Male aging and obesity effects on sperm methylome and consequences for the next generation

Alters- und Adipositas-Effekte auf das Spermien-Methylom und die Konsequenzen für die nächste Generation

Doctoral thesis for a doctoral degree at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section: Biomedicine

Submitted by

Ramya Sri Krishna Potabattula

From Tanuku, Andhra Pradesh, India

Würzburg 2018

Submitted on:

Members of the *Promotionskomitee*:

Chairperson:

Primary Supervisor:
Prof. Dr. Thomas Haaf
Director of Human Genetics Department
Julius-Maximilians-Universität Würzburg
Biocenter, Am Hubland
97074 Würzburg
mail: thomas.haaf@uni-wuerzburg.de

Supervisor (Second):
Prof. Dr. Helga Stopper
Chair of Toxicology Department
Julius-Maximilians-Universität Würzburg
Versbacherstrasse 9
97078 Würzburg
mail: stopper@toxi.uni-wuerzburg.de

Supervisor (Third):
Prof. Dr. Georg Krohne
Senior Professor at Microscopy Department
Julius-Maximilians-Universität Würzburg
Biocenter, Am Hubland
97074 Würzburg
mail: krohne@biozentrum.uni-wuerzburg.de

Date of Public Defence:

Date of Receipt of Certificates:

For Science

Affidavit

I hereby confirm that my thesis entitled "**Male aging and obesity effects on sperm methylome, and consequences for the next generation**" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 2018 Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "**Alters- und Adipositas-Effekte auf das Spermien-Methylom und die Konsequenzen für die nächste Generation**" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 2018 Unterschrift

Acknowledgments

First and foremost, I would like to express my deepest gratitude and utmost respect to my primary supervisor, Prof. Dr. Thomas Haaf, for his time, support, and guidance all through my PhD work.

I am grateful to my co-supervisors, Prof. Dr. Helga Stopper and Prof. Dr. Georg Krohne, for being part of my doctoral thesis committee. I would like to sincerely thank Dr. Nady El Hajj for his knowledgeable mentoring, stimulating scientific discussions, and valuable suggestions.

I am indebted to my beloved family – my father Mr. Prasad Potabattula, who has been a constant source of motivation and encouragement, and my mother Mrs. Tirumala Potabattula for showering her invaluable love, affection, and empathy, and my brother Mr. Chaitanya Potabattula for his moral and emotional support through various phases of my personal and professional life. I wholeheartedly dedicate the progress of my career to three of them.

I owe a deep sense of gratitude to all my friends and extended family members for their unconditional support and care.

I am very fortunate to have worked with an excellent research team in the Institute of Human Genetics, University of Wuerzburg who has made my time here truly memorable.

I feel very much obliged to the Graduate School of Life Sciences (University of Würzburg) for the fellowship, and allowing me to be a part of the structured doctoral program.

Summary

Besides a growing tendency for delayed parenthood, sedentary lifestyle coupled with overnutrition has dramatically increased worldwide over the last few decades. Epigenetic mechanisms can help us understand the epidemics and heritability of complex traits like obesity to a significant extent. Majority of the research till now has focused on determining the impact of maternal factors on health and disease risk in the offspring(s).

This doctoral thesis is focused on deciphering the potential effects of male aging and obesity on sperm methylome, and consequences/transmission via germline to the next generation. In humans, this was assessed in a unique cohort of ~300 sperm samples, collected after *in vitro* fertilization/intracytoplasmic sperm injection, as well as in conceived fetal cord blood samples of the children. Furthermore, aging effect on sperm samples derived from a bovine cohort was analyzed.

The study identified that human male aging significantly increased the DNA methylation levels of the promoter, the upstream core element, the 18S, and the 28S regions of ribosomal DNA (rDNA) in sperm. Prediction models were developed to anticipate an individual's age based on the methylation status of rDNA regions in his sperm. Hypermethylation of alpha satellite and LINE1 repeats in human sperm was also observed with aging. Epimutations, which are aberrantly methylated CpG sites, were significantly higher in sperm of older males compared to the younger ones. These effects on the male germline had a negative impact on embryo quality of the next generation. Consistent with these results, DNA methylation of rDNA regions, bovine alpha satellite, and testis satellite repeats displayed a significant positive correlation with aging sperm samples within the same individual and across different age-grouped bulls.

A positive association between human male obesity/body mass index (BMI) and DNA methylation of the imprinted *MEG3* gene and the obesity-related *HIF3A* gene was detected in sperm. These BMI-induced sperm DNA methylation signatures were transmitted to next generation fetal cord blood (FCB) samples in a gender-specific manner. Males, but not female offsprings exhibited a significant positive correlation between father's BMI and FCB DNA

methylation in the two above-mentioned amplicons. Additionally, hypomethylation of *IGF2* with increased paternal BMI was observed in female FCB samples. Parental allele-specific in-depth methylation analysis of imprinted genes using next generation sequencing technology also revealed significant correlations between paternal factors like age and BMI, and the corresponding father's allele DNA methylation in FCB samples.

Deep bisulphite sequencing of imprinted genes in diploid somatic cord blood cells of offspring detected that the levels of DNA methylation signatures largely depended on the underlying genetic variant, i.e. sequence haplotypes. Allele-specific epimutations were observed in *PEG1*, *PEG5*, *MEG3*, *H19*, and *IGF2* amplicons. For the former three genes, the non-imprinted unmethylated allele displayed more epimutations than the imprinted methylated allele. On the other hand, for the latter two genes, the imprinted allele exhibited higher epimutation rate than that of the non-imprinted allele.

In summary, the present study proved that male aging and obesity impacts the DNA methylome of repetitive elements and imprinted genes respectively in sperm, and also has considerable consequences on the next generation. Nevertheless, longitudinal follow-up studies are highly encouraged to elucidate if these effects can influence the risk of developing abnormal phenotype in the offspring during adulthood.

Zusammenfassung

Weltweit kann ein Trend zur späteren Elternschaft sowie die dramatische Zunahme eines bewegungsarmen Lebensstils in Kombination mit einer Überernährung beobachtet werden. Die Verbreitung sowie die Vererbung derartig komplexer Merkmale oder Erkrankungen lassen sich oftmals unter Berücksichtigung epigenetischer Mechanismen besser verstehen. Vorangegangene Studien untersuchten hierbei jedoch primär den Einfluss maternaler Faktoren auf die Gesundheit und das Krankheitsrisiko der nächsten Generation.

Diese Doktorarbeit hat das Ziel potentielle Altersinduzierte sowie Adipositas-bedingte Effekte auf das Methylom von Spermien sowie die damit einhergehenden Konsequenzen für die nächste Generation zu identifizieren. Hierfür stand eine humane Kohorte bestehend aus ~300 Spermienproben, welche nach einer *in vitro* Fertilisation bzw. Intrazytoplasmatischen Spermieninjektion gesammelt wurden, und das entsprechende Nabelschnurblut der mit Hilfe dieser Behandlungen gezeugten Kinder zur Verfügung. Ergänzend dazu wurde der Alterseffekt in bovinen Spermienproben analysiert.

Im Rahmen dieser Arbeit konnte eine signifikante Korrelation des männlichen Alterungsprozesses mit einer erhöhten DNA-Methylierung der ribosomalen DNA (rDNA) in der Promotorregion, dem strangaufwärts gelegenen Promotorelement (*Upstream Core Element)*, der 18S- sowie der 28S-Region von Keimzellen aufgezeigt werden. Hierauf basierend wurde ein Vorhersageprogramm entwickelt, welches es ermöglicht anhand des Methylierungslevels der rDNA in den Spermien das Alter des entsprechenden Mannes zu berechnen. Weiterhin konnte beobachtet werden, dass mit dem Alterungsprozess in Spermien eine Hypermethylierung der *Long Interspersed Nuclear Elements* (LINE-1) und der Alpha-Satelliten-DNA einhergeht.

Des Weiteren zeigte der Vergleich von Spermien alter und junger Männer auf, dass Epimutationen, welche als fehlerhaft methylierte CpG-Stellen definiert werden, signifikant häufiger in Spermien alter Männer detektierbar sind. Allgemein zeigte sich zudem, dass diese altersabhängigen Effekte auf die Keimzellen sich negativ auf die Qualität der Embryonen auswirken. Übereinstimmend mit diesen Ergebnissen konnte auch in Spermien von Rindern

dieser Alterseffekt nachgewiesen werden. Es zeigte sich mit zunehmendem Alter eine signifikant erhöhte Methylierung der rDNA-Regionen und der Alpha-/ Testis-Satelliten-DNAs sowohl innerhalb eines Individuums, als auch zwischen verschiedenen altersgruppierten Individuen.

Weiterhin konnte eine positive Korrelation des Body-Mass-Index (engl. b*ody mass index*; kurz: BMI) eines Mannes und der Methylierung in *MEG3* und *HIF3A* in dessen Spermien detektiert werden. Anhand der untersuchten Nabelschnurblute konnte aufgezeigt werden, dass diese BMI-induzierten Methylierungsveränderungen in einem geschlechter-spezifischen Kontext an die nächste Generation weitergegeben werden. So war ausschließlich bei männlichen Nachkommen eine signifikant positive Assoziation des väterlichen BMIs und der DNA-Methylierung dieser beiden Gene im Nabelschnurblut nachweisbar. Eine mit dem väterlichen BMI einhergehende Hypomethylierung des Gens *IGF2* war hingegen ausschließlich in den Nabelschnurblut-Proben weiblicher Nachkommen messbar. Die Auftrennung der Allele nach ihrer parentalen Herkunft, zeigte eine signifikante Korrelation zwischen den paternalen Faktoren, wie Alter und BMI, und dem Methylierungslevel des korrespondierenden paternalen Allels im Nabelschnurblut.

Mit Hilfe der *Deep Bisulfite* Sequenzierung diploider Zellen aus dem Nabelschnurblut der Nachkommen wurde die DNA-Methylierung geprägte Gene untersucht. Hierbei konnte aufgezeigt werden, dass das Methylierungslevel stark von der zu Grunde liegenden genetischen Variante (Haplotyp) abhängt. Allel-spezifische Epimutationen wurden in den untersuchten Amplikons der Gene *PEG1*, *PEG5*, *MEG3*, *H19* und *IGF2* nachgewiesen. Bei *PEG1*, *PEG5* und *MEG3* waren mehr Epimutationen der nicht-geprägten unmethylierten Allele, als der geprägten methylierten Allele detektierbar. Konträr hierzu war bei den Genen *H19* und *IGF2* eine erhöhte Epimutationsrate der geprägten und somit methylierten Allele nachweisbar.

Diese Doktorarbeit offenbart altersinduzierte sowie Adipositas-bedingte Einflüsse auf die DNA-Methylierung repetitiver Elemente und geprägter Gene in männlichen Keimzellen, sowie damit einhergehende potentielle Konsequenzen für die nächste Generation. Um schlussendlich jedoch eine Aussage treffen zu können, ob aufgrund dieser Effekte eine erhöhte Prädisposition der nächsten Generation zur Entwicklung anormaler Phänotypen besteht, müssen longitudinale Folgestudien durchgeführt werden.

Table of contents

1 Introduction

1.1 Missing heritability and epigenetics

Numerous genome-wide association studies (GWAS) of complex traits reliably showed that the associated genetic variants such as DNA mutations or single nucleotide polymorphism only account for a minor part of disease inheritance [1]. One example is that of the population stature in humans. Despite identifying a convincing fraction of height determining genes by GWAS studies [2-4], the associated polymorphisms could explain only a trivial portion of stature heritability. This disparity is known as "missing heritability". Epigenetic mechanisms, which would methodically be missed by traditional DNA sequence-based analysis, might play an important role in explaining this discrepancy [5]. The word epigenetics was coined by the embryologist Conrad Waddington in 1942 and contemporarily it is defined as the study of alterations in gene expression in a heritable (mitotically or trans-generationally) manner without changes in the underlying DNA sequence. Therefore, in addition to the Mendelian heredity of DNA sequences, a model of non-DNA sequence-linked information might effectively contribute to decoding missing heritability [6]. Some mechanisms which underlie epigenetic transmission include modifications of DNA (cytosine methylation, hydroxy-methylation), or of histones (methylation, acetylation, ubiquitination, sumoylation, phosphorylation), nucleosome positioning, specific RNAs (miRNAs, non-coding RNAs, siRNAs, piRNAs) and prions. Some of these processes are mechanistically linked and regulate the activity of genes by determining DNA's accessibility to the transcription machinery [7].

1.2 DNA methylation

One of the best characterized epigenetic mechanisms is DNA methylation. It is a covalent modification of DNA wherein a methyl group is added to the fifth carbon position of cytosine pyrimidine ring mostly within the context of cytosine-phosphate-guanine (CpG) dinucleotides in mammals. A methyl group is added from S-adenosine methionine (SAM) to cytosine to form 5-methylcytosine. DNA methyltransferases (DNMTs) are the set of enzymes that catalyze this modification. Amongst them, the most abundant one in a mammalian cell is DNMT1 which acts

as a maintenance methyltransferase by copying methylation patterns from hemimethylated DNA strands during DNA replication. The other known members of the DNMT3 family are DNMT3A, 3B, and 3L which are responsible for *de novo* methylation [8]. DNA methylation is associated with many key biological phenomena like X-chromosome inactivation, genomic imprinting, transposable elements' repression etc. [9].

