Genome-wide analysis on content and function of residual nucleosomes in sperm: studies in bovine and human

Inaugural Dissertation
submitted to the
Faculty of Medicine
in partial fulfillment of the requirements
for the PhD-Degree
of the Faculties of Veterinary Medicine and Medicine
of the Justus Liebig University Giessen

by
Yang Yang
of
People's Republic of China

Gießen (2013)

From the Department of Urology, Pediatric Urology and Andrology,

Section Molecular Andrology

Director / Chairman: Prof. Dr. Klaus Steger

of the Faculty of Medicine of the Justus Liebig University Giessen

First Supervisor and Committee Member: Prof. Dr. Klaus Steger

Second Supervisor and Committee Member: Prof. Dr. Rainer Renkawitz

Committee Members:

Date of Doctoral Defense:

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and referenced all text passages that are derived literally from or are based on the content of published or unpublished work of others, and all information that relates to verbal communications. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University of Giessen in carrying out the investigations described in the dissertation.

Gießen, August 2013

_____

Yang Yang

Summary

It is suggested that at fertilization, spermatozoa transfer both genetic and epigenetic information into the oocyte, e.g. the DNA methylation pattern and various histone modifications. However, the relevance of the sperm epigenome for fertilization and embryogenesis is, so far, for the most part unknown.

During spermiogenesis, replacement of DNA-binding histones by protamines is followed by a condensation of the nuclear chromatin resulting in a 10-fold-less volume of the original paternal genome. Interestingly, depending on species, 1% to 15% of nucleosomes remain within the sperm.

This thesis aims to analyze whether nucleosome retention in sperm from various species displays general regularities. For this purpose, we analyzed two mammalian species, man and bovine. In addition, we intend to understand the biological impact of nucleosome retention in sperm by applying a genome-wide analysis.

In contrast to data available from the literature, we observed only 4.8% nucleosomes in man and 14% nucleosomes in bovine. For the first time, we demonstrated that sperm nucleosomes were prominently associated with repetitive DNA elements, with a significant enrichment in heterochromatic centromere repeats and in retrotransposons, within intergenic and intron sequences. In contrast, nucleosome depletion could be observed predominantly in exon sequences, 5'-and 3'-UTRs and gene promoters, and was associated to simple sequence repeats, low complexity repeats and DNA transposons. Furthermore, nucleosome-retaining genes (preferably in gene bodies) were associated with biological functions, such as RNA- and protein-processing, calcium-ion transport, cell-cell adhesion, membrane and cytoskeletal organization. *HOX*-genes and genes important for morphogenesis and organ development exhibited nucleosome depletion.

Our data revealed a comparable distribution pattern of sperm nucleosomes in both analyzed species and support the hypothesis that sperm nucleosomes might play a role for gene expression in the developing embryo.

Zusammenfassung

Es wird vermutet, dass bei der Befruchtung die Spermatozoen nicht nur genetische, sondern auch epigenetische Information auf die Eizelle übertragen, wie zum Beispiel das Methylierungsmuster der DNA oder verschiedene Histonmodifizierungen. Die Bedeutung des Spermien-Epigenoms für die Befruchtung und Embryonalentwicklung ist bislang jedoch noch weitgehend ungeklärt.

Während der Spermiogenese kommt es durch den Austausch von Histonen gegen Protamine zu einer Kondensation des Kernchromatins auf etwa ein Zehntel des ursprünglichen Volumens. Interessanter Weise verläuft dieser Austausch unvollständig, so dass speziesabhängig 1-15% an Nukleosomen im Spermium verbleiben.

Im Rahmen der vorliegenden Dissertation wurde untersucht, inwieweit der Verbleib von Nukleosomen in Spermien verschiedener Spezies allgemeinen Regeln gehorcht. Zu diesem Zweck wurden die beiden Säugetierspezies Rind und Mensch analysiert. Weiterhin sollte anhand einer genomweiten Analyse die funktionelle Bedeutung der im Spermium verbleibenden Nukleosome aufgeklärt werden.

Im Gegensatz zu Daten aus der Literatur, konnte beim Menschen nur ein Nukleosomgehalt von 4,8% festgestellt werden (Bulle: 14%). Erstmals konnte zudem nachgewiesen werden, dass Nukleosome in Spermien vornehmlich in repetitiven DNA-Elementen (Zentromer-spezifische Repeats und Retrotransposons) vorzugsweise innerhalb intergenischer und Intron-Sequenzen angereichert sind. Demgegenüber wurde eine signifikante Nukleosom-Abreicherung in kodierenden und funktionellen Bereichen des Spermien-Genoms (5´- und 3´-UTRs, Genpromotoren) festgestellt. Die Nukleosom-Abreicherung konnte zu einfachen Repeats und Repeats mit niedriger Komplexität, sowie DNA-Transposons assoziiert werden. Weiterhin konnte festgestellt werden, dass für RNA- und Protein-Prozessierung, Kalziumionen-Transport, Zell-Zell-Adhäsion sowie Membran- und Zytoskelettorganisation relevante Gene häufig in ihrer Sequenz Nukleosome aufweisen. Dagegen besaßen *HOX*-Gene und Gene, die relevant für Morphogenese und Organentwicklung sind, kaum Nukleosome in ihrer Sequenz.

Die Ergebnisse der vorliegenden Arbeit zeigen ein vergleichbares Verteilungsmuster von Nukleosomen in den beiden untersuchten Spezies Mensch und Rind und

unterstützen die Hypothese, dass Spermien-Nukleosome eine Rolle bei der Genexpression in dem sich entwickelnden Embryo spielen könnten.

List of abbreviations

The following table describes the significance of various abbreviations and acronyms used throughout the thesis.

Abbreviation	Meaning
APS	ammonium persulfate
BME	basal medium eagle
Bp	base pair
BSA	Bovine serum albumin
ChIP	chromatin immunoprecipitation
CENP-A	centromere-specific H3-like protein
COBRA	combined bisulphite restriction analysis
CTCF	CCCTC-binding factor
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DTT	dithiothreitol
ECS	estrus cow serum
EDTA	ethylenediaminetetraacetic acid
F	forward primer
FDR	false discovery rate
FISH	fluorescence in situ hybridization
FSH	follicle-stimulating hormone
GO	gene ontology
H3K4me3	histone 3 tri-methylated at residual lysine 4
H3K4me2	histone 3 di-methylated at residual lysine 4
H4K12ac	histone 4 acetylated at residual lysine 12
HOX	homeobox

IGV	Integrative Genomics Viewer
ILF2	interleukin enhancer binding factor 2
IVF	in vitro fertilization
LB Amp	Lysogeny Broth (LB) containing ampicillin
LCRs	low complexity repeats
LINEs	long interspersed nuclear elements
LH	luteinizing hormone
LTRs	long terminal repeats
MARs	matrix attachment regions
MEM	eagle′s minimum essential medium
miRNAs	micro RNAs
MPM	modified parkers medium
M.SssI	CpG methyltransferase
PBS	phosphate buffered saline
PCDHGC3	gamma protocadherin C3
PGCs	primordial germ cells
Pol	polymerase
PVDF	polyvinylidene difluoride
PVP	polyvinylpyrrolidone
R	reverse primer
RNA	ribonucleic acid
RT	room temperature
SAM	S-adenosylmethionine
SDS	sodium dodecyl sulfate
SOF	synthetic oviduct fluid
SSR	simple sequence repeats
Ta	annealing temperature
TEMED	tetramethylethylenediamine
TPs	transition proteins
TSS	transcription start site
TTS	transcription termination sites

UTR	untranslated region

Table of contents

1. Introduction

1.1 Overview

Compared with agamogenetic creatures, gamogenetic creatures are much more evolved and therefore have their own ways to keep the stability of species and meanwhile diversify the hereditary basis of gametes to meet the adaptability requirement of evolution. This requires both parental sides to produce haploid gametes with various hereditary information. For the paternal side, this process is defined as spermatogenesis. Spermatogenesis is the process by which male germ cells differentiate from spermatogonia to mature spermatozoa. It consists of three major stages: pre-meiotic (or spermatogoniogenesis), meiotic (or spermatocytogenesis) and post-meiotic (or spermiogenesis). During spermatogenesis, the paternal genome undergoes both dramatic genetic and epigenetic changes.

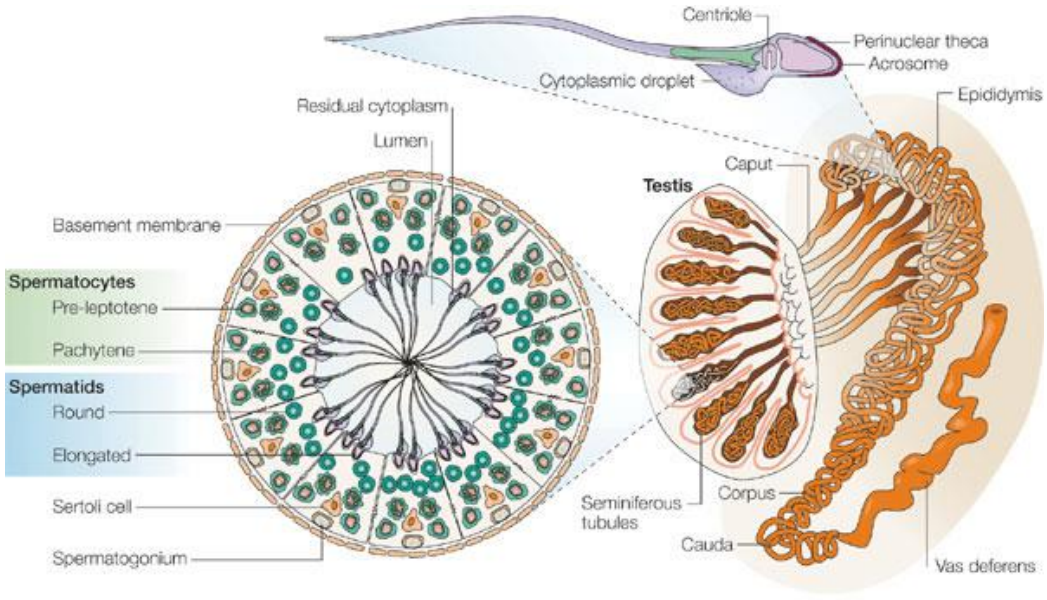


Figure 1 Graphic depiction of human testis and cross section of a seminiferous tubule (Krawetz, 2005). Spermatogonia at the basement membrane go through meiosis into round spermatids, which then go through spermiogenesis into elongated spermatids and finally though the maturation steps in epididymis, into a mature and motile spermatozoa.

For the genetic part, meiotic recombination and chromosome segregation directly affect genetic information. A-type spermatogonia proliferate (themselves) by mitosis. Some of

them differentiate into B-type spermatogonia and spermatocytes, which enter meiosis. Primary spermatocytes replicate DNA during the preleptotene stage and subsequently go through leptotene, zygotene, pachytene and diplotene stages of the first meiotic prophase. In the pachytene stage, homologous chromosome synapsis and recombination occur, exchanging DNA segments through a process of meiotic crossing-overs. In the second meiotic division, secondary spermatocytes divide into haploid round spermatids which differentiate into transcription-inactive elongated spermatids.

Besides genetic changes, which are unique in the development of gamogenetic creatures, dramatic epigenetic modifications, including histone modification, DNA methylation, chromatin remodeling and replacement of histones by protamines, also accompany with genetic changes throughout spermatogenesis. These epigenetic modifications interact and interdepend on each other, resulting in the production of mature spermatozoa with highly-condensed heterochromatin and more importantly, a sound paternal epigenome.

1.2 Contributions of sperm

Accordingly, the contribution of mature spermatozoa to early embryogenesis also includes both genetic and epigenetic factors. Genetic contributions include a haploid genome with intact coding regions and regulatory regions for essential genes. Spermatozoal DNA must contain the proper copy number of essential genes, and cannot have increased single- or double-stranded DNA breaks. There is evidence that mature sperm delivers much more than just the paternal genome into the zygote, within which mainly cover epigenome (Carrell, 2008; Miller et al., 2010). As illustrated by Figure 2, the epigenetic information transmitted into the oocyte by the fertilizing spermatozoon mainly include sperm-born RNAs, DNA methylation, modified sperm histones and other proteins (such as nuclear matrix proteins and perinuclear theca proteins).

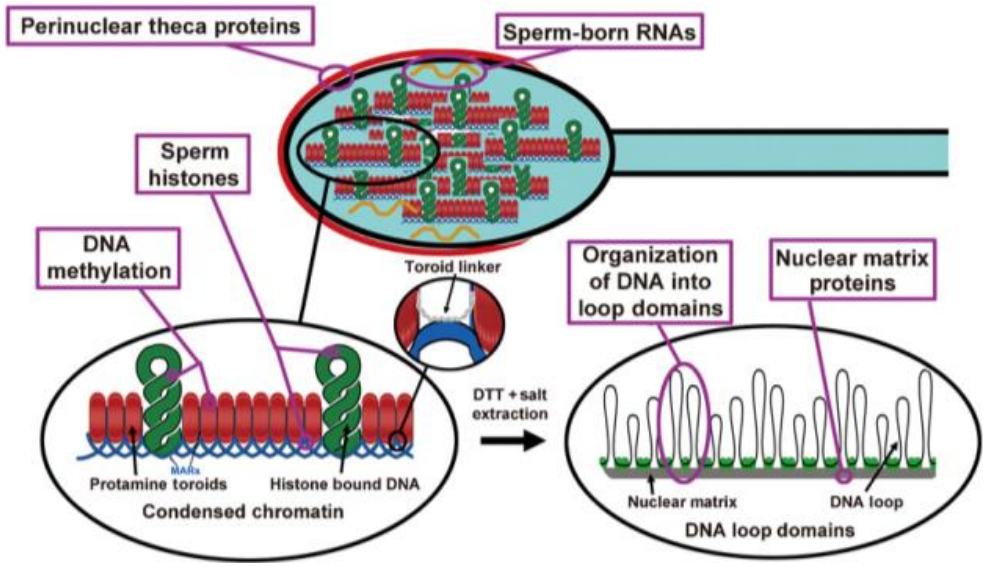


Figure 2 Sperm-derived epigenetic information transmitted to the embryos (Yamauchi et al., 2011). DNA methylation is the best-known example of non-DNA sequence information that is required for embryogenesis. Sperm DNA is tightly condensed by protamines into toroids (lower left inset), but some histones remain bound to the chromatin. The DNA is organized into loop domains that are required for DNA replication in the oocyte. Proteins of the nuclear matrix and perinuclear theca are also delivered to the oocyte. MARs: matrix attachment regions; DTT, dithiothreitol.

The spermatozoon provides mRNAs and micro RNAs (miRNAs), which may contribute to the embryonic transcriptome and regulate embryonic gene expression. Microarray analyses performed on sperm from fertile and infertile men without basic sperm analysis abnormalities have revealed a significantly different transcriptome (Garrido et al., 2009), suggesting that infertility from patients with normal semen parameters may be due to a lack of factors involved in correct sperm function. Transcriptome analysis in semen of low-fertile and high-fertile bulls by differential transcript profiling further revealed a significant difference in a portion of transcripts associated with metabolism, signal transduction, translation, glycosylation and protein degradation (Lalancette et al., 2008). On the other hand, miRNAs are short non-coding RNAs which are approximately 21 nucleotides in length and are recognized as regulators of post-transcriptional translation at every stage of spermatogenesis, the function of which becomes prominent particularly during spermiogenesis, when the compacting sperm nucleus becomes transcriptionally inactive (Dadoune, 2009; Krol et al., 2010). An extended number of

miRNAs has been observed to be differentially expressed in asthenozoospermic and oligoasthenozoospermic patients when compared with normozoospermic males (Abu-Halima et al., 2013). In the bovine system, similar results have been found. The miRNA profiling in bovine spermatozoa from high and low fertility bulls has shown significant differences (Govindaraju et al., 2012).

DNA methylation is another significant issue regarding the sperm epigenome. Specifically, DNMT1 (DNA methyltransferase 1) provides maintenance of the DNA methylation pattern throughout spermatogenesis (Jaenisch and Bird, 2003). Spermatogonia proliferate themselves by mitosis and then enter meiosis. Within the subsequent meiotic I division prophase, DNMT 3a together with its isoform DNMT 3L, help to reestablish the de novo methylation from the leptotene to the pachytene stages (Chedin et al., 2002). After homologous chromosome synapsis and recombination, secondary spermatocytes are produced and then divide into haploid round spermatids through meiotic II division. The spermatids undergo a global remodeling of its nucleus with chromatin condensation, packing the majority of the hypermethylated paternal genome into protamine toroids. Fertilization is followed by the first global demethylation in mammalian development. The paternal pronucleus quickly undergoes an active global demethylation, whose mechanism still remains unclear, while the maternal genome go through a passive demethylation, mainly due to the loss of DNMT1 during cell cleavage (Bestor, 2000). However, the paternal-specific imprinting marks escape this demethylation wave and manage to maintain themselves throughout embryogenesis. DNMT 3a and DNMT 3ab then take over and establish the de novo methylation according to cell differentiation and tissue-specific methylation changes occur throughout development. On the other hand, during the specification of primordial germ cells (PGCs) from epiblast cells at embryonic day (E) 6.5, PGCs undergo the second global demethylation in mammalian development. This time, both parental imprinting marks are erased, which aims to retain totipotency, and the imprinting marks in the developing gametes are reset subsequently. This erasure and resetting process ensures the establishment of correct imprinting marks in coming gametes. PGCs then start their migration toward the genital ridge, reside in and arrest in G1 phase of the cell

cycle and wait for the start signal of spermatogenesis at puberty. The whole DNA methylation reprogramming process is schematically presented in Figure 3.

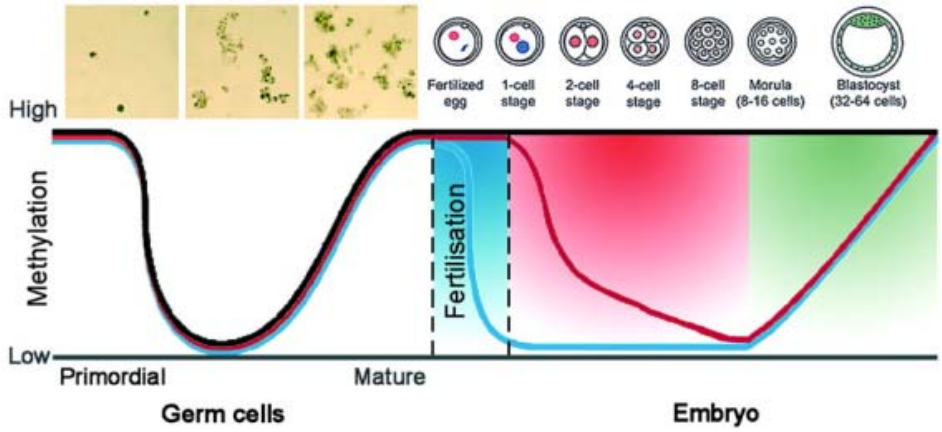


Figure 3 Methylation reprogramming in the mouse germ line and preimplantation embryo (Santos and Dean, 2004). Black line: methylation level in methylated imprinted genes; Red/ blue line: methylation level in maternal/paternal non-imprinted genetic sequences. Highly methylated primordial germ cells enter the germinal ridge and undergo loss and reacquisition of methylation during their expansion phase. Examples of these cells (day 11.5, 13.5 and 14.5) stained for alkaline phosphatase, a PGC marker, are pictured above. The horizontal time axis and the vertical axis indicating the relative methylation levels are not to scale.

A correct sperm methylome is very important for embryogenesis. It has been reported recently (Jiang et al., 2013) that the DNA methylome in early embryos of zebrafish was solely inherited from sperm, not from oocyte or a combination of both parental sides. The aberrant sperm methylome has also been associated with male infertility. An abnormal methylation level of imprinted genes has been observed in patients with oligozoospermia (Marques et al., 2008), obstructive azoospermia (Minor et al., 2011) and idiopathic infertility (Poplinski et al., 2010).

In somatic cells, DNA is organized in loop domains, with a length of 60 to 100 kb. This organization of DNA loop domains is also preserved in sperm, with a shorter length of about 20–50 kb. These loops are attached at their base to a proteinaceous structure termed nuclear matrix, which is also the structural element of the nucleus (Choudhary et al., 1995; Kalandadze et al., 1990; Kaplan et al., 1987; Ward et al., 1989). The importance of sperm nuclear matrix has been emphasized these years. As is known in

somatic cells, the DNA replication is initiated on the nuclear matrix (Dijkwel and Hamlin, 1995; Jackson and Cook, 1986). A similar finding is observed in the paternal nucleus in the one-cell embryo in mouse (Shaman et al., 2007). It has been suggested that this paternal loop attachment structure is inherited from sperm (Wakayama et al., 1998; Wilmut et al., 1997). When original attachment configuration of DNA to the nuclear matrix was disrupted, the DNA could not replicate anymore (Yamauchi et al., 2007). When entering the oocyte, the sperm also carries the perinuclear theca. A set of proteins is contained in this structure, including an extranuclear located histone H2B (Aul and Oko, 2002). These proteins are likely to be incorporated into the paternal pronucleus and may also be counted as part of the epigenetic inheritance from the spermatozoon.

1.3 Chromatin remodeling and histone modifications

While the impact of mRNAs, DNA methylation and nuclear matrix proteins in fertility is relatively well recorded, the relevance of the spermatozoal epigenome and epigenetic marked gene regions during fertilization and early embryogenesis is still for the most part unknown. However, a unique chromatin condensation process during mammalian spermatogenesis might give us some hints regarding this issue. Spermatozoa in mammals are known to have to march a painstaking long distance to reach their ultimate goal, the oocyte. To facilitate the transportation and more importantly, to protect the correct paternal hereditary information, spermatids have to undergo a sophisticated heterochromatinization process, resulting in a 10-fold-less size of the original paternal genome, before they become mature spermatozoa. This is known as spermiogenesis and the genome-wide chromatin condensation is achieved by the replacement of histones by transition proteins (TPs) and finally, by protamines. Protamines are small basic arginine-rich proteins around which DNA could be tightly wrapped and can therefore significantly reduce the chromatin size (Balhorn, 2007).

Histones, on the other hand, are the main chromatin proteins in nearly all types of eukaryotic cells, other than spermatids and mature spermatozoa. H1/H5, H2A, H2B, H3 and H4 (Bhasin et al., 2006) constitute the histone family, each of which contains several variants expressed across different tissues and species. In mice and humans,

the variants of H2A, H2B, H3 and H1 are expressed in testis, especially at the very beginning of spermiogenesis. Nucleosomes in spermatozoa are comprised of remnant canonical histones (H2A, H2B, H3 and H4), as well as testis-specific histone variants (e.g. hTSH2B/TH2B, H2BFWT and H1t) (Boulard et al., 2006; Gatewood et al., 1990; Kimmins and Sassone-Corsi, 2005; Li et al., 2005). Interestingly, studies both in mouse and human (Gonzalez-Romero et al., 2008; Govin et al., 2007) have shown that nucleosomes containing these variants were significantly less stable than those composed of canonical histones. Incorporation of histone variants are therefore believed to open chromatin and to form unstable nucleosomes which then constitute preferential targets for nucleosome disassembly and histone displacement. The process acting together with histone variants incorporation is global histone hyperacetylation. Acetylation of core histones plays an important role for the replacement of histones by protamines, as the addition of acetyl groups to lysine residues located at the amino-terminal end of histones turns the basic state of histones into a neutral one and thus decreases the affinity of histones to DNA and allows protamines to interact with DNA (Oliva et al., 1987). In elongating spermatids, where DNA replication and transcription is totally inactive, the histones are highly acetylated. Additional evidences have shown that with the appearance of TP1 and TP2, the acetylation signals gradually disappear during the course of histone replacement (Hazzouri et al., 2000). Therefore, histone hyperacetylation seems to be tightly linked to histone replacement (see Figure 4).

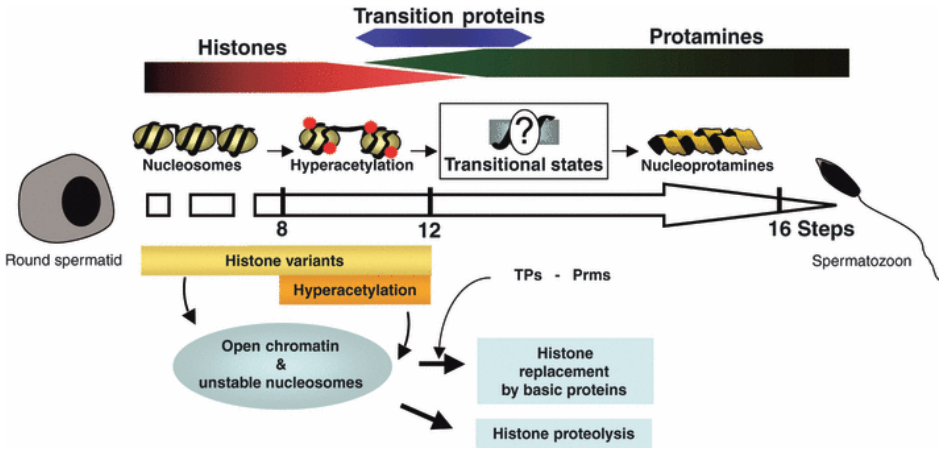


Figure 4 Major steps and various factors involved in histone replacement (Gaucher et al., 2010). Extensive incorporation of histone variants and global histone hyperacetylation prior to their replacement

create open chromatin domains containing unstable nucleosomes. The presence of highly basic small DNA-packaging proteins such as transition proteins (TPs) could facilitate histone eviction and a shift from a nucleosomal-based genome organization to non-histone protein-based DNA packaging. Prms: Protamines.

With the facilitation of histone variants and hyperacetylation, spermiogenesis carries on and a highly condensed haploid paternal genome is achieved. However, the replacement of histones by protamines is incomplete in many mammalian species. Residual nucleosomes are retained in mature ejaculated spermatozoa, ranging from 1% of the whole nuclear protein in the mouse (Balhorn et al., 1977) to over 50% in some marsupial species (Soon et al., 1997). Human sperm contains approximately 15% nucleosomes (Gatewood et al., 1990). Notably, the remaining histones, especially H3 and H4 with long amino acid tails protruding from the nucleosome, exhibit various covalent modifications at several positions, commonly at residual lysines. Modifications of the tails include methylation, acetylation, phosphorylation, ubiquitination and many more (Bannister and Kouzarides, 2011), as can be seen from Figure 5. Only methylation is known to occur in more than one copy per residue, while other modifications show only one copy at a specific position.

