Nevertheless, in Figure1a, in the repeat part of the genome, rRNA is excluded. Therefore, having more reads at repetitive part of the genome in sperm can not be attributed to rRNA. It is currently difficult to say whether the read count increase we see in the repetitive part of the genome in sperm could be due to new transcription, RNA stability or something completely technical (Supplementary Note 1 and Supplementary Figure 1).

A zoom into known Ensembl transcripts showed that for any given cell type, approximately 95 % of the reads were obtained for the protein-coding genes. In the remaining part of the Ensembl transcripts, we determined that the levels of miscellaneous RNA (misc_RNA), processed transcripts and microRNAs (miRNAs) increase as the cells differentiate into sperm. On the other hand, the levels of pseudogenes and small nuclear RNAs (snRNAs) go down during this differentiation process. Nevertheless, given the fact that annotation of the non-coding part of the genome is still not fully characterized, relative increase or decrease in the levels of non-coding transcripts need to be treated with caution.

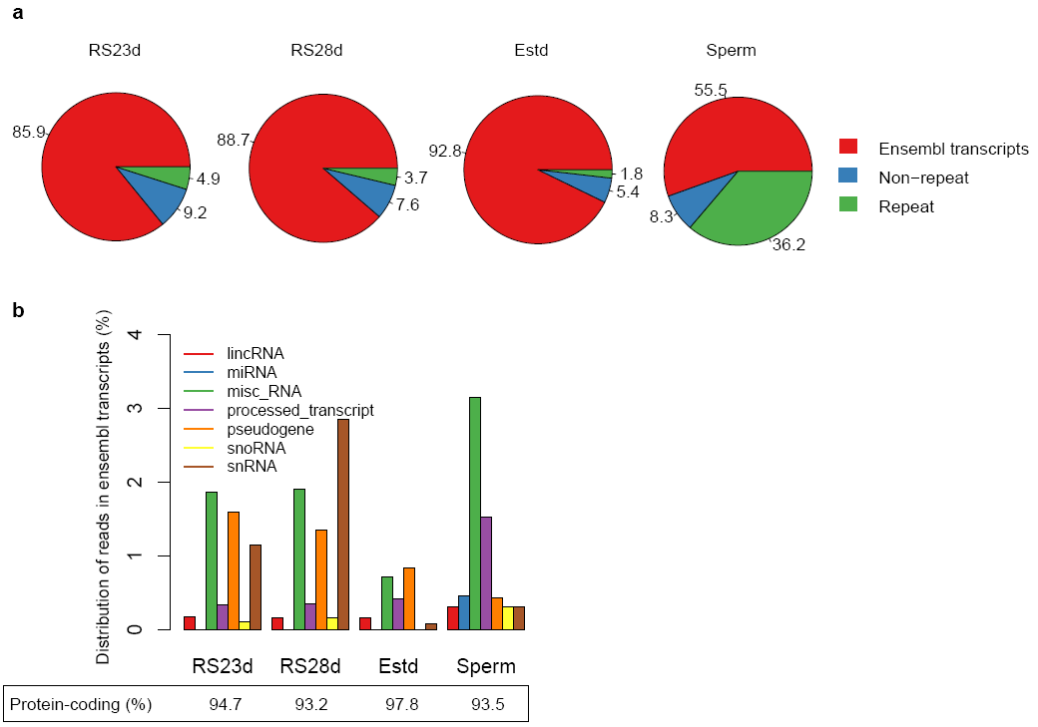


Figure 1: Classification of RNA-seq reads per sample. (A), Pie-charts show the distribution of reads in Ensembl annotation, repetitive and non-repetitive parts of the genome in 23 day round spermatids (RS23d), 28 day round spermatids (RS28d), elongating spermatids (Estd), and sperm, respectively. **(B)**, Zoom into Ensembl transcripts. Barplot shows the distribution of reads in major Ensembl transcripts.

Gene expression dynamics during spermiogenesis

After determining that the great majority of reads map to protein coding-genes, we checked differential gene expression among the 4 stages for Ensembl protein-coding genes. Initially, pair-wise scatter plot comparisons showed that transcript levels for the different samples and their replicates are highly correlating; 23 day and 28 day round spermatids being most similar, and towards differentiation into sperm a slight changes in expression (Supplementary Figure 2). Performing a k-means clustering showed that 80% of the transcripts are stable over the 4 stages (class a). Moreover, we identified transcripts with higher levels in round spermatids (class b), significantly lower levels in sperm (class c), specifically high only in elongating spermatids (class d), and finally with elevated levels in sperm (class e). The presence of increased levels of transcripts in elongating spermatids and sperm stages suggests that the condensing phase of spermiogenesis may not be transcriptionally inactive as originally thought. Furthermore, early (23 day) and late (28 day) round spermatids have remarkably similar transcript dynamics. In addition, profiles of sample replicates nicely illustrate the reproducibility of our results.

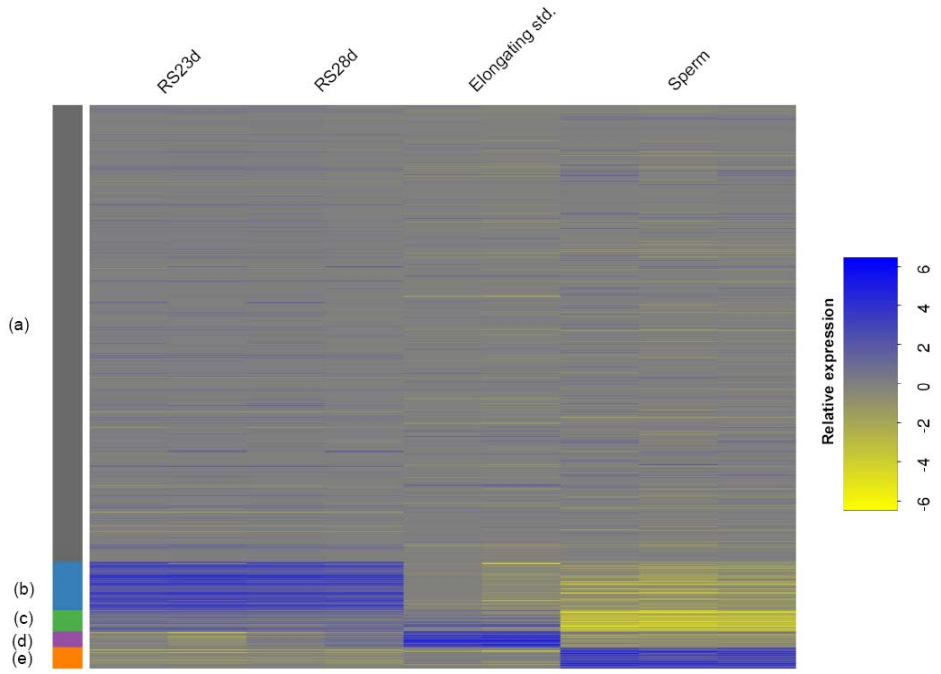


Figure 2: Differential gene expression analysis during spermiogenesis. Heatmaps shows the relative transcript levels (protein-coding genes) for the 4 different cell types (RS23d-2 replicates, RS28d-2 replicates, Elongating std.-2 replicates and sperm-3 replicates) compared the to median level of expression for all samples analyzed.