One of the striking features in a vertebrate's genome is the occurrence of a high number of CpGs at the 5' ends of the genes. These islands (CpGIs) are rich in GC sequences and span a length of approximately 300-3000 base pairs (bp) and they specifically occur within the promoter region of a gene [10]. CpGIs located close to gene promoters and transcriptional start sites (TSS), when methylated, mostly tend to be associated with repression of gene expression [11]. On the other hand, active genes are characterized by unmethylated/low methylated levels in gene promoters and/or in TSS enabling the accessibility of the transcriptional machinery to the DNA. In contrast, DNA methylation in gene body tends to maintain a positive correlation with gene transcription [10]. Overall, this very key process of DNA methylation can regulate the activity of the underlying DNA segment and plays an important role in developmental programming.

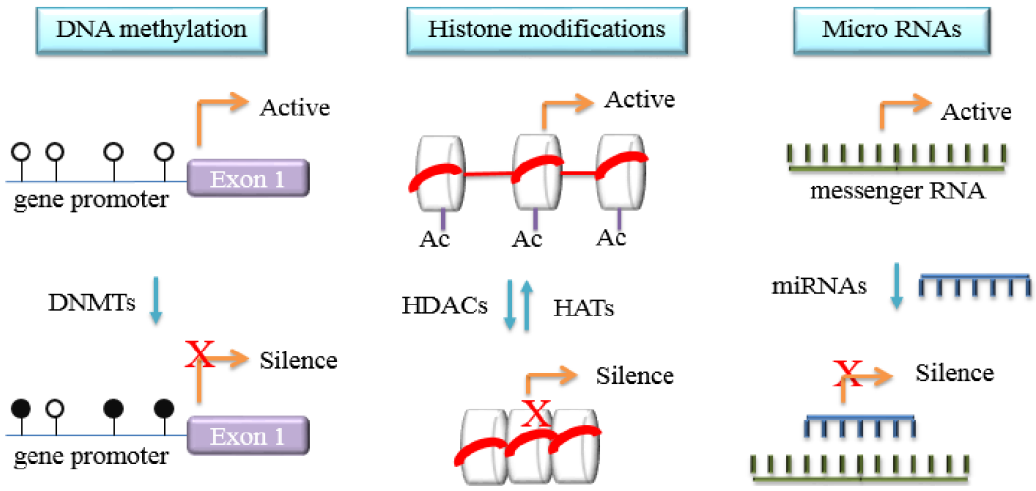


Figure 1. Some epigenetic modifications regulating gene expression

DNA methyltransferase (DNMTs) are involved in DNA methylation. Black lollipops: methylated CpG sites; uncolored ones: unmethylated CpGs (left). Acetylation of histones by histone acetyltransferases (HATs) forms relaxed chromatin leading to gene transcription; On the other hand, deacetylation by histone deacetylases (HDACs) results in closed chromatin and gene silencing. Ac represents acetylation (middle). Micro RNAs play a major role in post-transcriptional regulation of genes (right).

1.3 Epigenetic reprogramming

In mammals, the word 'epigenetic reprogramming' indicates erasure and re-establishment of epigenetic modifications between generations. Amongst many different mechanisms, alterations in DNA methylation that are best characterized, undergo two waves of reprogramming; one during gametogenesis and the other one in early embryogenesis.

1) Primordial germ cells (PGC) as they proliferate, migrate and enter the genital ridge, a broad range of demethylation events occur resulting in < 10% of genome-methylation to be retained. Since embryonic cells that have been adapted to somatic destiny give rise to PGCs, a large-scale reprogramming of methylation marks in order to meet the requirements of a germ cell is highly crucial. During this process, the methylation marks of imprinted genes are also excised [12]. This comprehensive de-methylation restores totipotency and provides an equivalent state of epigenetic alterations in germ cells of both sexes. A new wave of methylation re-establishment occurs by *de novo* methylation during germ cell development. Interestingly, this occurs in a gender-specific manner [13]. New *de novo* methylation landscape is re-established before meiosis in mitotically arrested G1 phase pro-spermatogonia cells and is completed before birth in male embryos. On the other hand, in female embryos, primary oocytes get arrested in the diplotene stage of prophase I of the first meiotic division. DNA methylation mechanism is only resumed after birth at puberty [14, 15].

2) After fertilization, the second wave of developmental reprogramming takes place. The paternal and maternal genomes are not affected in the same way by the de-methylation wave. Methylation signatures in male germ cells are rapidly lost by an active de-methylation mechanism in the zygote whereas maternal DNA methylation landscapes are passively deleted [16]. Certain epigenetic marks like imprinted genes escape this second reprogramming process enabling parent-specific activities of genes [13, 17]. The *de novo* methylation in the genome is established by DNMT3A and 3B at the blastocyst stage. Since the aim is to establish totipotency of the zygote at this stage, re-establishment enables differentiation into various embryonic cell lineages.

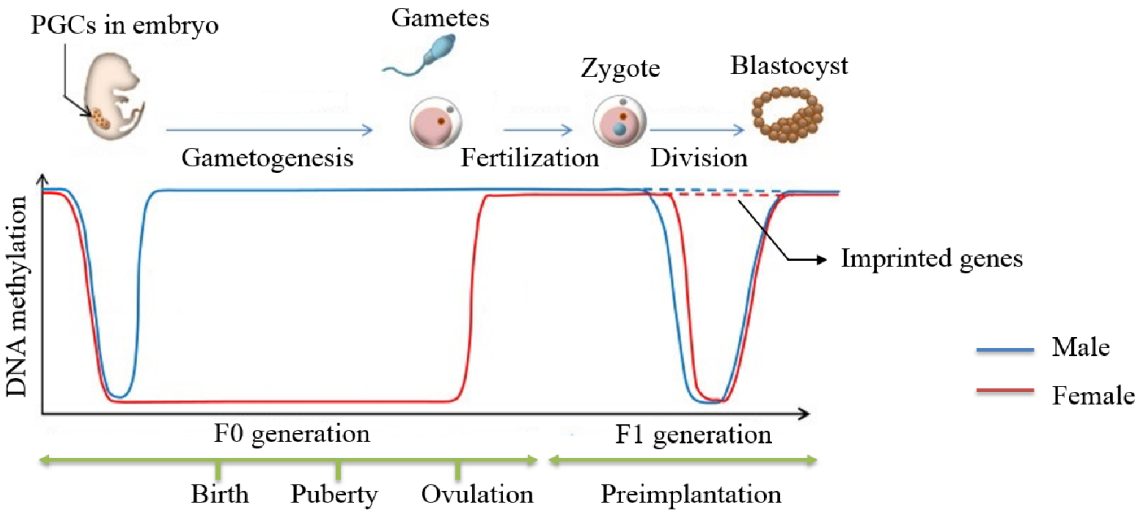


Figure 2. Two waves of genome-wide developmental epigenetic reprogramming phases in mammals
In primordial germ cells (PGCs) of F0 generation, a global erasure of DNA methylation occurs. Then, *de novo* methylation takes place in a sex-specific way. After fertilization in F1 generation zygote, a new wave of DNA de-methylation occurs. Methylation marks of imprinted genes escape this post-zygotic erasure (dotted lines). At the blastocyst stage, new methylation landscape is re-established. Blue and red lines represent male and female genomes respectively. The figure is taken from [18] and modified.

1.4 Fetal programming (Intra-uterine reprogramming of propensity to disease)

The Developmental Origins of Health and Disease (DOHaD) hypothesis associates adverse environmental exposures in the periconceptional and intrauterine period with increased susceptibilities for complex, in particular, metabolic and cardiovascular diseases later in life [19, 20]. In other words, the developing fetus senses its surroundings during the definite window of susceptibility, and reframes its metabolic responses by reprogramming its (epi) genome. Although Barker's hypothesis describes the 'uterine period' as an influential time for developmental plasticity [20], it is now being evident that pre-conception and early post-natal life also play a major role. 'The Dutch hunger winter' refers to a series of well-documented studies wherein most of the Dutch population survived on <1000 calories/day from September 1944 to May 1945. This famine became uniquely significant as pregnant mothers and fetuses were also exposed to malnutrition in 1st, 2nd, 3rd trimesters and even postnatally. The Dutch famine birth cohort study established that the offsprings of those exposed pregnant women were smaller and had a higher risk for obesity, diabetes, and other health problems during adulthood [21, 22].

Despite being extremely tragic, this investigation grants a perfectly designed human experiment wherein the effects of intrauterine growth restriction (IUGR) caused by maternal mal-nutrition or placental insufficiency on subsequent adult health is noted. Another example for fetal reprogramming is that of the observed lower birth weight of offsprings born with the help of *in vitro* fertilization technique when compared to those conceived spontaneously [23]. In this study, data supported the hypothesis that stress during embryo culture at pre-implantation period might be correlated with differences in metabolic profiling.

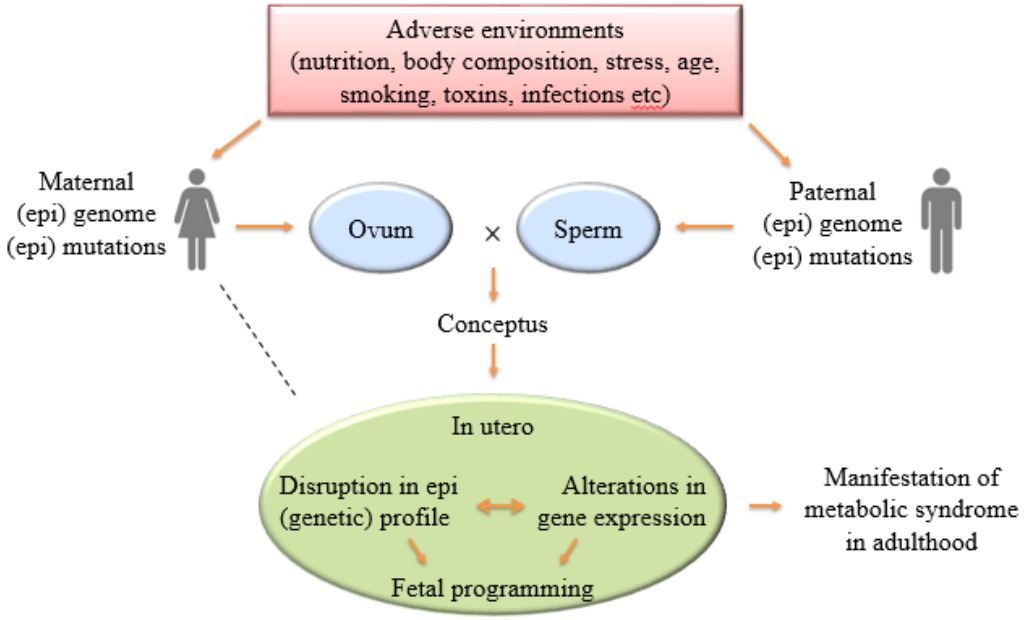


Figure 3. Fetal programming hypothesis

It proposes that adverse nutritional and/or environmental exposures during periods of sensitivity in mammalian development alter fetal (epi) genome and might manifest as metabolic syndrome during adulthood.

The most likely mechanism for translating the effects of environmental factors during germ-cell and embryonal/fetal development into long-term disease susceptibilities in adulthood is persisting epigenetic dysregulation of underlying genes and pathways. Several human epidemiological studies support the effect of DOHaD at the level of DNA methylation. El Hajj *et al*. (2013) reported changes in DNA methylation levels of the imprinted gene mesoderm specific transcript (*MEST*) between cord blood and placenta samples from offspring exposed or not to gestational diabetes mellitus (GDM). In addition, a similar directional methylation change in *MEST* gene was

observed in blood samples from normal weight and morbidly obese adults [24]. This supports the idea that offspring's methylome might be impacted by GDM exposure (fetal epigenome reprogramming) and those epi-variations could be correlated with obesity in adulthood. New knowledge on other epigenetic mechanisms like histone modifications and micro RNAs with respect to fetal reprogramming may soon emerge. There is also growing evidence that the effects which persist till adulthood might have an impact on the following generations resulting in trans-generational epigenetic inheritance [25, 26].

1.5 Trans-generational epigenetic inheritance

To be justifiable, the term 'epigenetic inheritance' can be broadly used in two different contexts; One to describe the transfer of epigenetic information from one cell to another during cell differentiation and body development (mitotic inheritance). The second is to represent the transmission of epigenetic signatures from one generation of an organism to the next generations (in simple words, parent to child transmission). In latter case, it can be appropriately termed as 'trans-generational epigenetic inheritance' which is the key to the model of 'missing heritability'. Despite the two levels being equivalent in the case of unicellular organisms, they might have evolutionary differences in multicellular organisms. At this point, it is very critical to distinguish the effects of *in utero* exposure of pregnant female (F0 generation) to specific environment on developing embryo (F1 generation) and its germline (F2 that will give rise to grandchildren), from genuinely trans-generational effects which are established in generations (from F3) that were unexposed to the initial environmental cue. On the other hand, in case of males, when an epigenetic change is triggered in F0 male and his germ line (future F1), only F2 and subsequent generations may be treated as an evidence of true trans-generational inheritance.

At the level of DNA methylation, although it was hypothesized that epigenetic marks are erased and re-established afresh at each generation (described in section 1.3), constant progress is being made in identifying specific loci in the (epi) genome that bypass methylation reprogramming mechanism. In other words, incomplete erasure of epigenetic landscape is required for the trans-generational inheritance which occurs in the case of imprinted genes and some repetitive elements [13, 17]. Generally, in sperm, histones are replaced by arginine-rich protamines. Despite

the belief that this replacement erases any epigenetic tags at histones and prevents transmission to the next generation, it is now becoming evident that some histone marks like H3K27me3 which are important for embryonic development are still retained in sperm and are inherited via the paternal line to the next generation [27]. Other mechanisms that might involve trans-generational epigenetic inheritance are transmission through RNAs like piRNAs, siRNAs etc.