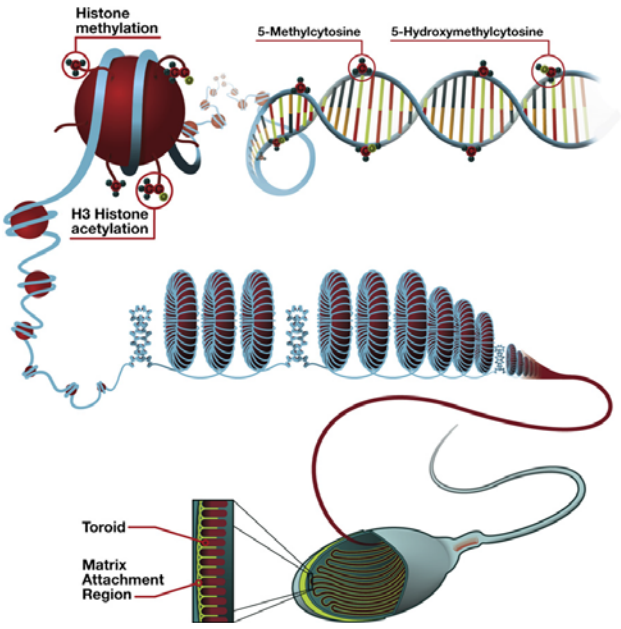


Figure 5 Chromatin structure of human spermatozoa (Carrell, 2012). DNA methylation is the first line of epigenetic modification of chromatin through methylation of position of cytosines found in CpG dinucleotides. DNA is bound to histones with various modifications that present a second level of regulation of gene transcription. Most histones are replaced with protamines that result in a higher order of DNA packaging and silence gene expression. Retained histones lies among the toroid of protamine-DNA package and may be bound to matrix attachment regions, which facilitates replication of loop domains in the embryo.

Remaining nucleosomes and associated DNA have triggered intense discussions in recent years. Using chromatin immunoprecipitation (ChIP) with antibodies against specific histone modifications, researchers have precipitated DNA fragments associated with modified histones in human sperm. After performing ChIP sequencing (ChIP-seq) and ChIP-on-chip methods, collective evidence have suggested that combinations of histone modifications constitute a specific, so-called "histone code". Therefore they may be involved in the establishment of adequate epigenetic information in the offspring and in starting early gene expression in the zygote and, thus, may be crucial for fertility. Arpanahi et al (Arpanahi et al., 2009) observed that in spermatozoal chromatin of both mouse and man, regions of increased endonuclease sensitivity (i.e. retained nucleosome regions) were closely associated with gene regulatory regions, including many promoter sequences and sequences recognized by the CCCTC-binding factor (CTCF). Hammoud et al (Hammoud et al., 2009) found that retained nucleosomes were significantly enriched at loci of developmental importance. Specifically, H3K4me3 (histone 3 tri-methylated at residual lysine 4) was significantly enriched in developmental promoters, regions in *HOX* clusters, certain non-coding RNAs, and generally to paternally-expressed imprinted loci. H3K4me2 and H3K27me3 were also enriched in certain developmental promoters. Gene ontology term analyses for genes associated with H3K4me3 revealed factors for changing nuclear architecture, RNA metabolism, spermatogenesis and a selected number of transcription factors important for embryonic development including *EVX1/2, ID1, STAT3, KLF4, FGF9* and *SOX7/9*. The majority of developmental and signaling transcription factors were significantly enriched with H3K27me3 and H3K4me2. The testis-specific histone H2B, which is incorporated late in spermatogenesis and comprises a large percentage of retained histones, was significantly enriched at genes for ion channels and genes involved in spermiogenesis, but not at promoters of developmental genes. Brykczynska et al (Brykczynska et al., 2010) also focused on methylated histones in human and mouse spermatozoa. Their results showed similarities between both species. H3K4me2 marked genes that were relevant in spermatogenesis and cellular homeostasis, while H3K27me3 marked developmental regulators. However, genes with extensive H3K27me3 coverage around transcriptional start sites in particular tended not to be

expressed during male and female gametogenesis or in preimplantation embryos. More recently, Paradowska et al (Paradowska et al., 2012) investigated the genome-wide promoter binding sites of an acetylated histone mark, H4K12ac, both in human and mouse spermatozoa. H4K12ac was found to be enriched predominantly between ± 2 kb from the transcription start site. Developmentally relevant promoters were also identified to be associated with H4K12ac. They also evaluated the potential function of H4K12ac-associated genes in mouse early embryos. Those genes revealed a weak correlation with genes expressed at 4-cell stage human embryos, while 23 genes were activated in 8-cell embryo and 39 in the blastocyst. Genes activated in 4-cell embryos were involved in gene expression, histone fold and DNA dependent transcription, while genes expressed in the blastocyst were classified as involved in developmental processes.

1.4 Repetitive sequences

Studies mentioned above have all focused on the nucleosome-retained genes. However, these genes count only for a minority fraction of the whole paternal genome. Generally speaking, a genome is comprised of both coding DNA (i.e. genes) and non-coding DNA sequences. Compared with genes, which code for proteins or RNAs, the non-coding DNA tends to have high copy numbers or repeat itself throughout the genome. DNA sequences with this characteristic are characterized as repetitive sequences (Flamm, 1972). Repetitive sequences could be characterized into four categories: simple repeats, tandem repeats, segmental duplications and interspersed repeats, the last of which further consists of short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs) and DNA transposons. LINEs and SINEs also form by far the biggest fraction of human interspersed repeats (Smit, 1996). Figure 6 shows structures of transposable elements which produce high copy numbers in mammalian interspersed repeats.

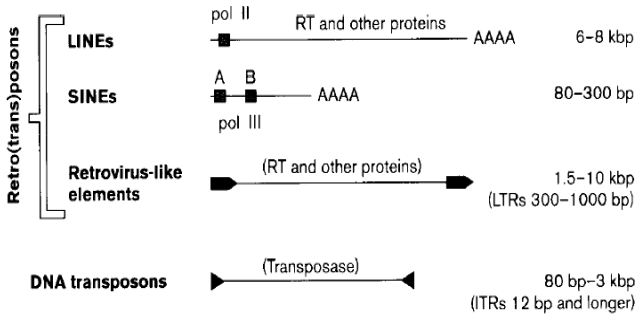


Figure 6 Schematic representation of major transposable elements (Smit, 1996). Shaded boxes: internal promoter sites; names in parentheses: only autonomous

elements code for these proteins. ITR: inverted terminal repeat; RT: reverse transcriptase; bp: base pair; kbp: kilo base pair.

Higher eukaryotic genome contains much more abundant repetitive sequences than coding DNA, and as evolution continues, the differences between those two components become more significant, ranging from almost equal in *Caenorhabditis elegans* (Stein et al., 2003) (16.5 % vs. 14%), to repetitive sequences taking predominance in mouse (Waterston et al., 2002) and human (Lander et al., 2001) (40% vs. 1.4%, over 50% vs. 1.2%, respectively). A recent study using a more sensitive strategy has suggested an even higher percentage of repetitive or repeat-derived sequences, up to 66%–69%, in human genome (de Koning et al., 2011). With repetitive sequences accounting for over 2/3 of the human genome, it could be speculated that they might overlap, to say the least, some nucleosome-retained DNA. On the other hand, studies regarding the function of these so-called “dark matter” of the genome have persisted for decades. Current opinions showed that some repetitive sequences were involved in regulation of gene expression. Tissue-specific transcription of SINE B2 repeat in mouse was required for gene activation of the growth hormone gene, by generating short, overlapping pol II-and pol III-driven transcripts, both of which were necessary and sufficient to enable a restructuring of the regulated locus into nuclear compartments (Lunyak et al., 2007). SINE B1 elements could influence the activity of downstream gene promoters, causing a repression effect (Estecio et al., 2012). LINE1 could be activated by satellite transcripts and lead to aberrant expression of neuroendocrine-associated genes proximal to LINE1 insertions (Ting et al., 2011). LINEs may also facilitate X chromosome inactivation by participating heterochromatin formation (Chow et al., 2010). However, the results remain fragmented and a clear panoramic functional view, as has been established regarding functional genes, is yet to be structured.

1.5 Aims

The two prominent aims of this study are

1) to understand the regularities of nucleosome-retention in mammalian sperm in a genome-wide manner and

2) to find out to which content this process underlies rules of great generality and has a biological function.

In order to get an overview on the nucleosome situation in sperm and to check the substance of previous theories, we will avoid ChIP-procedures, but sequence the whole 146 bp mononucleosomal DNA fraction isolated from sperm.

Two mammalian models, human and bovine, will be used to analyze all putative nucleosome-binding sites in their sperm genomes considering the following aspects: repetitive DNA elements, non-coding DNA (intergenic DNA and intron areas), coding DNA (exon areas) and known functional DNA elements (gene promoter, 5'-UTR and 3'-UTR).

According to sequencing results, the corresponding genes and other DNA elements in bovine oocytes and early embryos (zygote, 2-, 4-, 8-cell stage embryo, morula and blastocyst) will be analyzed considering their mRNA expression and promoter methylation, in order to determine the transcriptional status and chromatin constitution of ´sperm derived´ contributions before (oocyte), immediately after (zygote) and later after (2-cell embryo to blastocyst) fertilization.

To get a clear hint about epigenetic events in individual embryo stages and help to characterize the epigenetic mechanisms, the mRNA expression of key chromatin modifiers, main pluripotency genes and imprinted genes in bovine sperm, oocytes and early embryos will be analyzed.

Finally, we hope that our study will provide a panorama view regarding the influence of paternal contributed genetic and epigenetic factors on fertilization and early developmental gene activation.

2. Material and Methods

2.1 Human and bovine ejaculates, bovine oocytes and early embryos collection

After signing informed consent, human ejaculates were obtained from healthy donors with normal fertility (Votum of the local ethics committee 146/06, confirmed on December 15th 2010 for the 2nd funding period of the KFO 181). Donors were asked to keep sexual abstinence for 4–6 days. Ejaculates were obtained by masturbation into a dry wide-mouth sterile plastic container. Bovine ejaculates were obtained from fertile-proved bulls.

Samples from both species were liquefied in 37 °C incubator, washed twice with PBS and sperm numbers were counted manually by microscopy afterwards. Washed samples were kept at -80 °C until used.

For oocytes, ovaries of slaughtered cows were collected. After aspirating follicles of 2-6 mm, oocytes were in-vitro matured for 22 h in modified parkers medium (MPM) supplemented with 5 % estrus cow serum (ECS) and 0.25 U/ml FSH (Sioux, Iowa, US) and 0.125 U/ml LH (Sioux, Iowa, US) at 39 °C in a maximum humidified atmosphere of 5 % CO_2 in air.

To obtain zygotes, mature oocytes were co-cultured with frozen-thawed semen (10^6 spermatozoa/ml; capacitated with swim-up procedure) of a regular breeding bull for 18 h (IVF day 0). After removal of cumulus cells by vortexing, denuded oocytes/zygotes were washed three times in PBS with 0.1% PVP and frozen in 10 µl RNA-Later, 10 µl Trizol or for DNA in 10 µl PBS.

For embryo production, the presumptive zygotes were denuded by vortexing and cultured in synthetic oviduct fluid (SOF) supplemented with 5 % (v/v) ECS, 40 µl/ml BME (Sigma-Aldrich, Seelze, Germany) and 10 µl/ml MEM (Sigma-Aldrich, Seelze, Germany) covered with mineral oil at 39 °C in a humidified atmosphere of 5% CO_2, 5% O_2 and 90 % N_2. 2-cell stage embryos (28-30 h post IVF), 4-cell stage embryos (44-48 h

post IVF) and 8-cell stage embryos (64-64 h post IVF) were collected, washed and frozen as described before. Morulaes were collected on day 5 and blastocysts on day 7 after IVF.

2.2 Micrococcal nuclease digestion

To obtain mononucleosomes, the method described by Zalenskaya et al (Zalenskaya et al., 2000) and modified by Hammoud et al (Hammoud et al., 2009) was used throughout but with some minor modifications. Briefly, semen sample was thawed and diluted with PBS-proteinase inhibitor solution (proteinase inhibitor cocktail tablets by Sigma-Aldrich, Seelze, Germany). 10^7 cells per tube of these sperm suspension were centrifuged at 3500 rpm for 10 min. Pelleted sperms were suspended in the above buffer supplemented with 0.1% Lysolecithin (Sigma-Aldrich, Seelze, Germany) and incubated 15 min on ice for cell lysis. Pellet was obtained by centrifugation at 3500 rpm for 5 min and washed once with PBS-proteinase inhibitor solution. Following incubation in the above buffer supplemented with 10 mM DTT at 37 °C for 30 min, $CaCl_2$ were added to a concentration of 0.6 mM and 20 units of micrococcal nuclease (New England Biolabs, Frankfurt am Main, Germany) were added for digestion. After incubation for 3 min at 37 °C, digestion was stopped by adding EDTA to a concentration of 5 mM. The digested nuclei were centrifuged at 10,000 rpm for 3 min to separate histone-related chromatin fraction (supernatant) from the protamine-related fraction (pellet). The complete fractionation of histone- and protamine- associated chromatin was confirmed by Western blot analysis. For histone-associated DNA analysis, proteinase K and SDS were added to a concentration of 200 µg/ml and 0.5% in supernatant, respectively. After incubation over night at 55 °C, DNA was extracted by phenol-chlorophorm method and spread on a 2 % agarose gel.

2.3 146bp DNA purification

DNA fragment associated with histone (146 bp) was then cut from the agarose gel and purified with NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). Briefly, the gel slice was transferred into a clean tube and weighted. For each 100 mg of agarose gel, 200 µl Buffer NTI were added. The sample was then incubated for 10min at 50°C and vortexed every 2-3min until the gel slice was

completely dissolved. A spin column was then placed into a collection tube and all the dissolved sample mixtures were loaded into the column. The column and collection tube were centrifuged for 30 sec at 11,000 x g and the flow-through were discarded. After placing the column back into the collection tube, 700 µl Buffer NT3 were added into the column and centrifuged for 30 sec at 11,000 x g again. An additional 1 min of centrifugation was then performed to remove all the residual Buffer NT3, and meanwhile the spin column should not come in contact with the flow-through while removing it from the centrifuge and the collection tube. The column was then put into a new 1.5 ml microcentrifuge tube and 15 µl ddH_2O were added. After incubation at room temperature (RT) for 1 min, another centrifugation for 1 min at 11,000 x g would yield the purified DNA. The purified 146 bp DNA was sent for sequencing and PCR analysis.

2.4 Cross-linking ChIP

The ChIP method described by Weber et al. (Weber et al., 2007) was used with several modifications. Briefly, semen sample was thawed and diluted with PBS. For each tube (4 ml, 2×10^7cells/tube), 108 µl formaldehyd (36%) were added and the mixture was incubated in RT for 10 min for crosslinking. Crosslinking was stopped by adding 400 µl (0.125M) glysin into each tube. After centrifugation at 2,000 rpm for 5 min, the pellet was washed twice by 1ml ice-cold PBS. Then, 400 µl lyse buffer were added and 10 min incubation on ice was followed. The mixture was then sonicated (power: 55%, time: 2x15 sec) and the suspension was centrifuged 10 min at 13,000 rpm (4°C). The supernatant was put into a 2 ml tube and supplemented with 1600 µl ChIP-dilution buffer. The mixture was then separated into two parts: 160 µl was put into a new tube, supplemented with 340 µl TE-buffer and kept in -20°C as input control, and the rest 1,840 µl was incubated with 50µl “salmon sperm-DNA/Protein A- agarose-50%-slurry” (Millipore, Schwalbach, Germany) for 30 min at 4°C. After incubation, the mixture was centrifuged 2 min at 1,000 rpm and the supernatant was equally divided into three parts: 600 µl as negative control and the other 2x 600 µl was supplemented with 5 µl anti-H3K9ac antibody (0,5µg/µl, Abcam, Cambridge, UK) each probe. All probes were kept at 4°C overnight for incubation with shaking. After overnight incubation, 80 µl “salmon sperm-DNA/Protein A- agarose-50%-slurry” (Millipore, Schwalbach, Germany) were

added into each probe and incubated 2 h at 4°C with shaking. Afterwards, the mixture was centrifuged at 1,000 rpm for 1 min and the supernatant was discarded carefully. The pellet then underwent washing steps as follows. For each step, 1 ml of washing buffer added, the mixture shake for 3 min at 4°C, centrifuge for 1 min at 1,000 rpm, supernatant discarded and the next washing buffer added.

1) low salt immune complex wash buffer
2) high salt immune complex wash buffer
3) LiCl immune complex wash buffer
4) 1xTBE (twice)

After washing steps, the pellet was resuspended in 250 µl elusion buffer and incubated at RT for 15 min. After centrifugation for 2 min at 1,000rpm, another 250 µl elusion buffer was added and incubated for 15 min again. Therefore after centrifugation, there was 500 µl elusion solution in total. Then 20 µl NaCl (5M) was added into the input control and the elusion solution and all probes were incubated at 65°C for 4 h to reverse the cross-linking. Followed by proteinase K digestion, the DNA was extracted by phenol-chlorophorm method. The immune-precipitated H3K9ac-associated DNA was sent for sequencing and PCR analysis.

2.5 Western blot analysis

The protein fractions (nucleosome and protamine) obtained from both species were determined by Western blot analysis to confirm the presence of corresponding proteins.

2.5.1 Protein measurement

Separated histone fragments were subsequently concentrated using amicon ultra-2 centrifugal filter device (Millipore, Schwalbach, Germany).

Protein concentrations were determined using bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, Dreieich, Germany).

Bovine serum albumin (BSA) standards were prepared using the same diluent as for the samples.100 µl of protein sample and standards replicate were pipetted into labeled tubes. 2 ml of BCA kit reagent solution A and B (50:1) were added into each tube.

Reactions were incubated at 37°C for 30 min and then cooled to room temperature. The absorbance was measured at 562 nm on a spectrophotometer. Protein concentration of the proteins samples was measured by standard curve. Then the volume of protein sample needed to reach 25 µg of total protein to be calculated.

2.5.2 Gel preparation

For gel preparation, the apparatus was assembled with the glass plates and spacers. For the resolving gel, the gel solution was prepared by mixing all reagents together except the TEMED and 10 % APS. Immediately prior to pouring the gel, 10% APS and TEMED were added and swirled gently to initiate polymerization. Pour the resolving gel to about 1 cm below the wells of the comb and seal with 1 ml water. After the resolving gel reached solidification (commonly after 45 min), when gel has set, pour off the water. The stacking gel (~3 ml) was poured and then the comb was inserted immediately. The gel was then polymerized for 35-40 min. Components of the resolving gel and the stacking gel were listed as follows.

Resolving gel (for 2 Mini Gels)

	15 %
Rotiphorese Gel 30	5,0 ml
Puffer B	2,5 ml
Water	2,5 ml
10 % APS	50 µl
TEMED	5 µl

Stacking gel

Reagent	Quantity taken
Rotiphorese Gel 30	720 µl
Puffer C	1,2 ml
Water	2,8 ml
10 % APS	20 µl
TEMED	8 µl

2.5.3 Sample loading

425 µl sample buffer were supplemented with 25 µl of 60% glycerol and 50 µl of β -mercaptoethanol to reach a total volume of 500 µl. The prepared solution was mixed by vortexing and 20 µl of it were added into each sample and mixed with a pipette. After short spinning for approximately 15 sec for proper mixing, samples were denaturated by heating at 95°C for 5 min and again spinned for 15 sec.

Then, 10 µl ladder (Qiagen, Hilden, Germany) and samples were loaded into the well, according to the concentration of the proteins. Run the gel at 100 volts through the stacking part of the gel and turn the volts up to 120 V after the proteins had gone through the stack and were migrating through the resolving gel. The running time was approximately 90 min.

2.5.4 Gel removal

The tank lid was removed and carefully lifted out of the inner chamber assembly. After discarding the running buffer, the electrode assembly was pulled out of the clamping frame and the gel cassette sandwiches were removed. Then, the gels were removed from the gel cassette sandwich by gently separating the two plates of the gel cassette. Finally, the gel was removed by floating it off the glass plate by inverting the gel and plate under fixative or transfer solution, agitating gently until it separated from the plate. Coomassie stain (ROTH, Karlsruhe, Germany) was used for visualization of proteins in gels.

2.5.5 Transfer

Wet transfer method was used. The PVDF- membrane and Whatman filter papers were cut to the dimensions of the gel. PVDF- membrane was activated by methanol for 1 min, rinsed with water and then soaked in transfer buffer. Cassette was prepared by loading up in the following order: case (clear side), sponge, Whatman paper, membrane, gel, Whatman paper, sponge, case (black side). Then place the cassette in the transfer apparatus (Bio-Rad, Munich, Germany) with black side facing black. Run the transfer at

100 V for 1 h. Ice-packs were used to cool down the apparatus. Further, the membrane was stained with 1x Ponceau S for 5-10 min on agitator to check the transfer status and destined by washing with ddH_2O.

Block the membrane directly after the blotting in blocking buffer (5% non-fat milk in 1xTBST) for 1 h at RT. The membrane was then incubated with polyclonal primary antibody raised in rabbit against histone H3 (Abcam, Cambridge, UK) in TBS/T-milk overnight at 4°C. Membrane was washed with TBST (3 × 15 min) and subsequently incubated with HRP-conjugated goat anti-rabbit secondary antibody in TBS/T-milk for 1h (ROTH, Karlsruhe, Germany). The membrane was again washed (3 × 15min) with TBST.

2.5.6 Development

After washing steps, the membrane was transferred onto a thin opened polyethylene film. 1 ml chemi-luminescent reagent A and 1 ml reagent B (1:1 ratio, Thermo Scientific, Dreieich, Germany) were mixed and added gently onto the membrane dropwise until the membrane was covered entirely. Then, after incubation for 5 min in darkness, the excess solution was removed and the blot was transferred into the Rontgen cassette. After exposure in X-ray for the required interval, the blot was placed into the developing solution for 2 min, and then transferred into the fixing solution for 5 min. It was then washed briefly in water for 5 min and dried to be visualized.

2.6 Sequencing

DNA fragment associated with sperm histone (146bp) was sent for sequencing with Illumina sequencing technology (Munich Gene Center). 100ng of histone DNA diluted in 10µl ddH_2O were applied for preparation of the DNA library.

2.6.1 Library preparation

2.6.1.1 Sample sonification

The total sample volume was added to 85µl. The Bioruptor (Diagenode, Liège, Belgium) was used for sonification and the program was set as: 30 sec on, 30 sec off, 15 cycles,

low power. Sonification time was 15 min and after adding more ice for cooling down, another 15 min sonification was performed.

2.6.1.2 End repair of fragmented DNA

Mix the following components in a sterile microfuge tube and incubate in a thermal cycler for 30 min at 20°C.

Fragmented DNA	85µl
NEBNext end repair reaction buffer(10X)	10µl
NEBNext end repair enzyme mix	5µl
Total volume	100µl

Purify DNA sample with Agencourt AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany) to get rid of residual primers. Add beads, mix with pipette and stay 10min for binding. Then, put the tube on magnetic stand, remove all liquid inside and wash beads with 80% ethanol 200 µl twice. Elute DNA in 42µl sterile dH_2O.

2.6.1.3 dA-tailing of end repaired DNA

Mix the following components in a sterile microfuge tube and incubate in a thermal cycler for 30 min at 37°C.

End repaired, blunt DNA	42µl
NEBNext dA-tailing reaction buffer(10X)	5µl
Klenow fragment (3→5 exo)	3µl
Total volume	50µl

Purify DNA sample with Agencourt AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany) again and elute DNA in 25µl of sterile dH_2O.

2.6.1.4 Adaptor ligation of dA-tailed DNA

Mix the following components in a sterile microfuge tube and incubate in a thermal cycler for 15 min at 20°C.

dA-Tailed DNA	25µl
Quick ligation reaction buffer(5X)	10µl
15uM DNA adaptors	10µl
Qucik T4 DNA ligase	5µl
Total volume	50µl

Purify DNA sample with Agencourt AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany) again and elute DNA in 30µl of sterile dH_2O.

2.6.1.5 Size select adaptor ligated DNA

Size select library fragments in the appropriate size range were performed using standard 2% agarose gels by cutting the right band and purify with Ultrafree-DA gelextraktionskit (Millipore, Schwalbach, Germany). Then elute purified DNA sample in 30ul of sterile dH_2O.

2.6.1.6 PCR enrichment adaptor ligated DNA

Mix the following components in a sterile microfuge tube and start a PCR reaction with the following program.

DNA	1µl
Primer 1 (25uM stock)	1µl
Primer 2 (25uM stock)	1µl
Phusion high-fidelity PCR master mix with HF buffer,2X	25µl
Sterile H_2O	22µl
Total volume	50µl

PCR Program

Cycle step	Temperature(°C)	Time(sec)	Cycles
Initial denaturation	98	30	1
Denaturation	98	10	12
Annealing	65	30	
Extension	72	30	
Final extension	72	300	1

Purify PCR products with Agencourt AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany) and elute in 50µl of sterile dH_2O.

2.6.2 Cluster generation

Before cluster generation, the concentration of DNA sample should be strictly controlled at 10nmol/L using Bioanalyzer (Agilent, Böblingen, Germany) to determine the size and concentration.

Mix the following components in a sterile microfuge tube and stay 5 min to denature. Take 10µl of the mixture, supplemented with 490µl hybridbuffer to reach a 1:50 dilution. Then take 50µl of this dilution, supplemented with another 150µl hybridbuffer. Finally, add 120µl of this final mixture into Cbot (Illumina, Eindhoven, The Netherlands) and the cluster generation was completed after 4h.

Prepared library DNA sample (10nmol/L)	3µl
NaOH (2mol/L)	1µl
10mM tris-buffer (PH8.9)	16µl

2.6.3 Sequencing

After cluster generation, sequencing was accomplished with the sequencing-by-synthesis technology using Genome analyzer IIx (Illumina, Eindhoven, The Netherlands). It took 4 days to complete sequencing.

2.7 Sodium bisulphite and CpG methyltransferase (M.SssI) treatment of sperm DNA

Bovine sperm DNA was extracted by phenol/chloroform method and treated as follows.