Transcription in round spermatids

RNA-seq analysis shows the relative abundance of a transcript in a given cell type. To understand whether transcripts we detected in round spermatids are a reflection of ongoing transcription in round spermatids or showing transcriptional activity from the earlier stages, we related RNA levels to RNA Polymerase II (RNAPII) occupancy around transcriptional start sites in late round spermatids. To measure RNAPII, we used a RNAPII antibody which recognizes both Ser5P phosphorylated and unphosphorylated forms of RNAPII. Enrichment level of this form of RNAPII around transcriptional start site of genes has been shown to be a good measure of gene expression in embryonic stem cells (Brookes et al., 2012). We identified that the great majority of the genes actually were not associated with RNAPII peak (Figure 3A). Genes which are started to be expressed from spermatogonia or spermatocyte stage onwards (as determined by the analysis of the microarray data for spermatogenic gene expression) (Brykczynska et al., 2010; Namekawa et al., 2006) have major contributions to RNA levels measured in round spermatids, which highly suggests that expression levels of the genes without

RNAPII are reflecting a history from previous stages (Figure 3B). In contrast, history from earlier stages has minor and irregular contributions to RNA levels for the genes with RNAPII (Supplementary Figure 3).

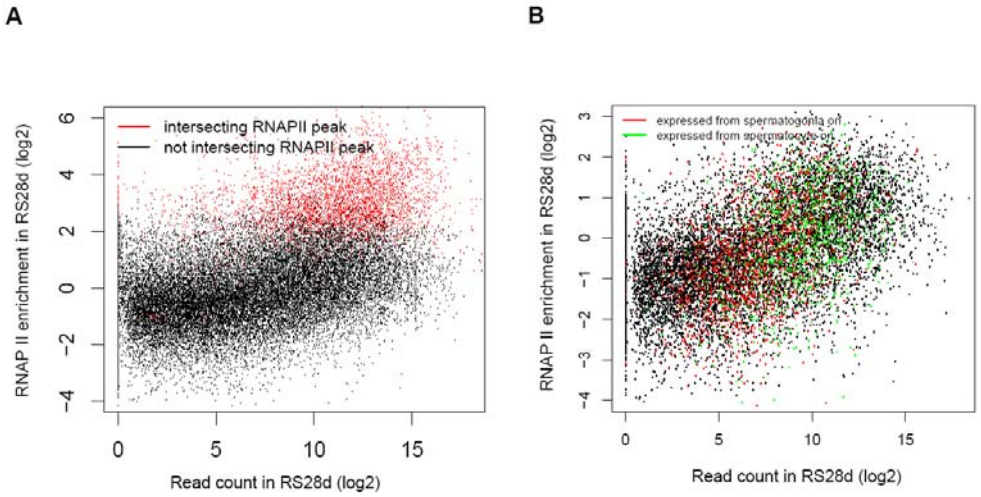


Figure 3: Comparison of RNA levels and RNAPII enrichments in 28 day round spermatids. (A) Scatter plot shows the correlation of transcript levels and RNAPII enrichments. Genes with promoter regions intersecting RNAPII peak are shown in red. (B) Scatter plot shows the RNA level and RNAPII enrichment comparisons only for the genes which are not associated with RNAPII. Red shows the genes which are started to be expressed from spermatogonia stage onwards, and green marks the genes which are started to be expressed from spermatocyte on (Brykczynska et al., 2010).

Annotation of genes according to chromatin states and function in relation to their RNAPII status

After determining that genes with detectable RNA are or are not associated with RNAPII in round spermatids, we analyzed the chromatin states of the genes with and without RNAPII. Our analysis showed that genes with RNAPII in round spermatids are almost devoid of H3K27me3 and they are exclusively marked by H3K4me3. Almost 80% of these genes have high CpG-containing (HCP) or intermediate CpG-containing (ICP) promoters (Figure 4).

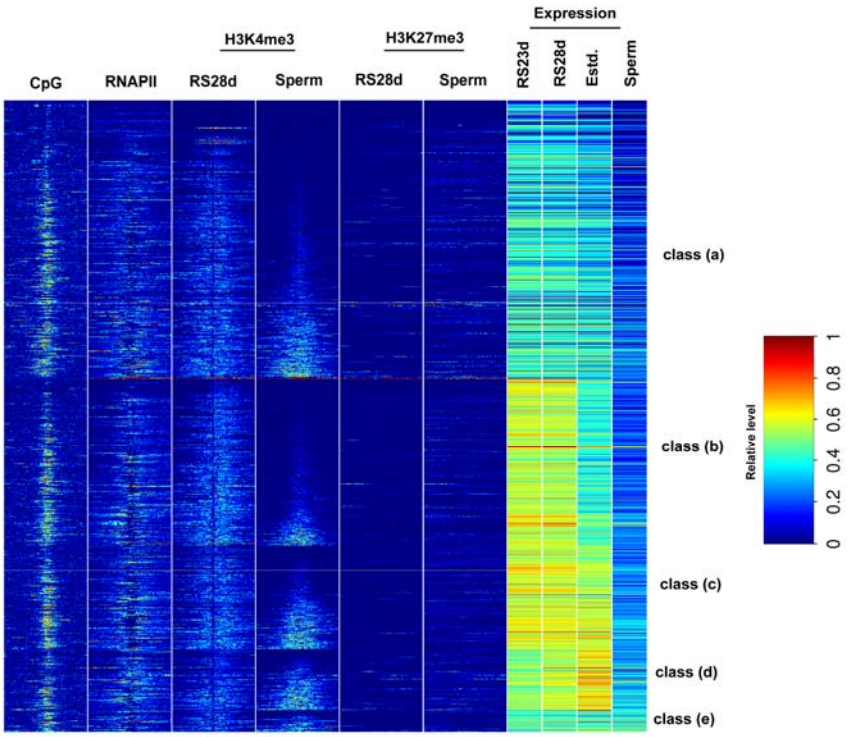


Figure 4: Chromatin states of genes associated with RNAPII. Heatmap of genes illustrating CpG density, RNAPII in 28 day round spermatids, H3K4me3 and H3K27me3 coverage in round spermatids and sperm around TSS (±3kb), and gene expression levels in 23 day round spermatids, 28 day round spermatids, elongating spermatids and sperm, respectively. Feature density shows the scaled read densities from ChIP-seq experiments or relative expression levels. Number of genes in each class is as follows: class a, 1750, class b, 1019, class c, 678, class d, 376, and class e, 170. 1000 genes were randomly selected for visualization.

On the other hand, 85 % of the genes which are not associated with RNAPII are in class (a), very low level expressed or not expressed and approximately 40 % have low CpG-containing (LCP) promoters (Figure 5).

Furthermore, we have identified that differentially expressed genes could be differentially enriched for a functional go term depending on their RNAPII status (Figure 6). Genes which are stably and low level expressed (class a) are not significantly associated with any term except RNA processing (genes having RNAPII). Genes with relatively higher levels of RNA in round spermatids (class b) are involved in metabolic functions and regulation of spermatogenesis. Genes with stable levels of expression in round and elongating spermatids but lower levels in sperm (class c) are associated with spermatid development related functions, and RNAPII associated genes are more enriched for this function.