1.6 Genomic imprinting

Genomic imprinting, also known as gametic or parental imprinting is a process wherein depending on the gene, either the paternal or the maternal allele is silenced resulting in parent-specific, mono-allelic expression of genes. It is an epigenetically inherited phenomenon which mainly involves processes like DNA methylation and histone modifications. These epigenetic tags are marked or imprinted in the germline (egg and sperm cells) of parents and they are retained through mitotic divisions in somatic cells of an individual or organism. Out of 25000 genes in the human genome, 100-200 are estimated to be imprinted [28]. They typically reside in clusters which contain differentially methylation regions (DMRs) between maternal and paternal chromosomes. The DMRs controlling the imprinting of one or more genes are called imprinting control regions (ICRs). Imprinted genes are good candidate markers to study trans-generational epigenetic inheritance since these regions are expected to be partially or completely resistant to methylation reprogramming event, thus ensuring parent-of-origin expression as these epigenetic signatures are passed on to the offspring.

Imprinted genes are mostly involved in early growth, development, metabolism, behavior, brain function and environmental adaptation etc. [29-31]. During growth, maternally and paternally expressed genes carry opposite functions because of sexual antagonism (sexual conflict) of each parent for evolutionary fitness of their genes. This is named as ‘parental conflict hypothesis’ or ‘kinship theory of genomic imprinting’. In simple words, paternally expressed genes (PEGs) tend to promote growth by extracting nutrients from the mother during pregnancy to maximize the success of offspring. On the other hand, maternally expressed genes (MEGs) are growth limiting and they restrain intrauterine growth to preserve resources for future pregnancies [32]. The phenomenon of genomic imprinting prevents the advantage of diploidy as a mutation in the

transcribed allele resulting in absence or dysfunctional gene product can't be balanced through the other mutation-free allele (since the latter is imprinted or silenced) leading to several imprinting disorders like Angelman syndrome, Prader-Willi syndrome, Beckwith-Wiedemann syndrome or transient neonatal diabetes mellitus etc. [33].

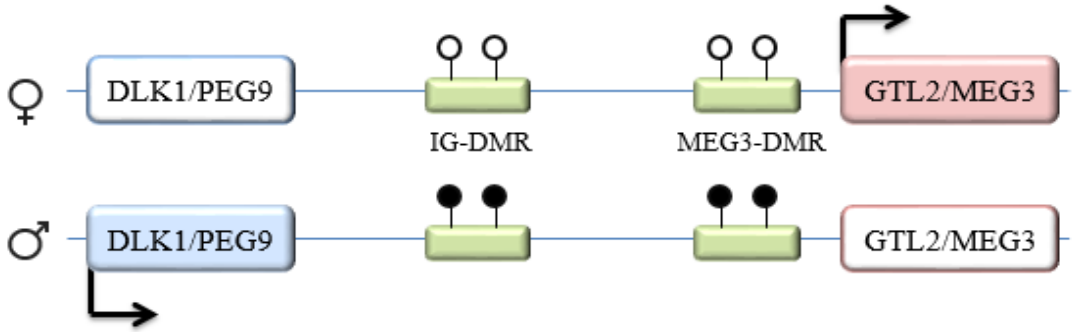


Figure 4. The imprinting cluster at *DLK1*/*MEG3* locus

It contains a primary ICR (IG-DMR; Intergenic DMR) and secondary ICR (*MEG3*-DMR) which overlaps with the *MEG3* promoter. Both ICRs are paternally methylated. Black colored lollipops represent methylated CpG sites whereas uncolored ones are unmethylated CpG sites. *PEG3*: Paternally expressed gene; *MEG3*: Maternally expressed gene. ♀ and ♂ are maternal and paternal alleles respectively. The figure is taken from [34] and modified.

1.7 Repetitive elements

Repetitive elements, also known as repeated sequences or repeats, are patterns of DNA sequences that are present in multiple copies in the genome. Over two-thirds of sequence contains repeats in human genomic DNA. Broadly, they are divided into two classes;

Tandem repeats: They represent patterns that lie adjacent to each other. Based on the length, they are divided into micro-satellites (also known as short tandem repeats; DNA motifs ranging in length of <10 base pairs, typically 5-50 times; includes telomeres), mini-satellites (ranging from 10-60 base pairs, typically 5-50 times), and satellite DNA (large arrays of non-coding DNA repeats). There are also multiple copy genes such as ribosomal RNA genes which fall under the category of tandem repetitive DNA sequences.

Interspersed repeats: They are also known as interspersed nuclear elements. They are dispersed throughout the genome and are non-adjacent. Based on the length, they are in turn divided into

short interspersed nuclear elements (SINEs – 50 to 500 base pair long) and long interspersed nuclear elements (LINEs - ~7000 base pair long). As both SINEs and LINEs do not contain long terminal repeats (LTRs), they are termed as non-LTR retrotransposons. Unlike LINEs, SINEs are non-autonomous since they rely on molecular machinery of LINE elements for proliferation and transposition throughout the genome [35]. SINEs and LINEs make up to 13% and 21.1% of human genome respectively.

In somatic tissues, DNA methylation of transposon-derived repetitive sequences is one of the key mechanisms for defending against their transposition activity and maintaining genomic stability in humans. However, most of these repeats display a low level of DNA methylation in sperm compared to oocytes and many other somatic cells [36]. Since methylation of repeats comprises more than one-third of the total genomic DNA methylation [37], they are often widely used as surrogate markers for analysis of global DNA methylation content in human studies.

1.8 The need to focus on fathers – Importance of studying male germline epigenome

The idea of fetal programming and/or trans-generational epigenetic inheritance has fascinated researchers for many decades. Nevertheless, studies mostly emphasized on the influence of adverse preconceptional and *in utero* exposures on the offspring via maternal line [22, 38]. These epidemiological data on mothers and their next generation contributed to the evidence that along with genetic damage, epigenetic signatures can also lead to heritable phenotypic modifications that may persist through many generations [39, 40]. On the other hand, for a long time, the male germ line epigenome was assumed to be of lesser importance for early embryo development probably because of the following reasons.

1) During the replacement of histones with strong alkaline protamines in sperm, it was assumed that nearly a complete layer of epigenetic information is lost. This replacement leads to 6-20x denser chromatin structure transforming spermatozoa into a small and fast courier of DNA [41]. However, histone to protamine transition is incomplete leaving 5-15% of histone-bound DNA in human sperm [42]. Thus, the retained histone marks may be involved in delivering epigenetic information to the oocyte and thus subsequently affect the offspring [27].

2) Concerning DNA methylation patterns in sperm, as described in section 1.6, imprinted genes are an exceptional landmark to capture and maintain the environmental cues, as they are protected from the large-scale DNA methylation erasure after fertilization. It is also known that these imprinted genes and the retained histone marks in sperm play a crucial role in early embryonic development [27].

3) Having a direct contact of a fetus to the mother during pregnancy and/or disproportion between oocyte and spermatozoon in their sizes and cellular components etc. might have also created an assumption that father's role in potential epigenetic inheritance is not very important. However, studies have also shown that sperm RNA molecules (piRNAs, sncRNAs, siRNAs and micro RNAs etc.) might act as potential mediators for (epi) gene-environmental interactions and that they can moderate non-Mendelian inheritance of acquired traits [43].

In line with this, constant progress is being made to determine the influence of adverse environmental exposures such as malnutrition, high fat diet, smoking, stress, aging, drugs, ionizing radiation etc. on altering the epigenome of sperm at the level of DNA methylation, RNA molecules etc. Therefore, it is very important to understand the male germ line epigenetics to investigate the paternal effects on embryogenesis and on the developing offspring.

1.9 Aging

Definition: Aging, which is the process of becoming older, represents the collection of changes in a human being over time. In spite of the advantages of an increase in knowledge of the world and wisdom etc., aging is among the strongest and greatest known risk factors for many diseases in humans. Of ~150,000 people who die every day across the world, ~ two-thirds of them die due to age-related causes. Theories such as accumulating DNA oxidative damage and changes in DNA methylation etc. are proposed as some contributing factors to the complex process of aging.

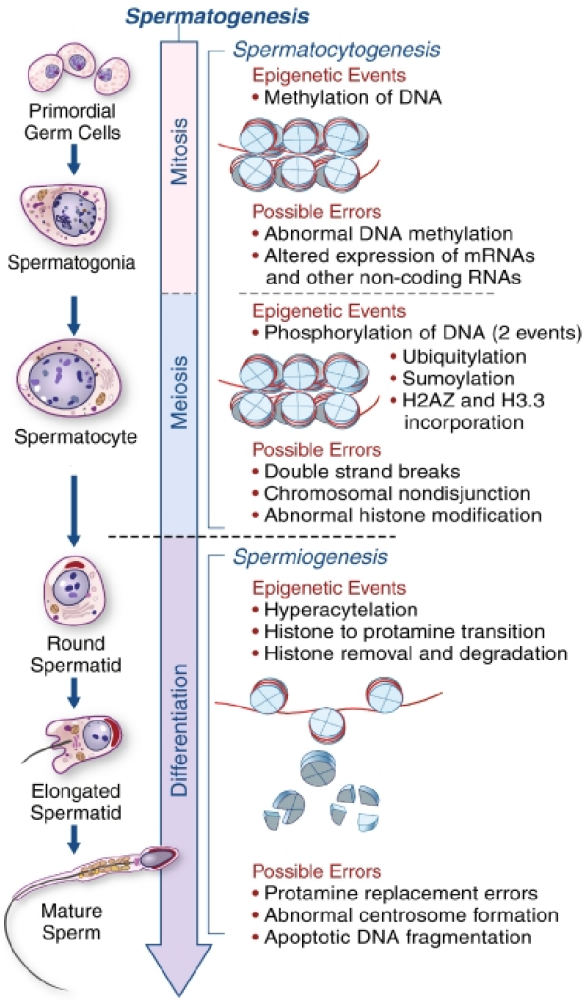
Nowadays, there is a growing trend for delayed parenthood amongst couples mainly because of the pressure for career establishment and economic reasons etc. Historically, the consequences of

this phenomenon are mostly related to maternal aging associating it with reduced fertility, increased risk of spontaneous abortions and chromosomal defects like down syndrome [44]. On the other hand, it is interesting to note that the miscarriage risk for a couple with a ≥35-year-old woman and ≥40-year-old man was ~two-fold higher than that of a couple with a ≥35-year-old woman and a younger male partner [45]. Not much attention has been paid to paternal aging effects probably due to the life-long spermatogenesis process and potential male fertility factor. However, recent accumulating evidence from epidemiological studies in humans and animal models have shown an association of advanced paternal age with an elevated risk for several neuro-developmental disorders like autism, bipolar disorder, and schizophrenia etc. in their offspring [46-48]. Additionally, paternal age is also shown to affect the learning and cognition in their corresponding children [49].

DNA methylation and aging: Alterations in DNA methylation patterns might be associated with advanced age in approximately every tissue of the body. For example, the 5-methylcytosine content of the DNA markedly reduced when diploid fibroblasts from humans, mice, and hamsters were cultured to senescence [50]. Another example was the occurrence of hypermethylation at promoter CpGs in tumor suppressor genes with an increase in age, ultimately contributing to the risk for cancer (age-related disease) [51]. Research also showed that frequently dividing cells seemed to be more accessible to changes in methylation patterns with increasing age than that of the less recurrently dividing cells [52]. Male gametogenesis represents one such system where the constantly dividing cellular system might be affected by errors with an increase in age.

Spermatogenesis: Among all the body cells, sperm cells have an exceptional status since they are highly specialized to overcome the different environments of both male and female reproductive tracts. Spermatogenesis is a life-long process which starts with primordial male germ cells undergoing ~30 mitotic divisions during embryogenesis producing spermatogonial stem cells after which the process is stopped until puberty. Unlike in females, even after puberty, the stem cell divides mitotically every 16 days producing two different types of cells in males (asymmetrical division): one stem cell (to replenish stem cell pool) and the other spermatogonium which further undergoes mitosis to produce two primary spermatocytes. Each of these primary spermatocytes undergoes two meiotic divisions ultimately giving rise to four

mature sperm cells. In an extrapolation, this whole mechanism leads to a fast increase of mitotic divisions with an increase in age in males. In detail, spermatogonial stem cells divide ~35 times at puberty and ~840 times in a 50-year-old [53].



The figure is taken from [54] and modified.

Figure 5. The possible epigenetic events during spermatogenesis

DNA methylation occurs to establish paternal-specific imprints; Phosphorylation to facilitate recombination and XY body formation; Hyperacetylation to assist in histone to protamine exchange. In addition to double-strand breaks, abnormal or altered histone, RNA and methylation modifications, spermatogenesis also gives rise to chromosomal non-disjunctions during meiosis. Factors like paternal aging and obesity might affect the process of spermatogenesis and subsequently impact the corresponding offspring's health.

Effect of paternal aging on sperm epigenome: Considering that spermatogenesis is a continuous process supports the fact that the sperm of older men are more prone to *de novo* mutations and their corresponding offspring are more susceptible to various genetic diseases [53]. In other words, during each cell division (both in mitosis and in meiosis), genetic aberrations might occur continually elevating the mutational load in sperm of older males. Recent evidence suggests that these altered clonal lines are positively selected and expanded within testis via a process analogous to oncogenesis [55]. This expansion of mutation-carrying spermatogonial cells, which

is described as 'selfish spermatogonial selection model' or 'germline selective advantage model' [55, 56], leads to the relative enrichment of mutant sperm over time with aging.

Not only the genetic information but also the epigenetic modifications should be copied to the daughter cells during each cell division. Since research showed that biochemical copying of epigenetic signatures is more prone to errors than genetic marks [57, 58], one can plausibly hypothesize that aging sperm cells acquire 10-100 times more epigenetic than DNA sequence changes. Epigenetic alterations, in particular, DNA methylation during aging have been widely studied [59]. Mouse studies elegantly displayed an association between age-related sperm methylation aberrations and changes in methylation, gene expression in offspring of older mice [60, 61]. Thus, it is important to understand the paternal age effects on sperm methylome and its consequences on embryogenesis and developing offspring in humans.