For bisulphite treatment, 2 µg DNA was denatured in 3 M NaOH at 37 °C for 10 min. 3.8 M sodium bisulphite and 0.1 M hydroquinone were then added and the mixture was incubated at 56 °C for 7 h. Bisulphite-treated DNA was purified using Wizard® DNA clean-up system (Promega, Mannheim, Germany). 1 ml of DNA clean-up resin was added to a 1.5 ml microcentrifuge tube and mixed with DNA sample. After assembling the syringe barrel and minicolumn together and attaching them into the vacuum manifold, the resin/DNA mixture was pipetted into the syringe barrel and a vacuum to draw the solution through the minicolumn was applied. 2 ml of 80 % isopropanol were added to wash the column and a vacuum was re-applied to draw the solution through the minicolumn. The column was dried by continuing to draw a vacuum for 30 sec after the solution had been pulled through the column. Remove the syringe barrel and transfer the minicolumn onto a 1.5ml microcentrifuge tube. After centrifugation at 10,000 x g for 2 min, transfer the minicolumn onto a new microcentrifuge tube. Apply 50 µl of prewarmed (65–70 °C) water to the minicolumn and wait for 1 min. Centrifuge the minicolumn for 20 sec at 10,000 x g to elute the bound DNA.

After bisulphite-treated DNA purification, desulphonation was performed by adding 3 M NaOH and incubated at 37 °C for 10 min. DNA was precipitated by 7.5 M ammonium acetate and 100% ethanol and dissolved in TE buffer. The treated DNA could be directly used for PCR amplification.

For M.SssI treatment, 2 µg sperm DNA was put together with M.SssI, buffer, SAM and filled up to 40 µl total reaction volume with double distilled water. The mixture was then incubated at 37°C for 15 min and DNA was extracted by phenol/chloroform method.

2.8 DNA methylation analysis (COBRA+ bisulphite sequencing)

2.8.1 COBRA (combined bisulphite restriction analysis)

COBRA PCR was performed using 200 ng bisulphite-treated DNA as template and 10 pmol special-designed COBRA primers together with rotor gene master mix (Qiagen, Hilden, Germany). PCR was performed in rotor gene cycler (Qiagen, Hilden, Germany). Amplified PCR products, together with M.sssI-treated DNA as positive control, were further digested with specific endonucleases (Taq I or Bsh123, Fermentas, Schwerte, Germany), according to their digestion sequence. Mock digestion was also included as negative control. Digested products were then spread on a 2% agarose gel to identify methylation status.

2.8.2 Bisulphite sequencing

For bisulphite sequencing, COBRA PCR products were purified using Nucleospin gel and PCR clean up kit (Macherey-Nagel, Düren, Germany). Purified PCR products were sub-cloned into pGEM-T vector (Promega, Mannheim, Germany) by mixing with the following reagents and incubating overnight at 4 °C.

Reagents	Quantity taken
2X rapid ligation buffer	5 µl
pGEM®-T vector	1 µl
PCR products	3 µl
T4 DNA ligase	1 µl
Final volume	10 µl

The plasmid DNA was transformed into Ca^{2+} competent cells. Firstly, all plasmid DNA of one ligation reaction was mixed with one vial of Ca^{2+} competent cells by gently tapping the bottom of the vial 2-3 times. The vial was incubated on ice for 30 min and then

heated shock for 30 sec at 42°C without shaking. After cooling down on ice for 2 min, 900 µl of prewarmed LB medium were added into each vial, followed by incubation at 37 °C for 1 h at 250-300 rpm in a shaking incubator. After centrifugation at 1,500rpm for 1 min, the supernatant (approximately 700 µl) were removed and the remaining 300 µl were cultivated on a prewarmed LB-Amp (100 µg/ml) plate with 40 µl of Xgal spread on top of it. The cells were cultivated overnight at 37 °C. Then, the positive colonies (color in white, instead of blue) were picked up and incubated in 5ml LB-Amp overnight at 37°C with shaking.

After overnight incubation, the plasmid DNA was isolated. Cells were pelleted by centrifugation for 30 sec at 11,000 x g. The supernatant was discarded and cells were lysed by adding 250 µl Buffer A1, resuspending thoroughly and then adding 250 µl Buffer A2. After gently mixing, the suspension was incubated at RT until lysate appeared clear. 300 µl Buffer A3 were added, followed by centrifugation for 5-10 min at 11,000 x g to clarify the lysate. The lysate was then pipetted onto a NucleoSpin® plasmid/plasmid (Macherey-Nagel, Düren, Germany) column and centrifuged for 1 min at 11,000 x g. The column was washed by adding 600 µl Buffer A4 and centrifugation for 1 min at 11,000 x g. The plasmid DNA was eluted by adding 50 µl ddH_2O, incubating for 1 min and centrifugation for 1 min at 11,000 x g.

Before sending the plasmid DNA for sequencing, the COBRA PCR products were further checked by performing a restriction digestion. The following reagents were mixed, incubated at 37°C for 1 h and loaded on 2 % agarose gel.

Reagent	Quantity taken
Tango buffer	2 µl
Sal1	0.5 µl
NcoI	0.5 µl
Plasmid DNA	7 µl
Total volume	10 µl

Positive probes were then sequenced using T7 primers.

2.9 RNA extraction, cDNA reverse transcription and qRT-PCR analysis

For spermatozoa, semen sample was thawed and diluted with PBS. 5×10^7 cells per tube of these sperm suspension were taken and washed twice in PBS. Spermatozoa were lysed using Ultra Turrax (IKA, Staufen, Germany) for 30 sec. 1 ml TriZol reagent and 25 µl DTT (0.1M) were added. After vortexing for 2 min and centrifugation at 4°C, 13000 rpm for 5 min, supernatant were mixed with 300 µl chloroform, incubated at room temperature for 10min and centrifuged at 13000 rpm, 4°C for 20 min. RNA was precipitated from upper phase. 1 µg of RNA were then mixed with 4 µl MMLV buffer, 8 ul dNTP, 1 µl Hexamera(10 pmol/L), 1 µl Poly dT (10 pmol/L) and added with water up to 20 µl. After incubation in 62 °C for 10 min, 0.5 µl of MMRL-RT and 0.5 µl of RNAsin were further added. The mixture was then incubated in 42°C for 1 h and in 90°C for 5 min, sequentially. The cDNA were kept at -20°C for storage.

Bovine oocytes, zygotes and 2-, 4-, 8-cell embryos, morulae and blastocysts were collected in 100 µl TriZol reagent and frozen in liquid nitrogen. Each probe contains 20 oocytes or 20 early embryos, respectively. Due to the low number of cells, cDNA preparation was achieved using FastLane cell cDNA kit (Qiagen, Hilden, Germany). Briefly, 500 µl Buffer FCW were added into the cells and discarded. Then, 200 µl Buffer FCP were added, followed by incubation for 5 min at room temperature. The cell lysate (containing RNA) were transferred into another tube, mixed with the following components and incubated for 5 min at 42°C to eliminate the genomic DNA.

Component	Volume/reaction	Final concentration
gDNA wipeout buffer, 7x	2 µl	1x
Cell lysate	4 µl	
RNase-free water	8 µl	
Total volume	14 µl	–

Then prepare the reverse-transcription reaction components according to the following table and add template RNA. After incubation for 30 min at 42°C and incubation for 3 min at 95°C, the cDNA was ready to be used for qRT-PCR analysis.

Component	Volume/reaction	Final concentration
Quantiscript reverse transcriptase	1 µl	
Quantiscript RT buffer	4 µl	1x
RT primer mix	1 µl	
Cell lysate	14 µl	
Total volume	20 µl	–

qRT-PCR was performed using rotor gene cycler (Qiagen, Hilden, Germany) and 50ng cDNA as template. Comparative quantification for expression was automatically calculated. Amplified PCR products were then spread on a 2% agarose gel.

PCR program

Cycle step	Temperature(°C)	Time(sec)	Cycles
Initial denaturation	95	60	1
Denaturation	95	30	40
Annealing	52-60*	30	
Extension	72	30	
Final extension	72	300	1

*Annealing temperature varies according to different primer pairs

3. Results

3.1 Characterization of the nucleosomal fraction isolated from human and bovine sperm

Experimental protocols established on sperm (Hammoud et al., 2009; Zalenskaya et al., 2000) allowed us a proper separation of nucleosomal from protamine-associated chromatin and a subsequent extraction of nucleosomal DNA and proteins, respectively (Figure 7, 8A and 8B). Analyzing the DNA isolated from the nucleosomal fraction of human and bovine sperm, we found that this DNA comprised not only the expected 146bp mono-nucleosomal DNA and its multiples (146x n), respectively, but contained also high- and low-molecular DNA fractions of uncertain origin (Figure 8B). Therefore, we cleaned up the 146 bp mono-nucleosomal DNA before its utilization for high throughput sequencing. Sequencing was performed by Illumina technique and all putative nucleosomal binding sites were analyzed in a genome-wide manner.

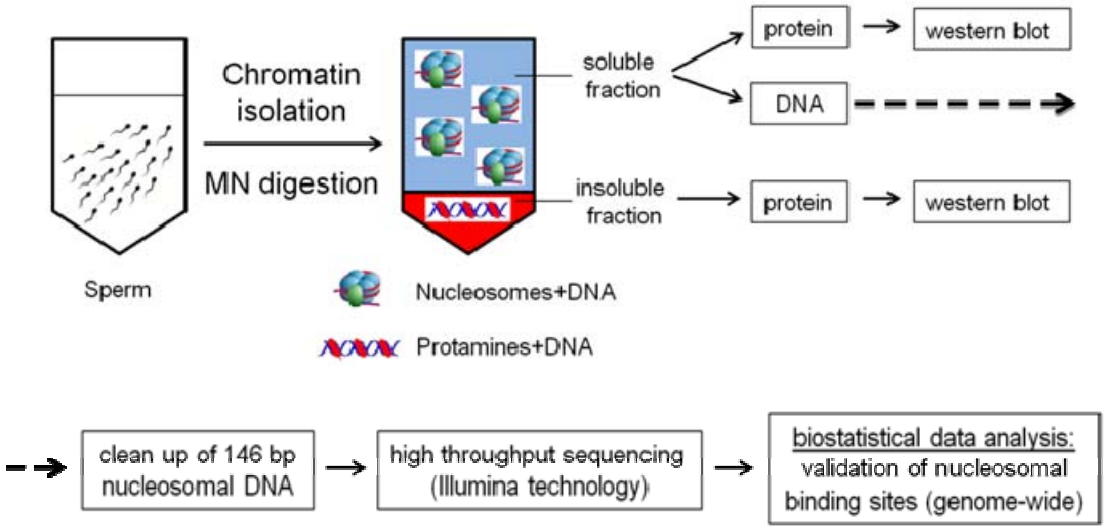


Figure 7 Preparation of the 146 bp mono-nucleosomal DNA from human and bovine sperm. Scheme of workflow for separation of soluble nucleosomal from insoluble protamine fraction from sperm with subsequent isolation and analysis of proteins and of the 146 bp mono-nucleosomal DNA by indicated procedures.

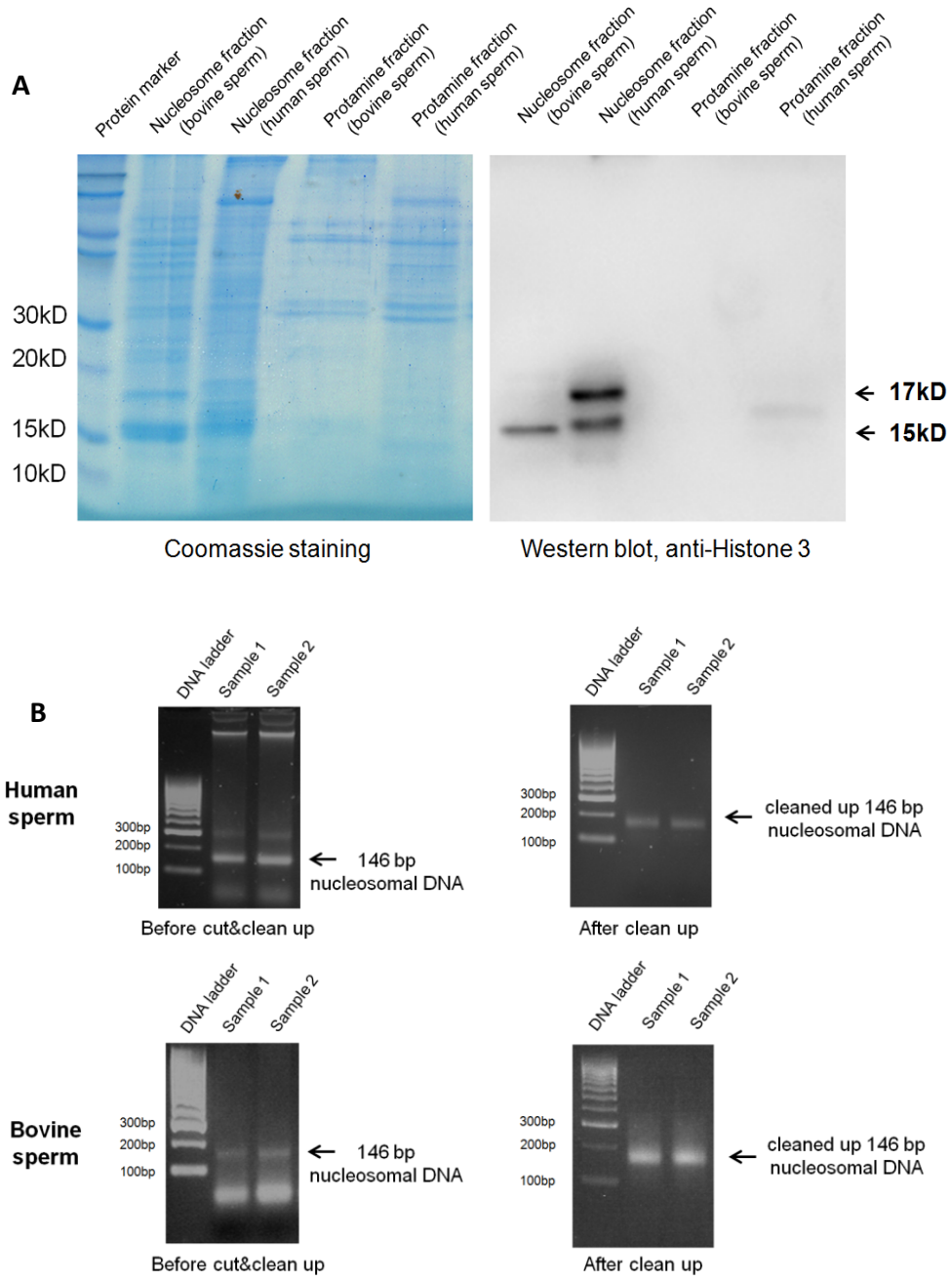


Figure 8 Confirmation of the 146 bp mono-nucleosomal DNA from human and bovine sperm. A) Successful preparation of the nucleosomal fraction from human and bovine sperm was confirmed by western blot analysis using antibody against histone 3; **B)** Total DNA from nucleosomal fraction prepared from human (upper panel) and bovine sperm (lower panel) was separated on 2% agarose gel, and the 146 bp mono-nucleosomal DNA fragment was cleaned up for further utilization in high throughput sequencing.

3.2 Distribution pattern of retained nucleosomes in sperm genome

In human sperm genome, we found 99,626 putative nucleosomal binding sites, which covered around 147 million nucleotides and comprised 4.8% of the paternal genome. In bovine sperm genome, 299,000 putative nucleosomal binding sites covered around 374 million nucleotides and comprised 14% of the paternal genome. Nucleosome-binding sites were evenly spread along the human and bovine chromosomes as exemplary shown in Figure 9A and 9B.

Next, we analyzed whether there was a difference between the chromosomes regarding the amount of retained nucleosomes. In human sperm, we found an enrichment of nucleosomes particularly in gonosomes (X-chromosome: randomly expected 4.2%, examined 7.3%; Y-chromosome: randomly expected 0.6%, examined 1.3%), whereas all autosomes exhibited values comparable with a random distribution (Figure 10 and 11). In bovine sperm, the X-chromosome exhibited also an enrichment of retained nucleosomes (randomly expected 2.7%, examined 3.8%). Interestingly, bovine Y-chromosome showed a strong nucleosome-depletion (randomly expected 3.5%, examined 0.2%).

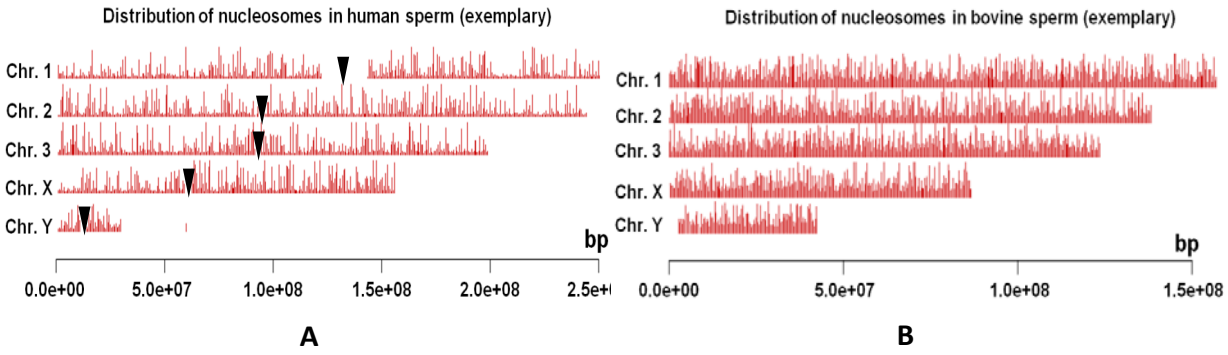


Figure 9 Distribution of retained nucleosomes along human and bovine sperm chromosomes. Sperm nucleosomes are scattered along the human **(A)** and bovine **(B)** chromosomes (chromosomes 1, 2, 3, X and Y are shown exemplarily; human centromeres are indicated with triangles; bovine centromere localization is currently unknown).

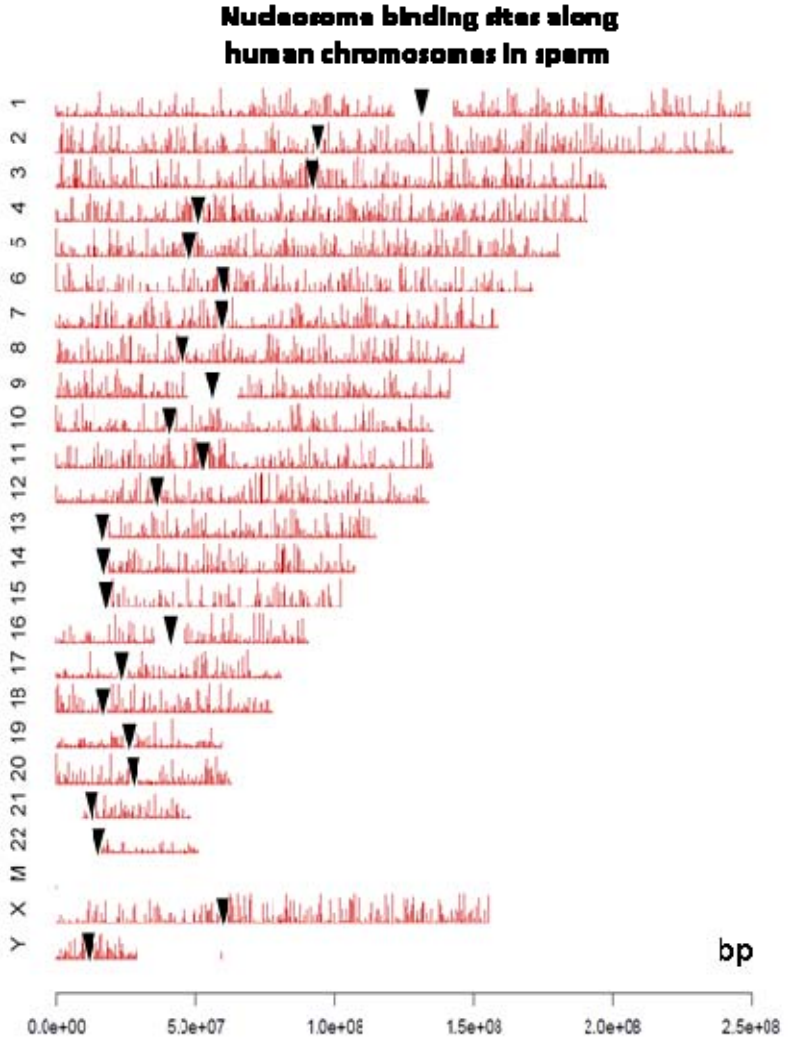


Figure 10 Distribution of retained nucleosomes along human chromosomes. In human sperm, both gonosomes show a high enrichment of nucleosomes, whereas the autosomes show values comparable with a random distribution. Centromere regions in 22 human autosomes and 2 gonosomes are indicated as black triangles.

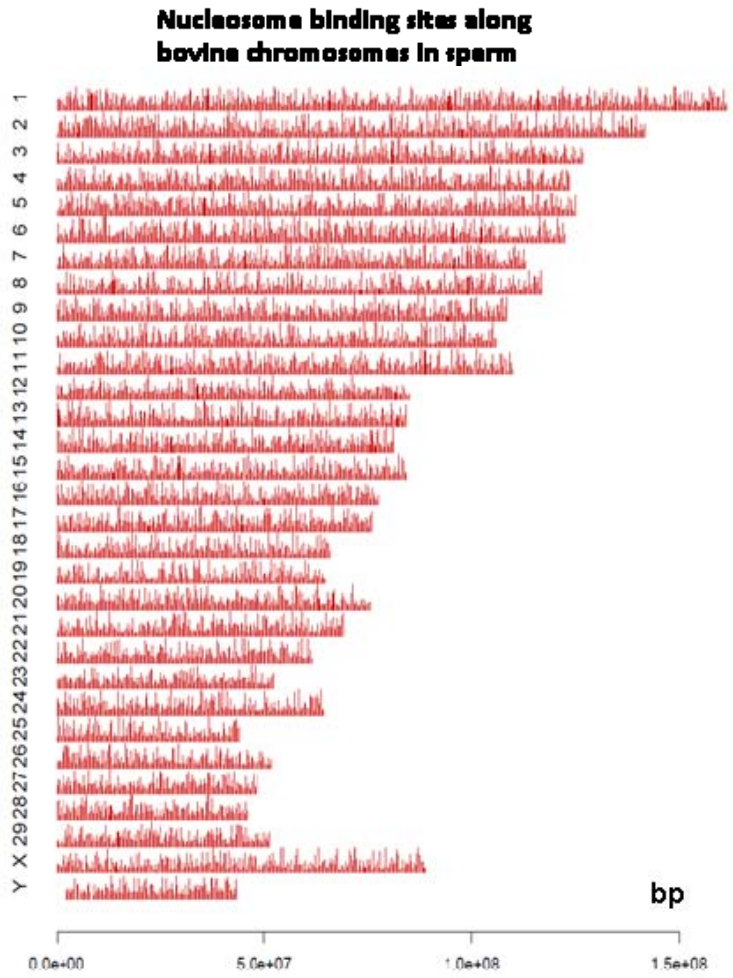


Figure 11 Distribution of retained nucleosomes along bovine sperm chromosomes. In bovine sperm, X-chromosome shows a nucleosome-enrichment and Y-chromosome - a strong nucleosome-depletion. Autosomes exhibit values comparable with a random distribution. Bovine centromere regions are not localized yet and thus, are not indicated.

3.3 Association of sperm nucleosomes to different types of repetitive DNA elements

The major part of nucleosome binding sites in sperm of both species was located in repetitive DNA elements (human: 84%; bovine: 85%). Repetitive DNA elements are scattered throughout the mammalian genome and are abundant in CpG-methylation, especially near retrotransposons. We supposed that the heterochromatic state of certain repetitive elements might affect the protamine incorporation into DNA and lead

to nucleosome retention. In this context we analyzed in sperm the distribution of nucleosome binding sites in different repetitive sequences (retrotransposons: LINEs, SINEs and LTRs; DNA transposons; low complexity repeats and simple sequence repeats).

Pre-eminently, a significant enrichment of nucleosomes could be detected in human sperm near centromere regions within so-called centromere-specific repeats as exemplarily shown for chromosomes 2, 5 and X (Figure 12). In bovine sperm chromosomes, we could also detect a remarkable nucleosome-enrichment especially in centromere repeats (Figure 13), which were scattered along the chromosomes. The exact localization of functional bovine centromeres is currently unknown and all the more interesting are our detected sites.

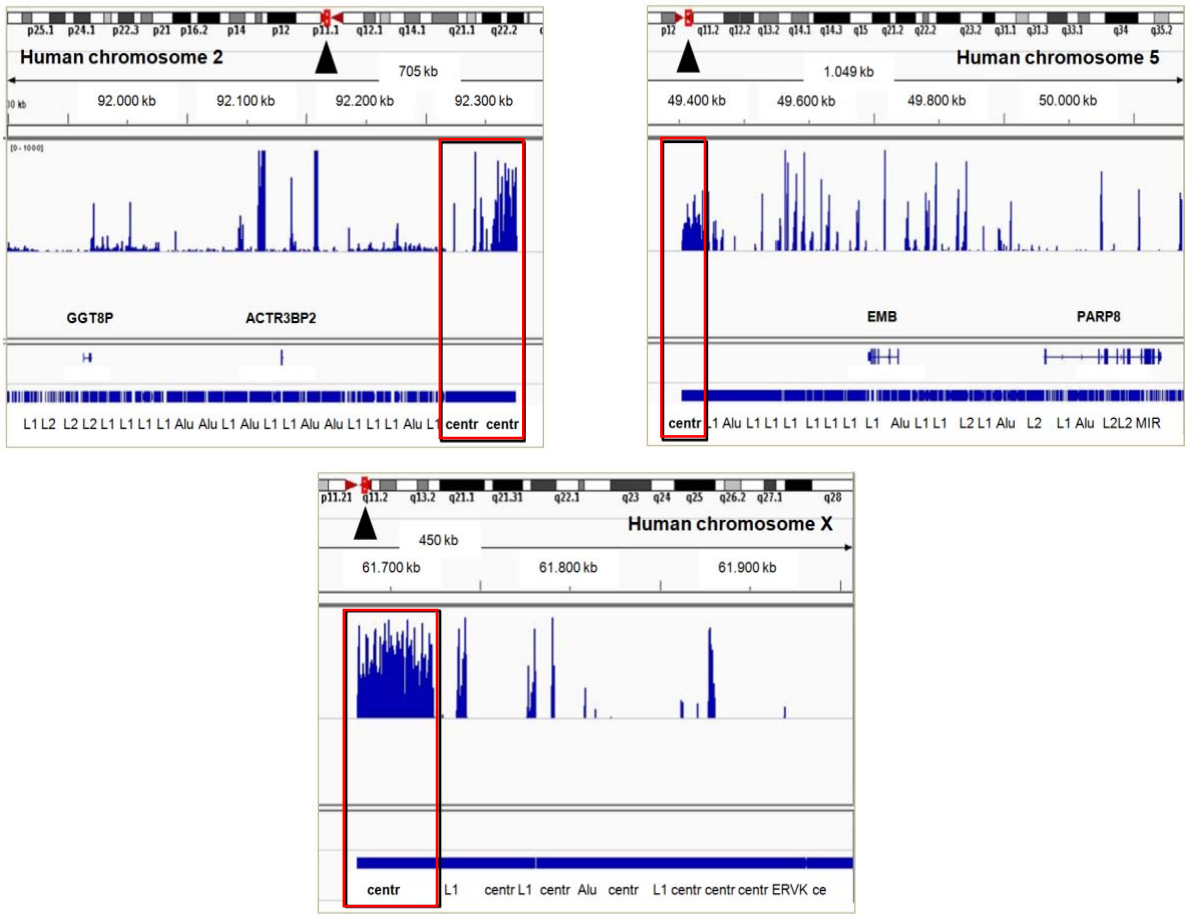


Figure 12 Distribution of retained nucleosomes in human centromere repeats. Exemplary IGV-screen shots of human centromeres in chromosomes 2, 5 and X representing an enrichment of sperm-

nucleosomes at centromere repeats (highlighted box) (triangles indicate screen shot region; retained nucleosomes, genes and repetitive sequences are shown). IGV: Integrative Genomics Viewer.