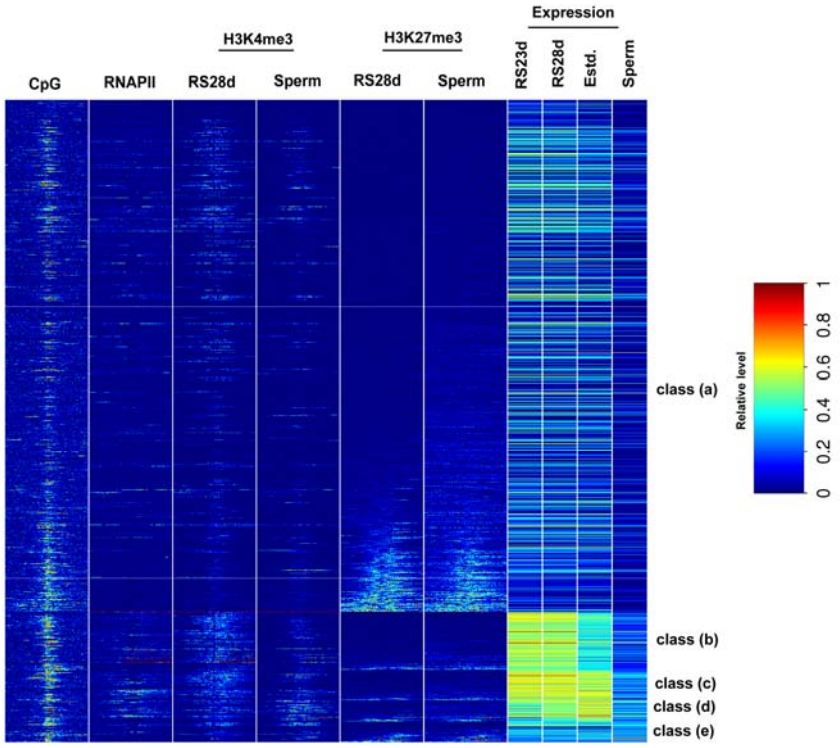


Figure 5: Chromatin states of genes which are not associated with RNAPII. Heatmap of genes illustrating CpG density, RNAPII in 28 day round spermatids, H3K4me3 and H3K27me3 coverage in round spermatids and sperm around TSS (±3kb), and gene expression levels in 23 day round spermatids, 28 day round spermatids, elongating spermatids and sperm, respectively. Feature density shows the scaled read densities from ChIP-seq experiments or relative expression levels. Number of genes in each class is as follows: class a, 27383, class b, 2192, class c, 989, class d, 628, and class e, 1276. 1000 genes were randomly selected for visualization.

The most remarkable difference between RNAPII associated versus non-associated genes is for the genes which have higher levels of RNA in elongating spermatids (class d). For this class, genes which are not associated with RNAPII show almost no enrichment for spermatogenesis related functions, which suggests that those genes need to be transcribed from round spermatids on. Genes with higher levels of RNA in sperm (class e) are involved in RNA processing functions or spermatid development if they are marked by RNAPII in round spermatids. Nevertheless, genes in class (e) which are not marked by RNAPII in round spermatids are associated with functions involved in cell activation and embryonic development. At this point, it could be proposed that these genes get RNAPII later in spermiogenesis and their expression might regulate early embryonic development.

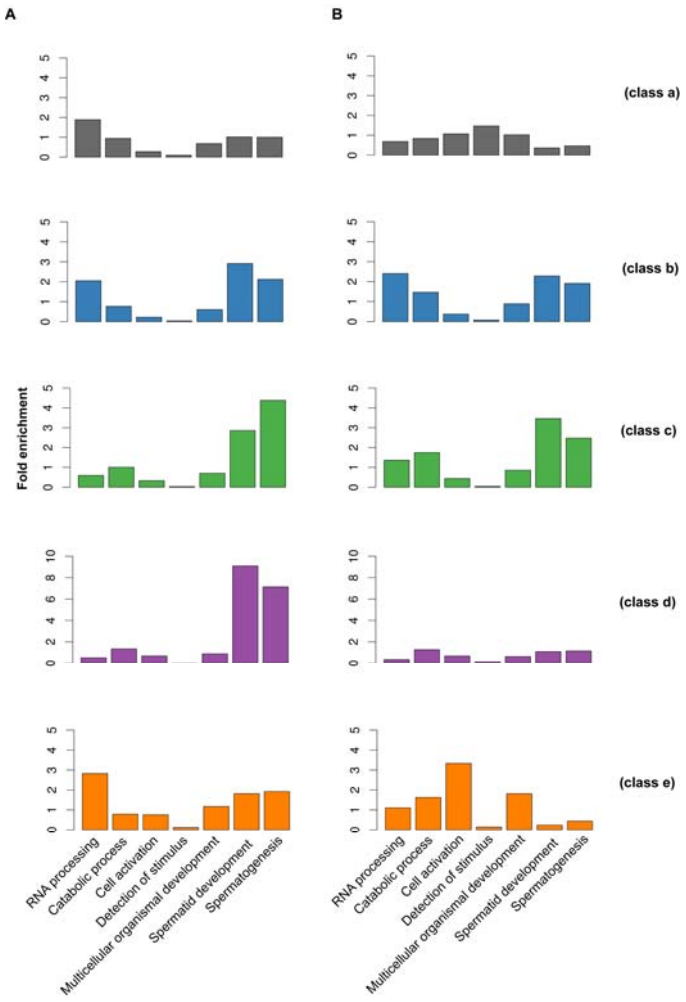


Figure 6: Go term analysis according to differential gene expression states and RNAPII association. (A) Genes with RNAPII. (B) Genes without RNAPII. Colors refer to classes identified in differential gene expression analysis (Figure 2).

H3K4me3 and gene expression dynamics during spermiogenesis

Our chromatin analysis for the different groups of genes showed that changes in H3K4me3 levels highly relates to gene expression changes during spermiogenesis (Figure 4 and 5). To analyze this observation in more detail, we related change in H3K4me3 and expression levels separately for each differentially expressed gene group and their RNAPII status. Enrichment level of H3K4me3 around promoter regions of genes has been identified to be a good predictor in estimation of transcriptional activity from chromatin states in the ENCODE project consortium (Dunham et al., 2012). We identified that the genes which are associated with RNAPII in round spermatids and have lower levels of expression in sperm (class b or

class c) lose their expression levels from round spermatids towards differentiation into sperm most likely as a result of shut-down of transcription as it is reflected in change in H3K4me3 levels. Nevertheless, genes which do not have RNAPII in round spermatids lose their expression towards differentiation into sperm most probably as a result of a post-transcriptional mechanism such as active degradation. For those genes changes in expression levels do not correlate to change in H3K4me3 levels (Figure 7A and B).