1.10 Obesity

Definition: Obesity can be termed as a medical condition wherein the excess fat in the body of an organism or an individual accumulates to such an extent that it might have a detrimental effect on health. It is generally defined by body mass index (BMI) or Quetelet index which is a measurement attained when dividing a person's body mass by the square of his/her body height and is universally expressed in kg/m^2. The table below presents the international classification of an adult underweight, normal weight, overweight, and obesity proposed by World Health Organization (WHO) in the year 2000 [62].

Table 1. Classification by WHO of overweight or obesity according to BMI

BMI (kg/m^2)	Classification
<18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Overweight / Pre-obese
30.0-34.9	Obese class 1 (Moderate)
35.0-39.9	Obese class 2 (Severe)
≥40.0	Obese class 3 (Very severe)

Occurrence and effects: According to WHO, since 1980, the worldwide prevalence of obesity has nearly doubled. In 2016, throughout the world, more than 1.9 billion adults were overweight of which over 650 million were obese. 41 million children (below 5 years of age) and over 340 million children and adolescents (5-19 years) were overweight or obese in 2016. It is well established that obesity is a key risk factor for the development of various diseases, particularly type 2 diabetes (T2D), cardiovascular problems, obstructive sleep apnea, osteoarthritis, depression and certain types of cancer [63]. It is also surprising to know that most of the world's population lives in those countries where obesity kills more people than underweight.

Causes: Obesity can be considered as a complex, multifactorial chronic condition where its susceptibility can be affected by independent and interacting effects of environment, genetic and epigenetic origins [64]. One of the fundamental causes is an energy imbalance between calories consumed and expended. Some obesogenic causes are listed as follows:

1) Environmental origins include excessive intake of high-fat, high-sugar, energy-dense foods, inadequate sleep, pollutants and pathogens that might interfere with lipid metabolism, use of certain medications, pregnancy at a late age, physical inactivity, sedentary lifestyle and certain illnesses etc. Studies have shown that when adequate food energy is present, polymorphisms and/or mutations in certain genes that control appetite and body metabolism can pre-dispose to obesity [65]. So far, over 40 of such common genetic markers in the human genome have been linked with obesity or BMI out of which some genes namely *FTO*, *POMC*, *BDNF*, *TMEM18*, *PCSK1*, *CTNNBL1*, and *MC4R* etc. are the strongest factors that appeared in many studies [63, 66]. In spite of the predicted heritability of obesity being ~75%, the associated DNA-sequence based signatures till date could explain only <30% of BMI variation [67].

2) Lately, researchers proposed that epigenetic tags could be considered as a further form of variables influencing the etiology, susceptibility, and heritability of obesity [64, 66, 68]. Recent advances in high-throughput technologies like epigenome-wide association studies (EWAS), chromatin immunoprecipitation-sequencing (ChIP-seq) and RNA-sequencing etc. are contributing to explore the epigenetic basis of metabolic disorders like obesity.

Role of DNA methylation in explaining obesity dynamics: One profound example indicating that epigenetic dysregulation (silencing) of an imprinted region on chromosome 15 resulting in obesity is the case of Prader-Willi syndrome [69]. This region harbors paternally-expressed imprinted genes namely *MKRN3*, *MAGEL2*, *NDN*, *SNURF*, *SNRPN*, and a family of six paternal-only expressed snoRNA genes, and preferentially maternally-expressed *UBE3A* and *ATP10A* genes [70]. Recently, a EWAS study has identified a positive correlation of BMI with methylation at human non-imprinted *HIF3A* gene loci in blood and adipose tissue [71]. Pre-natal and/or peri-conceptional maternal under-nutrition showed epigenetic alterations in DNA methylation of genes like *hIGF2*, *hGNASAS*, *hMEG3*, *hIL10*, and *hLEP* [72, 73] . Surgery-induced weight loss and acute or long-term exercise resulted in modified DNA methylation in skeletal muscle and adipose tissue in humans [74-76]. Additionally, many studies on animal systems including mice, drosophila, rat, macaque, sheep, and pig showed that obesity influences epigenome at the level of DNA methylation and histone modifications of not only the exposed experimental model, but also their subsequent generation(s) proposing the evidence of fetal re-programming (intra-uterine exposure) and/or trans(inter)-generational epigenetic inheritance [77]. These obesogenic changes are not restricted to maternal influences alone as there is a growing epidemiological evidence that paternal dietary exposures also influence his metabolic phenotype and sperm epigenome by a greater extent and thus impact the next generation(s) [78-80].

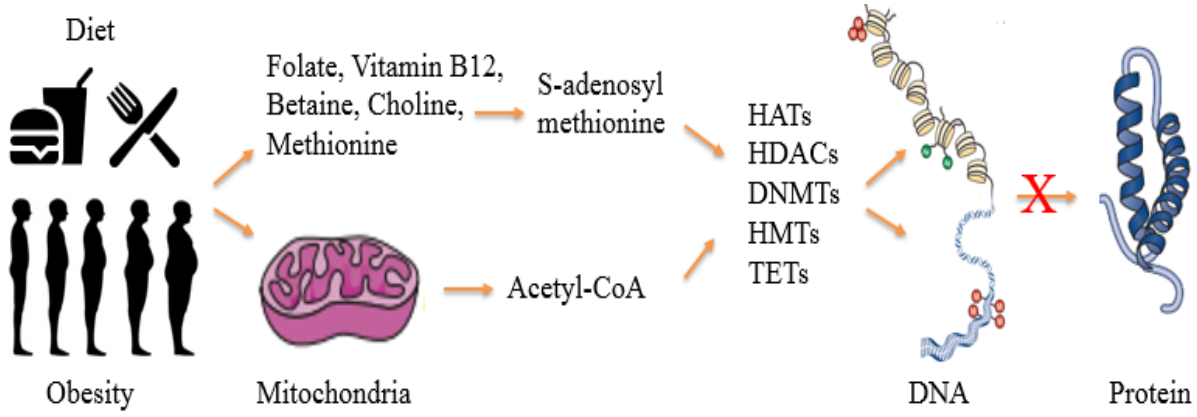


Figure 6. The effect of dietary factors and obesity on the epigenome

Methyl and acetyl groups required for epigenetic processes are obtained from the diet. HATs: Histone acetyltransferases, HDACs: Histone deacetylases, DNMTs: DNA methyltransferases. HMTs: Histone methyltransferases, TETs: Ten-eleven translocation enzymes. M stands for methylation (red) on DNA strand and Ac stands for acetylation (green) of histone molecules. The figure is taken from [81] and modified.

Effect of obesity on sperm epigenome: Obesity-induced epigenetic alterations in somatic cells of an organism or an individual is shown to influence its (his) germline [82]. Studies showed that sperm samples exhibited altered DNA methylation patterns in male rats fed a high-fat diet compared to controls [83]. Differences in RNA content and chromatin packaging were also observed in sperm samples of mice fed a low-protein diet [84]. Additionally, obesity-associated increase in histone modifications, in particular, acetylation in spermatids was reported. Indeed, receiving high-fat diet for male mice was prone to affect the fertility, increase sperm DNA damage, decrease in the number of motile spermatozoa and generation of unbalanced reactive oxygen species [85].

Recent reports in human studies also proved that paternal obesity can reprogram male germ line (sperm) at DNA methylation level [86]. For example, weight loss after gastric bypass surgery was established to re-model sperm methylome especially in gene regions that play a major role in the control of appetite [82]. In line with this, there is a growing interest among researchers to find out if this obesity or higher body mass index induced epigenetic alterations in sperm samples will have an impact on their next generation offspring in humans.

1.11 Aim of the thesis

- To study the effect of male aging on human and bovine sperm DNA methylation at specific repetitive elements in the genome. The impact of human paternal age effects on next generation embryo grading was also explored.

- To examine the influence of male obesity/body mass index on human sperm DNA methylation and to detect the epigenetic transmission into the next generation fetal cord blood samples. In-depth methylation analysis at the regulatory regions of 8 imprinted genes and one obesity-related gene was performed.

To study the methylation effects, techniques namely pyrosequencing and deep bisulphite sequencing on Illumina MiSeq platform were applied. The effect of promoter CpG methylation on expression of the gene was examined by functional analysis using Luciferase Reporter Assay.

1.12 Regions of interest

1.12.1 Repetitive Elements: As discussed in section 1.7, repetitive DNA sequences/elements comprise over two-thirds of the human genome [87]. They are used as surrogate signatures for measuring genome-wide DNA methylation content in humans [88] since their methylation constitutes more than one-third of the total genomic DNA methylation [89]. The repeats studied in this present work are summarized in the below figure.

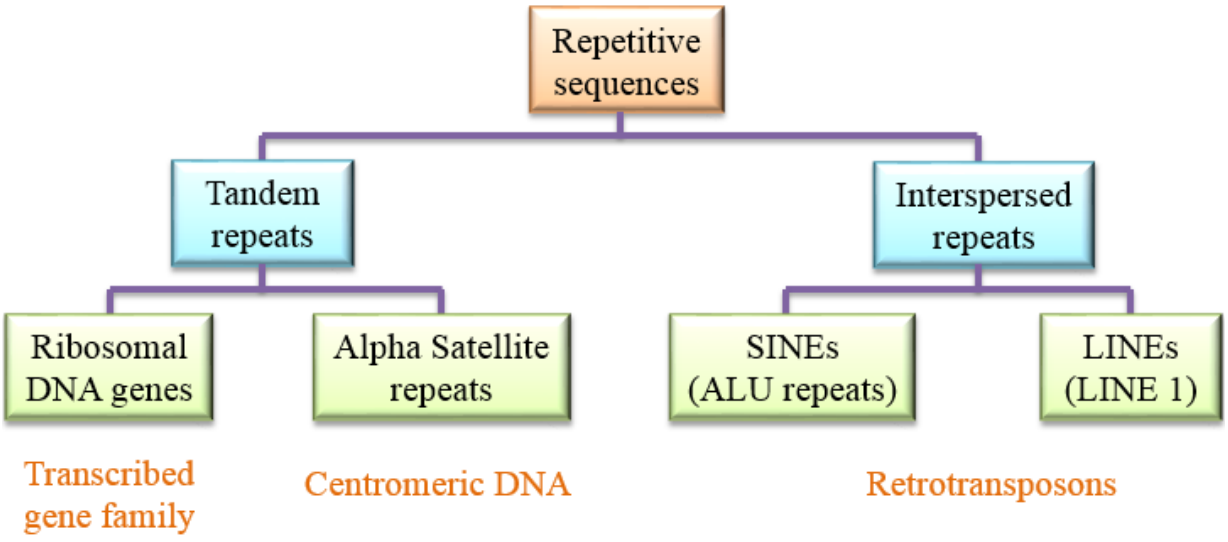


Figure 7. Classification of the studied repetitive elements

They can be divided into tandem and interspersed repeats as described under section 1.7. Ribosomal DNA genes and alpha-satellite elements come under tandem repeats and ALU and LINE1 come under interspersed repeats.

Ribosomal DNA (rDNA): Ribosomal DNA is a DNA sequence that codes for ribosomal RNA (rRNA, the most abundant form of RNA). rRNA molecules along with proteins assemble to form ribosome(s), a complex molecular machinery which serves as the site of biological protein synthesis. In the human genome, there are ~300-400 copies of rDNA genes which are organized in tandemly arrayed repeats within nucleolar organizing regions (NORs). These NORs are located on the short arm of five acrocentric chromosomes 13, 14, 15, 21, and 22.

rDNA genes are initially transcribed as 45S pre-RNA that are then spliced into 5.8S, 18S, and 28S rRNA transcripts which are assembled in the nucleolus as respective ribosomal subunits

[90]. 5S (coded separately), 5.8S, and 28S rRNA molecules are present in the larger subunit (60S) of eukaryotic ribosome whereas 18S molecules are in the smaller subunit (40S) [91].

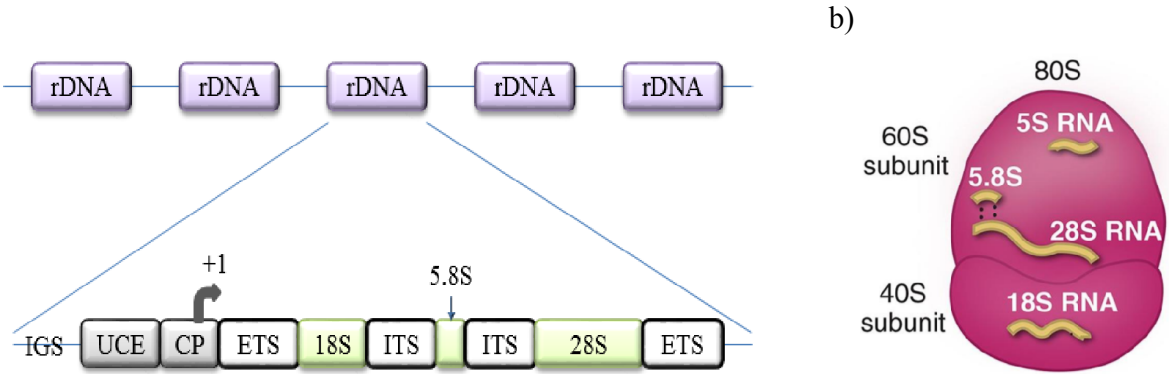


Figure 8. The structural organization of human rDNA repeating unit

IGS: intergenic spacer; UCE: upstream control element; CP: core promoter; ETS: external transcribed spacer; ITS: internal transcribed spacer (left). Eukaryotic ribosome showing larger subunit 60S with 5S, 5.8S, and 28S rRNA molecules and smaller subunit 40S with 18S rRNA molecules (right). Figure 8a is taken from [92] and modified.