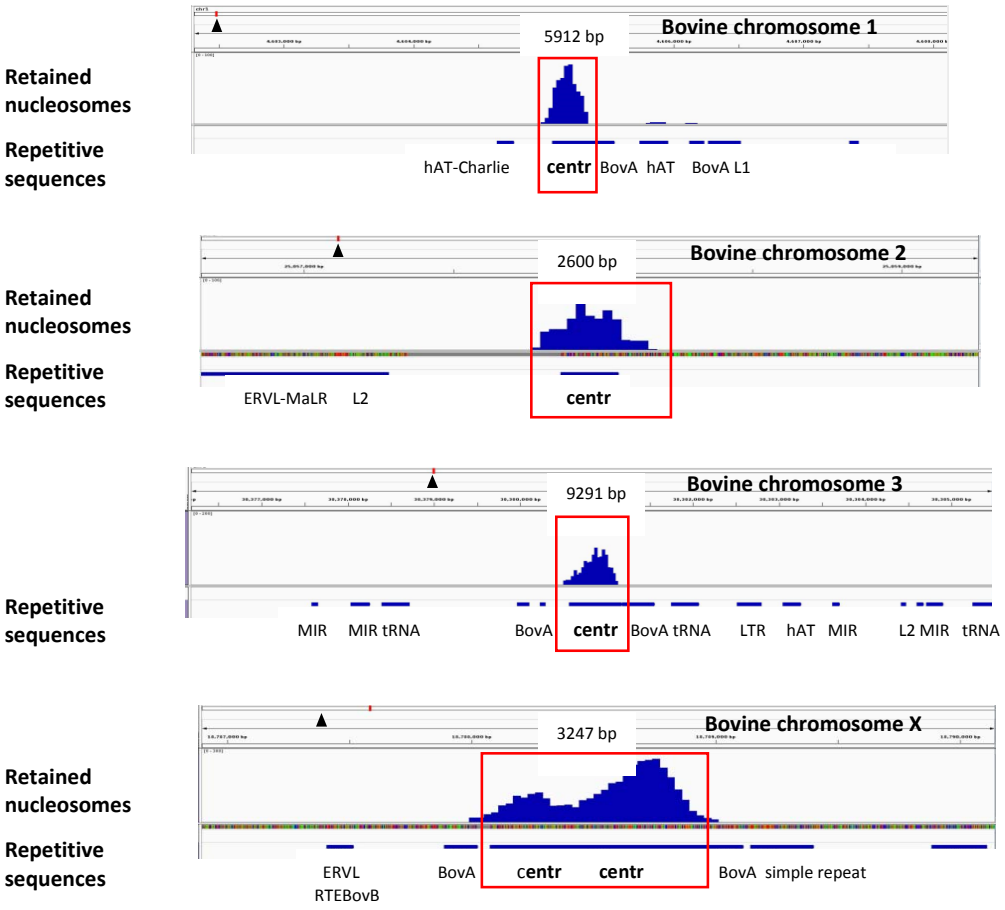


Figure 13 Distribution of retained nucleosomes in bovine centromere repeats. Exemplary IGV-screen shots in bovine chromosomes 1, 2, 3 and X demonstrating exemplary an enrichment of nucleosome binding sites in bovine sperm specifically in centromere repeats (triangles indicate screen shot region; retained nucleosomes, genes and repetitive sequences are shown). IGV: Integrative Genomics Viewer.

In sperm of both species, genome-wide evaluated, we found a significant nucleosome-enrichment in LINEs, especially in LINE1, and SINEs (Figure 14A.1 and 14A.2). Analysis of intragenic LINEs and SINEs showed the same pattern (Figure 14B.1 and 14B.2). In contrast to LINEs and SINEs, which were generated by amplification of themselves, the retrotransposable LTR elements with a viral origin were, in human sperm, nucleosome-depleted. LTRs in bovine sperm showed no clear tendency regarding their association to nucleosomes. When analyzed genome-wide, we detected

a slight depletion, whereas in intragenic LTRs, we found a slight enrichment. Massive nucleosome-depletion could be further examined in human as well as in bovine sperm particularly within low complexity repeats (LCRs) and DNA transposons (Figure 14A.1, 14A.2, 14B.1 and 14B.2). In human sperm, we observed additionally a noticeable nucleosome-depletion in simple sequence repeats (SSRs, also called microsatellites).

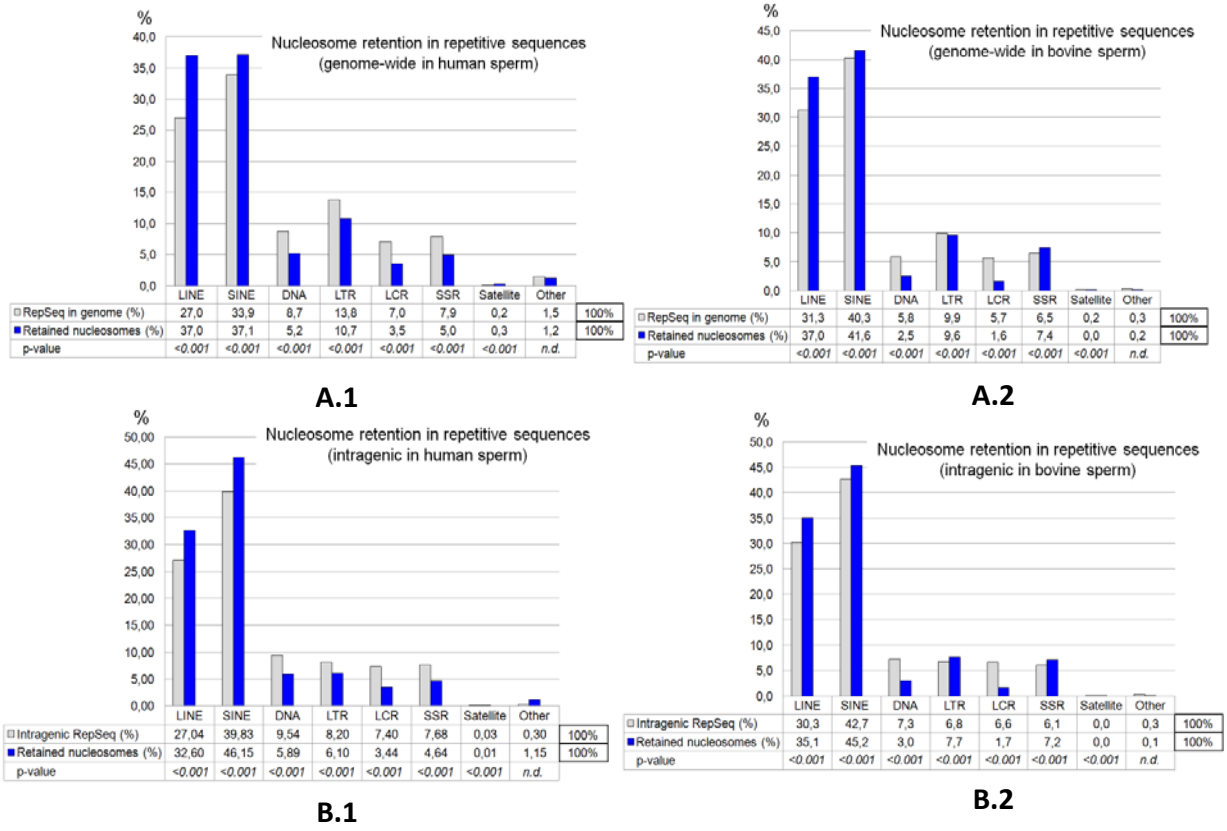


Figure 14 Association of retained nucleosomes in human and bovine sperm to different types of repetitive DNA elements. A) Proportions of common types of repetitive DNA elements (light grey bars) in whole human **(A.1)** and bovine genome **(A.2)** and proportions of repetitive DNA elements exhibiting retained nucleosomes (dark blue bars) (p-values of nucleosome-enrichment and nucleosome-depletion are indicated; RepSeq: repetitive sequence; LINE/SINE: long/short interspersed elements; DNA: DNA transposons; LTR: long terminal repeats; LCR: low complexity DNA repeats; SSR: simple sequence repeats); **B)** Proportions of common types of repetitive DNA elements (light grey bars) in intragenic area of human **(B.1)** and bovine genome **(B.2)** and proportions of repetitive DNA elements in there exhibiting retained nucleosomes (dark blue bars) (p-values of nucleosome-enrichment and nucleosome-depletion are indicated).

3.4 CpG-methylation analysis of sperm repetitive DNA elements

Based on our hypothesis that the heterochromatic state of certain repetitive elements might affect the protamine incorporation into DNA and lead to nucleosome retention, we further checked the CpG-methylation status of sperm nucleosomes-associated repetitive DNA elements in bovine sperm. A frequent methylation was observed in LINEs and SINEs (Figure 15, 3 out of 7 randomly selected LINEs and SINEs were hypermethylated). The same frequent methylation was observed in LTRs (Figure 15, 3 out of 4 randomly chosen LTRs were hypermethylated). Interestingly, neither LCRs, nor SSRs showed a CpG-methylation (Figure 15, all seven analyzed LCRs and SSRs were CpG-unmethylated).

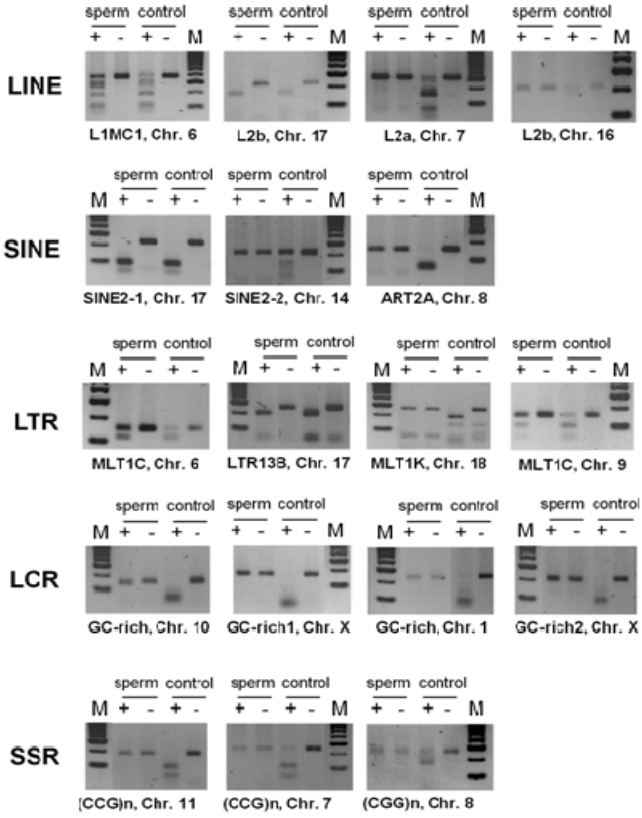
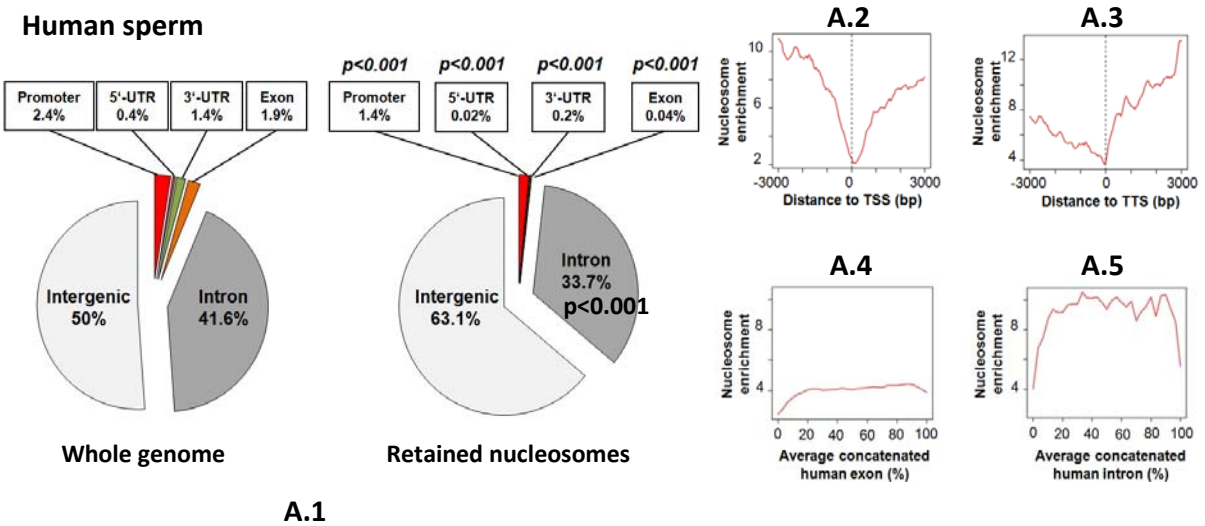


Figure 15 Analysis of CpG-methylation in different types of repetitive DNA elements in bovine sperm by COBRA technique. Analyzing several randomly chosen repetitive elements we revealed that CpGs in LINEs, SINEs and LTRs are often methylated. In contrast, all analyzed LCRs and SSRs were unmethylated; **M**: 100 bp marker; **„+“** digestion with CG-specific enzyme; **„-“** mock digestion; **LINE**: long interspersed nuclear elements; **SINE**: short interspersed nuclear elements; **LTR**: long terminal repeats; **LCR**: low complexity DNA repeats; **SSR**: simple sequence repeats (microsatellites).

3.5 Nucleosome occurrence in non-coding DNA and their absence in functional genome areas

To reveal whether sperm nucleosomes remain preferably in functional or non-functional genome areas we analyzed the distribution of nucleosome binding sites in exons, 5'-UTR, 3'-UTR, gene promoters (-3000 bp from TSS), introns and in intergenic sequences and compared the values to whole genome data (Figure 16A.1 and 16B.1). Remarkably, most sperm-nucleosomes (96.8% in human and 98.1% in bovine) were located in non-coding intergenic and intron DNA. An intense nucleosome-depletion was observed in sperm of both species particularly within exons, 5'-UTR and 3'-UTR, and in human sperm also within promoters (Figure 16A.1 and 16B.1). Human sperm exhibited a clear nucleosome-enrichment in intergenic area, which was accompanied by nucleosome-depletion in introns. In bovine sperm we observed a slightly different tendency, which was probably attributable to incomplete validated non-coding bovine DNA. Furthermore in both species, transcription start sites (TSS) and transcription termination sites (TTS) lacked nucleosomes (Figure 16A.2, 16A.3, 16B.2 and 16B.3). Average concatenated human and bovine exons exhibited nucleosome-depletion profiles, whereas average concatenated introns showed intense nucleosome-enrichment profiles (Figure 16A.5 and 16B.5).



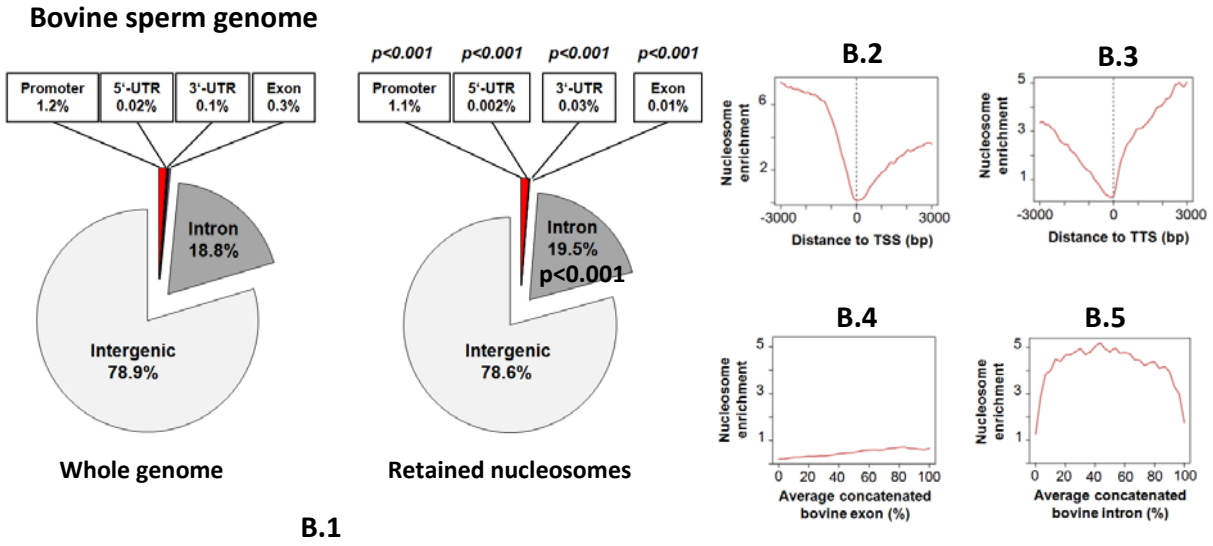


Figure 16 CEAS-distribution of retained nucleosomes in coding and non-coding areas of human and bovine sperm genome. **A.1** and **B.1**, left) Percentage of DNA in different human and bovine genome regions: intergenic, intron, exon, promoter (-3000 bp), 5'-UTR and 3'-UTR; A.1 and B.1, right) Distribution of retained nucleosomes in human and bovine sperm genome revealing their accumulation in intergenic and intron regions. Nucleosome-depletion was observed in functional genome areas, especially in exons, in 5'-UTR, 3'-UTR and slightly in gene promoters (CEAS-evaluated p-values are indicated); **A.2** and **B.2**) Average profile for nucleosome occurence at transcriptional start sites (TSS) in human and bovine sperm; **A.3** and **B.3**) Average profile for nucleosome occurence at transcriptional termination sites (TTS) in human and bovine sperm; **A.4** and **B.4**) Average concatenated human and bovine exon exhibiting a nucleosome-depletion profile; **A.5** and **B.5**) Average concatenated human and bovine intron exhibiting a nucleosome-enrichment profile.

3.6 Analysis of nucleosome-free genes in human and bovine sperm considering their functional impact

By performing direct sequencing of mono-nucleosomal DNA without antibody-precipitation, we observed a nucleosome binding scenario in prompter regions, which was different from previous studies. Among 43.8% human and 25% bovine genes in sperm exhibiting complete nucleosome-free promoters (-3000bp) and gene bodies (Figure 17, table), we surprisingly found a significant enrichment of all *HOX* genes (Table S1, GO term enrichment). IGV-screen shots confirmed the absence of nucleosome-binding sites in entire *HOX*A, *HOXB*, *HOXC* and *HOXD* clusters in human

as well as in bovine sperm genome (Fig. 18A: exemplary screen shot for human *HOXA* cluster; see also Figure S1 for all four human *HOX* clusters and Figure S2 for all four bovine *HOX* clusters). The SSRs and LCRs abundances of *HOX* genes, which had been shown by other studies (Mainguy et al., 2007; Huang et al., 2009) were also proved by our sequencing data (see Figure 18A). Interestingly, among human and bovine overlapping nucleosome-free genes (n=1665) we further found an enrichment of factors, which were also in part highly conserved (Elsik et al., 2009), namely genes relevant for organ development, morphogenesis, regulation of biosynthetic and metabolic processes, and response to different stimuli (Figure 17, left TOP5 box; see also Table S1 for enriched GO terms). Considering the functional impact of nucleosome-free genes, we suggest them to represent "Genes for embryogenesis executive program" and to be functional relevant in post-implantation embryogenesis.

Comparison of human and bovine homologous genes

	human	**bovine**
Nucleosome-free genes	43.8%	25%
Nucleosomes in gene body	44%	38.7%
Nucleosomes in promoter and gene body	7.4%	25.7%
Nucleosomes in promoter	5.3%	10.4%

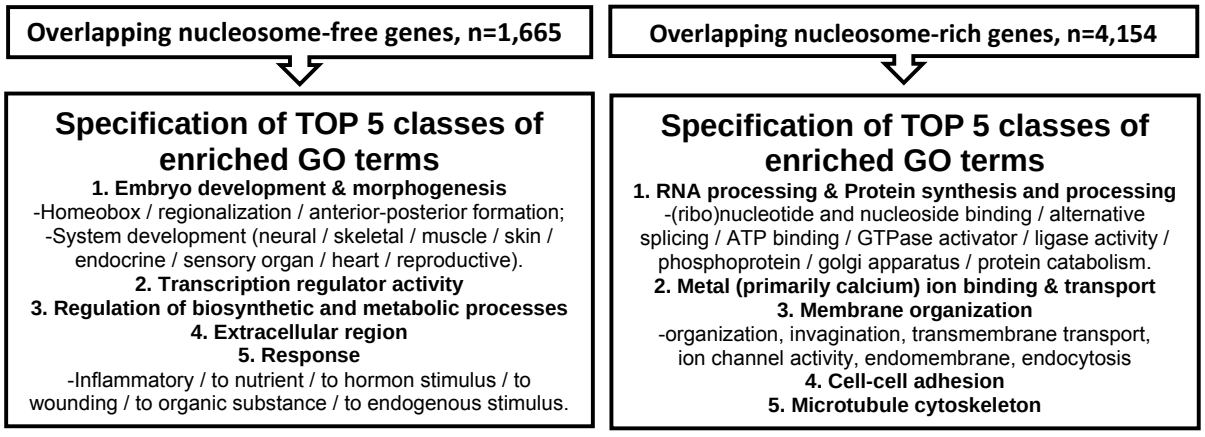


Figure 17 Intragenic distribution of retained nucleosomes in human and bovine sperm genome considering gene promoters (-3,000bp) and gene bodies (exon and intron). Table demonstrates the proportions of genes in human and bovine sperm genomes with and without retained nucleosomes. Comparison of human and bovine data revealed a functional overlap, i.e. an enrichment of same GO

terms, regarding nucleosome-free and nucleosome-rich genes. Specifications of TOP 5 classes of enriched GO terms for overlapping nucleosome-free and nucleosome-rich genes are shown in boxes.

Figure 18 Exemplary gene with nucleosome depletion and enrichment in promoter region and gene body. A) IGV-screen shot of *HOXA* gene cluster in human sperm demonstrating nucleosome-depletion along the whole cluster (genes and associated repetitive sequences are indicated) (see also Figure S1 and S2); **B)** IGV-screen shot of *ILF2* gene in bovine sperm demonstrating nucleosome enrichment in gene body and promoter (associated repetitive sequences are indicated).

3.7 Analysis of nucleosome-rich genes in human and bovine sperm considering their functional impact

A relative big part of genes (human: 44%, bovine: 38.7%) exhibited scattered nucleosomes exclusively in their gene bodies (Figure 17, table). 7.4% of human and 25.7% of bovine genes comprised nucleosomes in both, promoter and gene body. We found that “nucleosome-rich” genes were abundant in LINEs and SINEs as shown exemplary in case of *ILF2* (Figure 18B). GO term analyses of overlapping human and bovine nucleosome-rich genes (n=4,154) revealed pre-eminently an enrichment of factors for RNA- and protein-processing, for metal (primarily calcium)-ion binding and transport, for membrane organization, for cell-cell adhesion and for microtubule cytoskeleton organization (Figure 17, right TOP5 box; see also Table S2 for enriched

GO terms). Further, a small proportion of genes (human: 5.3%, bovine: 10.4%) exhibited nucleosome binding sites exclusively in their promoters (-3,000 bp) (Figure 17). Functional annotation of overlapping nucleosome-free promoters (n=286) showed an enrichment of factors crucial for cell-cell adhesion, calcium ion binding and RNA-processing (Figure 19, TOP3 box; see also Table S3 for enriched GO terms). Calcium-dependent cell adhesion events with participation of Cadherins coordinate the cellular allocation and spatial segregation of inner cell mass in blastocyst, and are crucial for early morphogenesis (Fleiming et al., 2001). However, we found the promoter of gamma protocadherin C3 (*PCDHGC3*) to be nucleosome-enriched in human as well as in bovine sperm. For "nucleosome-rich" genes we suggest the umbrella term "Genes for embryogenesis initializing program". These genes are probably those, which have to be euchromatic and active shortly after fertilization to start the paternal transcriptional and translational machinery and to ensure the basic requirements in forming of an early embryo. Their abundance on retrotransposons might reflect their predisposition to evolutionary variances.

Comparison of human and bovine homologous genes

Nucleosome-rich promoter (human sperm) n=2,386		Nucleosome-rich promoter (bovine sperm) n=4,091	
CpG promoter	Non CpG promoter	CpG promoter	Non CpG promoter
41%	59%	39%	61%

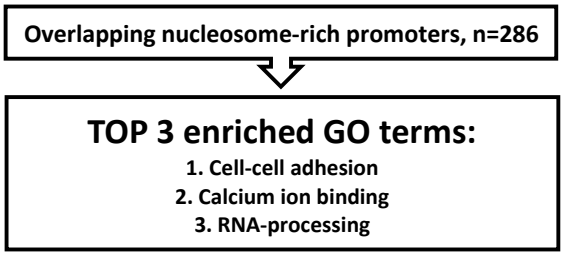


Figure 19 Intragenic distribution of retained nucleosomes in human and bovine sperm genome considering gene promoters (-3,000bp) and gene bodies (exon and intron). Table shows the proportions of CpG- and non-CpG-promoters among nucleosome-rich promoters in human and bovine sperm. Corresponding genes were analyzed and overlapping GO terms were evaluated (see TOP 3 enriched GO terms).