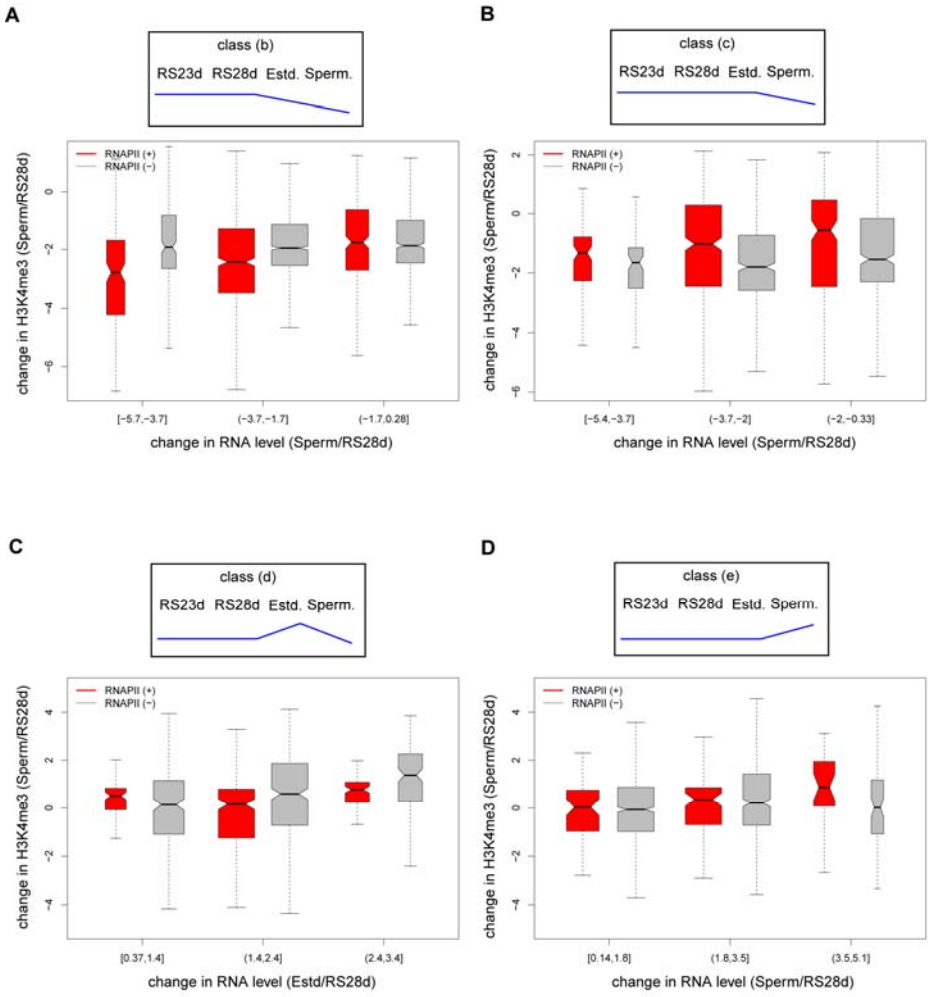


Figure 7: Relationship of change in H3K4me3 levels to change in RNA levels during spermiogenesis. For each class of the genes, changes in RNA levels were binned into 3 groups. Boxplots show the change in H3K4me3 levels for the different groups of change in RNA levels. Red color specifies the genes which are associated with RNAPII in round spermatids. Gray color specifies the genes which are not associated with RNAPII in round spermatids. At the top of the boxplots, an image representing the gene expression class is shown. (A) class (b) genes. (B) class (c) genes. (C) class (d) genes. (D) class (e) genes.

Changes in the expression level of the genes with higher levels in elongating spermatids (class d) correlate to changes in H3K4me3 for both RNAPII associated and non-RNAPII associated genes. This finding highly suggests that there is ongoing transcriptional activity in elongating spermatids and part of this could be achieved by *de novo* marking of some genes by RNAPII in elongating spermatids (Figure 7C). Genes which are upregulated in elongating spermatids but not associated initially with RNAPII spermatids are involved in ketone body metabolism (Fisher's exact test P-value=0.0016). Ketone bodies have been shown to be involved in sperm motility and a testis specific isoform of succinyl CoA transferase (SCOT-t) localizes to mitochondria of elongating spermatids and sperm (Tanaka et al., 2004). In this respect, *de novo* transcription of ketone metabolism genes in elongating spermatids might be required for the energy supply of sperm.

Regarding the genes with elevated levels in sperm (class e), the relationship between gene expression and H3K4me3 levels is a bit more complicated (Figure 7D). In this case, increased level of expression in sperm might be both achieved by *de novo* targeting of RNAPII or RNA stability.

Potential function of sperm RNA for embryonic development

We have performed RNA expression analysis during spermiogenesis and confirmed the presence of RNA in sperm. At this point, one of the critical questions is that whether the transmission of RNA to early embryo could have a function. To understand this, we first determined the genes with transcripts detected in 2-cell embryos, but not expressed or low level expressed in oocytes and at the same time are not de novo transcribed in 2-cell embryos (α-amanitin insensitive) (Posfai et al., 2012). Then, we checked the expression level of those genes in sperm. If sperm RNA is potentially regulating embryonic development, it will more likely to have an effect with paternal transmission only. We determined that almost 80% of the transcripts which are not maternal, not zygotically transcribed but present in early embryo had "medium" or "high" expression levels in sperm. Among those genes, we found a significant number involved in chromatin organization (Fisher`s exact test P-value=5.8E-09). These findings highly suggest for a potential function of sperm RNA in early embryonic development (Supplementary Figure 4). Nevertheless, given the little amount of RNA sperm could have, comparison of absolute levels of RNA for the respective genes in oocytes versus sperm is required to support this hypothesis.

Discussion

Sperm is produced as a result of complex chromatin remodeling and nuclear reorganization events. Nevertheless, still the details of these events and control mechanisms are not fully characterized. Because of its compact chromatin structure, sperm is thought to be inert. Here, we investigated the transcript dynamics during later stages of spermatogenesis to understand gene expression regulation in different stages of spermatids and mature sperm and potential of sperm RNA pool.

We have determined that the transcriptome of early round, late round and elongating spermatids are very similar. Sperm transcriptome showed an increased abundance for the transcripts associated with repetitive elements. Nevertheless, significance of this increase needs to be determined. Overall, transcript levels for the protein-coding part of the genome are quite stable. In addition to this stable expression pattern, our analysis revealed that genes especially the ones involved in regulation of spermatogenesis showed dynamic expression patterns.

We identified that actually the majority of the genes with detectable levels of RNA in round spermatids were not associated with RNAPII and detected RNA most probably reflect transcriptional history from spermatocytes and spermatogonia. Furthermore, genes depending on their association with RNAPII showed differential chromatin states and functional annotation.

Although transcription is thought to be shut-down after round spermatid stage, our results for the first time strongly argue that there is ongoing transcriptional activity during later stages of spermatogenesis. We have determined that changes in transcript levels highly relate to changes in H3K4me3 levels. H3K4me3 has been determined to be one of the main histone modifications in the analysis inferring transcriptional activity from chromatin states (Dunham et al., 2012). Therefore, we propose that change in chromatin states is reflected in transcriptional outputs during spermiogenesis. This hypothesis could be further supported by experiments showing the association of RNAPII with genes in elongating spermatids and maybe even in sperm. Nevertheless, compact structure of elongating spermatids and sperm could bring some technical difficulties to fixed ChIP experiments in these cells. In addition, labeling of RNA for instance by using 5-ethynyl uridine (EU) in specific stages of spermiogenesis and analysis of labeled RNA could show ongoing transcriptional activity in a more quantitative manner. This approach requires injection of labeling molecule into testis at specific time points and careful characterization of turnover rates from clean population of cells.

Presence of RNA has been already shown in human sperm. Actually many studies related the content of human sperm RNA pool to fertility. A study by profiling RNA from 24 fertile donors showed that although there might be slight variations

within individual donors, overall fertile donors had stable abundance of certain transcripts (Lalancette et al., 2009). In addition, comparison of fertile donor transcripts with the ones from teratozoospermic individuals indicated the differential presence of transcripts involved in ubiquitin-proteosome pathyway between fertile versus teratozoospermic individuals (Platts et al., 2007). Here, we also showed that mouse sperm has a defined RNA content as it has been indicated by the three replicates we have. Importantly, transcripts with elevated levels in sperm are involved in functions related to cell activation and early embryonic development. Furthermore, the transcripts which are more likely to be transmitted paternally are involved in chromatin organization which could have an impact also on the genome organization of early embryo. Still the remaining issue with sperm transmitted RNA is its relative effect compared to the large pool of maternal transcripts. Human sperm has been determined to have 10-400 fg of RNA (Hamatani, 2012). In contrast, human oocyte is estimated to contain 1500 pg of RNA (Elder and Dale, 2011). Although there seems to be a huge difference between oocyte and sperm RNA amounts, maybe quality of RNA rather than the actual quantity might be more important in certain cases. One of the best examples to this situation might be the unusual accumulation of *Kit* m-RNA in sperm, which causes a degradation product and lead to a paramutation at *Kit* locus (Rassoulzadegan et al., 2006).