Alpha-Satellite repeats: These are large arrays of tandemly organized repeating, non-coding DNA. These repeats makeup to ~5% of the human genome and are principally positioned at the centromeric and pericentromeric regions of each human chromosome. Because the repetitions of DNA sequences tend to yield a different frequency of four bases and varying density from that of the bulk DNA, formation of second or satellite band occurs when genomic DNA is separated on a density gradient. Hence the name 'satellite DNA'.

They are composed of repeating monomeric sequences of ~171 base pair length that differ by as much as 40% [93]. Several monomers give rise to a higher order repeat which reiterates 100-1000 times and in a multimeric organization so that arrays are highly homogenous spanning 100 kilobases to several megabases. Alpha satellite DNA plays an important role in cell division and kinetochore formation. The latent transcriptional potential of pericentromeric elements is repressed by DNA methylation which is essential for appropriate chromosome alignment, segregation, and integrity during mitosis [94]. Studies showed that this repeat plays an important role in maintaining chromatin in a particular conformation [95].

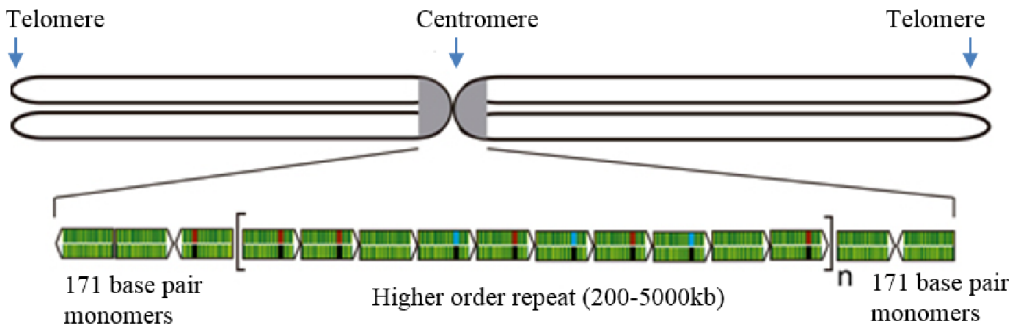


Figure 9. Structure of centromeric alpha-satellite repeats

171 base pair monomeric unit, in turn, gives rise to higher order repeat which spans 100 kilobases to several megabases. The figure is taken from [96] and modified.

Long interspersed nuclear element-1 (LINE-1): LINEs are non-long terminal repeat (non-LTR) retrotransposons which are ~ 6-8 kb in length. With more than 500,000 copies, the most abundant group is LINE-1 repeats comprising ~17% of human genome. The consensus sequence of L1 repeats consists of two non-overlapping open reading frames (ORF) encoding an RNA-binding protein and a reverse transcriptase [89]. LINE1 RNAs are also abundant in spermatogenic cells and preimplantation embryos. They can be transmitted into the oocyte by the spermatozoa at fertilization. This proposes a certain role for these repeats during spermatogenesis or early embryogenesis [97]. Active transposable elements generally target protein-coding genes for insertion, cause chromosomal breakage, genome rearrangement, and incorrect recombination. Though some repeats retain their ability to retro-transpose, the majority of the repeats are inactive in human genome so that they can maintain genomic stability.

ALU elements: One class of short interspersed nuclear elements is ALU repeats which are originally characterized by *Arthrobacter luteus* (ALU) restriction endonuclease. They are a group of non-autonomous and one of the most abundant transposable elements consisting of one million copies interspersed throughout the human genome thereby covering more than 10%. The full-length ALU repeats have a size of around 300 base pairs; hence they are classified under short interspersed nuclear elements (SINEs). They are usually located in introns, 3' untranslated regions of genes, and intergenic regions [98]. ALU elements are associated with most gene promoters and may play a role in gene regulation [99]. ALU repeats display lower methylation levels in sperm compared to oocytes and various other somatic tissues [36]. Since ALU elements

contain an abundant amount of CpG dinucleotides, they serve as a site for methylation contributing up to ~30% of the total methylation sites in the human genome.

1.12.2 Imprinted Genes: As discussed in section 1.6, these set of genes are ideal markers to investigate trans-generational effects since they resist the global epigenetic reprogramming mechanism after fertilization thus inheriting alterations to methylation signatures from parents.

MEST/PEG1: Paternally expressed gene 1 (*PEG1*), also known as the mesoderm-specific transcript (*MEST*) is the first imprinted gene mapped on the human genome. It is a maternally imprinted gene located on chromosome 7q32 in humans [100]. The gene uses two distinct promoters encoding two different isoforms in humans; Isoform 1 is expressed only from paternal allele whereas isoform 2 from both the alleles [101]. The *MEST* promoter studied in this work was methylated on the maternal allele (isoform 1) and was demonstrated to play a role in prenatal and postnatal development in various tissues. In mice, with genetically caused or diet-induced obesity, *Mest* gene expression was clearly enhanced in white adipose tissue leading to the enhancement of adipocytes and increased expression of metabolic related genes [102]. In addition, loss of *Mest* function was associated with intrauterine growth restriction [103]. A study in our group showed that the *MEST* methylation in peripheral blood of obese adults was significantly lower compared to normal weight controls [24].

PW1/PEG3: Paternally expressed gene 3 (*PEG3*) is a maternally imprinted gene located on chromosome 19q13 in humans [104]. The protein encoded by this gene acts as a zinc finger transcription factor. Alternative splicing leads to multiple transcript variants encoding distinct isoforms. It is mainly expressed during embryogenesis but also in adult tissues like brain, reproductive organs, and muscles. In mice, *Peg3* was involved in a physiological pathway regulating obesity, feeding behavior, and body temperature [105]. $Peg3^{+/-}$ mutant mice developed an excess intra-scapular, abdominal, and subcutaneous fat, and this mutation was shown to regulate multiple events correlating to energy homeostasis [105]. This gene was also known to play an important role in mammalian reproduction, fetal growth rates, and nurturing behaviors [106].

SNRPN/PEG4: Paternally expressed gene 4 (*PEG4*), also known as small nuclear ribonucleoprotein polypeptide N (*SNRPN*) is a maternally imprinted gene located on chromosome 15q11 in humans. The protein encoded by this gene plays an important role in tissue-specific alternative splicing and in pre-mRNA processing. Deletion or mutation in the paternally expressed allele of this gene leads to imprinting disorder named Prader-Willi syndrome which is characterized by obesity, hyperphagia, and hypogonadism [107]. A recent study in Spanish population showed an association between genetic variability in *SNRPN* gene and the risk for developing obesity [108].

NNAT/PEG5: Paternally expressed gene 5 (*PEG5*), also known as Neuronatin (*NNAT*) is a maternally imprinted gene located on chromosome 20q11 in humans. The protein encoded by this gene is a proteolipid and is involved in mammalian brain development by maintaining and forming the structure of the nervous system. Increased expression of the neuronatin protein in pancreatic islet beta cells caused cells to undergo apoptosis ultimately leading to diabetes mellitus [109]. Studies showed that during fasting condition and under defective leptin signaling in rodents, mRNA expression of this gene was declined in various key appetite-related neuropeptides [110]. At the genetic level, single nucleotide polymorphisms (SNPs) in human *NNAT* gene were associated with severe childhood and adult obesity [111].

SGCE/PEG10: Paternally expressed gene 10 (*PEG10*), also known as epsilon-sarcoglycan (*SGCE*) is a maternally imprinted gene located on chromosome 7q21 in humans [112]. This gene is expressed in adult and embryonic tissue, most prominently in the placenta. It plays a major role in cell differentiation, proliferation, and apoptosis. A knockout model of this gene in mice reported early embryonic lethality along with placental defects [113]. It is known to be one of the genes expressed at early stages of adipogenesis [114].

GTL2/MEG3: Maternally expressed gene 3 (*MEG3*), also known as gene trap locus 2 (*GTL2*) is a paternally imprinted gene located on chromosome 14q32 in humans, in the imprinted *DLK1-MEG3* locus. As shown in fig.4, the cluster which acts as imprinting control center contains an intergenic DMR (IG-DMR; primary imprinting control region), and *MEG3*-DMR (secondary ICR) in which the latter overlaps with the *MEG3* promoter [34]. Due to alternative splicing, there

are at least 12 different isoforms of this gene, all of which are long non-coding RNAs (lnc RNAs) [115]. In the present study, we assessed the DNA methylation of IG-DMR. Hypermethylation of IG-DMR was related to loss of *MEG3* expression [116]. A decreased expression of *MEG3* gene and dramatic down-regulation of non-coding RNAs of this locus was found in human islets from type 2 diabetes mellitus patients [117].

H19: It is a paternally imprinted/silenced (maternally expressed) gene located on chromosome 11p15 in humans which codes for a long non-coding RNA. It is known to be strongly expressed during embryogenesis and in mammary glands and uterus during pregnancy in females. Like many other imprinted genes, *H19* also plays a major role in early growth and development. Mice with deficient *H19* function revealed an overgrowth phenotype whereas its over-expression caused lethal effects [118]. Hypermethylation of *H19* maternal copy was associated with over/biallelic expression of *IGF2* gene, ultimately leading to Beckwith-Wiedemann syndrome with overgrowth phenotype [119]. On the other hand, hypomethylation of *H19* was linked to *IGF2* under-expression resulting in Silver-Russell syndrome with primordial dwarfism [120]. Research showed that *H19* CpG methylation was associated with childhood obesity suggesting its role in the regulation of body weight [121].

IGF2: Insulin-like growth factor 2 (*IGF2*) gene is located on chromosome 11p15 (like *H19* gene) in humans. It is an imprinted gene with expression occurring from paternally inherited allele. Transcription factor named zinc finger protein or CCCTC-binding factor (CTCF) is involved in repressing this gene's expression by binding to *H19* imprinting control region (ICR) and DMR1, and thus ultimately limiting the downstream 'enhancer' access to *IGF2* locus. This is applicable only in the case of the maternal allele; whereas, on the paternal allele, CTCF is no longer bound allowing enhancer to turn on the paternal *IGF2* promoter [34]. DMR0 between exon 2 and 3 at P0 promoter, and DMR2 between exon 8 and 9 are other DMRs of *IGF2* gene in humans. However, *IGF2* transcription is principally regulated by *H19* DMR1 in the imprinting control region and *IGF2* DMR0.

IGF2 has a major role in cell division and differentiation. It also participates in metabolic regulation, fetal growth, and development during gestation. The epigenetic alterations in this

region were associated with Wilms tumor, Beckwith-Wiedemann, and Silver-Russell syndromes [122]. It was hypothesized to be correlated with diabetes mellitus since *IGF2* enhanced the development of fetal pancreatic beta cells [123]. In the present study, *IGF2* and *H19* regions methylation levels were measured at DMR0 and *H19* DMR respectively.

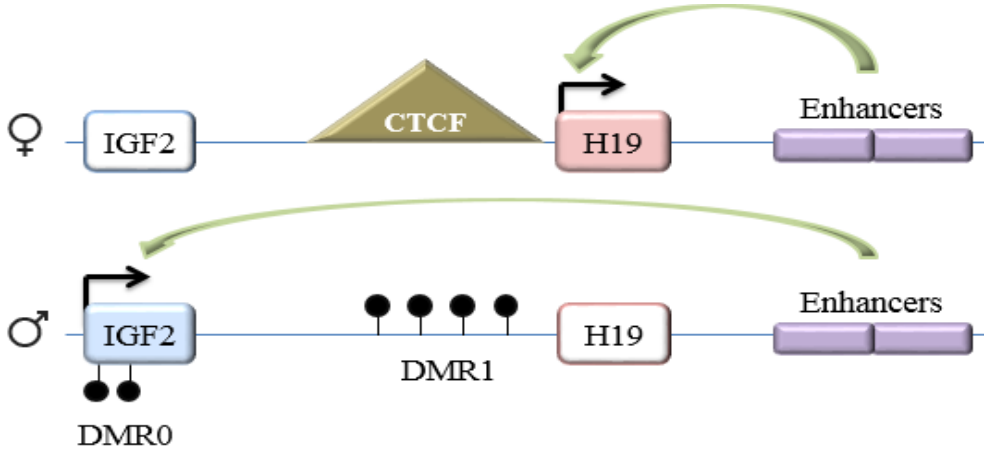


Figure 10. The imprinting cluster at H19/IGF2 locus

It consists of paternally silenced ICR. On the maternal allele, ICR is bound by CTCF (insulator protein) thus preventing the availability of *IGF2* promoter to downstream enhancers. Methylation at ICR on the paternal allele does not allow CTCF protein to bind, enabling *IGF2* expression. Black colored lollipops represent methylated CpG sites whereas the uncolored ones are unmethylated CpG sites. ♀ and ♂ represent maternal and paternal alleles respectively. DMR stands for the differentially methylated region. The figure is taken from [124] and modified.

1.12.3 Non-imprinted HIF3A: Hypoxia-inducible factor 3 alpha is a non-imprinted gene located on chromosome 19q13 in humans. The protein encoded by *HIF3A* gene is alpha 3 sub-unit amongst various α/β hetero-dimeric transcriptional factors which collectively control adaptive responses to decrease oxygen tension (hypoxia). It was hypothesized that factors containing alpha 3 subunits act as negative regulators of hypoxia and was shown in the case of rats [125]. On the other hand, *HIF3A* is involved in cellular response to insulin and glucose and functions as an accelerator of adipogenesis [126]. It enhances adipocyte differentiation by inducing the expression of several kinds of adipocytes-related genes. A genome-wide methylation analysis using Infinium Human Methylation 450 array revealed an association between body mass index and methylation at the *HIF3A* locus [71]. The same study showed a significant inverse correlation between methylation and expression of *HIF3A* probes in adipose tissue.