3.8 Nucleosome-occurrence in CpG-promoters within the sperm genome

Here we analyzed, whether nucleosome-free, i.e. protamine-occupied, CpG-promoters differed from nucleosome-rich CpG-promoters regarding their DNA-methylation status. We examined eight totally nucleosome-free gene promoters and six promoters, with the highest amount of retained nucleosomes (>70% of promoters was occupied by nucleosomes). Using COBRA technique, we found that CpG-promoters were in general hypomethylated regardless of their occupancy with protamines or nucleosomes (Figure 20). This result was further confirmed by our bisulfite-sequencing results (Figure 23).

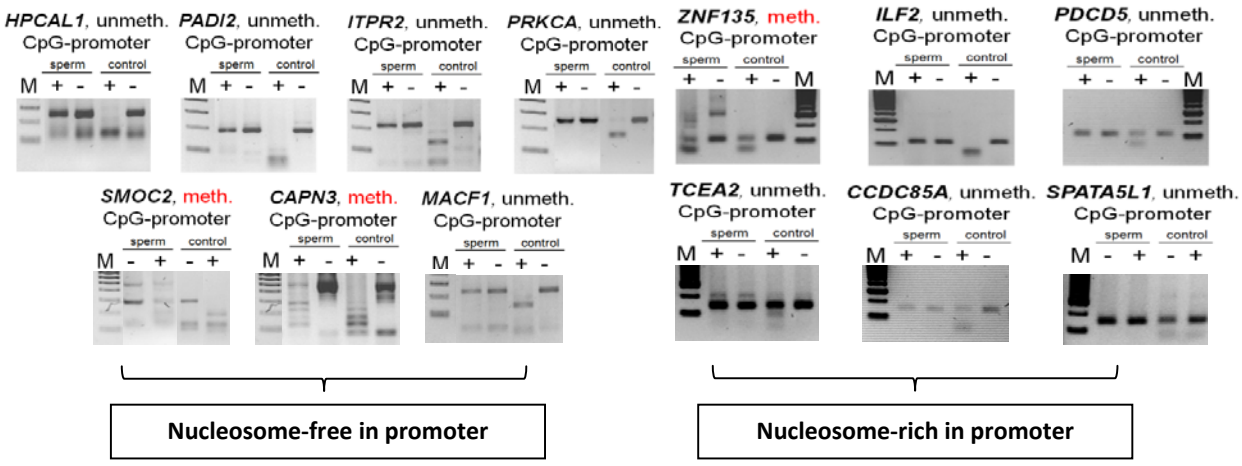


Figure 20 COBRA analysis of genes exhibiting nucleosome-free CpG-promoters (left) and nucleosome-rich CpG-promoters (right) in bovine sperm. (**M**: 100 bp marker, **„+“** digestion with CG specific enzyme, **„-“** mock digestion. Some gels were cut and put together to make a full picture, e.g. the marker of *SMOC2* and *CAPN3*, and the control lanes of *PRKCA*).

Next, we analyzed five paternal imprinted and two maternal imprinted bovine genes with nucleosome-free CpG-promoters. In accordance with generally accepted knowledge, our results showed that only maternal imprinted genes exhibited intense methylated CpG-promoters, whereas paternal imprinted genes were all unmethylated (Figure 21; see also Figure 23). Moreover, there was no preference in nucleosome retention regarding CpG- or non-CpG-promoters (Figure 19, table).

Furthermore, five *HOX* genes and five pluripotency genes, which in sperm exhibited completely nucleosome-free gene bodies and promoters (Figure 17), were analyzed by

COBRA technique. The result showed that their CpG-promoter were also in general hypomethylated (Figure 22).

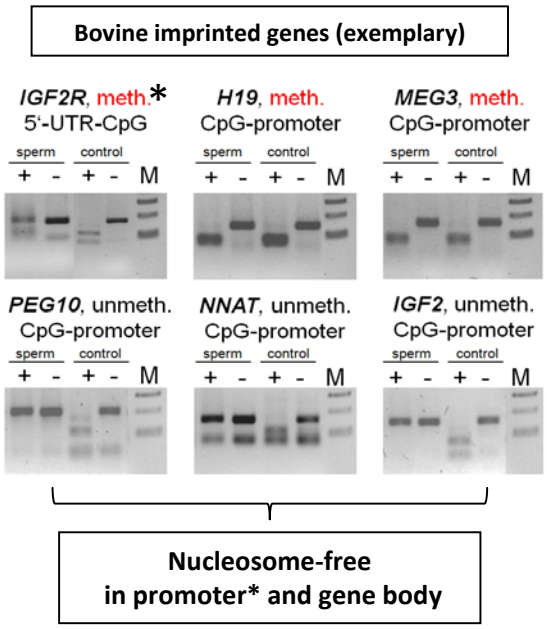


Figure 21 COBRA analysis of bovine imprinted genes in sperm. CpG-promoter methylation was analyzed in three maternal (upper panel) and three paternal imprinted genes (lower panel) (**M**: 100 bp marker, **„+"** digestion with CG specific enzyme, **„-"** mock digestion. Some gels were cut and put together to make a full picture, e.g. the sperm lanes of *IGF2R*); *All bovine imprinted genes with CpG-promoters were nucleosome-free in their promoters, except *IGF2R*.

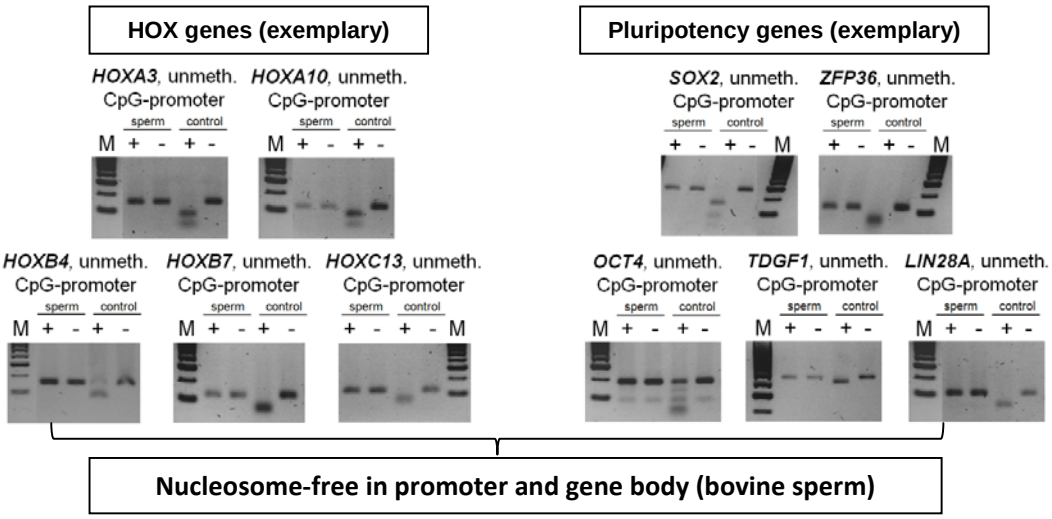


Figure 22 COBRA analysis of *HOX* and pluripotency genes in bovine sperm. CpG-promoter methylation was analyzed exemplary in five *HOX* genes (left) and five pluripotency genes (right) without retained nucleosomes (**M:** 100 bp marker, **„+“** digestion with CG specific enzyme, **„-“** mock digestion. Some gels were cut and put together to make a full picture, e.g. the marker of *HOXA3, HOXA10, SOX2 and LIN28A*).

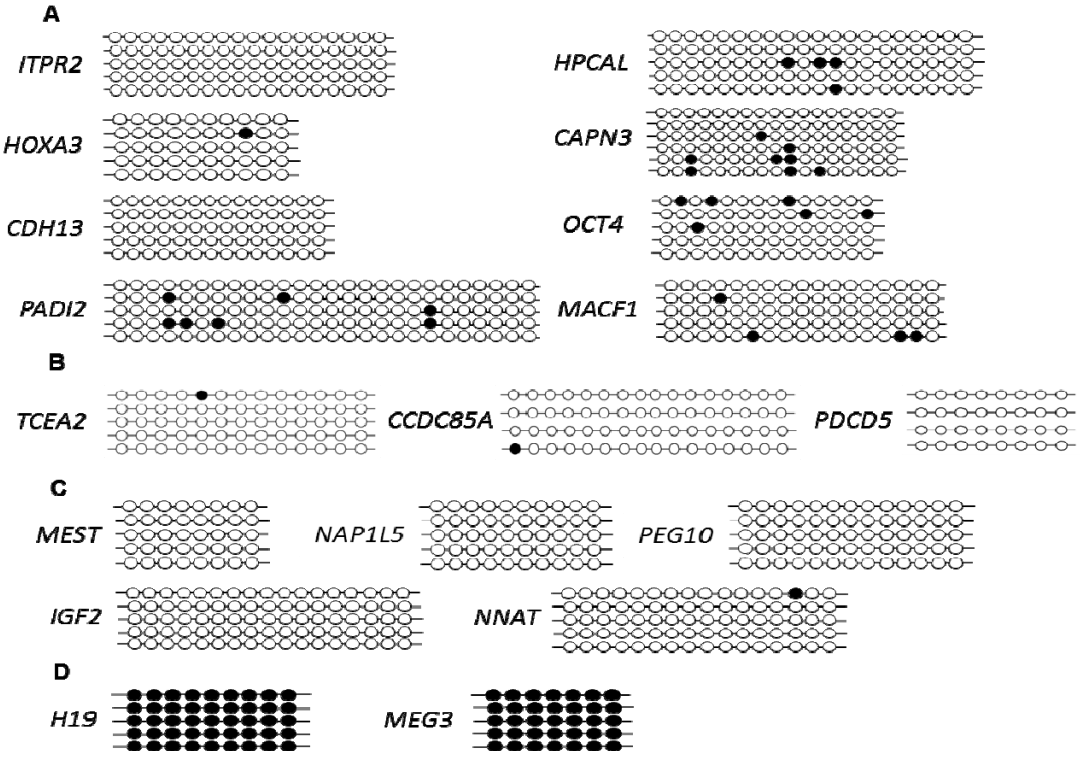


Figure 23 DNA methylation analyses in bovine sperm by bisulfite-sequencing in different CpG-promoters (-3,000bp) with different content of retained nucleosomes. A) Randomly selected genes with nucleosome-free CpG-promoters, i.e. solely protamine-occupied promoters, show hypomethylated DNA; **B)** Randomly selected genes with nucleosome-rich CpG-promoters (>70% of CpG-promoter is nucleosome-occupied) show also hypomethylated DNA; **C)** CpG-promoters of paternal imprinted genes show hypomethylated DNA; **D)** Only CpG-promoters of maternal imprinted genes show highly methylated DNA.

3.9 Impact of sperm-derived nucleosomes in post-fertilization activation of genes

In all animals, the initial events of embryogenesis are controlled by maternal descendant proteins and RNAs, and the major embryonic genome activation in mammals is supposed to start during 2-cell up to 16-cell stages (Bensaude et al., 1983; Misirlioglu et al., 2006). The question, whether in early embryos the sperm-contributed nucleosomes, i.e. specific histone-modifications in “developmental” promoters as suggested before, are responsible for expressional activation of corresponding genes is difficult to analyze and still a matter of debate. Here, we examined whether there was a measurable difference regarding the post-fertilization expression between sperm-derived nucleosome-free and nucleosome-rich promoters. We utilized RNA samples isolated from bovine sperm, oocytes and early embryos (zygote, 2-cell, 4-cell stage, morula and blastocyst). We analyzed exemplarily the relative mRNA-levels of (1) six randomly selected genes with 65% up to 100% of promoter occupied by nucleosomes and seven genes with >30% of gene body occupied by nucleosomes (Figure 24), (2) thirteen known bovine imprinted genes (Tveden-Nyborg et al., 2008), which were all in sperm nucleosome-free in their gene bodies and promoters, except *IGF2R*-promoter (Figure 25), and (3) five *HOX* genes and five pluripotency genes (Figure 26), which in sperm exhibited completely nucleosome-free gene bodies and promoters.

In sperm, all analyzed transcripts were absent, whereas in oocytes several transcripts were stored. At a first glance, we could not detect obvious differences between the expression pattern of sperm-derived nucleosome-free and nucleosome-rich promoters. However, clear evident was the fact that regardless of nucleosome-association of respective promoter in the sperm, a transcript was detectable in early embryos up to morula and blastocyst stage mostly, when it was already present in the oocyte (e.g. *PRKCA* vs. *HPCAL1* in Figure 24; *CCDC85A* vs. *SPATA5L1* in Figure 25; *HOXA3* vs. *HOXA10* in Figure 26).

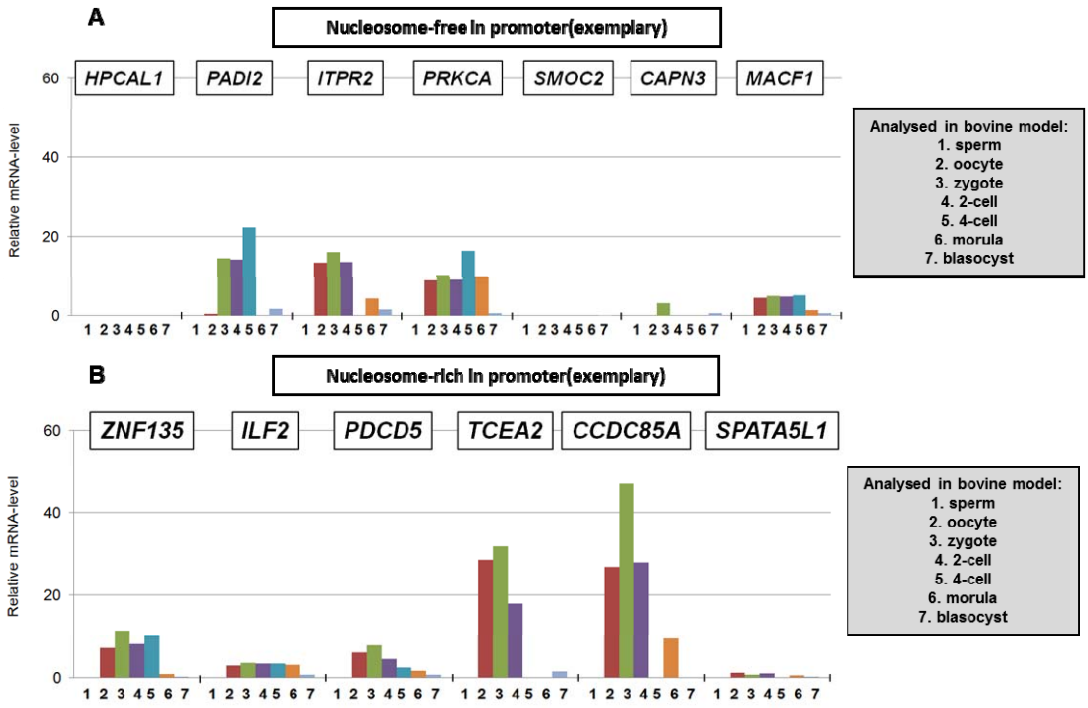


Figure 24 Expression analysis of genes with nucleosome-free promoters and nucleosome-rich promoters in bovine model. mRNA-level of corresponding genes in sperm, oocyte and in early embryo stages were comparatively quantified by qRT-PCR analyses. **A)** genes with nucleosome-free promoters, **B)** genes with nucleosome-rich promoters (65% up to 100% of promoter occupied by nucleosomes)**.**

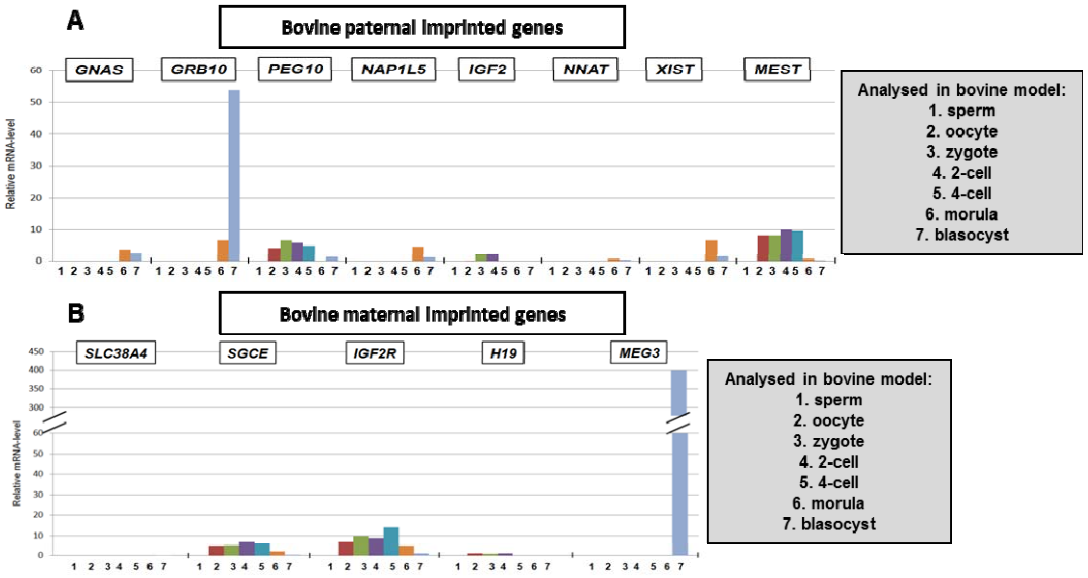


Figure 25 Expression analysis of bovine imprinted genes in bovine germ cells and early embryos. mRNA-level of corresponding genes were comparatively quantified by qRT-PCR analyses. **A)** paternal

imprinted genes in bovine germ cells and in early embryo stages; **B)** maternal imprinted genes in bovine germ cells and early embryo stages.

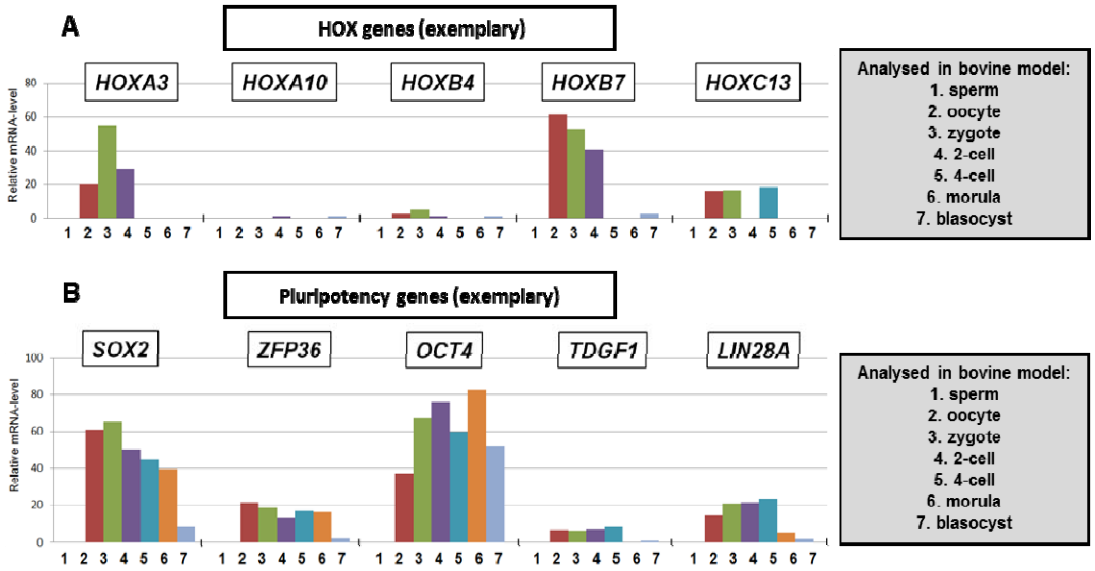


Figure 26 Expression analysis of *HOX* and pluripotency genes (all with completely nucleosome-free gene bodies and promoters) in bovine model. mRNA-level of corresponding genes in sperm, oocyte and in early embryo stages were comparatively quantified by qRT-PCR analyses. **A)** selected five *HOX* genes; **B)** selected five pluripotency genes.

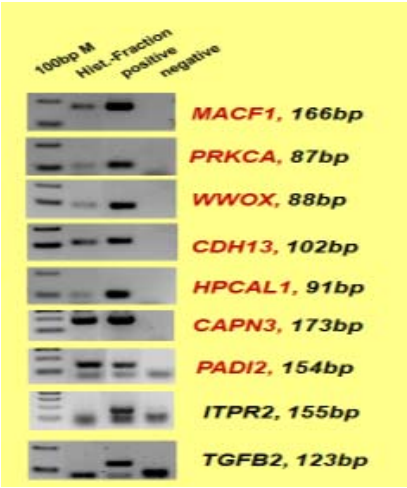
3.10 Difference between 146 bp nucleosome direct sequencing data and cross-linking ChIP-Seq data

To validate the direct sequencing data of 146 bp nucleosome, immuno-precipitated H3K9ac-associated bovine DNA was sent for Illumina sequencing. Based on ChIP-Seq results, 12 enriched genes (Figure 27A) were selected. The cross-linking ChIP-PCR results showed 9/12 genes with CpG promoter were associated with H3K9ac in the promoter region (-3000bp from TSS) (Figure 27B). However, the 146 bp nucleosome sequencing data showed significant difference with ChIP-PCR results. Only one gene, *CDH13*, was associated with nucleosomes in the promoter region. All other genes were either associated with nucleosomes in the gene body (1/3-3/3 gene) or had no association with nucleosomes at all (Figure 27C).

A

H3K9ac-enriched genes with CpG promoter	H3K9ac-enriched genes without CpG promoter
HPCAL1, CDH13, PADI2, MACF1, ITPR2, SMOC2, CAPN3, WWOX, TGFB2, PRKCA	*SLC24A1, LCP1*

B



C

Gene	-3,000bp TSS	1/3 gene	2/3 gene	3/3 gene
TGFB2	0.12	0.03	0.09	0.11
WWOX	0	0.07	0.06	0.07
PRKCA	0	0.11	0.08	0.11
LCP1	0.49	0	0.07	0.07
CDH13	0.23	0.1	0.14	0.12
PADI2	0	0.18	0.21	0.02
HPCAL1	0	0.1	0.03	0.14
ITPR2	0	0.07	0.12	0.12
CAPN3	0	0.14	0.35	0.09
MACF1	0	0.07	0.15	0.05
SLC24A1	0	0	0.11	0
SMOC2	0	0	0	0

Figure 27 Comparison of 146 bp nucleosome seq and cross-linking ChIP-Seq data. A) list of 12 genes enriched in H3K9ac-associated DNA and their CpG promoter conditions. **B)** ChIP-PCR results of H3K9ac-associated DNA, genes with association to H3K9ac were marked with red color. **C)** 146 bp nucleosome sequencing results regarding percentage of different gene positions associated with nucleosome (e.g. *CDH13* showed 0.23 in "-3,000bp TSS", which meant 23% of the promoter region of *CDH13* was associated with nucleosome).

4. Discussion

4.1 Pitfalls of current theories on histone-code

Based on data of recent published studies (Arpanahi et al., 2009; Brykczynska et al., 2010; Hammoud et al., 2009; Paradowska et al., 2012), there seemed to be a consensus regarding modified somatic-like histones in spermatozoa, that the DNA fragment associated with them contained mainly development relevant genes. Nevertheless, this theory somehow contains parts which need to be further clarified. Indeed, if developmental genes were retained in residual nucleosomes, they might bypass the chromatin decondensation steps after fertilization and start transcription immediately. However, both theoretical and practical problems will emerge based on this situation. As the importance of heterochromatinization in spermiogenesis lies in facilitating sperm transportation and above all, the protection of paternal genome and epigenome during the long and painstaking way to the oocyte, it is illogical to speculate that those genes, very important for early embryogenesis, would risk themselves the long way through female genital duct with high vulnerability, regardless of various uncertainties including hormone, pH change, immune factors and so on. If so, the whole chromatin compaction steps would seem pointless. On the other hand, whether those development relevant genes are really activated at the very beginning of early embryogenesis is still questionable. Studies have shown that the activation of embryonic genome in mouse occurs at the late two-cell stage (Bensaude et al., 1983; Misirlioglu et al., 2006), however, the studies from Paradowska et al (Paradowska et al., 2012), the only above study investigated the relevance of histone-related genes in early embryonic stages, showed H4K12ac-associated genes in mouse correlated weak with genes expressed at 4-cell stage, only until 8-cell and blastocyst stage the number of genes activated were increased (23 and 39 genes, respectively). Therefore, while the embryonic genome activation in mouse has already started, those H4K12ac-associated genes yet still waited to be transcribed, generating self-contradiction of this “histone code” theory.

4.2 From remnant nucleosomes to LINE1 and CENP-A – a function chain

Our sequencing results of 146 bp nucleosome DNA showed that the majority of nucleosome binding sites in both species (human: 84%; bovine: 85%) were located in repetitive DNA elements and evenly spread along the chromosomes. This finding is in accordance with a recent study by Meyer-Ficca et al (Meyer-Ficca et al., 2013). Using fluorescence in situ hybridization (FISH), they found that a large fraction of the histone-associated sperm genome was repetitive in nature, while a smaller fraction was associated with unique DNA sequences. More specifically, we further identified that the remnant nucleosomes in human as well as in bovine sperm were significantly enriched in LINEs (especially in LINE1) and SINEs, i.e. the non-LTR (long terminal repeat) retrotransposons. LINE1 (L1) retrotransposon belongs to the only active subfamily of LINEs and is a major group of interspersed repetitive elements that comprise 17% of the human genome. Active transcription and translation of L1 retrotransposons have been detected in a variety of cell types and implicated to be a potential regulator for cellular processes. The function of L1 has been investigated in Drosophila. Chueh et al (Chueh et al., 2009) found a significant enrichment of L1 retrotransposons (containing primate-specific L1 (L1-P1) subfamily) within the CENP-A (centromere-specific H3-like protein)-binding domain and also the 10q25 neocentromere. Given the L1P subfamily included active full-length L1 (FL-L1) retrotranposons and L1 RNAs were actively transcribed from full-length elements that contain an internal promoter, they analyzed the expression of six FL-L1s at the 10q25 neocentormere chromatin and found only one actively transcribed: FL-L1b. They found FL-L1b single-stranded RNA transcripts were incorporated as part of the ribonucleoprotein component of the CENP-A-associated domain. RNAi knockdown of FL-L1b transcripts reduced mitotic stability, level of CENP-A protein and transcriptional activities of two genes within and/or neighboring the CENP-A-associated chromatin (ATRNL1 and TRUB1). Thus, LINE retrotransposon RNA is an essential structural and functional component of the neocentromeric chromatin. Also, given many long single-stranded centromeric RNA transcripts were found in diverse species (in Zea mays, Arabidopsis thaliana, Palorus ratzeburgi and human), it indicated that a pool of single-stranded RNA could be directly transcribed from the satellite repeats (and centromere-specific retrotransposons) of the normal

centromeres or the L1 retrotransposon of a neocentromere and subsequently incorporated into the core centromeric/neocentromeric chromatin.