Materials and Methods

Isolation of spermatids and sperm

Mouse sperm were collected from C57BL/6J mice by using swip-up procedure as described (Brykczynska et al., 2010). To isolate early or late round spermatids, testicular cells were prepared from 23 day or 28 day C57BL/6J mice. Isolated cells were subjected to Hoechst (Invitrogen, catalog number 33342) staining for 30 min at 37°C and round spermatids were collected via Fluorescent Activated Cell Sorter (FACS) with 90 - 95 % purity. Elongating spermatids were isolated by collecting testicular cells from a transgenic mouse line which expresses Protamine1-GFP. Elongating spermatids were isolated based on GFP and Hoechst staining intensity via FACS with 90 % purity.

RNA isolation and library preparation

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. Before preparing RNA-seq libraries, except for sperm

RNA (given the very low amounts of RNA that could be isolated), rRNA from all spermatid sample RNAs 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).

RNAPII ChIP in round spermatids

28 day round spermatids were collected by FACS (around 10 million cells). Fixed ChIP was performed as described in (Weber et al., 2007) with slight modifications. Fixed round spermatid chromatin was sonicated by using Bioruptor in 10 cycles with 30 sec on and 45 sec off settings. Pre-cleared chromatin was incubated 5 ul of RNAPII antibody (Covance CTD4H8) overnight. Then, antibody-chromatin complex were coupled to Dynabeads 280 (Invitrogen). After elution of beads and reversal of crosslink, immunoprecipitated DNA was isolated by phenol/chloroform extraction.

Processing and alignment of the reads

Filtering, alignment and processing of the reads for RNA-seq were done as described (Stadler et al., 2011).

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). All annotated transcript analysis is based on Ensembl65 gene annotation (Mus_musculus.NCBIM37.65.gtf). For protein-coding genes, differential gene expression analysis was performed by selecting the transcript with most extreme coordinates per gene.

Analysis of genome-wide distribution of reads

Number of weighted alignments are assigned to each group according to the following hierarchy: rRNA, Ensembl65 transcript, repeat and non-repeat. RNA-seq library preparation for sperm did not include rRNA depletion. Therefore, in sperm approximately 95 % of reads map to rRNA. Pie chart showing the distribution of the reads in the genome (Figure 1A) is done by excluding the reads mapping to rRNA to make all spermatid and sperm samples comparable.

Differential gene expression analysis

Differential gene expression analysis (Figure 2) was done for protein-coding Ensembl transcripts by selecting one transcript per gene with the most extreme coordinates. First, simply number of reads mapping to each transcript was counted per sample and counts were normalized for library size differences among the samples. By using the normalized counts, a “pseudosample”, which has the median level of expression for all samples analyzed, is defined. Differential gene expression for any given cell type per gene was calculated in the following way: log2((datNorm + pscnt) / (Med + pscnt)), where datNorm is the normalized read count for the given cell type, Med is the median level of read count in all cell types analyzed, and pscnt is a constant number (8), which was used to stabilize values based on low read counts. 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. Results were visualized by using heatmap.2 function of R (www.r-project.org).

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 a range 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 was quantified as log2 (read count per transcript).

GO-term analysis

GO-term analysis was performed by using Bioconductor package topGO (Alexa et al., 2006). Enrichment tests were done by using Fisher`s exact test.

RNAPII peak finding

RNAPII peak finding was performed by using MACS peak finder (Zhang et al., 2008) with the default parameters except –mfold=5 and –tsize=51.

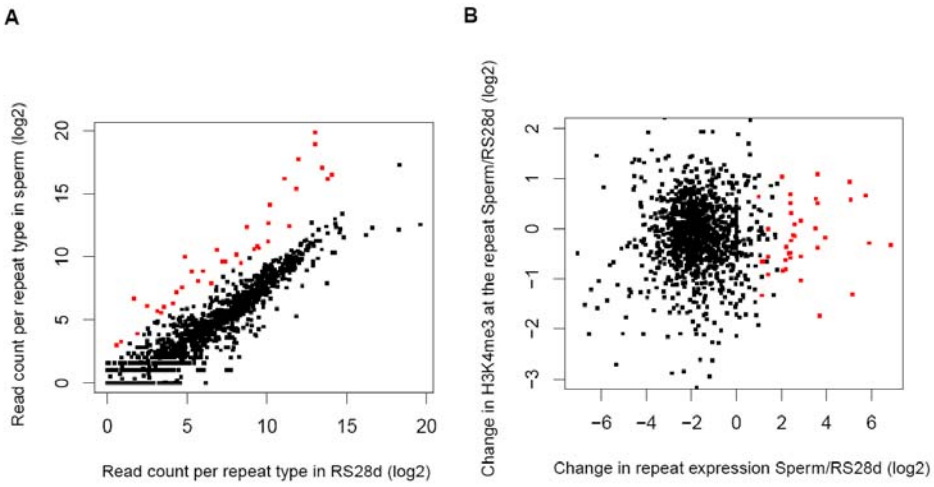
Classification of gene expression in sperm

Sperm expression data was quantified by summing the total number of reads mapping to Ensembl transcripts. Concerning the classification of the expression status, transcripts without any aligned reads were classified as “not detected”, and the remaining transcripts were classified on the basis of increasing expression values into three equally sized groups termed “low, “medium” and “high”.

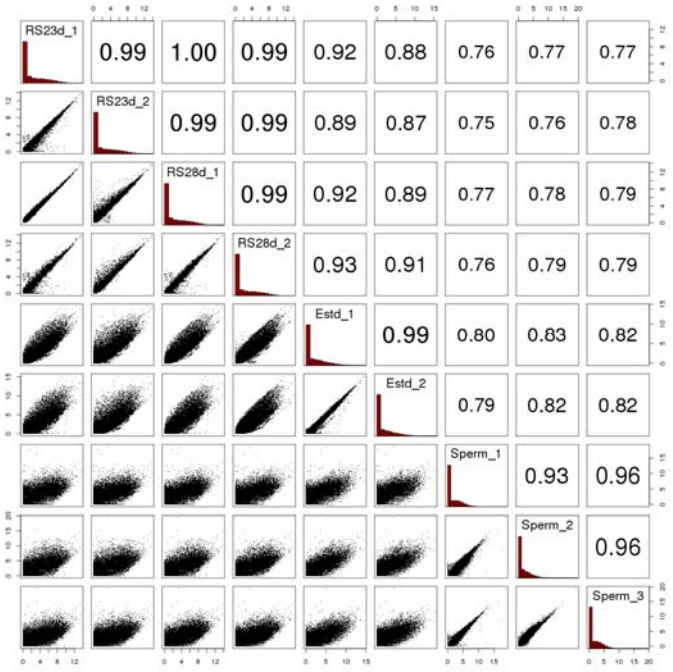
Supplementary Figures, Tables and Notes

Supplementary Note 1

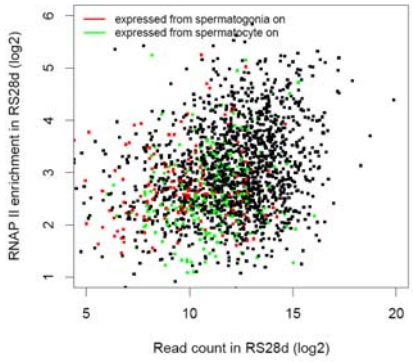
We determined that the proportion of the reads mapping to the repetitive part of the genome is increased in sperm compared to spermatids samples. To tackle this question in more depth, we compared read counts per repeat type between 28 day round spermatids and sperm and related this comparison to H3K4me3 levels (as a marker of transcription) in these cell types. Pair-wise comparison of read counts showed that there are repeat elements with more abundant read counts in sperm (these include CG-rich repeats, tRNA repeats, some LINEs and SINEs) (Supplementary Figure 1a). Nevertheless, for these repeat elements a corresponding increase in H3K4me3 levels was not clearly apparent (Supplementary Figure 1b).