2 Materials and Methods

2.1 Materials

All general chemicals, materials, and solutions were purchased from diverse providers or prepared in-house and are not listed here.

Ethics declaration

The ethics committee at the medical faculty of the University of Würzburg approved the study on human sperm and fetal cord blood samples (no. 117/11 and 212/15). Written informed consent was obtained from couples undergoing treatment at the Fertility Center Wiesbaden. All the collected samples were pseudonymized and frozen at -80°C until further use.

Samples

2.1.1 Human sperm samples: All analyzed samples were provided by the Fertility Center "Kinderwunschzentrum" in Wiesbaden, Germany. After *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI), the left-over swim-up sperm fractions (excess material) were collected and frozen at -80°C. The human sperm cohort in this work consisted of a total number of 294 samples. Out of these, the corresponding fetal cord blood samples from 109 sperm undergoing IVF/ICSI were analyzed in this work. Sperm cohort consisted of information such as paternal age (ranging from 25-65 years), body mass index (ranging from 17-40 kg/m^2), and spermiogram parameters like sperm concentration. Additionally, twenty fertile male sperm samples who conceived naturally were obtained from the Center of Reproductive Medicine and Andrology, Muenster through Prof. Dr. Jörg Gromoll.

2.1.2 Human fetal cord blood samples: Following live birth after successful IVF/ICSI of 109 previously mentioned sperm samples, 113 fetal cord blood (FCB) samples (including some twins) were obtained (analyzed in this study) from collaborating obstetric clinics throughout Germany. For FCB samples, information about gender, birth weight, and gestational week of the child, mother's age and BMI along with the above mentioned paternal parameters were registered.

2.1.3 Bull sperm samples: A total of 36 semen samples were obtained from 15 different bulls, out of which 12 bulls had their sperm collected at 2 or 3 different ages. All the samples were collected in duplicates; one for methylation analysis and the other to carry out functional studies in embryos. All the samples were obtained through Prof. Dr. Heiner Niemann from the Institute of Farm Animal Genetics, Friedrich Loeffler Institute, Mariensee.

2.2 Methods

2.2.1 Sperm purification

Human sperm samples: The gently thawed sperm sample was purified by PureSperm 80 and 40 (Nidacon, Mölndal, Sweden) two-phase density gradient to eliminate the contamination by bacteria, lymphocytes, epithelial and other somatic cells. 2mL of PureSperm 80 (lower layer) was added to a 15mL glass tube and later 2mL of PureSperm 40 (upper layer) was carefully layered on the top of it. Up to 1.5mL of liquefied semen was pipetted on it without any interference to the gradient. The tubes were centrifuged at 300 x g for 20 minutes. All the layers were recovered by carefully aspirating in circular movements, except 4-6 mm 80 fraction of the remaining pellet (the lowermost part after centrifugation) that contained the purified spermatocytes.

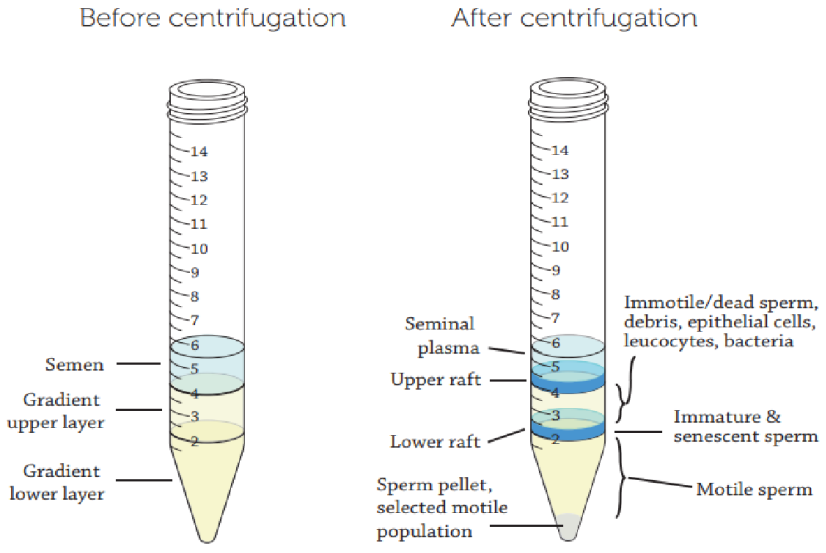


Figure 11. Two-phase density gradient centrifugation for purification of human sperm samples

The figure is taken from Nidacon PureSperm density gradient preparation protocol.

Bull sperm samples: The purification was done by BoviPure and BoviDilute (Nidacon, Mölndal, Sweden). The 40 fractions were prepared by combining 200µL BoviPure and 300µL BoviDilute and 80 fractions were made by 400µL BoviPure and 100µL BoviDilute. 500µL each of 80 and 40 fractions over which 250µL of semen was layered in a 1.5mL tube in case of bull sperm. The Eppendorf tubes were centrifuged at 300 x g for 15 minutes at room temperature by not using any brakes. The supernatant fractions were carefully collected in separate tubes leaving only the pellet which contained the purified spermatocytes.

2.2.2 Sperm DNA extraction (Human and Bull)

The bottom 0.5mL or pellet was supplemented with 300µL of sperm resuspension buffer (5mL of 5 M NaCl, 5mL of 1 M Tris-HCl (pH 8), 5mL of 10% SDS (pH 7.2), 1mL of 0.5 M EDTA (pH 8), 1mL of 100% β-mercaptoethanol, and 33mL of dH_2O), and 100µL Proteinase K (Qiagen, Hilden, Germany), and the samples were incubated for 2 hours at 56^{o}C on a thermomixer. After 2 hours, 20µL of Proteinase K was added and incubated for 2 more hours. DNA isolation was done using the DNeasy Blood and Tissue kit (Qiagen). After addition of 400µL of Buffer AL and 400µL of absolute ethanol (Sigma Aldrich, München, Germany), the lysate was mixed thoroughly by vortexing to yield a homogenous solution. Initially, 650µL of this mixture was transferred to spin columns (provided in the kit), centrifuged at 8000rpm for 1 minute, and the flow through was discarded. This above step was repeated until the whole mixture was processed. Subsequently, two washing steps were performed by adding 500µL of the AW1 buffer, centrifuging at 8000rpm, discarding supernatant followed by adding 500µL of the AW2 buffer, centrifuging at 14000rpm for 3 minutes, and discarding the supernatant. 100µL of AE buffer was pipetted directly onto the DNeasy membrane, incubated at room temperature for 1 min, and the solution was eluted by centrifugation at 8000rpm for 1 min. Elution with an additional 100µL AE buffer was also done in a separate 1.5mL tube.

2.2.3 Fetal Cord Blood (FCB) DNA extraction

Initially, FCB samples were quickly thawed in a water bath at 37^{o}C after which DNA extraction was done using FlexiGene kit (Qiagen). 500µL of the sample was mixed with 1250µL of the FG1 buffer in a 2mL tube, and mixture was inverted 5 times. Two centrifugation steps were carried

out (by adding 1000µL of FG1 buffer second time) at 10000g for 1 min wherein the supernatant was discarded each time, the tube left inverted onto a clean piece of absorbent paper for 2 min, taking care that the pellet remains in the tube. A mix of 500µL FG2 and 2.5µL Qiagen Protease was prepared, 250µL of this mixture was added to the dried pellet and the tube was vortexed to yield a homogeneous solution which was heated for 5 min at 65°C. After mixing the solution with 250µL of 100% isopropanol, DNA precipitate became visible. After centrifuging for 3 min at 10000g and discarding the supernatant, the pellet was vortexed for 5 seconds with 250µL of ethanol (70%). After removing the supernatant by centrifugation, the tube was inverted again on a clean piece of absorbent paper for 5 minutes to air dry the pellet. 200µL of FG3 buffer was used to dissolve the DNA by incubating at 1 hour at 65°C and then overnight at 37°C at 250 rpm.

2.2.4 DNA quantification

DNA concentration and purity was measured by NanoDrop 2000c spectrophotometer (peqLab, Erlangen, Germany) by loading 1µL of each sample. AE buffer or water was used as a blank. A 260/280 ratio of ~1.8 and a 260/230 ratio of 1.8-2.2 was considered as pure DNA.

2.2.5 Bisulphite conversion

Sodium bisulphite treatment of DNA leads to the conversion of unmethylated cytosine residues to uracil whereas the methylated cytosines remain unchanged. In this way after this step, unmethylated and methylated cytosines can be distinguished from each other.

	Original sequence	After bisulphite conversion
Unmethylated sequence	C-G-C-G-C-G-C-G-C-G	U-G-U-G-U-G-U-G-U-G
Methylated sequence	C-G-C-G-C-G-C-G-C-G	C-G-C-G-C-G-C-G-C-G

EpiTect Fast 96 DNA Bisulphite kit (Qiagen) was used for bisulphite conversion of all the samples. The isolated and quantified genomic DNA was thawed, and ~1µg from each sample was added to a 200µL PCR tube. Pipetting scheme and the thermocycler conditions are listed in table 2 and 3 respectively. Upon completion of the reaction, the converted DNA was cleaned. Initially, the EpiTect 96 plate was placed on the top of an S Block (provided in the kit). Each well of the

Epitect plate was filled with 310µL of BL buffer, the total volume of bisulphite reaction and 250µL of ethanol (96-100%), and the whole reaction solution was mixed well by pipetting up and down. The plate was centrifuged at 5800g for 1 minute at room temperature and the supernatant was discarded. After a washing step with 500µL of BW buffer, an incubation step with 250µL of BD buffer for 15 min followed. Subsequently, two wash steps each with 500µL of BW, and one wash step with 250µL of ethanol (96-100%) were done. Finally, the plate was centrifuged for 15 min at 5800g to ensure the evaporation of any residual ethanol. Each sample was eluted with 70µL of EB buffer into an EpiTect elution plate, and the final purified plate was stored at -20°C.

Table 2. Pipetting scheme for bisulphite conversion of genomic DNA for each sample

Component	Volume for 1ng-2µg DNA	Volume for 1ng-500ng DNA
DNA	Variable (maximum 20 µL)	Variable (maximum 40µL)
Bisulphite solution	85µL	85µL
DNA protect buffer	35µL	15µL

Table 3. Thermocycler conditions for bisulphite conversion of genomic DNA

Step	Time	Temperature
Denaturation	5 min	95°C
Incubation	20 min	60°C
Denaturation	5 min	95°C
Incubation	20 min	60°C
Hold	Indefinite	20°C

2.2.6 Polymerase Chain Reaction (PCR)

It is a technique used to amplify DNA fragments at several orders of magnitude generating millions of copies of the desired amplicon. It was first developed by Kary Mullis in 1983. It is an enzymatic reaction which involves Taq DNA polymerase along with four deoxyribonucleotide triphosphates (dNTPs), PCR buffer, and $MgCl_2$. The template DNA along with primers should be incubated at specific conditions that allow amplification. In this work, Fast Start Taq DNA polymerase (Roche, Mannheim, Germany) enzyme was used. Both for methylation analysis and genotyping, bisulphite converted DNA served as a template. A gradient PCR was performed to

determine the specific PCR conditions for each analyzed gene. The different steps involved in PCR are listed below.

a) Denaturation: It takes place at 95ºC and allows double-stranded DNA to melt or denature by breaking the hydrogen bonds and thus yielding single-stranded DNA molecules.
b) Annealing: The temperature in this step largely depends on the melting temperature of the primers and can range between 40ºC to 65ºC. This step allows the binding of oligonucleotide primers to the targeted sequence.
c) Extension/Elongation: It takes place at 72ºC where the enzyme Taq polymerase catalyzes the primer extension by adding free dNTPs from the reaction mixture in 5' to 3' direction. All the three steps are repeated for several cycles in order to obtain a large amount of PCR product.

Table 4. Pipetting scheme for polymerase chain reaction for each sample

Reagent	**Volume for one reaction (μL)**
10xPCR buffer with MgCl2	2.5
deoxyribonucleotide triphosphates (dNTPs)	0.5
10 μM Forward Primer	1.25
10 μM Reverse Primer	1.25
Fast Start Taq DNA Polymerase	0.2
DNA template	1
Water	18.3
Total	25

Table 5. Thermocycler conditions for PCR to amplify the desired region of interest

Step	**Temperature**	**Time**	
Initial denaturation	95ºC	5 minutes	1 cycle
Denaturation	95ºC	30 seconds	35-40 cycles
Primer annealing	varies	30 seconds	
Elongation	72ºC	45 seconds	
Final elongation	72ºC	10 minutes	1 cycle
	4ºC	∞	

2.2.6.1 Gel visualization

Agarose gel electrophoresis was performed to check successful and specific amplification of the PCR. 4µL of PCR product was diluted with 6µL of 1x gel loading dye (Thermo Fisher Scientific, Waltham, USA) and the mixture was loaded on a 1.5% agarose gel. The whole set up was run in an electrophoresis chamber containing 1x TAE buffer (50x TAE was prepared by 242g Tris, 100mL 0.5 M EDTA (pH 8), 57.1mL acetic acid and make up the volume to 1000mL with dH_2O). The product length was measured by comparing with a 100 bp gene ruler DNA ladder mix (Thermo Fisher Scientific) and by visualizing the gel under ultraviolet light in gel imager (Intas Science Imager Instruments GmbH, Gottingen, Germany).