One of the central observations of our study was the significant enrichment of nucleosomes in human as well as in bovine sperm particularly in centromere repeats. From the aspect of the bovine system, this finding was coincident with an earlier observation (Palmer et al., 1990), suggesting CENP-A was quantitatively retained in mature bovine spermatozoa. Using indirect immunofluorescence staining, they also showed that CENP-A was retained in sperm nuclei in a discrete manner, rather than being dispersed throughout the sperm head. Functionally, CENP-A is essential for centromere specification, kinetochore formation, and chromosome segregation during cell division (Regnier et al., 2005). The functional mechanism of CENP-A has been highlighted in Drosophila, targeting its homologue of human CENP-A named CID. Dunleavy et al (Dunleavy et al., 2012) found that in male meiosis, CID was loaded in two phases, during the first stages of meiosis I and after the second meiotic division. They also reported a novel drop in CID levels after meiosis I and before meiosis II, which correlated with the timing of kinetochore reorientation. RNAi depletion of CAL1 and CENP-C resulted in reduced CID localization at centromeres, suggesting they were necessary for CID assembly and chromosome segregation, and their levels at centromeres decreased as meiosis progresses and eventually disappeared in mature sperm. The function of CID in early embryogenesis was further discovered by Raychaudhuri et al (Raychaudhuri et al., 2012). They found that of the three centromere protein described in Drosophila, CID, CENP-C and CAL1, only CID was present in mature sperm and remained associated with paternal centromeres during chromatin remodeling and male pronucleus formation, followed by equal distribution onto sister centromeres during the first S phase. Fertilization with sperm was still possible after CID elimination. However, the development after fertilization was abnormal. Paternal centromeres could not acquire maternally derived CID-EGFP after degradation of CID-EGFP during spermatogenesis, resulting in gynogenetic haploid embryos. Increased or decreased CID levels on paternal centromeres appeared to be maintained throughout development of the next generation and could not be recovered to normal level. This result supported the hypothesis that the centromere during cell proliferation was

‘‘template-governed.’’ After random distribution of centromeric CenH3 nucleosomes during chromosome replication onto the two sister chromatids, these old nucleosomes might act as a template, allowing the local stoichiometric loading of new CenH3 nucleosomes during each cell cycle.

These findings above provided a function chain of centromeric protein from spermatogenesis till early embryogenesis in Drosophila. With the help of CAL1 and CENP-C, CID contributes to the normal progression through male meiosis. As meiosis progresses, CAL1 and CENP-C decrease and disappear in mature sperm. Only CID is present in mature sperm and remains associated with paternal centromeres during chromatin remodeling and male pronucleus formation. Taking the findings regarding FL-L1b (suggesting LINE retrotransposon RNA is an essential structural and functional component of CENP-A), LINE1 might be important for correct cell division both in spermatogenesis and early embryo stages by contributing to CENP-A. This is supported by a very recent study (Fadloun et al., 2013), showing LINE1 retrotransposons became reactivated from both parental genomes after fertilization. Transcriptional activation of LINE1 occurred as early as in zygotes, but the expression of LINE1 strongly decreased between 2-cell and 8-cell stages. This initial reactivation and subsequent suppression as the development progresses of LINE1 further confirmed our finding that LINE1 was bound to nucleosomes, which was considerably easier to access and be activated. As proved by our study, both CENP-A and LINE1 originate from paternal remnant histones, with their significant enrichment in nucleosomes. Therefore, we suggest that the remaining nucleosomes are designated to deliver CENP-A and LINE1, which are essential for male pronucleus formation at the initial stage after fertilization, and govern as templates during chromosome replication for all the daughter cells of zygote.

4.3 Nucleosome-free and nucleosome-rich genes implicate diverse functions

Regarding the genes significantly enriched in nucleosome- and protamine-bound sequences, our findings could be classified into two categories: nucleosome-free genes and nucleosome-rich genes. For nucleosome-free genes, homeobox (*HOX*) genes were a major highlight of our findings. *HOX* genes are characterised by the conserved DNA homeobox, which encodes a DNA-binding protein domain. Many homeobox genes play

important roles in embryonic patterning and cell differentiation (Booth and Holland, 2007). Previous reports emphasized the role of sperm-specific histone-modifications, e.g. in *HOX*-promoters, as epigenetic marks for early development (Arpanahi et al., 2009; Brykczynska et al., 2010; Hammoud et al., 2009; Paradowska et al., 2012). This statement implicates that e.g. *HOX*-promoters are bound to nucleosomes at least in some parts of their sequence. However, performing a different method (direct sequencing of mono-nucleosomal DNA without antibody-precipitation procedure) we observed a different scenario. Among 43.8% human and 25% bovine genes exhibiting in sperm completely nucleosome-free promoters (-3,000bp) and gene bodies we found, surprisingly, a significant enrichment of all *HOX* genes. *HOX* genes are known to be abundant in SSRs (simple sequence repeats) and LCRs (low complexity repeats) (Huang et al., 2009; Mainguy et al., 2007) and have an evolutionary conserved sequence. It has been suggested (Huang et al., 2009) that the upstream regions of *HOX* genes containing high-density repeats were critical regulatory regions which had avoided transposable element insertion events during a long period of evolution. These regions were also a prolific source for evolution. SSRs often serve to modify genes, and affect gene regulation, transcription, and protein function. From an evolutionary perspective, the high density repeats could have a positive role in adaptation. Besides *HOX* genes, significant enrichment also lay in genes relevant to organ development, morphogenesis, regulation of biosynthetic and metabolic processes (genes regarded as highly conserved metabolic genes (Elsik et al., 2009)) and response. Thus, nucleosome-free genes are classifiable as “Genes for embryogenesis executive program” and are very likely to act in post-implantation embryogenesis.

On the aspect of nucleosome-rich genes, different function categories have been observed according to the binding positions of scattered nucleosomes. A relative big part of genes (human: 44%, bovine: 38.7%) exhibited nucleosomes-bound exclusively in their gene bodies. 7.4% of human and 25.7% of bovine genes comprised nucleosomes in both promoter and gene body. Regarding the gene function of this category, GO term analyses revealed an enrichment of factors for RNA- and protein-processing, metal (primarily calcium)-ion binding / transport, for membrane organization, cell-cell adhesion and microtubule cytoskeleton organization. As mentioned before, we

found that the majority of nucleosome were bound to repetitive sequences such as LINEs and SINEs. Therefore it was in accordance with expectations when we found those nucleosome-rich genes were also abundant in LINEs and SINEs, as shown exemplarily in case of *ILF2* (interleukin enhancer binding factor 2) gene. Genomic analyses and experimental evidence have demonstrated that retrotransposons could function as regulatory units for host genes and appeared to contribute to many mammalian gene regulatory sequences (Jordan et al., 2003; van de Lagemaat et al., 2003). Lunyak et al (Lunyak et al., 2007) showed in their study that tissue-specific transcription of SINE B2 repeat in murine was required for gene activation of growth hormone gene, by generating short, overlapping Pol II-and Pol III-driven transcripts, both of which are necessary and sufficient to enable a restructuring of the regulated locus into nuclear compartments. Estecio et al (Estecio et al., 2012) also suggested SINE B1 elements could cause epigenetic reprogramming of surrounding gene promoters by influencing the activity of downstream gene promoters, with acquisition of DNA methylation and loss of activating histone marks. Kunarso et al (Kunarso et al., 2010) further quantified transposable elements, suggesting that they made up to 25% contribution of the bound sites in humans and mice by wiring new genes into the core regulatory network of embryonic stem cells. It indicated that species-specific transposable elements had substantially altered the transcriptional circuitry of pluripotent stem cells. Peaston et al (Peaston et al., 2004) also showed retrotransposons could regulate host genes in mouse preimplantation embryos by providing an alternative 5' exon to many transcripts in early embryo stages. We therefore suggest that the retrotransposons within nucleosome-rich genes would function as regulatory elements and are involved in their activation and expression in early embryo stages.

Especially, a small proportion of genes (human: 5.3%, bovine: 10.4%) exhibited nucleosome binding sites exclusively in their promoters (-3,000 bp). Functional annotation showed an enrichment of cell-cell adhesion factors, calcium ion binding and RNA-processing factors. Calcium-dependent cell adhesion events, with participation of Cadherins, coordinate the cellular allocation and spatial segregation of inner cell mass in blastocyst, and are crucial for early morphogenesis (Fleming et al., 2001). However,

we found the promoter of gamma protocadherin C3 (*PCDHGC3*) to be nucleosome-enriched in human as well as in bovine sperm. For nucleosome-rich genes we suggest the umbrella term “Genes for embryogenesis initializing program”. These genes are probably those, which have to be euchromatic/active shortly after fertilization to start the paternal transcriptional and translational machinery and to ensure the basic requirements in forming of an early embryo. Their abundance on retrotransposons might reflect their predisposition to evolutionary variances. Moreover, it has been suggested (Gardiner-Garden et al., 1998) that genes contained in nucleosomal chromatin underwent earlier transcriptional activation in contrast to genes that were contained in protamine-based chromatin. Therefore we propose that nucleosome-rich genes are initially activated genes after fertilization and before implantation, whose functions serve as prerequisite for the subsequent pronucleus formation, cell division and the de novo activation of paternal genome packed within protamine heterochromatin. With both species showing the same tendency, an evolutionary consistency could also be expected here.

4.4 CpG-promoters are hypomethylated regardless of their occupancy with protamines or nucleosomes

Besides the regulation mechanisms through nucleosome binding or protamine package, DNA methylation status of both repetitive sequences and functional genes remain major concern directly regarding the following transcriptional activities in early embryogenesis. The importance of DNA methylome in sperm has been highlighted by a recent study on zebrafish (Jiang et al., 2013), showing that the embryos inherited the DNA methylome solely from the paternal side. For repetitive elements, as they were known to be scattered throughout the mammalian genome and were abundant in CpG-methylation, especially near retrotransposons, we supposed that the heterochromatic state of certain repetitive elements might affect the protamine incorporation into DNA and lead to nucleosome retention. As our findings showed, in both species the repetitive elements exhibited often CpG-hypermethylation. We therefore hypothesize that sperm-derived nucleosomes scattered genome-wide within LINEs and SINEs are required for post-fertilization decondensation of paternal chromatin in a comprehensive manner.

For genes bound to nucleosomes and protamines, previous studies emphasized that particularly “developmental” histone-associated promoters were hypomethylated in human sperm (Arpanahi et al., 2009; Hammoud et al., 2009). Here we analyzed whether nucleosome-free CpG-promoters differed from nucleosome-rich CpG-promoters regarding the DNA-methylation. Using bovine sperm DNA, we found that CpG-promoters were in general hypomethylated regardless of their occupancy with protamines or nucleosomes, except strong methylated CpG-promoters of maternal expressed genes. Moreover, there was no preference in nucleosome retention regarding CpG- or non-CpG-promoters.

4.5 Expression of genes in early embryos is not correlated with nucleosome-association of promoters

In all animals, the initial events of embryogenesis are controlled by maternal proteins and RNAs that are deposited into the developing oocyte, and the major embryonic genome activation is supposed to start during 2-cell stage (mice, rats), 4-8-cell-stage (humans, pigs) up to 8-16-cell stage (cattle, sheep) (Bensaude et al., 1983; Misirlioglu et al., 2006). Paternal chromatin undergoes post-fertilization a genome-wide demethylation and decondensation, whereby the protamines are removed and exchanged with maternal-descendant nucleosomes (Cantone and Fisher, 2013; Gu et al., 2011; Reik, 2007). The question, whether in early embryos the sperm contributed nucleosomes, i.e. specific histone-modifications, are responsible for activation of corresponding gene promoter from the paternal allele is difficult to analyze and still a matter of debate. By comparing the nascent RNA synthesis in parthenogenetic, androgenetic and normally fertilized embryos, one study (Bui et al., 2011) have shown that paternal chromatin was important in the regulation of transcriptional activity during mouse preimplantation development and that this capacity was acquired during spermiogenesis. Here we examined whether there was a measurable difference regarding the post-fertilization expression between sperm-derived nucleosome-free and nucleosome-rich promoters. We used RNA isolated from bovine sperm, oocytes and early embryos (zygote, 2-cell, 4-cell stage, morula and blastocyst). We analyzed exemplary five *HOX* genes and five pluripotency genes (all nucleosome-free in gene

body and promoter), thirteen bovine imprinted genes (Tveden-Nyborg et al., 2008) (all nucleosome-free in gene body and promoter, except *IGF2R*-promoter), seven randomly selected genes with >30% of gene body occupied by nucleosomes and six randomly selected genes with 65% up to 100% of promoter occupied by nucleosomes. In sperm, all analyzed transcripts were absent, whereas in oocytes several transcripts were stored. We saw neither a decisive difference between the expression-pattern of sperm-derived nucleosome-free and nucleosome-rich promoters nor between "developmental" and "non-developmental" genes as suggested before. However, clear evident was the fact that regardless of nucleosome-association of respective promoter in the sperm, a transcript was detectable in early embryos up to morula and/or blastocyst stage mostly when it was already present in the oocyte.

To accomplish our point of view, we found in human as well as in bovine sperm-genome a significant depletion of nucleosomes in exons, 5'-UTR, 3'-UTR and promoters, and their enrichment in introns and intergenic area. The fact that important regulatory regions and coding DNA are more thoroughly freed from nucleosomes and thus, are more safely packaged in protamines is logic and reflects for our opinion exactly the meaning of nucleosome-retention. Nucleosomes which are scattered within the sperm-genome in non-coding regions are not critical for paternal genetic integrity, and nevertheless can facilitate the post-fertilization decondensation / activation of paternal chromatin in a genome-wide and effective manner, and moreover are utilized for molecular recognition and function of paternal centromere.

4.6 Direct sequencing without antibody reveals nucleosome content and avoids false positive signals

Hitherto there has been no report suggesting the proportion of retained histone in bovine sperm, whereas it has been reported that in human sperm, about 15% of histones remained. However, reviews (Carrell et al., 2008; Miller et al., 2010) and studies (Arpanahi et al., 2009; Hammoud et al., 2009) all seemed to cite the same paper by Gatewood et al (Gatewood et al., 1987). The authors cautiously suggested that the percentage of DNA associated with histone was about 10 to 15%, meanwhile admitting the quantitation of the relative distribution of DNA in histone and protamine

fraction was not accurate. Later they (Gatewood et al., 1990) continued research by using high performance liquid chromatography and managed to identify the histone variants in human sperm. This time they suggested that histone proteins were a minor component of each mature spermatozoon, without giving the accurate histone proportion. Later studies (Bench et al., 1996) also just suggested an estimation of remaining histone to be as much as 15%, while some more recent studies (van der Heijden et al., 2008) even claimed that about 15% to 30% of the DNA in human sperm was packed in nucleosomes. Therefore a more accurate examination of remnant histone content and the percentage of paternal genome packed inside of them in human sperm is advised. Our study has provided results by the approach of micrococcal nuclease digestion and direct sequencing without antibody precipitation, showing for the first time that remaining nucleosome packed 14% of paternal genome in bovine sperm, and 4.8% in human sperm.

We also tried to validate our direct sequencing data by comparing with H3K9ac-ChIP-seq data, as using ChIP with antibody was the common method in previous studies. However, as shown in Figure 27, the ChIP-PCR results based on cross-linking ChIP using anti-H3K9ac antibody differed a lot with direct sequencing results of 146 bp nucleosome fraction. Only one gene out of 12 was consistent between the two methods. We believe that direct sequencing of the 146 bp DNA-fragment was the best way to ensure the work with solely nucleosomal DNA and to avoid false positive signals common in ChIP procedures (e.g. due to unspecific bond of antibodies, or due to enrichment-effect during ChIP: when an antibody captures a promoter fragment, which is representative only for the minority of the analyzed cell population, the experiment can lead to misinterpretations regarding the functional impact of this promoter).

5. Conclusion

Our direct sequencing results of sperm nucleosome-bound DNA in bovine and human demonstrate in a genome-wide manner that nucleosome retention in sperm-chromatin is oriented on repetitive DNA elements, especially in heterochromatic centromere repeats and retrotransposons. Repetitive DNA elements are the root for segmentation of paternal chromatin into nucleosome-poor and nucleosome-rich regions differing in nature of covered DNA and comprised genes.

Compared with previous studies, we suggest an alternative point of view concerning the biological impact of sperm-derived nucleosomes in post-fertilization processes. Non-coding DNA in sperm is nucleosome-rich, whereas functional DNA is nucleosome-poor. On the gene level, nucleosome-rich and nucleosome-free genes implicate diverse functions, serving the pre-implantation and post-implantation embryogenesis, respectively.

Our results demonstrate a global concordance between mammals concerning the pattern of nucleosome retention in sperm and suggest how such a genome-wide comprehensive process like protamine-assembly can be guided and stable maintained through the evolution. Our findings strongly suggest that the evolutionary importance of remnant sperm-nucleosomes might lie in their contribution to the post-fertilization paternal centromere function and therefore, the activation of paternal chromatin within pre-implantation embryogenesis and initial cell divisions of zygote. Development-relevant genes packed with protamines, on the other hand, were well-preserved for post-implantation embryogenesis. Future research could continue to investigate the function of paternal-derived repetitive DNA elements and nucleosome-bound genes on early embryogenesis, and further to reveal the etiology of idiopathic male infertility.

6. References

Abu-Halima, M., Hammadeh, M., Schmitt, J., Leidinger, P., Keller, A., Meese, E., and Backes, C. (2013). Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. Fertil Steril *99*, 1249-1255.

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

Aul, R.B., and Oko, R.J. (2001). The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome during spermiogenesis. Dev Biol *239*, 376-387.

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

Balhorn, R., Gledhill, B.L., and Wyrobek, A.J. (1977). Mouse Sperm Chromatin Proteins - Quantitative Isolation and Partial Characterization. Biochemistry-Us *16*, 4074-4080.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Bench, G.S., Friz, A.M., Corzett, M.H., Morse, D.H., and Balhorn, R. (1996). DNA and total protamine masses in individual sperm from fertile mammalian subjects. Cytometry *23*, 263-271.

Bensaude, O., Babinet, C., Morange, M., and Jacob, F. (1983). Heat shock proteins, first major products of zygotic gene activity in mouse embryo. Nature *305*, 331-333.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet *9*, 2395-2402.

Bhasin, M., Reinherz, E.L., and Reche, P.A. (2006). Recognition and classification of histones using support vector machine. J Comput Biol *13*, 102-112.

Booth, H.A.F., and Holland, P.W.H. (2007). Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. Gene *387*, 7-14.

Boulard, M., Gautier, T., Mbele, G.O., Gerson, V., Hamiche, A., Angelov, D., Bouvet, P., and Dimitrov, S. (2006). The NH2 tail of the novel histone variant H2BFWT exhibits properties distinct from conventional H2B with respect to the assembly of mitotic chromosomes. Mol Cell Biol *26*, 1518-1526.

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.

Bui, H.T., Wakayama, S., Mizutani, E., Park, K.K., Kim, J.H., Van Thuan, N., and Wakayama, T. (2011). Essential role of paternal chromatin in the regulation of transcriptional activity during mouse preimplantation development. Reproduction *141*, 67-77.

Cantone, I., and Fisher, A.G. (2013). Epigenetic programming and reprogramming during development. Nat Struct Mol Biol *20*, 282-289.

Carrell, D.T. (2008). Contributions of spermatozoa to embryogenesis: assays to evaluate their genetic and epigenetic fitness. Reprod Biomed Online *16*, 474-484.

Carrell, D.T. (2012). Epigenetics of the male gamete. Fertil Steril *97*, 267-274.

Carrell, D.T., Emery, B.R., and Hammoud, S. (2008). The aetiology of sperm protamine abnormalities and their potential impact on the sperm epigenome. Int J Androl *31*, 537-545.

Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. P Natl Acad Sci USA *99*, 16916-16921.

Choudhary, S.K., Wykes, S.M., Kramer, J.A., Mohamed, A.N., Koppitch, F., Nelson, J.E., and Krawetz, S.A. (1995). A Haploid Expressed Gene-Cluster Exists as a Single Chromatin Domain in Human Sperm. J Biol Chem *270*, 8755-8762.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E.*, et al.* (2010). LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. Cell *141*, 956-969.

Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H.A., and Wong, L.H. (2009). LINE Retrotransposon RNA Is an Essential Structural and Functional Epigenetic Component of a Core Neocentromeric Chromatin. Plos Genet *5,* e1000354.
Dadoune, J.P. (2009). Spermatozoal RNAs: what about their functions? Microsc Res Tech *72*, 536-551.
de Koning, A.P.J., Gu, W.J., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. Plos Genet *7,* e1002384.
Dijkwel, P.A., and Hamlin, J.L. (1995). Origins of replication and the nuclear matrix: The DHFR domain as a paradigm. Int Rev Cytol *162A*, 455-484.
Dunleavy, E.M., Beier, N.L., Gorgescu, W., Tang, J., Costes, S.V., and Karpen, G.H. (2012). The Cell Cycle Timing of Centromeric Chromatin Assembly in Drosophila Meiosis Is Distinct from Mitosis Yet Requires CAL1 and CENP-C. Plos Biol *10,* e1001460.
Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Abatepaulo, A.R.R., Abbey, C.A., Adelson, D.L., Aerts, J., Ahola, V., Alexander, L.*, et al.* (2009). The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science *324*, 522-528.
Estecio, M.R.H., Gallegos, J., Dekmezian, M., Lu, Y., Liang, S.D., and Issa, J.P.J. (2012). SINE Retrotransposons Cause Epigenetic Reprogramming of Adjacent Gene Promoters. Mol Cancer Res *10*, 1332-1342.
Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol *20*, 332-338.
Flamm, W. (1972). Highly repetitive sequences of DNA in chromosomes. Int Rev Cytol *32,* 1-51.
Fleming, T.P., Sheth, B., and Fesenko, I. (2001). Cell adhesion in the preimplantation mammalian embryo and its role in trophectoderm differentiation and blastocyst morphogenesis. Front Biosci *6*, D1000-D1007.
Gardiner-Garden, M., Ballesteros, M., Gordon, M., and Tam, P.P.L. (1998). Histone- and protamine-DNA association: Conservation of different patterns within the beta-globin domain in human sperm. Mol Cell Biol *18*, 3350-3356.

Garrido, N., Martinez-Conejero, J.A., Jauregui, J., Horcajadas, J.A., Simon, C., Remohi, J., and Meseguer, M. (2009). Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. Fertil Steril *91*, 1307-1310.

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.

Gatewood, J.M., Cook, G.R., Balhorn, R., Schmid, C.W., and Bradbury, E.M. (1990). Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. J Biol Chem *265*, 20662-20666.

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.

Gonzalez-Romero, R., Mendez, J., Ausio, J., and Eirin-Lopez, J.M. (2008). Quickly evolving histones, nucleosome stability and chromatin folding: All about histone H2A.Bbd. Gene *413*, 1-7.

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

Govindaraju, A., Uzun, A., Robertson, L., Atli, M.O., Kaya, A., Topper, E., Crate, E.A., Padbury, J., Perkins, A., and Memili, E. (2012). Dynamics of microRNAs in bull spermatozoa. Reprod Biol Endocrin *10,* 82.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L.Y., He, X.Y., Jin, S.G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-U136.

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.

Hazzouri, M., Pivot-Pajot, C., Faure, A.K., Usson, Y., Pelletier, R., Sele, B., Khochbin, S., and Rousseaux, S. (2000). Regulated hyperacetylation of core histones during

mouse spermatogenesis: involvement of histone deacetylases. Eur J Cell Biol *79*, 950-960.
Huang, Q.L., Guo, J.M., Ge, Q., Li-Ling, J., Chen, X.P., and Ma, F. (2009). Comparative analysis of distinct non-coding characteristics potentially contributing to the divergence of human tissue-specific genes. Genetica *136*, 127-134.
Jackson, D.A., and Cook, P.R. (1986). Replication Occurs at a Nucleoskeleton. Embo Journal *5*, 1403-1410.
Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33*, 245-254.
Jiang, L., Zhang, J., Wang, J.-J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J.*, et al.* (2013). Sperm, but Not Oocyte, DNA Methylome Is Inherited by Zebrafish Early Embryos. Cell *153*, 773-784.
Jordan, I.K., Rogozin, I.B., Glazko, G.V., and Koonin, E.V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet *19*, 68-72.
Kalandadze, A.G., Bushara, S.A., Vassetzky, Y.S., Jr., and Razin, S.V. (1990). Characterization of DNA pattern in the site of permanent attachment to the nuclear matrix located in the vicinity of replication origin. Biochem Biophys Res Commun *168*, 9-15.
Kaplan, J.G., Brown, D.L., Chaly, N., Greer, W.L., Prasad, K.V., Severini, A., and Sahai, B.M. (1987). Structural and evolutionary implications of the packaging of DNA for differentiation and proliferation in the lymphocyte. J Mol Evol *26*, 173-179.
Kimmins, S., and Sassone-Corsi, P. (2005). Chromatin remodelling and epigenetic features of germ cells. Nature *434*, 583-589.
Krawetz, S.A. (2005). Paternal contribution: new insights and future challenges. Nat Rev Genet *6*, 633-642.
Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597-610.
Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X.Y., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet *42*, 631-U111.

Lalancette, C., Thibault, C., Bachand, I., Caron, N., and Bissonnette, N. (2008). Transcriptome analysis of bull semen with extreme nonreturn rate: Use of suppression-subtractive hybridization to identify functional markers for fertility. Biol Reprod *78*, 618-635.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Li, A., Maffey, A.H., Abbott, W.D., Conde e Silva, N., Prunell, A., Siino, J., Churikov, D., Zalensky, A.O., and Ausio, J. (2005). Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). Biochemistry-Us *44*, 2529-2535.

Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X.Y.*, et al.* (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science *317*, 248-251.

Mainguy, G., Koster, J., Woltering, J., Jansen, H., and Durston, A. (2007). Extensive Polycistronism and Antisense Transcription in the Mammalian Hox Clusters. Plos One *2,* e356.

Marques, C.J., Costa, P., Vaz, B., Carvalho, F., Fernandes, S., Barros, A., and Sousa, M. (2008). Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. Mol Hum Reprod *14*, 67-74.