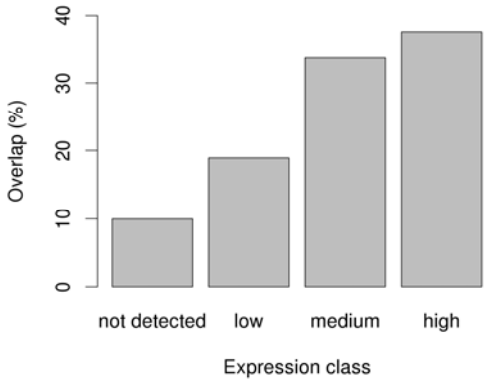
Supplementary Figure 1: Repeat element expression in 28 day round spermatids and sperm. (A) Comparison of read counts per repeat type (repeat masker rename) in round spermatids and sperm. Repeats deviating from the regression line (residuals >3) are shown in red. (B) Comparison of change in repeat expression to change in H3K4me3.



Supplementary Figure 2: Comparison of transcript levels in spermatids and sperm. Pair-wise scatter plot shows the correlation of read counts (log2) across the samples for Ensembl protein-coding transcripts.



Supplementary Figure 3: Comparison of RNA and RNAPII levels in 28 day round spermatids for the genes with RNAPII. Red shows the genes which are started to be expressed from spermatogonia stage onwards, and green marks the genes which are started to be expressed from spermatocyte on (Brykczynska et al., 2010).



Supplementary Figure 4: Relation of gene expression levels in sperm to embryonic gene expression. First the genes with transcripts detected in 2-cell embryos, but not expressed or low level expressed in oocytes and at the same time are not de novo transcribed in 2-cell (α-amanitin insensitive) were determined (Posfai et al., 2012). Barplot shows the expression class of the resulting genes in sperm.

Supplementary Table 1: Table shows the read counts mapping to different annotations for different samples and their replicates.

class	RS23d_1	RS23d_2	RS28d_1	RS28d_2	Estd_1	Estd_2	Sperm_1	Sperm_2	Sperm_3
rRNA+	237666	1146333	291127	800790	231884	327046	33473706	118673348	145713708
rRNA-	48221	145821	49542	121908	70738	158671	137	439	1093
Ensembl+	25147815	108603180	25332883	107995312	29177922	100094143	1256429	2528708	3231193
Ensembl-	139986	564579	153982	447836	127151	326343	13202	25624	31889
non-repeat	2918833	11664220	3256594	9310394	2425081	5843058	205599	381587	403205
repeat	1437148	6265262	1675088	4549303	948137	1965864	204506	1667408	2092001
Total	29929669	128389395	30759216	123225543	32980913	108715125	35153579	123277114	151473089

References

Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600-1607.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N.*, et al.* (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157-170.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schubeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol *17*, 679-687.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R.*, et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Elder, K., and Dale, B. (2011). In-Vitro Fertilization.

Hamatani, T. (2012). Human spermatozoal RNAs. Fertil Steril *97*, 275-281.

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.

Kang, M.K., and Han, S.J. (2011). Post-transcriptional and post-translational regulation during mouse oocyte maturation. BMB Rep *44*, 147-157.

Kierszenbaum, A.L., and Tres, L.L. (1975). Structural and transcriptional features of the mouse spermatid genome. J Cell Biol *65*, 258-270.

Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., and Krawetz, S.A. (2009). Identification of human sperm transcripts as candidate markers of male fertility. J Mol Med (Berl) *87*, 735-748.

Namekawa, S.H., Park, P.J., Zhang, L.F., Shima, J.E., McCarrey, J.R., Griswold, M.D., and Lee, J.T. (2006). Postmeiotic sex chromatin in the male germline of mice. Curr Biol *16*, 660-667.

Platts, A.E., Dix, D.J., Chemes, H.E., Thompson, K.E., Goodrich, R., Rockett, J.C., Rawe, V.Y., Quintana, S., Diamond, M.P., Strader, L.F., and Krawetz, S.A. (2007). Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum Mol Genet *16*, 763-773.

Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T.C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M.*, et al.* (2012). Polycomb function during oogenesis is required for mouse embryonic development. Genes Dev *26*, 920-932.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Schmidt, E.E., and Schibler, U. (1995). High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids. Development *121*, 2373-2383.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Steger, K. (2001). Haploid spermatids exhibit translationally repressed mRNAs. Anat Embryol (Berl) *203*, 323-334.

Tanaka, H., Takahashi, T., Iguchi, N., Kitamura, K., Miyagawa, Y., Tsujimura, A., Matsumiya, K., Okuyama, A., and Nishimune, Y. (2004). Ketone bodies could support the motility but not the acrosome reaction of mouse sperm. Int J Androl *27*, 172-177.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Chapter 3: Concluding remarks and discussion

3.1 DNA sequence-encoded regulation mechanisms

As I already discussed in introduction the term “epigenetics” changed definitions over the years with the new discoveries. Currently, a clear definition which is interpreted in the same manner by everybody is not available. Nevertheless, I would like to discuss about “independency of epigenetics from DNA sequence” which is present in the recent definition of epigenetics - “the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence”- (Wu and Morris, 2001).

In my project studying the molecular determinants of nucleosome retention in mouse sperm, we identified that the CpG dinucleotide is the most important factor specifying the genomic regions keeping nucleosomes. Furthermore, CpG specific nucleosome retention is histone variant type sensitive. We see retention of H3.3 containing nucleosomes but depletion of H3.1/H3.2 containing nucleosomes at CGIs. If the transmission of sperm histones and their modifications are potentially regulating early embryonic gene expression, then we could argue that this regulation will be sequence-specific. As it will be discussed below, actually, there are already many known examples showing how important DNA sequence is in directing “epigenetic” regulation mechanisms.

DNA in a eukaryotic cell is wrapped around histone proteins to form the fundamental unit of chromatin, nucleosomes (Luger et al., 2012). Nucleosomes could be still thought as the primary structures which change the nuclear environment with interaction with other protein/RNA complexes. Even this primary level of organization depends on DNA sequence (Kaplan et al., 2009; Ramirez-Carrozzi et al., 2009; Tillo and Hughes, 2009).

DNA methylation is currently being regarded as a system which could be truly called “epigenetic”, given the fact that CpGs are symmetrically methylated and such DNA methylation pattern is maintained during cellular division (Goll and Bestor, 2005). However, there have been several studies which question the independency of DNA methylation from DNA sequence. A recent study showed that DNA methylation patterning in the mouse genome is mainly shaped by transcription factor binding, and pattern of DNA methylation for a given cell type is also dependent on cell type specific expression of transcription factors (Stadler et al., 2011). Furthermore, several studies showed that Kruppel-associated box-containing zinc finger protein Zfp57 was involved in maintenance of DNA methylation at imprinted loci via recognizing a methylated hexanucleotide sequence (Li et al., 2008; Quenneville et al., 2011; Zuo et al., 2011).

In the context of deposition of histone marks, H3K4me3 is deposited at almost all of high CpG-content promoters of the genes independent of their expression (Greer and Shi, 2012; Zhou et al., 2011). In *Drosophila*, Polycomb group proteins (PcG) binds to Polycomb response elements (PREs) within the genomic sequence. In mammals specific DNA elements required for PcG binding are only identified for a few loci (Sing et al., 2009; Woo et al., 2010), and genome-wide studies showed that PcG was mainly targeted to CpG islands (Lanzuolo and Orlando, 2012). A recent study actually demonstrated that CXXC-domain containing lysine demethylase KDM2B targets PRC1 to CpG islands (Farcas et al., 2012).