2.2.6.2 Primer Design

Forward and reverse primers for PCR, and sequencing primers for methylation analysis and genotyping were designed using the PyroMark Assay Design Software (Biotage, Uppsala, Sweden). Primers were ordered from Metabion (Martinsried, Germany) and they were delivered lyophilized. To get the primer concentration to 100pmol/µL, they were dissolved in a volume of desalted water indicated by the manufacturer. For all the performed experiments, primers were diluted to 10pmol/µL using desalted water. The primers (repetitive elements [human and bull] and imprinted genes) used in the present study for methylation analysis by pyrosequencing are listed in table 6, 7, and 8 respectively. The primers for genotyping and deep bisulphite sequencing are given in table 9, 10 and 11.

Table 6. Repetitive DNA primers used for methylation by pyrosequencing in humans

Assay	Primer	Sequence (5'-3')[a]	no. of CpGs	A.T.
rDNA Promoter	Forward	*AAAAAAACRTCCCCAACCTCC		60°C
	Reverse	GTTTTYGTTGTGAGTTAGGTAGAGTTT		
	Sequencing	GGTTTATGTGGGGGAGAGGTTGT	9	
Upstream	Forward	GTGTTTTTGGGTTGATTAGAGG		60°C

Core Element	Reverse	*CATCCCAAAACCCAACCTCTCC		
	Sequencing 1	GGTTGATTAGAGGGTT	5	
	Sequencing 2	TTTTGGGGATAGGTG	13	
	Sequencing 3	TTYGGGGGAGGTATATTTT	8	
18S rDNA	Forward	AGGTTTGTGATGTTTTTAGATGT		60ºC
	Reverse	*AAAACCTCACTAAACCATC		
	Sequencing 1	TTTGTGATGTTTTTAGATGTT	4	
	Sequencing 2	ATTAAGTTTTTGTTTTTTGTATATA	4	
28S rDNA	Forward	GGTTTTAAGTAGGAGGTGTTAGAAAAG		60ºC
	Reverse	*CAACCAAACACATACACCAAATATCT		
	Sequencing 1	GGATAATTGGTTTGTGG	7	
	Sequencing 2	GTTGGATTGTTTATTTATTAATAGG	3	
Alpha Satellite Repeat	Forward	*TGTAAGTGGATATTTGGATTATTGG		57ºC
	Reverse	AACAATTTCAAAACTACTCCATCAA		
	Sequencing	CTCAAAAATTTCTAAAAATACTTCTC	4	
LINE1	Forward	TTTTGAGTTAGGTGTGGGATA		60ºC
	Reverse	*CTCACTAAAAAATACCAAACAA		
	Sequencing	GTTAGGTGTGGGATATAGTT	4	
ALU	Forward	GGGACACCGCTGATCGTATATTTTTATT AAAAATATAAA		52ºC
	Reverse	CCAAACTAAAATACAATAA		
	Uni Biotin	*GGGACACCGCTGATCGTATA		
	Sequencing	AATAACTAAAATTACAAAC	3	

[a] Primers indicated by a star are biotinylated at the 5' end. A.T. stands for annealing temperature for each primer pair.

Table 7. Repetitive DNA primers used for methylation by pyrosequencing in bulls

Assay	Primer	Sequence (5'-3')[a]	no. of CpGs	A.T
18S rDNA	Forward	AGGTTTGTGATGTTTTTAGATGT		60ºC
	Reverse	*AAAACCTCACTAAACCATC		
	Sequencing 1	TTTGTGATGTTTTTAGATGTT	4	
	Sequencing 2	ATTAAGTTTTTGTTTTTTGTATATA	4	
28S rDNA	Forward	GGTTTTAAGTAGGAGGTGTTAGAAAAG		60ºC
	Reverse	*CAACCAAACACATACACCAAATATCT		
	Sequencing 1	GGATAATTGGTTTGTGG	7	
	Sequencing 2	GTTGGATTGTTTATTTATTAATAGG	3	
Bovine	Forward	TGTTTTGTTTGGAAGGGGTTT		58ºC
Alpha	Reverse	*AACTATATTTAAAACCAAAAATTTTTCC		
Satellite	Sequencing 1	GTTTGGAAGGGGTTTT	3	
	Sequencing 2	GTGGGTGGTTTTATATT	3	
	Sequencing 3	GGTTTTTTTTTGATAAGAATT	3	
	Sequencing 4	GAAGGGGTAGTTTTT	3	
Bovine	Forward	TTGGGTTTGGTGTATTGGAAGA		58ºC
Testis	Reverse	*ACTCCACCCCTATAAATACAAT		
Satellite	Sequencing 1	GTAGGGTATTTTTGATTTTAGA	2	
	Sequencing 2	GGGTTGAGGTATGGAA	5	
	Sequencing 3	GGAATTTGGGGTTTTTT	2	

[a] Primers indicated by a star are biotinylated at the 5' end. A.T. stands for annealing temperature for each primer pair.

Table 8. Primers (imprinted genes+*HIF3A*) for methylation by pyrosequencing in humans

Gene	Primer	Sequence (5'-3')[a]	no. of CpGs	A.T
MEST	Forward	GATTTAAAGGATAGGTTTTAGTAT		58ºC
	Reverse	*AACCAAAATAAACAATCCCTAC		
	Sequencing	GTATTTTTTAGATTTTAGTAATAAG	4	
PEG3	Forward	GGTGTAGAAGTTTGGGTAGTTG		57ºC
	Reverse	*CTCACCTCACCTCAATACTAC		
	Sequencing	TGTTTATTTTGGGTTGGT	8	
SNRPN	Forward	*AGGGAGTTGGGATTTTTGTATT		56ºC
	Reverse	CCCAAACTATCTCTTAAAAAAAAC		
	Sequencing 1	ACACAACTAACCTTACCC	3	
	Sequencing 2	CCAACCTACCTCTAC	3	
PEG5	Forward	GGATTTTGTTTTAAAATGGAGGGGTAT		58ºC
	Reverse	*CCATCCATCCCCAAAATAAACTT		
	Sequencing	TTTAAAATGGAGGGGTATT	2	
PEG10	Forward	*GTGTTAAGGAGTTGGGAGGA		57ºC
	Reverse	TCTACAACCCTATAACAACCAATCTCA		
	Sequencing	CCTAATATACCTTCTCTA	3	
H19	Forward	*TGGGTATTTTTGGAGGTTTTTTT		54ºC
	Reverse	ATAAATATCCTATTCCCAAATAA		
	Sequencing	ATATATAAACCTACACTACC	3	
GTL2	Forward	AGGGTTAGGAAGTTTAGTAGGTTA		60ºC
	Reverse	*ACTACTCCTTAAACAAAAAAACACATAAT		
	Sequencing	GTAGTAAATTAAAGTGTATTAGAGA	5	

IGF2	Forward	GGAGGGGGTTTTAGTAAAAGTTAT		55°C
	Reverse	*TCCCAACCTCCCTAACACAAA		
	Sequencing	AGTAAAAGTTATTGGATATATAGT	3	
HIF3A	Forward	TGGTTGAAGGGTTATTTAGGG		60°C
	Reverse	*ACTCTATCCCACCCCTTTT		
	Sequencing 1	TTTAGGGGGTGTAGG	7	
	Sequencing 2	GGTGAGATGATTTTATAGGAA	1	
	Sequencing 3	GTTAAGAGGGGTTTTTATT	3	

[a] Primers indicated by a star are biotinylated at the 5' end. A.T. stands for annealing temperature for each primer pair.

2.2.7 Genotyping

In case of imprinted genes, to differentiate between the parental alleles and to analyze the transmission through maternal and paternal germlines separately, fetal cord blood samples that were heterozygous for a single nucleotide polymorphism (SNP) with highest minor allele frequency (MAF) were identified by pyrosequencing. Genotyping assays were developed on bisulphite converted DNA for eight imprinted genes. In case of PEGs, fully methylated allele (after deep bisulphite sequencing run) was assigned as maternal and un-methylated allele as paternal in each sample and the opposite was applicable for MEGs. The resulting heterozygous FCB samples were analyzed using deep bisulphite sequencing. All the primers required for genotyping and the SNPs for each amplicon are listed in table 9. Protocols described in section 2.2.6 and 2.2.8 were followed.

Table 9. Primers and analyzed SNPs for genotyping in humans on bisulphite converted DNA

Gene	Primer	Sequence (5'-3')[a]	SNP	A.T
MEST	Forward	GTTAGTAGGTAAAATGGGTTAAGTAAATAT		57°C
	Reverse	*AATTTCAAAACCCAAATATCAC		
	Sequencing	ATAATAAGTAAATTTTAGTAATAAA	rs3778859	

PEG3	Forward	*GGTGTAGAAGTTTGGGTAGTT	56ºC
	Reverse	CCACCAACCCAAAATAAACA	
	Sequencing	CCAACCCAAAATAAACATC	rs2302376
SNRPN	Forward	*ATTTATATTTATTTTTAGGTTTTTTTATGT	56ºC
	Reverse	AACTTTTCCAAACCAACTTT	
	Sequencing 1	CTTTTCCAAACCAACTTTTT	rs220030
PEG5	Forward	*AGGGAGTTGAGTTAGAGTATTG	56ºC
	Reverse	CCAAACATATTCCAACCACTTCT	
		ATCCTTCCTATATCCCT	rs6019062
PEG10	Forward	*ATTGTGGGTTTGTTGAGTATGG	56ºC
	Reverse	AAATCCCCCCTCCCAAAC	
	Sequencing	CCAAACCCCTCCCAC	rs3814105
H19	Forward	GGGAGGTAATTGTTAGTTTAGTAAAAG	56ºC
	Reverse	*ACATCTATCCCAAAACACAACT	
	Sequencing	TGGGGGTTGTTTTTAGA	rs11042167
GTL2	Forward	*GGGGTATTTGTGTGAAATAATGTTAA	58ºC
	Reverse	ACCAAAATAAAAATCTAAAACAATTCCTAC	
	Sequencing	ACTACAAAATTTCAACAAATACT	rs12437020
IGF2	Forward	*GGGGGTTTTAGTAAAAGTTATTGG	57ºC
	Reverse	ACCCTAAACCCAAACTCCTC	
	Sequencing	AATAATATCTATAAAAAAAAAATTC	rs3741210

[a] Primers indicated by a star are biotinylated at the 5' end. A.T. stands for annealing temperature for each primer pair.

2.2.8 Pyrosequencing (both for methylation analysis and genotyping)

2.2.8.1 Principle of Pyrosequencing

It is a fast and effective DNA sequencing technique based on the sequencing-by-synthesis principle. To quantify the DNA methylation and genotype status with this technique, one of the PCR primers should be biotinylated at the 5' end (Table 6, 7, 8 and 9). The principle of pyrosequencing is described in the below figure.

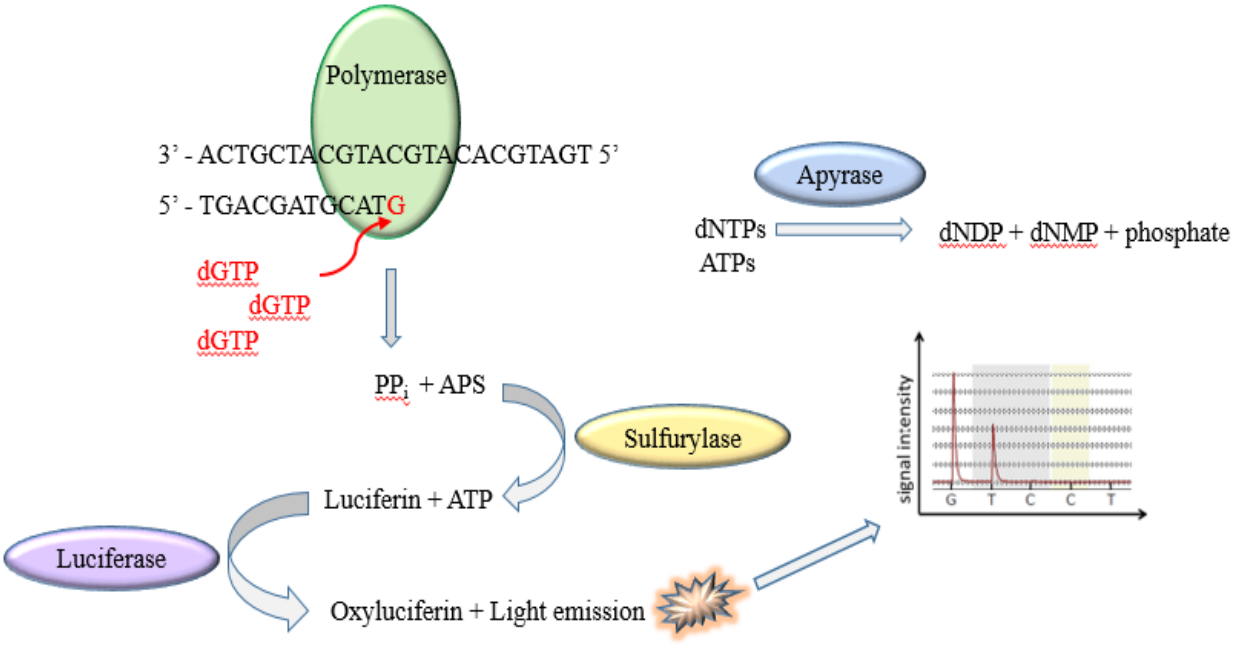


Figure 12. The principle of pyrosequencing

When a nucleotide that is complementary to the analyzed sequence is incorporated by a polymerase, a pyrophosphate (PP_i) molecule is released. Sulfurylase uses this molecule to convert adenosine 5' phosphosulfate (APS) to ATP that ultimately provides energy for luciferase to transform luciferin to oxyluciferin. Oxyluciferin emits a light signal which is detected by CCD camera and a pyrogram is generated. Unincorporated nucleotides and excess ATP are degraded by apyrase. The figure is taken from [127] and modified.