Meyer-Ficca, M.L., Lonchar, J.D., Ihara, M., Bader, J.J., and Meyer, R.G. (2013). Alteration of poly(ADP-ribose) metabolism affects murine sperm nuclear architecture by impairing pericentric heterochromatin condensation. Chromosoma *122*, 319-335.

Miller, D., Brinkworth, M., and Iles, D. (2010). Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. Reproduction *139*, 287-301.

Minor, A., Chow, V., and Ma, S. (2011). Aberrant DNA methylation at imprinted genes in testicular sperm retrieved from men with obstructive azoospermia and undergoing vasectomy reversal. Reproduction *141*, 749-757.

Misirlioglu, M., Page, G.P., Sagirkaya, H., Kaya, A., Parrish, J.J., First, N.L., and Memili, E. (2006). Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. P Natl Acad Sci USA *103*, 18905-18910.
Oliva, R., Bazett-Jones, D., Mezquita, C., and Dixon, G.H. (1987). Factors affecting nucleosome disassembly by protamines in vitro. Histone hyperacetylation and chromatin structure, time dependence, and the size of the sperm nuclear proteins. J Biol Chem *262*, 17016-17025.
Palmer, D.K., O'Day, K., and Margolis, R.L. (1990). The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma *100*, 32-36.
Paradowska, A.S., Miller, D., Spiess, A.N., Vieweg, M., Cerna, M., Dvorakova-Hortova, K., Bartkuhn, M., Schuppe, H.C., Weidner, W., and Steger, K. (2012). Genome wide identification of promoter binding sites for H4K12ac in human sperm and its relevance for early embryonic development. Epigenetics *7*, 1057-1070.
Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell *7*, 597-606.
Poplinski, A., Tuttelmann, F., Kanber, D., Horsthemke, B., and Gromoll, J. (2010). Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. Int J Androl *33*, 642-649.
Raychaudhuri, N., Dubruille, R., Orsi, G.A., Bagheri, H.C., Loppin, B., and Lehner, C.F. (2012). Transgenerational propagation and quantitative maintenance of paternal centromeres depends on cid/cenp-a presence in Drosophila sperm. PLoS Biol *10*, e1001434.
Regnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W., and Brown, W. (2005). CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. Mol Cell Biol *25*, 3967-3981.
Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.
Santos, F., and Dean, W. (2004). Epigenetic reprogramming during early development in mammals. Reproduction *127*, 643-651.

Shaman, J.A., Yamauchi, Y., and Ward, W.S. (2007). The sperm nuclear matrix is required for paternal DNA replication. J Cell Biochem *102*, 680-688.
Smit, A.F. (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev *6*, 743-748.
Soon, L.L., Ausio, J., Breed, W.G., Power, J.H., and Muller, S. (1997). Isolation of histones and related chromatin structures from spermatozoa nuclei of a dasyurid marsupial, Sminthopsis crassicaudata. J Exp Zool *278*, 322-332.
Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A.*, et al.* (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol *1*, E45.
Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S.*, et al.* (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. Science *331*, 593-596.
Tveden-Nyborg, P.Y., Alexopoulos, N.I., Cooney, M.A., French, A.J., Tecirlioglu, R.T., Holland, M.K., Thomsen, P.D., and D'Cruz, N.T. (2008). Analysis of the expression of putatively imprinted genes in bovine peri-implantation embryos. Theriogenology *70*, 1119-1128.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet *19*, 530-536.
van der Heijden, G.W., Ramos, L., Baart, E.B., van den Berg, I.M., Derijck, A.A., van der Vlag, J., Martini, E., and de Boer, P. (2008). Sperm-derived histones contribute to zygotic chromatin in humans. BMC Dev Biol *8*, 34.
Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369-374.
Ward, W.S., Partin, A.W., and Coffey, D.S. (1989). DNA loop domains in mammalian spermatozoa. Chromosoma *98*, 153-159.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.
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.
Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.
Yamauchi, Y., Shaman, J.A., Boaz, S.A., and Ward, W.S. (2007). Paternal pronuclear DNA degradation is functionally linked to DNA replication in mouse oocytes. Biol Reprod *77*, 407-415.
Yamauchi, Y., Shaman, J.A., and Ward, W.S. (2011). Non-genetic contributions of the sperm nucleus to embryonic development. Asian J Androl *13*, 31-35.
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.

7. Supplementary Figures and legends

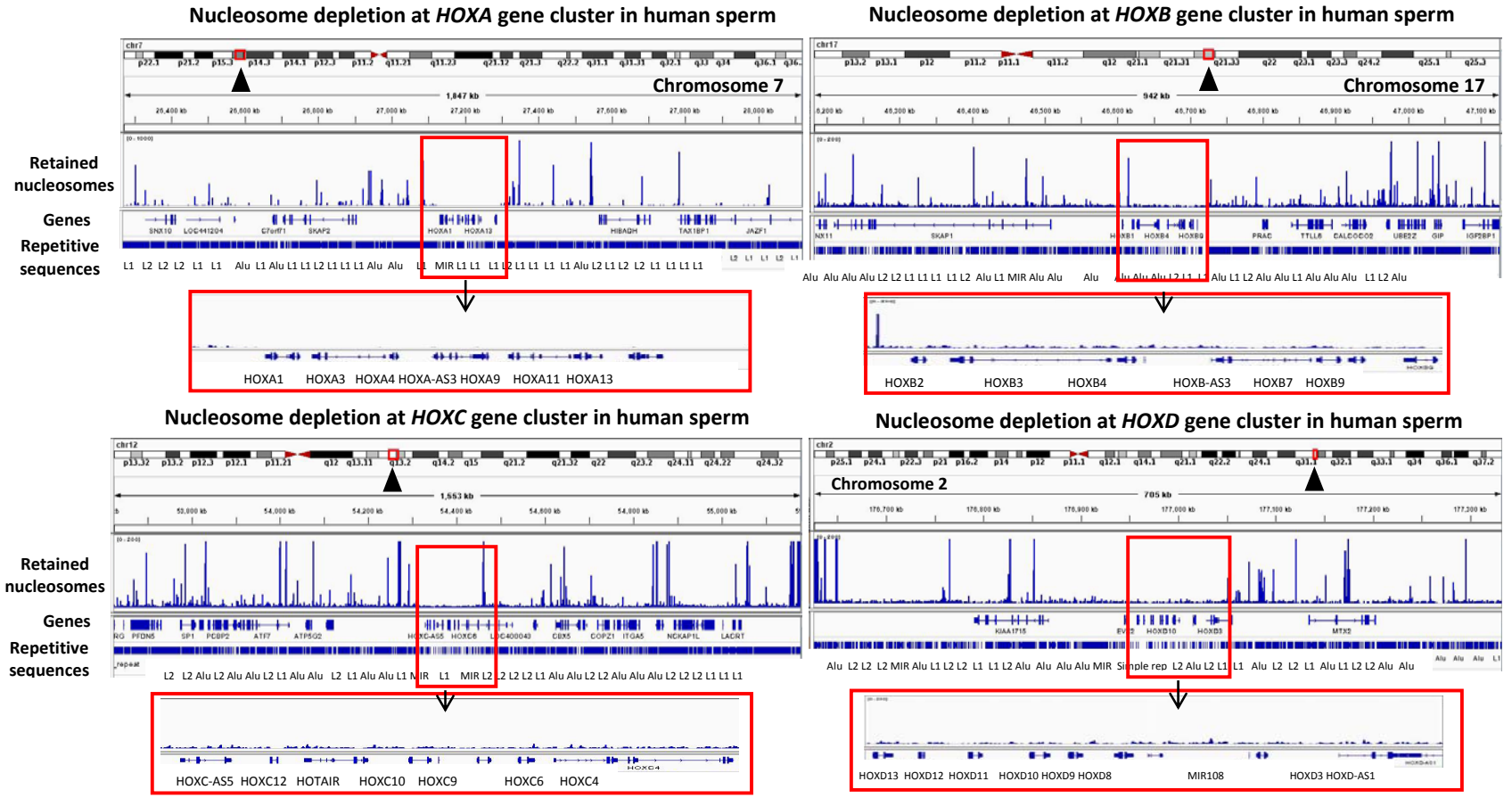


Figure S1 IGV-screen shots of all four *HOX* gene clusters (*HOXA*, *HOXB*, *HOXC* and *HOXD*) in human sperm reveal strong nucleosome depletion in these genome loci (highlighted boxes). Single *HOX* clusters with comprised single genes are presented in detail in zoomed screen shots (arrows). Chromosome locations of *HOX* clusters are marked with black triangles. Peaks representing binding sites of retained nucleosomes, affected genes and repetitive sequences are shown.

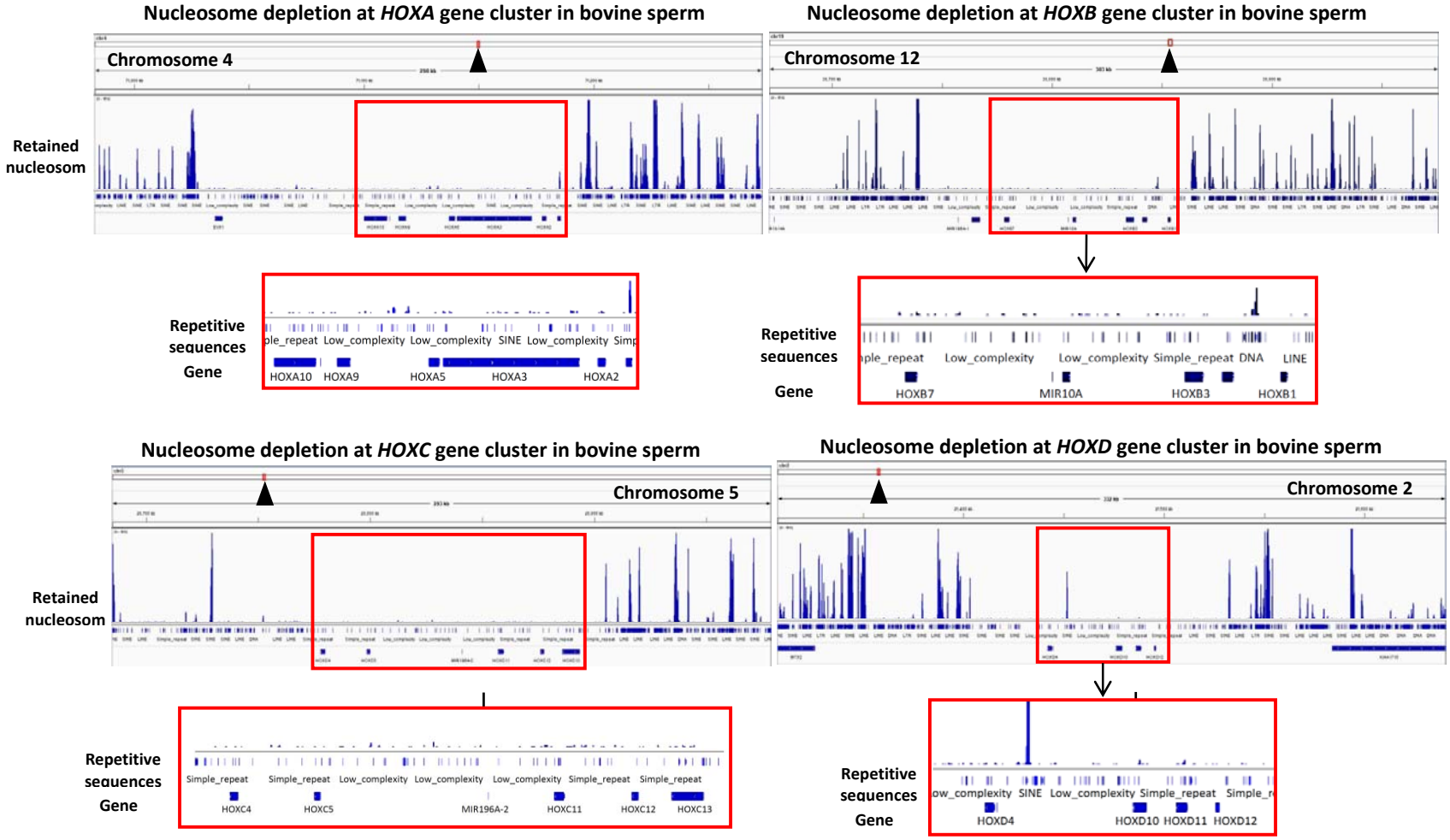


Figure S2 IGV-screen shots of all four *HOX* gene clusters (*HOXA, HOXB, HOXC* and *HOXD*) in bovine sperm reveal strong nucleosome depletion in these genome loci (highlighted boxes). Single *HOX* clusters are additionally presented in detail in zoomed screen shots (arrows). Chromosome locations of *HOX* clusters are marked with black triangles. Peaks representing binding sites of retained nucleosomes, affected genes and repetitive sequences are shown.

Table S1 Overlapping nucleosome-free genes in human and bovine sperm (enriched GO terms)

Term (DAVID)		Count	%	P-Value	List Total	Pop Hits	Pop Total	Fold Enrichment	FDR (up to 5%)
IPR017970	**Homeobox, conserved site**	**55**	**3,38045482**	**1,23E-18**	**1416**	**93**	**8047**	**3,36086052**	**2,09E-15**
IPR001356	**Homeobox**	**54**	**3,31899201**	**4,17E-18**	**1416**	**92**	**8047**	**3,33562085**	**7,07E-15**
IPR012287	**Homeodomain-related**	**55**	**3,38045482**	**8,71E-18**	**1416**	**96**	**8047**	**3,25583363**	**1,48E-14**
GO:0003700	transcription factor activity	139	8,54333128	1,97E-13	1167	421	6318	1,78748115	3,17E-10
GO:0043565	sequence-specific DNA binding	98	6,02335587	6,28E-11	1167	283	6318	1,87477177	1,01E-07
GO:0030528	transcription regulator activity	195	11,9852489	1,07E-10	1167	700	6318	1,50815277	1,73E-07
GO:0005576	extracellular region	237	14,5666872	2,66E-09	1150	943	6385	1,3954032	3,88E-06
GO:0003677	DNA binding	236	14,5052243	2,87E-09	1167	916	6318	1,3948429	4,62E-06
GO:0006355	regulation of transcription, DNA-dependent	194	11,9237861	4,25E-09	1232	728	6720	1,45354645	7,88E-06
GO:0051252	regulation of RNA metabolic process	199	12,2311002	6,80E-09	1232	755	6720	1,43768814	1,26E-05
GO:0000786	nucleosome	20	1,2292563	8,19E-09	1150	28	6385	3,96583851	1,20E-05
O:0006333	chromatin assembly or disassembly	29	1,78242163	4,00E-08	1232	55	6720	2,87603306	7,43E-05
GO:0005615	extracellular space	112	6,88383528	7,70E-08	1150	388	6385	1,60268938	1,12E-04
GO:0045449	regulation of transcription	265	16,287646	8,91E-08	1232	1095	6720	1,32004981	1,65E-04
GO:0006334	nucleosome assembly	22	1,35218193	9,46E-08	1232	36	6720	3,33333333	1,75E-04
GO:0048568	**embryonic organ development**	**36**	**2,21266134**	**1,15E-07**	**1232**	**80**	**6720**	**2,45454545**	**2,13E-04**
GO:0031497	chromatin assembly	22	1,35218193	1,81E-07	1232	37	6720	3,24324324	3,36E-04
GO:0009952	**anterior/posterior pattern formation**	**31**	**1,90534726**	**2,21E-07**	**1232**	**65**	**6720**	**2,6013986**	**4,09E-04**
GO:0048598	**embryonic morphogenesis**	**51**	**3,13460356**	**2,22E-07**	**1232**	**136**	**6720**	**2,04545455**	**4,12E-04**
GO:0003002	**regionalization**	**37**	**2,27412415**	**2,92E-07**	**1232**	**86**	**6720**	**2,34672304**	**5,42E-04**
GO:0005179	hormone activity	33	2,02827289	3,05E-07	1167	72	6318	2,48136247	4,90E-04

GO:0065004	protein-DNA complex assembly	22	1,35218193	3,36E-07	1232	38	6720	3,15789474	6,22E-04
GO:0044421	extracellular region part	138	8,48186847	3,45E-07	1150	517	6385	1,48201161	5,04E-04
GO:0007389	pattern specification process	44	2,70436386	5,26E-07	1232	113	6720	2,12389381	9,75E-04
GO:0048562	**embryonic organ morphogenesis**	**29**	**1,78242163**	**6,44E-07**	**1232**	**61**	**6720**	**2,59314456**	**1,19E-03**
GO:0006952	defense response	99	6,08481868	7,90E-07	1232	340	6720	1,58823529	1,46E-03
GO:0051173	positive regulation of nitrogen compound metabolic process	98	6,02335587	8,33E-07	1232	336	6720	1,59090909	1,54E-03
GO:0030182	neuron differentiation	65	3,99508297	8,88E-07	1232	197	6720	1,79972312	1,65E-03
GO:0006357	regulation of transcription from RNA polymerase II promoter	109	6,69944683	1,20E-06	1232	387	6720	1,53629316	0,00223255
GO:0031328	positive regulation of cellular biosynthetic process	104	6,39213276	1,41E-06	1232	366	6720	1,54992548	0,00261833
GO:0005198	structural molecule activity	95	5,83896742	1,77E-06	1167	326	6318	1,57766493	0,00285708
GO:0009891	positive regulation of biosynthetic process	105	6,45359557	2,30E-06	1232	374	6720	1,53135634	0,00426226
GO:0009991	response to extracellular stimulus	53	3,25752919	2,64E-06	1232	154	6720	1,8772137	0,00489107
GO:0010557	positive regulation of macromolecule biosynthetic process	99	6,08481868	2,87E-06	1232	349	6720	1,54727794	0,00531304
GO:0034728	nucleosome organization	22	1,35218193	2,96E-06	1232	42	6720	2,85714286	0,00548916
GO:0045892	negative regulation of transcription, DNA-dependent	61	3,74923171	3,08E-06	1232	187	6720	1,77929023	0,00571304
GO:0048706	**embryonic skeletal system development**	**20**	**1,2292563**	**3,11E-06**	**1232**	**36**	**6720**	**3,03030303**	**0,00577089**

GO:0045935	positive regulation of nucleobase metabolic process	93	5,71604179	3,51E-06	1232	324	6720	1,56565657	0,00650989
GO:0031667	response to nutrient levels	49	3,01167793	3,94E-06	1232	140	6720	1,90909091	0,00730828
GO:0051253	negative regulation of RNA metabolic process	61	3,74923171	5,46E-06	1232	190	6720	1,75119617	0,010125
GO:0010628	positive regulation of gene expression	87	5,3472649	5,64E-06	1232	301	6720	1,57656297	0,0104598
GO:0045596	negative regulation of cell differentiation	39	2,39704978	9,65E-06	1232	105	6720	2,02597403	0,01788699
GO:0030900	**forebrain development**	**30**	**1,84388445**	**1,03E-05**	**1232**	**72**	**6720**	**2,27272727**	**0,01917779**
GO:0048663	**neuron fate commitment**	**14**	**0,86047941**	**1,16E-05**	**1232**	**21**	**6720**	**3,63636364**	**0,02149796**
GO:0000122	negative regulation of transcription from RNA polymerase II pr	48	2,95021512	1,17E-05	1232	141	6720	1,85686654	0,02177666
GO:0000785	chromatin	33	2,02827289	1,31E-05	1150	85	6385	2,15554987	0,01908514
GO:0032526	response to retinoic acid	15	0,92194222	1,39E-05	1232	24	6720	3,40909091	0,02582965
GO:0045941	positive regulation of transcription	83	5,10141364	1,43E-05	1232	290	6720	1,56112853	0,02658626
GO:0010604	positive regulation of macromolecule metabolic process	124	7,62138906	1,53E-05	1232	476	6720	1,42093201	0,02840007
GO:0032993	protein-DNA complex	20	1,2292563	1,72E-05	1150	40	6385	2,77608696	0,02516263
GO:0006954	inflammatory response	63	3,87215734	1,77E-05	1232	205	6720	1,67627494	0,03286697
GO:0006350	transcription	197	12,1081746	3,48E-05	1232	834	6720	1,28842381	0,06448109
GO:0033273	response to vitamin	22	1,35218193	3,99E-05	1232	48	6720	2,5	0,07392695
GO:0007584	response to nutrient	36	2,21266134	4,85E-05	1232	100	6720	1,96363636	0,08998869
GO:0042127	regulation of cell proliferation	121	7,43700061	5,91E-05	1232	475	6720	1,38947368	0,10956354
GO:0043009	**chordate embryonic development**	**53**	**3,25752919**	**7,11E-05**	**1232**	**171**	**6720**	**1,69059011**	**0,1318496**

GO:0009792	**embryonic development ending in birth or egg hatching**	**53**	**3,25752919**	**8,43E-05**	**1232**	**172**	**6720**	**1,6807611**	**0,15614346**
GO:0035270	**endocrine system development**	**19**	**1,16779348**	**9,03E-05**	**1232**	**40**	**6720**	**2,59090909**	**0,16735603**
GO:0006323	DNA packaging	23	1,41364474	9,77E-05	1232	54	6720	2,32323232	0,18097037
GO:0045893	positive regulation of transcription, DNA-dependent	70	4,30239705	1,04E-04	1232	247	6720	1,5458226	0,19188581
GO:0001501	**skeletal system development**	**55**	**3,38045482**	**1,26E-04**	**1232**	**183**	**6720**	**1,63934426**	**0,2339673**
GO:0051254	positive regulation of RNA metabolic process	70	4,30239705	1,35E-04	1232	249	6720	1,53340635	0,25083041
GO:0033189	response to vitamin A	15	0,92194222	1,37E-04	1232	28	6720	2,92207792	0,25332808
GO:0007423	**sensory organ development**	**36**	**2,21266134**	**1,50E-04**	**1232**	**105**	**6720**	**1,87012987**	**0,27804646**
GO:0005125	cytokine activity	35	2,15119852	2,14E-04	1167	102	6318	1,85770452	0,3436394
GO:0048704	**embryonic skeletal system morphogenesis**	**15**	**0,92194222**	**2,20E-04**	**1232**	**29**	**6720**	**2,82131661**	**0,40758488**
GO:0016481	negative regulation of transcription	68	4,17947142	2,57E-04	1232	245	6720	1,51391466	0,47476082
GO:0000904	**cell morphogenesis involved in differentiation**	**35**	**2,15119852**	**2,82E-04**	**1232**	**104**	**6720**	**1,83566434**	**0,52199982**
GO:0051094	**positive regulation of developmental process**	**50**	**3,07314075**	**2,97E-04**	**1232**	**167**	**6720**	**1,63309744**	**0,5493967**
GO:0045165	**cell fate commitment**	**26**	**1,59803319**	**3,09E-04**	**1232**	**69**	**6720**	**2,05533597**	**0,57164375**
GO:0048545	response to steroid hormone stimulus	43	2,64290104	3,38E-04	1232	138	6720	1,69960474	0,62438969
GO:0048666	**neuron development**	**46**	**2,82728949**	**3,55E-04**	**1232**	**151**	**6720**	**1,66164961**	**0,65572935**
GO:0031327	negative regulation of cellular biosynthetic process	79	4,85556238	3,75E-04	1232	298	6720	1,44600366	0,69254197
GO:0001709	**cell fate determination**	**12**	**0,73755378**	**4,33E-04**	**1232**	**21**	**6720**	**3,11688312**	**0,80029514**

GO:0009611	response to wounding	85	5,22433927	4,34E-04	1232	327	6720	1,41784821	0,80186211
GO:0003007	**heart morphogenesis**	**18**	**1,10633067**	**4,68E-04**	**1232**	**41**	**6720**	**2,39467849**	**0,8638779**
GO:0045944	positive regulation of transcription from RNA polymerase II pr	55	3,38045482	4,79E-04	1232	192	6720	1,5625	0,88368233
GO:0007517	**muscle organ development**	**40**	**2,4585126**	**5,30E-04**	**1232**	**128**	**6720**	**1,70454545**	**0,97845387**
GO:0010558	negative regulation of macromolecule biosynthetic process	77	4,73263675	5,94E-04	1232	293	6720	1,4334471	1,09567474
GO:0021537	**telencephalon development**	**14**	**0,86047941**	**5,95E-04**	**1232**	**28**	**6720**	**2,72727273**	**1,09807174**
GO:0002700	regulation of production of molecular mediator of immune resp	14	0,86047941	5,95E-04	1232	28	6720	2,72727273	1,09807174
GO:0009725	response to hormone stimulus	62	3,81069453	6,12E-04	1232	225	6720	1,5030303	1,12830014
GO:0014706	**striated muscle tissue development**	**25**	**1,53657037**	**6,31E-04**	**1232**	**68**	**6720**	**2,00534759**	**1,16358181**
GO:0045934	negative regulation of nucleobase metabolic process	72	4,42532268	8,23E-04	1232	273	6720	1,43856144	1,5159816
GO:0001890	placenta development	16	0,98340504	9,41E-04	1232	36	6720	2,42424242	1,73138177
GO:0032583	regulation of gene-specific transcription	29	1,78242163	1,02E-03	1232	86	6720	1,83932347	1,86768668
GO:0051172	negative regulation of nitrogen compound metabolic process	72	4,42532268	1,02E-03	1232	275	6720	1,42809917	1,87909737
GO:0007409	**axonogenesis**	**26**	**1,59803319**	**1,02E-03**	**1232**	**74**	**6720**	**1,91646192**	**1,88233603**
GO:0009890	negative regulation of biosynthetic process	79	4,85556238	1,08E-03	1232	308	6720	1,39905549	1,9841974
GO:0001893	maternal placenta	8	0,49170252	1,13E-03	1232	11	6720	3,96694215	2,07065277

	development								
GO:0010629	negative regulation of gene expression	70	4,30239705	1,17E-03	1232	267	6720	1,43003064	2,1396581
GO:0009719	response to endogenous stimulus	66	4,05654579	1,22E-03	1232	249	6720	1,44578313	2,23162583
GO:0048667	**cell morphogenesis involved in neuron differentiation**	**28**	**1,72095882**	**1,26E-03**	**1232**	**83**	**6720**	**1,84008762**	**2,3096053**
GO:0060537	**muscle tissue development**	**25**	**1,53657037**	**1,27E-03**	**1232**	**71**	**6720**	**1,9206146**	**2,329533**
GO:0045095	**keratin filament**	**13**	**0,79901659**	**1,32E-03**	**1150**	**27**	**6385**	**2,67326892**	**1,90647603**
GO:0010605	negative regulation of macromolecule metabolic process	101	6,20774431	1,32E-03	1232	414	6720	1,33069829	2,42620206
GO:0016564	transcription repressor activity	47	2,8887523	1,51E-03	1167	164	6318	1,55153928	2,40566522
GO:0048729	**tissue morphogenesis**	**28**	**1,72095882**	**1,87E-03**	**1232**	**85**	**6720**	**1,79679144**	**3,41220086**
GO:0021871	**forebrain regionalization**	**7**	**0,4302397**	**1,89E-03**	**1232**	**9**	**6720**	**4,24242424**	**3,44854101**
GO:0003006	**reproductive developmental process**	**40**	**2,4585126**	**1,90E-03**	**1232**	**136**	**6720**	**1,60427807**	**3,46328237**
GO:0010033	response to organic substance	109	6,69944683	1,91E-03	1232	457	6720	1,30097474	3,47484764
GO:0045137	**development of primary sexual characteristics**	**23**	**1,41364474**	**1,96E-03**	**1232**	**65**	**6720**	**1,93006993**	**3,5670295**
GO:0006325	chromatin organization	48	2,95021512	2,09E-03	1232	172	6720	1,52219873	3,80428452
GO:0031175	**neuron projection development**	**33**	**2,02827289**	**2,26E-03**	**1232**	**107**	**6720**	**1,68224299**	**4,11754157**

GO: gene ontology; **DAVID**: Database for Annotation, Visualization and Integrated Discovery; **FDR**: false discovery rate; Highlighted GO terms belong to the major enriched group “Embryo development & morphogenesis”.