All the examples above highlight the importance of DNA sequence in many genome regulation mechanisms (used by many people as “epigenetic”). Personally, I think that DNA in a eukaryotic cell is in interaction with histones and other protein complexes, and depending on the presence/concentration of all these components, different regulation programs could arise. In this sense, definition and interpretation of epigenetics should be revisited.

3.2 Molecular mechanism behind specific nucleosome retention versus eviction

Differentiation of haploid round spermatids into sperm consists of highly complicated chromatin remodeling events. At the end of these sophisticated remodeling processes, a genome packing structure arise which is actually not present in any other cell types. Although a number of studies showed the role of incorporation of histone variants, global hyperacetylation of histones, and some other chromatin proteins in remodeling of the sperm chromatin (Gill et al., 2012), still the details of the remodeling events and the actual role of already identified proteins/complexes remains as a mystery. For instance, transition proteins are thought to first replace the histones and finally replaced by protamines. Nevertheless, this argument currently mainly relies on the dynamics of these proteins levels during spermiogenesis. Regarding the role of chromatin remodelers in this process, actually there are not many studies addressing this question. I checked the expression of main chromatin remodelers in several stages of spermatids and sperm by using RNA-seq data. I found out that based on gene expression data it is a bit difficult to say which chromatin remodeler might be involved in organization of the sperm chromatin based on the fact that many of them are expressed during spermiogenesis.

In my project, we mainly determined the principles specifying selective nucleosome retention versus eviction from a genomics perspective. Our results are

not directly linked to the molecular mechanism of this process. Nevertheless, our genomic approach provided important hints about how the selective remodeling process is achieved, which could enable the design of functional experiments for the future.

We have identified that overall selective nucleosome retention occurs as a result of combinatorial action of many factors. However, the major finding from my project is that nucleosomes are retained selectively at CpG islands. The big question is how this sequence specific packaging is achieved. We have determined that in round spermatids, there is high nucleosome turnover as a result of high eviction of canonical H3 variants H3.1/H3.2 at CpG islands. These sites correspond to the regions where we have the high enrichment of the replacement variant H3.3 in sperm. In this respect, I propose that specific enrichment of the nucleosomes at CpG islands is coupled to a mechanism which does not evict nucleosomes at those sites because of dynamic chromatin environment and accessibility issues, resulting in H3.3 fill in at CpG islands. The same mechanism enables H3.1/H3.2 retention at CpG islands most probably linked to low turnover of H3K27me3 marked nucleosomes at CpG islands. This model assumes that H3.3 and H3.1/H3.2 are present at the same CpG rich regions occasionally. Nevertheless, as the majority of H3.1/H3.2 is replaced with H3.3 in round spermatids, CpG specific retention of nucleosomes in sperm is best reflected with H3.3 retention.

One of the main questions concerning this dynamic chromatin environment-coupled retention of nucleosomes at CpG islands is that whether this is a general mechanism which means that presence of any transcription/ chromatin factor creates a dynamic chromatin environment and protection against exchange with protamines is achieved passively. Alternatively, the presence of a specific CpG binding protein could actively prevent the replacement of nucleosomes by protamines. Even in an another theory, protamines may not simply have affinity for CpG-rich sequences and do not remodel nucleosomes at CpG islands. These possibilities might be clarified by the identification of protein complexes interacting with nucleosomes in round and elongating spermatids. In addition, experiments measuring the affinity of the protamines for a range of DNA sequences with different composition might be designed.

In addition to specific retention of nucleosomes at CpG islands, occasionally incomplete remodeling of nucleosomes by protamines results in random retention of nucleosomes. This random retention theory should be equally true for the retention of both H3.3 and H3.1/H3.2 containing nucleosomes. Nevertheless, as the majority of

H3.1/H3.2 containing nucleosomes is evicted from CpG islands in round spermatids, globally H3.1/H3.2 retention appears as more irregular.

3.3 Potential function of chromatin states brought by sperm in early embryonic development

Sperm has been thought for many years as inert and not carrying any information for embryonic development. This could be mostly attributed to little amount of histone retention and presence of very small amount of RNA in mature sperm.

Upon fertilization protamines are rapidly exchanged by maternally provided histones. Retained histones in paternal genome are not readily detected by immonofluorescence staining. For instance, H3K27me2/3 becomes only visible during first DNA replication (Albert and Peters, 2009). Nevertheless, genome-wide studies in human sperm (Brykczynska et al., 2010; Hammoud et al., 2009) and data from this study in mouse sperm showed the existence of modified histones in mature sperm. This example basically highlights the importance of use of an high-resolution method in making conclusions about a cellular system.

After fertilization information carried by sperm and oocyte encounter a critical decisive process: reprogramming or inheritance? At this point, most of the ideas are based on DNA methylation patterns. Initially, it was thought that genome-wide DNA methylation landscapes were reprogrammed upon fertilization. Nevertheless, a number of genome-wide studies showed the existence of escapers from this reprogramming process (Borgel et al., 2010; Kobayashi et al., 2012; Smallwood et al., 2011; Smith et al., 2012). For chromatin marks, reprogramming versus inheritance issue is more under debate since it is very difficult to perform ChIP experiments in oocyes and early embryos and compare histone modifications to the ones obtained with sperm given the little amount of material that could be obtained with early embryos and oocytes.

Several studies investigating the role of paternal epigenetic inheritance pointed out the potential role of sperm-transmitted information to the next generations although the molecular mechanisms linking the epigenome of sperm to the phenotypes observed are mostly unclear (Carone et al., 2010; Zeybel et al., 2012).

We have identified that histones are retained in sequence and histone variant specific manner in mouse sperm. Although, we do not have functional data yet, we identified that the majority of the genes which have repressive H3K27me3 were never expressed during early embryogenesis, on the other hand, genes which are marked by active H3K4me3 in sperm were associated with embryonic gene activity.

Furthermore, histone modification states in sperm are highly similar to histone modifications identified in embryonic stem cells. These data highly suggests that retained histones in sperm are involved in the regulation of gene activity in early embryonic development.

We have determined that histone H3 variants are differentially retained in mature sperm. In this respect, transmission of information on H3.3 and H3.1/H3.2 might be highly different. We could say that H3.3 has more potential to transmit epigenetic information than H3.1/H3.2 and the genomic regions passing information on H3.3 to the next generation will be more specific (mostly CpG islands). However, transmission of H3.1/H3.2 containing nucleosomes might be more subject to variegation as its retention is most probably connected to incomplete remodeling events.

Transmission of specific chromatin structure of the sperm to the early embryo might be critical for the correct patterning of embryogenesis. Although it is not highly efficient, normally round spermatid injection (ROSI) into oocyte gives rise to offspring without the need of a mature sperm. It has been shown that embryos obtained with ROSI have abnormal levels of DNA methylation at the end of first DNA replication compared to embryos obtained with intracytoplasmic sperm injection (ICSI). It was identified that H3K9me3 was present in centromeric heterochromatin region in round spermatids but not in mature sperm. Given the close association between DNA methylation and H3K9me3, it has been suggested that presence of H3K9me3 might be the cause of abnormal DNA methylation in early embryos obtained with ROSI (Kishigami et al., 2006). Fittingly, we have observed that H3K9me3 is not retained in mature sperm. It mainly marks the repetitive part of the genome in round spermatids (data not shown). Overall, these findings highly suggest that improper retention of nucleosomes in mature sperm actually can impair proper chromatin organization of the early embryo.

Although it is not completely clear, a study identified that in infertile men histones were more randomly retained and levels of H3K4me3 and H3K27me3 at developmentally important loci were slightly reduced compared to fertile men (Hammoud et al., 2011). With respect to our finding that histones are retained in histone variant specific manner, in future it would be interesting to check whether differential histone variant retention is associated with fertility status of individuals.

In conclusion, currently functional data which shows the necessity and function of retained nucleosomes for early embryonic development does not exist. Nevertheless, our results and data from other studies indicate the potential of sperm in this respect. At the end, even if the absence or the presence of a chromatin mark

in sperm would not be critical for the full competence of early embryo, it could have a role in fitness of an organism.

Cited literature

Albert, M., and Peters, A.H. (2009). Genetic and epigenetic control of early mouse development. Curr Opin Genet Dev *19*, 113-121.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schubeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol *17*, 679-687.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D.*, et al.* (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell *143*, 1084-1096.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W.*, et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. elife *1*, e00205.

Gill, M.E., Erkek, S., and Peters, A.H. (2012). Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? Curr Opin Cell Biol *24*, 387-396.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet *13*, 343-357.

Hammoud, S.S., Nix, D.A., Hammoud, A.O., Gibson, M., Cairns, B.R., and Carrell, D.T. (2011). Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. Hum Reprod *26*, 2558-2569.

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.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kishigami, S., Van Thuan, N., Hikichi, T., Ohta, H., Wakayama, S., Mizutani, E., and Wakayama, T. (2006). Epigenetic abnormalities of the mouse paternal zygotic

genome associated with microinsemination of round spermatids. Developmental biology *289*, 195-205.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y.*, et al.* (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet *8*, e1002440.

Lanzuolo, C., and Orlando, V. (2012). Memories from the polycomb group proteins. Annu Rev Genet *46*, 561-589.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell *15*, 547-557.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol *13*, 436-447.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., and Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114-128.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell *138*, 885-897.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet *43*, 811-814.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC bioinformatics *10*, 442.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell *140*, 99-110.

Wu, C., and Morris, J.R. (2001). Genes, genetics, and epigenetics: a correspondence. Science *293*, 1103-1105.

Zeybel, M., Hardy, T., Wong, Y.K., Mathers, J.C., Fox, C.R., Gackowska, A., Oakley, F., Burt, A.D., Wilson, C.L., Anstee, Q.M.*, et al.* (2012). Multigenerational epigenetic adaptation of the hepatic wound-healing response. Nat Med *18*, 1369-1377.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7-18.

Zuo, X., Sheng, J., Lau, H.T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., and Li, X. (2011). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem *287*, 2107-2118.

Acknowledgements

First of all, I would like to thank Antoine for giving me an opportunity to work in his lab and especially in a project which involves integrative data analysis, his valuable support and many discussions during my PhD.

I am grateful to Michael for his great help and supervision with the bioinformatics analysis and also for fruitful discussions which helped me a lot for deciding about my future career..

I would like thank Mizue for establishing chromatin immunoprecipitation experiments in sperm, sharing her expertise with me, and her nice friendship.

I would like to thank my thesis committee for guiding me throughout my PhD and helpful discussions.

Thanks to Lily and Mark for the nice discussions and help in experiments.

Special thanks to all Peters group for the nice and friendly atmosphere in the lab!

I would like to thank Sophie and Tim for great help with preparing our sequencing libraries.

Thanks to Lukas and Dimos for help and a lot of suggestions with data analysis.

Thanks to Hubertus for many FACS experiments..

Thanks to Dirk Schuebeler group for the joint lab meetings..

Thanks to whole FMI for creating a pleasant atmosphere to work..

Last but not least to my family and friends who supported me all the time...

Serap ERKEK

Home Address	Feldbergstrasse 140 4057 Basel Switzerland	**Work Address**	Maulbeerstrasse 66 4058 Basel Switzerland
E-mail Address	serap.erkek@fmi.ch	**Telephone**	0041786851540

Education

2008 - 2013 **PhD "An epigenomic approach to understanding the mechanism of nucleosome retention in mouse spermatozoa"**
Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
Supervised by Prof. Antoine Peters

2006 - 2008 **M.Sc. in Molecular Biology of Genetics**

M.Sc. Thesis: "HANEIN-1, a novel conserved eukaryotic protein ubiquitously expressed in human tissues"
Bilkent University, Ankara, Turkey
Supervised by Assoc. Prof. Dr. Uygar Tazebay

2002 - 2006 **B.Sc. in Chemical Engineering**
B.Sc. Minor in Biology
Middle East Technical University, Ankara, Turkey

Research Experience

PhD Studies **Sequence and histone variant specific packaging of mouse sperm genomic DNA**

- High-throughput sequencing to map nucleosomes and histone variants in mouse sperm
- Modeling ChIP-seq data to identify mechanism involved in specific retention of nucleosomes in sperm

Histone methylation and gene expression dynamics during sperm development

- ChIP-seq analysis to identify histone methylation states in mouse sperm and preceding round spermatids
- RNA-seq analysis to quantify expression during late spermatogenesis

M.Sc. Studies **Characterization of HANEIN-1, a novel conserved eukaryotic protein ubiquitously expressed in human tissues**

- Sequence similarity analysis
- Biochemical assays to characterize HANEIN-1 protein

Teaching Experience

2007 Spring	"Molecular Genetics" course, laboratory assistantship, Bilkent University, Ankara
2008 Spring	"Molecular Biology of the Cell" course, laboratory assistantship, Bilkent University, Ankara

Computational Skills

- R programming language: Analysis of ChIP-seq and RNA-seq data sets

Conferences and Courses

- **TriRhena Chromatin and Transcription Club Meeting**, Freiburg, Germany, 17 October 2012 (talk).
- **FMI Annual Meeting**, Grindelwald, Switzerland, 19 - 21 September 2012 (talk).
- **Where would epigenetics meet systems biology?**, Weizmann Institute, Rehovot, Israel, 03 - 06 June 2012 (poster presentation).
- **Swiss Institute of Bioinformatics (SIB) days**, Biel, Switzerland, 30 - 31 January 2012.
- **Gordon Research Conference on Epigenetics**, Stonehill College, Easton, USA, 7 - 12 August 2011 (poster presentation).
- **9th EMBL Conference on Transcription and Chromatin**, Heidelberg, Germany, 28 - 31 August 2010 (poster presentation).

Honors and Awards

- **Boehringer Ingelheim Fonds PhD fellowship**, August 2009 - September 2011.
- **M.Sc. Bursary TUBITAK** (Scientific & Technical Research Council of Turkey), September 2006 - August 2008.
- **B.Sc. in Chemical Engineering degree honor**: 3rd rank out of 132.

Memberships

- Member of Swiss Institute of Bioinformatics (SIB) PhD Training Network.

Publications

- Erkek S., Hisano M., Liang C.-Y., Gill M., Murr R., Dieker J., Schübeler D., van der Vlag J., Stadler M.B., and A.H.F.M. Peters. Density and methylation state of CpG dinucleotides define histone variant specific retention of nucleosomes in mouse spermatozoa. (Submitted)
- Gill M. E., Erkek S., Peters A. H. F.M. (2012). Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Curr Opin Cell Biol.*, 3, 387 - 96.
- Brykczynska U., Hisano M., Erkek S., Ramos L., Oakeley E. J., Roloff T. C., Beisel C., Schübeler D., Stadler M. B., Peters A. H. F. M. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol.*, 6, 679 - 87.