Table S2 Overlapping genes in human and bovine sperm with retained nucleosomes in gene body and gene promoter (enriched GO terms)

Term (DAVID)		Count	%	P-Value	List Total	Pop Hits	Pop Total	Fold Enrichment	FDR (up to 5%)
SP_PIR_KEYWORDS	alternative splicing	1758	43,33251	4,97E-26	3983	3402	8916	1,156763	7,68E-23
SP_PIR_KEYWORDS	phosphoprotein	1877	46,26571	2,32E-20	3983	3730	8916	1,126458	3,59E-17
GO:0005524	ATP binding	428	10,54967	1,16E-10	2826	770	6318	1,242683	1,98E-07
GO:0001882	nucleoside binding	471	11,60956	1,26E-10	2826	858	6318	1,227273	2,15E-07
GO:0001883	purine nucleoside binding	468	11,53562	1,30E-10	2826	852	6318	1,228043	2,22E-07
GO:0030554	adenyl nucleotide binding	458	11,28913	1,81E-10	2826	833	6318	1,229215	3,09E-07
GO:0032559	adenyl ribonucleotide binding	433	10,67291	1,82E-10	2826	782	6318	1,237909	3,11E-07
GO:0017076	purine nucleotide binding	547	13,48287	3,76E-09	2826	1031	6318	1,18614	6,44E-06
GO:0032553	ribonucleotide binding	521	12,842	4,82E-09	2826	978	6318	1,190985	8,24E-06
GO:0032555	purine ribonucleotide binding	521	12,842	4,82E-09	2826	978	6318	1,190985	8,24E-06
GO:0000166	nucleotide binding	624	15,38082	6,04E-08	2826	1208	6318	1,154849	1,03E-04
GO:0043167	ion binding	955	23,53956	2,17E-06	2826	1945	6318	1,097719	0,003707
GO:0046872	metal ion binding	922	22,72615	7,57E-06	2826	1884	6318	1,094101	0,012942
GO:0043169	cation binding	930	22,92334	8,20E-06	2826	1902	6318	1,09315	0,01403
GO:0016192	vesicle-mediated transport	192	4,732561	1,41E-05	2995	343	6720	1,255971	0,026807
GO:0007156	homophilic cell adhesion	41	1,010599	2,01E-05	2995	55	6720	1,672606	0,038152
GO:0005509	calcium ion binding	249	6,13754	1,65E-04	2826	470	6318	1,184429	0,28176
GO:0019941	modification-dependent protein catabolic process	185	4,56002	2,17E-04	2995	341	6720	1,217278	0,411026
GO:0043632	modification-dependent macromolecule catabolic process	185	4,56002	2,17E-04	2995	341	6720	1,217278	0,411026
GO:0005096	GTPase activator activity	60	1,478925	2,96E-04	2826	94	6318	1,427023	0,50425
GO:0051603	proteolysis	194	4,7818	2,96E-	2995	361	6720	1,20577	0,560

	involved in cellular protein catabolic process		**59**	**04**				**7**	**865**
GO:0044257	**cellular protein catabolic process**	**194**	**4,781859**	**2,96E-04**	**2995**	**361**	**6720**	**1,205777**	**0,560865**
GO:0030695	**GTPase regulator activity**	**98**	**2,415578**	**4,01E-04**	**2826**	**168**	**6318**	**1,30414**	**0,683583**
GO:0005083	**small GTPase regulator activity**	**67**	**1,651467**	**4,02E-04**	**2826**	**108**	**6318**	**1,386943**	**0,686057**
GO:0060589	**nucleoside-triphosphatase regulator activity**	**101**	**2,489524**	**4,13E-04**	**2826**	**174**	**6318**	**1,297716**	**0,70327**
GO:0016879	**ligase activity, forming carbon-nitrogen bonds**	**81**	**1,996549**	**4,21E-04**	**2826**	**135**	**6318**	**1,341401**	**0,717282**
GO:0010324	membrane invagination	78	1,922603	4,83E-04	2995	130	6720	1,346244	0,913378
GO:0006897	endocytosis	78	1,922603	4,83E-04	2995	130	6720	1,346244	0,913378
GO:0016044	membrane organization	126	3,105743	6,18E-04	2995	226	6720	1,250934	1,168565
GO:0044265	**cellular macromolecule catabolic process**	**227**	**5,595267**	**8,11E-04**	**2995**	**435**	**6720**	**1,170871**	**1,529689**
GO:0005794	**Golgi apparatus**	**252**	**6,211486**	**8,17E-04**	**2812**	**493**	**6385**	**1,160644**	**1,237261**
GO:0006796	**phosphate metabolic process**	**274**	**6,753759**	**0,001011**	**2995**	**535**	**6720**	**1,14913**	**1,903335**
GO:0006793	**phosphorus metabolic process**	**274**	**6,753759**	**0,001011**	**2995**	**535**	**6720**	**1,14913**	**1,903335**
GO:0016877	**ligase activity, forming carbon-sulfur bonds**	**14**	**0,345083**	**0,001016**	**2826**	**15**	**6318**	**2,086624**	**1,723674**
GO:0030163	**protein catabolic process**	**197**	**4,855805**	**0,001031**	**2995**	**374**	**6720**	**1,181863**	**1,939823**
GO:0046578	regulation of Ras protein signal transduction	48	1,18314	0,001143	2995	75	6720	1,435993	2,150164
GO:0051056	**regulation of small GTPase mediated signal transduction**	**58**	**1,429628**	**0,001157**	**2995**	**94**	**6720**	**1,384435**	**2,175995**
GO:0005216	ion channel activity	109	2,686714	0,001205	2826	194	6318	1,256123	2,041041
GO:0016337	cell-cell adhesion	77	1,897954	0,001216	2995	131	6720	1,318839	2,284705
GO:0005254	chloride channel activity	30	0,739463	0,001217	2826	42	6318	1,596906	2,061289
GO:0035091	phosphoinositide binding	35	0,862706	0,001245	2826	51	6318	1,534283	2,108188
GO:0015630	microtubule cytoskeleton	158	3,894503	0,001293	2812	298	6385	1,203889	1,950488
GO:0008509	anion	54	1,3310	0,0016	2826	87	6318	1,38765	2,722

	transmembrane transporter activity		33	13				6	828
GO:0070647	**protein modification by small protein conjugation or removal**	**60**	**1,478925**	**0,001734**	**2995**	**99**	**6720**	**1,359842**	**3,243549**
GO:0019898	extrinsic to membrane	135	3,327582	0,00181	2812	252	6385	1,216407	2,720606
GO:0034702	ion channel complex	66	1,626818	0,001833	2812	112	6385	1,338047	2,755199
GO:0005099	Ras GTPase activator activity	29	0,714814	0,001916	2826	41	6318	1,581327	3,227502
GO:0008047	enzyme activator activity	89	2,193739	0,00196	2826	156	6318	1,275478	3,300103
GO:0016881	**acid-amino acid ligase activity**	**68**	**1,676115**	**0,002216**	**2826**	**115**	**6318**	**1,321961**	**3,723771**
GO:0045202	synapse	102	2,514173	0,002724	2812	186	6385	1,245182	4,068753
GO:0007229	integrin-mediated signaling pathway	26	0,640868	0,002196	2995	36	6720	1,620479	4,090553
GO:0012505	endomembrane system	239	5,891053	0,002859	2812	474	6385	1,144894	4,266077
GO:0005262	calcium channel activity	24	0,59157	0,003232	2826	33	6318	1,625941	5,385791

GO: gene ontology; **DAVID**: Database for Annotation, Visualization and Integrated Discovery; **FDR**: false discovery rate; Highlighted GO terms belong to the major enriched group “RNA processing & Protein synthesis, processing and catabolism”.

Table S3 Overlapping promoters in human and bovine sperm genome with retained nucleosomes (enriched GO terms)

Term (DAVID)		Count	%	P-Value	List Total	Pop Hits	Pop Total	Fold Enrichment	FDR (up tp 5%)
UP_SEQ_FEATURE	**domain:Cadherin 1**	**19**	**6,375839**	**3,63E-15**	**294**	**48**	**8914**	**12,00156**	**5,61E-12**
GO:0007156	**homophilic cell adhesion**	**19**	**6,375839**	**1,02E-14**	**203**	**55**	**6720**	**11,43574**	**1,68E-11**
GO:0016337	**cell-cell adhesion**	**20**	**6,711409**	**9,49E-09**	**203**	**131**	**6720**	**5,053962**	**1,56E-05**
GO:0007155	**cell adhesion**	**24**	**8,053691**	**1,42E-04**	**203**	**332**	**6720**	**2,39302**	**0,232872**
GO:0022610	**biological adhesion**	**24**	**8,053691**	**1,48E-04**	**203**	**333**	**6720**	**2,385834**	**0,24358**
GO:0005509	calcium ion binding	30	10,06711	1,71E-04	193	470	6318	2,089516	0,24278
GO:0006396	RNA processing	20	6,711409	0,002821	203	315	6720	2,101806	4,540713
GO:0006766	vitamin metabolic process	7	2,348993	0,002933	203	48	6720	4,827586	4,717049

GO: gene ontology; **DAVID**: Database for Annotation, Visualization and Integrated Discovery; **FDR**: false discovery rate; Highlighted GO terms belong to the major enriched group “Cell-cell adhesion”.

Table S4 Primer list for CpG-promoter methylation analyses in selected genes

Gene	Primer Sequence (5'-3') for COBRA (bovine)	Product (bp)	Ta (°C)	Restriction enzyme*
IGF2R	F: GTTTAYGTGATYGGGTTTGG; R: ACAAAAACRCRAAAAACCAC	188	58	*Bsh1236I*
H19	F: GTATTGYGGTTYGGGAGTTA; R: CRCCCCACCTAACCTAATCT	154	58	*TaqI*
MEG3	F: TGTTTGGTTTGGGGTGAGTT; R: CRACCCCRTAATCRAAATAA	152	58	*TaqI*
PEG10	F: GYGGTTTGTTGGTTTTAGGA; R: RAAAACRACCRACCTAACCA	183	58	*Bsh1236I*
IGF2	F: GGTTTTTAGTTTYGYGGTGA; R: AAACRCTTAACCCCCRTTAT	229	58	*Bsh1236I*
NNAT	F: TATTTAAGGYGYGGTTATYG; R: ACAATTCAACCRACRCTACC	153	58	*TaqI*
PRKCA	F: GGAYGTGGTTAATYGTTTYG; R: AATCCAAAAACRACRCACAC	156	60	*Bsh1236I*
WWOX	F: GYGGAGTTTTGGGTTAGGAT; R: TTCRCAACTACCRAAACAAA	194	60	*Bsh1236I*
HPCAL1	F: TATYGTGTGTGAGYGGTTGT; R: ACTATCCCCAAAACCCRAAC	178	60	*Bsh1236I*
CAPN3	F: ATGYGYGGTTAGGTGTTTAG; R: TCATAAAAATAAACCRACAAAAATCA	445	60	*TaqI*
PADI2	F: GGGGTTTATTTGYGGTAGGT; R: ATTTCRAACCCAACCAACAA	212	60	*TaqI*
ITPR2	F: ATTTTTTYGAGGGGATTTAAAGGTTTTGTA; R: CTCTCCRACAAATTTCCTATTCCTTTCAA	177	58	*Bsh1236I*
MACF1	F: YGGGATTAGGGATGAGGAGT; R: CTACCCCCRAATCCTTTCTC	199	60	*TaqI*
OCT4	F: GAGTGGGGTYGGTGTTTTTA; R: AACTCACTCRCCTCCTCAAA	175	60	*TaqI*
TDGF1	F: TTGAGYGTTGGGAGATTGTT; R: TTAAACCCRTACTCCRAAAA	418	60	*Bsh1236I*
ZFP36	F: GTTTAGTTTTTYGGYGYGTA; R: AACTAACCCCCTCCCCTCTT	116	60	*TaqI*
LIN28A	F: GGCGTAGGAGTACGAGAGGTT; R: TACCAACTCCGACCAATTCC	135	60	*TaqI*
ILF2	F: AYGTTATGYGGTAAGGGTTG; R: TTCCRTACAACCCCCAAATA	126	60	*TaqI*
SMOC2	F:GTAYGGTTTTGGGGGTTTTT; R:TTCCCRAATCCTCCTCTACC	172	58	*Bsh1236I*
SPATA5L1	F: AGAGTYGYGTGGTGGTTTAG; R: TCRAACCTCCCAAATCTACR	130	60	*Bsh1236I*
CCDC85A	F: GYGGTATTYGTTTTTYGGTA; R: RCRCTAAACRACCATCCTAC	144	60	*Bsh1236I*
TCEA2	F: YGTTYGGAGGTTGGATAAGA; R: RAATAATCACCRCRACCCTA	134	60	*TaqI*
ZNF135	F: GGAGAAGGGGAYGATAGAGG; R: CCAAACCTCCCTATCCATCA	104	60	*TaqI*
PDCD5	F: YGGGGTGAGYGAGTTTAATA; R: CCRAAACAAAAACRCCTAAA	110	60	*TaqI*
HOXB4	F: GTTAGYGTTGTGAGGYGATT;	158	60	*Bsh1236I*

	R: AAAATACCCACRCACTCACC			
HOXA10	F: GYGGGTTTGATTTTTGAGTT; R: CAAAACCCCAACCRAATTTA	129	60	*TaqI*
HOXA3	F: GAGTYGTGAATYGGGTTTGT; R: TCCTACRCCCAAACTTCAAC	150	60	*TaqI*
HOXB7	F: TGAAAAGGGYGGAAGAGTTA; R: CRACTCCCCCAACAAATCTA	109	60	*TaqI*
HOXC13	F: GGAGGAGGAGTAGGGATTYG; R: TCTCCCTCATACCACRTTCC	105	60	*Bsh1236I*
SOX2	F: GYGGYGTAAGATGGTTTAAG; R: TACTATTACCRCCCRAAACR	149	60	*TaqI*

COBRA: Combined bisulfite restriction analysis; **Ta**: annealing temperature; bp: base pair; **F**: forward primer; **R**: reverse primer.

*Restriction recognition sites (5'-3): *Bsh1236I* (CG..CG); *TaqI* (T..CGA).

Table S5 Primer list for CpG methylation analyses in repetitive sequence

Repetitive sequence (type)	Chromosome: location	Primer sequence (5'-3') for COBRA (bovine)	Product (bp)	Ta (°C)	Restriction enzyme*
LINE (L1MC1)	Chr6: 118118407-118118586	F:TGATTAGYGGGTATYGGTTT; R: CCCCATAATCAACTCCCAAA	353	60	*Bsh1236I*
LINE (L2b)	Chr16: 46708789-46708869	F: TTTTGTTYGGYGGTGTTAGT; R: RCRCRACAAAATTACCCTAA	205	60	*Bsh1236I*
LINE (L2a)	Chr7: 43820332-43820373	F:AGYGTGGGTGTGTATTTTYG; R:AAACTTACTACACCRCRAAACC	270	60	*TaqI*
LINE (L2b)	Chr17: 75868601-75868727	F: TTTAGTTTTYGGGGTTTTYG; R: CTAAACCCRCCTCCCTAAAC	281	60	*TaqI*
SINE (SINE2-1)	Chr17: 75275300-75275415	F:AGYGAGGGGTAGGAGATGTT; R: ATCCRTCCTCCRTTATCRTT	222	60	*TaqI*
SINE (SINE2-2)	Chr14: 80452749-80452871	F: AGGTGTYGYGTGTAGGTTTT; R: RCRTCRCTACCCTAATTCCT	229	60	*Bsh1236I*
SINE (ART2A)	Chr8: 67456144-67456183	F:GGTAAAGGGGGYGTTATAGG; R: AACRCCRAAAACRACTCTAA	168	60	*TaqI*
LTR (MLT1C)	Chr6: 120182703-120182809	F: GGYGTAGGAGGGTTTTGTTT; R: AAAAACCCCTAAACCCCRTA	144	60	*Bsh1236I*
LTR (LTR3B)	Chr17: 75771942-75772092	F: GTTTTGYGGGTGAAGTTGTT; R: CCCRAACRACTCCATCTAAC	235	60	*Bsh1236I*
LTR (MLT1K)	Chr18: 6615801-6615912	F: TAGGTAGGAGTTYGYGTGGA; R: TACCTCCRACCAATTTACRC	280	60	*TaqI*
LTR (MLT1C)	Chr9: 105744901-105745047	F:TATTTTYGGGTTTGYGTTTY; R: CRACCTCRTCTTCCCAATTA	159	60	*Bsh1236I*
LCR (GCrich)	Chr10: 1888613-1888660	F: GYGGTYGTTGGTTGTTAGTY; R: AAACAACAAAACRCRAAACC	143	60	*Bsh1236I*
LCR (GCrich1)	ChrX: 87011110-87011145	F: TTYGGGGTTGTAGYGTAGAY; R: CTCRAAACRCAAACAAAACA	187	60	*Bsh1236I*
LCR (GCrich)	Chr1: 157742-157792	F: GYGTTTCGGAYGTTTTAGG; R: AACRAAACRCAAAACCAAAC	229	60	*Bsh1236I*
LCR (GCrich2)	ChrX: 86358798-86358837	F: GGYGTAGAAATYGGTGTTTG; R: CAAACCCCCTCRTTCCTAAC	213	60	*Bsh1236I*
SSR (CCG)n	Chr11: 19928558-19928591	F: TGTTGYGTTATYGYGTTTAT; R: TCCRATTCAACCCAAACTTC	193	60	*Bsh1236I*
SSR (CCG)n	Chr7: 43799254-43799308	F: TYGATGGTTYGGAAGAAGTY; R: TCAAACRCACTTCRAACAAC	222	60	*Bsh1236I*

SSR (CGG)n	Chr8: 83260603-83260647	F: GGTAGYGYGAGTTGTTGTTT; R: RCRAACCRACRAACTAAAAA	233	60	*Bsh1236I6I*

COBRA: Combined bisulfite restriction analysis; **Ta**: annealing temperature; **bp**: base pair; **F**: forward primer; **R**: reverse primer; **LINE**: long interspersed nuclear element; **SINE**: short interspersed nuclear element; **LTR**: long terminal repeat; **LCR**: low complexity repeat; **SSR**: simple sequence repeat; Restriction recognition sites (5'-3): *Bsh1236I* (CG..CG); *TaqI* (T..CGA).

Table S6 Primer list for mRNA-expression analyses in selected genes

Gene	Primer sequence (5'-3') for real-time RT-PCR (bovine)	Product (bp)	Ta (°C)
ACTB	F: AATCTTCGCCTTAATACTTGTT; R: AAAGCCTTCATACATCTCAAG	103	60
SLC38A4	F: AATTCCAAGCATCACTAACC; R: GAATCTTCCTGTTACTTCCTATG	130	60
MEST	F: CGGCTTACAATCAAGAGTC; R: GAGAATACGAAGAAGTTCATCA	143	60
SGCE	F: ACCTGGTGAGATTAGTAATGAT; R: CGCTATATGGTGTTCTTTGG	104	60
IGF2R	F: CATTCTGTGGGTGACTCT; R: GAAGGTGATGCTACTCTGA	111	60
H19	F: CGCACAGAGGGATATGATA; R: CGTCAGGAGACTAAAGGAA	119	60
MEG3	F: TAATCTTCGCTTGCCTCC; R: GAACTACCCATCATTATTGCTAA	131	60
GNAS	F: GCAGCCTATAGATTAAGATTAAGA; R: GCACAACACGATATTTATTTCAT	135	60
GRB10	F: ATAGAGAGGAATTTCTTTGTACGA; R: CGGAACCTGCACCTAATC	100	60
PEG10	F: CAACTACCCAGCCTTCAT; R: TCATCTGGAAAGCATTAGAGTA	137	60
NAP1L5	F: GATCTCTTCTGTGAGGACTA; R: CCTAGTGCGATACTGTGA	105	60
IGF2	F: TCCAGCGATTAGAAGTGAG; R: GACGGTACAGGGATTTCA	119	60
NNAT	F: AAACGAATCCCATCTTTATCAA; R: GCAATTACAATTAGCAATTACCA	119	60
XIST	F: TTGTGTGAGTGGACCTAC; R: ACCTTCCTAGTGATACTTAGC	150	60
PRKCA	F: TAAAGGACCCGACACAGA; R: TGCACGTTCATATCACAGG	152	60
WWOX	F: GCGAGACCCTTCACCAAGT; R: AGTCTTCGCTCTGGGCTTC	155	60
HPCAL1	F: CGACGGCAACGGCTACAT; R: GTTTCTCCGGCGTGGACT	113	60
PADI2	F: TCTGGACCGACGTCTACAGC; R: CTTGCCTGGCTCATGGTG	183	60
ITPR2	F: TCAACACGCTGGGACTGG; R: CGGTTCATAGGGCACACCTT	115	60
CAPN3	F: GTCATTAGCGCCTCTGTG; R: TGCTCGAATGTCTTCTCTT	181	60
SMOC2	F: AGATGTTGCATCACGGTACCC; R: ACTCGGGGATGACCACGTT	158	60
MACF1	F: CCTGGAATCTGGAAAAGC; R: GAAATGGGTTTTACACGC	175	60
OCT4	F: CCCAACGTGAGGATTTTG; R: GAAGAGTACAGAGTAGTGAAGT	130	60
TDGF1	F: GTCTCTCTAATGTCCCAACT; R: GGTGCTTCAAGGAAATCTTT	100	60
ZFP36	F: GCCTCTTCTCAAACTTCAC;	108	60

	R: CTCCAATCACCAGACACT		
LIN28A	F: TGCAGAAACGCAGATCAAAG; R: TTCTTCCTCCTCCCGAAAGT	203	60
ILF2	F: ACTGCTGAAGGACCTGAGGA; R: GTCTGGTGGGGTTGTTCATC	104	60
SPATA5L1	F: GACCGAGAGGTTGTCATTGG; R: GCCAACTGTCATTTCTGCAA	126	60
CCDC85A	F: TGCTGTTTCCTGGATGATGA; R: TTCACCTCCAGCTCCTTCAG	137	60
TCEA2	F: CCTGTCACACTGCATCTGCT; R: CATCCAGGAGCTTCTTCCAG	133	60
ZNF135	F: TGGACAAAGGGAGAATCCAG; R: GTGTTCGGTGGTGTTCAATG	128	60
PDCD5	F: AAAGCACAGGGAAGCAGAAA; R: GTCCATACCGTGCCATCTGT	149	60
HOXB4	F: TACAACCGCTACCTGACACG; R: GTTGGGCAACTTGTGGTCTT	129	60
HOXA10	F: AGTTTCATCCTGCGGTTCTG; R: CCCTACACCAAGCACCAGAC	149	60
HOXA3	F: CTTCAGTCTCCCCACCTCAG; R: GCCGAGACTCTTTCATCCAG	123	60
HOXB7	F: GAGCAGAGGGACTCGGACTT; R: CAGCTCCAGGGTCTGATAGC	123	60
HOXC13	F: GCAAGAAACGAGTGCCCTAC; R: TCCGAGAGGTTTGTGGTAGC	121	60
SOX2	F: CACAACTCGGAGATCAGCAA; R: CGGGGCCGGTATTTATAATC	140	60

Ta: annealing temperature; **bp**: base pair; **F**: forward primer; **R**: reverse primer.

Acknowledgements

I would like to thank Prof.Dr. Klaus Steger and PD Dr. Undraga Schagdarsurengin for their great supervisorship. By their abundant help and continuously providing invaluable assistance, support and guidance, they taught me a lot. This project and thesis would not have been possible without their support. Also, my PhD career in their lab was a great, unforgettable time.

I would like to thank Dr. Birgit Samans from Research Centre for BioSystems, LandUse and Nutrition, JLU, for her bioinformatic support, great help and expertise, which greatly contributed to my thesis.

I would like to thank all the present and past members of Prof. Steger´s lab for creating a great, stimulating and joyful working atmosphere. Especially, I thank Mr. Gaurav Sarode and Mrs. Carmen Selignow with whom I collaborated on my project and who greatly contributed to the results presented in this thesis.

I would like to thank Prof. Dr. Axel Wehrend, Head of Clinic for Obstetrics, Gynecology and Andrology of Large and Small Animals with Ambulatory Service, JLU, for provision of bovine ejaculates and a very effective collaboration. I would like to thank Prof.Dr. Eckhard Wolf and Dr. Myriam Reichenbach from Molecular Animal Breeding and Biotechnology, Gene Center, Ludwig-Maximilians-University (LMU) Munich, for provision of bovine early embryos. Also I would like to thank Dr. Helmut Blum and Dr. Stefan Krebs from Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, LMU, for their great help of high throughput sequencing techniques.

I acknowledge all the excellent facilities in Justus-Liebig-University Giessen (JLU Giessen)and our Biomedizinisches Forschungszentrum Seltersberg (BFS).

I acknowledge German Research Foundation (DFG) for providing funding for this project.

I acknowledge China Scholarship Council (CSC) for providing myself with funding.

I acknowledge International Giessen Graduate Centre for the Life Sciences (GGL) for organizing great workshops and seminars, from which I benefit a lot.

I thank my parents and friends for supporting me during these three years.

我要衷心感谢三年来我的父母、家人和朋友对我的学习、工作以及生活的倾力支持。不论远隔万里还是近在眼前，你们的期待永远是我的精神支柱！我永远为你们而战！

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.