The PCR amplified biotinylated DNA product is immobilized by streptavidin sepharose beads (GE Healthcare, München, Germany), denaturated with NaOH (AppliChem GmbH, Darmstadt, Germany) to yield single strand and is washed using a vacuum prep workstation. The sequencing

primer is annealed to the captured biotinylated strand and is sequenced on Pyromark MD system (Qiagen) in a pyro-sequencing plate. During this step, the PCR amplicon is initially incubated with enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with substrates like adenosine-5-phosphosulfate, and luciferin. According to the dispensation order that depends on the analyzed sequence, one of the four nucleotides is sequentially added at each point of time. When the added dNTP is complementary to the template strand base, it is incorporated into the DNA strand. The incorporation of a nucleotide is catalyzed by the DNA polymerase and leads to the release of pyrophosphate (PPi) that is converted to ATP. This ATP provides energy for the conversion of luciferin to oxyluciferin via a luciferase-catalyzed reaction. Oxyluciferin generates light that is detected by a charge coupled device (CCD) camera and is displayed as peaks in a pyrogram.

Though the technique of pyrosequencing is fast and less expensive, some of the setbacks are as following:

a) It can only quantify the methylation levels of CpGs located within 30-50 base pairs 3' from the sequencing primer. In other words, smaller amplicon can only be sequenced.
b) It is not possible to measure methylation at single molecule (allele) level since this technique provides mean methylation of all the molecules in a sample at particular CpG site(s).
c) Epimutations (alleles showing >50% aberrant (de) methylated CpG sites) can't be determined with this technique.

2.2.8.2 Protocol of Pyrosequencing

a) For each sample, 11.5µL of annealing buffer and 0.5µL of pyrosequencing primer were pipetted into the pyrosequencing plate. The corresponding assay was loaded onto the sequencer for each well of the plate.
b) A mixer containing 2µL of streptavidin sepharose beads, 40µL of binding buffer, and 28µL of water was prepared for each reaction. To this, 10µL of PCR product was added. After sealing the 96 well plate containing many samples, the plate was vortexed for 5 minutes at 1400 rpm.

c) After vortexing, the microtiter plate was moved onto the vacuum prep tool wherein the head of the vacuum was lowered into the plate to capture the beads.
d) Then the head was placed in 70% ethanol, denaturation buffer (8g NaOH in 1-liter distilled water), and finally in wash buffer (1.21g Tris in 1-liter distilled water (pH 7.6)) for 10 seconds each.
e) To release the beads into the mixture containing sequencing primer, the head was placed over the pyrosequencing plate and vacuum was turned off. The plate was heated for 2 minutes at 80^{o}C and allowed to cool for 5 minutes at room temperature.
f) The enzyme and substrate mixture along with all four nucleotides were pipetted into the dispensing tips. A dispensation test was performed to check the proper dispensation of all the chemicals.
g) The plate was placed in PyroMark 96 MD sequencer and the run was started. PyroMark Q-CpG and PyroMark MD software (Qiagen) were used for data analysis.

2.2.8.3 Statistical analyses

Statistical analysis was performed with IBM SPSS version 24. Male/paternal factors such as age and body mass index were correlated with the methylation of corresponding gene/amplicon of interest at individual CpG level. Sperm samples which might probably be contaminated by somatic cells (based on the methylation levels of imprinted genes) were eliminated from the analysis for all the amplicons. Depending on the data distribution, either Pearson or Spearman correlations were applied. For comparison between the groups, either parametric T-tests or nonparametric Mann Whitney U tests were implemented. Possible confounding factors such as sperm concentration and male fertility status in case of sperm analysis were adjusted. The methylation levels of all imprinted genes in fetal cord blood (FCB) samples displayed approximately 50%. Maternal age and BMI were also considered as potential confounding factors in the case of FCB samples and so were adjusted. Linear regression models were developed for amplicons that showed a strong association between the sperm DNA methylation and the male factor. This enabled us to devise a prediction tool/formula so as to interpret the probable level/value of male factor based on the methylation of the considered region. A p-value <0.05 was considered as significant throughout the analyses.

2.2.9 Deep Bisulphite Sequencing (DBS)

DBS is a next generation sequencing (NGS) technique that enables one to examine DNA methylation level of the target amplicon at single allele (molecule) level. It also allows one to study epimutations (allele methylation errors) and to differentiate between paternal and maternal allele in fetal cord blood (FCB) samples through an informative heterozygous single nucleotide polymorphism (SNP). For unmethylated and methylated alleles, the molecules displaying >50% and <50% methylation values respectively were considered as epimutations. Epimutation rate (ER) was subsequently calculated by the formula:

(Number of epimutations/total number of reads for that sample) * 100

DBS on Illumina MiSeq Platform

This NGS method involves paired-end sequencing technology wherein both ends of DNA fragments are sequenced and forward and reverse reads are aligned as read pairs. A total number of 48 multiplex identifiers (MIDs) were used for barcoding the amplicon pools and NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) kit was used for barcoding each sample. BisPCR^2 protocol was adapted to sequence the amplicons on Illumina MiSeq system [128]. The template-specific primers for deep bisulphite sequencing are given in table 10 and 11.

Table 10. Template specific ribosomal DNA region primers for DBS on MiSeq platform

Region	Primer	Sequence (5'-3')	Varia-nt	no. of CpGs	A.T
rDNA Assay1	Forward	TATTYGGAGGTTTAATTTTTTTAG	A/G	25	56^{o}C
	Reverse	TATATCCTAAAATTAACCAAAAAACCCC			
rDNA Assay2	Forward	GGAGTTAGYGGGGTGGGGTTGT	A/G	38	56^{o}C
	Reverse	ACTAAAAAAATTAAACCTCC			

Table 11. Template specific imprinted gene primers for DBS on Illumina MiSeq platform

Gene	Primer	Sequence (5'-3')	no. of CpGs	A.T
MEST	Forward	TGGTTAAAATTTTAATTATTTGATGAGTTA	32	57ºC
	Reverse	ACTTAAAAAATAAACCCCAACA		
PEG3	Forward	TTTTGGGGGTTGTATTTTTATTATTTAATT	26	56ºC
	Reverse	ACTCACCTCACCTCAATACTAC		
SNRPN	Forward	TAGGTTGTTTTTTGAGAGAAGT	23	56ºC
	Reverse	ATTTAACTTTTCCAAACCAACT		
PEG5	Forward	TTGTTTTAAAATGGAGGGGTATT	19	56ºC
	Reverse	AATCCTTCCTATATCCCTCC		
PEG10	Forward	TGGAAATTGTGGGTTTGTTGA	15	57ºC
	Reverse	CAACTCACCTCAAAATAAAATAACCA		
H19	Forward	GGGAGGTAATTGTTAGTTTAGTAAAAG	25	56ºC
	Reverse	AACACATAAATATTTCTAAAAACTTCTCCT		
GTL2	Forward	AGATGTTAATTATTTTTTGGATAAGAGAGT	8	58ºC
	Reverse	AAATAAAAATCTAAAACAATTCCTACTACA		
IGF2	Forward	GGGAAGAAGTGGTGAGAAG	10	57ºC
	Reverse	TCCCAACCTCCCTAACAC		

A.T. stands for annealing temperature for each primer pair.

Protocol

2.2.9.1 Target or amplicon enrichment (PCR 1) and pooling

In the first PCR step, bisulphite converted DNA was amplified by template enrichment primers (Table 10 and 11) that were modified by adding the below partial adapter overhangs:

Forward Primer Overhang: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
Reverse Primer Overhang: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′

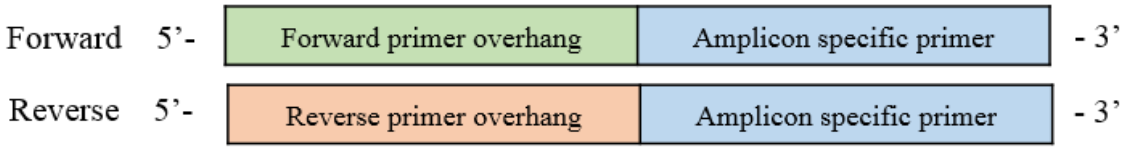


Figure 13. Schematic illustration of primers used for first round PCR in DBS run
Forward and reverse primers were tailed with overhangs at the 5' end of the sequence.

DNA standards of 0%, 50%, and 100% methylation were included. The annealing temperature for primers and the number of CpGs sequenced are given in Table 10 and 11. The pipetting scheme from Table 12 was employed for all the amplicons and the mixture was incubated in a thermocycler as described in Table 13.

Table 12. Pipetting scheme for PCR1 of DBS on bisulphite converted DNA of each sample

Reagent	**Volume for one reaction (µL)**
10xPCR buffer with MgCl2	5
deoxyribonucleotide triphosphates (dNTPs)	1
10 µM Forward Primer	2.5
10 µM Reverse Primer	2.5
Fast Start Taq DNA Polymerase	0.4
DNA template	2
Water	36.6
Total	50

Table 13. Thermocycler conditions for amplification of assays in PCR1 of DBS run

Step	Temperature	Time	
Initial denaturation	95°C	5 minutes	1 cycle
Denaturation	95°C	30 seconds	45 cycles
Primer annealing	Varies (Table 9)	30 seconds	
Elongation	72°C	60 seconds	
Final elongation	72°C	10 minutes	1 cycle
	4°C	∞	

4µL PCR product with 6µL gel loading dye was applied to a 1.5% agarose gel (17 volts for 15 minutes) and visualized. Once the samples showed the band at desired length without primer dimers, they were cleaned with Agencourt AMPure XP Beads (Beckmann Coulter GmbH, Krefeld, Germany) on JANUS automated workstation. In brief, the remaining 46µL PCR product was mixed with 27µL beads by snapping and incubated for 15 minutes at room temperature and then for 5 minutes on the magnet. The clear phase was rejected and beads were washed with 200µL of 75% ethanol. After washing the beads twice, they were dried till little cracks appeared. 30µL of elution buffer was given to the mixture, incubated for 2 minutes outside the magnet and then 5 minutes on the magnet. Finally, the clear phase was eluted. The below figure illustrates the position of PCR plate, elution plate etc. on the JANUS workstation.

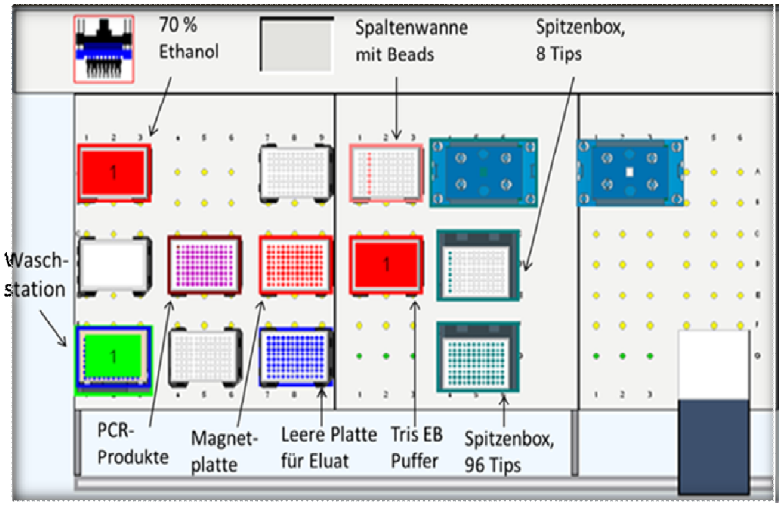


Figure 14. Illustration of set up on JANUS automated workstation

Position of ethanol tray, PCR plate, magnetic plate, and plate with elution buffer on JANUS automated workstation.

The cleaned PCR product was quantified using Qubit dsDNA BR Assay System kit (Thermo Fisher Scientific, Waltham, USA) on the Qubit Fluorometer (Invitrogen, Darmstadt, Germany) as per the manufacturer's instructions. All samples were diluted to 0.2ng/µL each with elution buffer by using the following formulae.

Pool final concentration (ng/µL) = number of assays * 0.2ng/µL

µL of dilution = measured concentration of samples / pool final concentration

µL of elution buffer for dilution = [(µL of dilution) – 1] * 3

Sample with 0.2ng/µL = (µL of elution buffer for dilution) + (3µL DNA)

3µL of each sample with 0.2ng/µL concentration from different amplicons were pooled together to 48 distinct pools that were later barcoded using 48 multiplex identifiers in PCR2. One pool contained various genes but with the same sample multiplex identifier (MID). For instance, 3µL of sample 1 of gene 1 was pooled with 3µL of sample 1 of gene 2 and so on.

2.2.9.2 Sample barcoding (PCR2) and pooling

Unique sequencing barcodes were joined into each sample of different amplicons by this step of PCR amplification. NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) and Illumina index primer rack were used. 1µL of forward primer (any of i501 – i508) and 1µL of reverse primer (any of i701 – i706) were used in different combinations for 48 pools. Touch down PCR thermocycler conditions were used to provide homogenous amplification of PCR templates of varying sizes. The program was in such a way that the temperature decreased by 1.2°C for every cycle. The primer sequences of i501 – i508 and i701 – i706, pipetting scheme for the PCR, thermocycler conditions were employed as per the tables 14, 15, and 16 respectively.

Table 14. Index primers that were included for barcoding each sample in PCR2 of DBS run

Forward Primers	Sequence	Reverse Primers	Sequence
NEBNext i501 Primer	TATAGCCT	NEBNext i701 Primer	ATTACTCG
NEBNext i502 Primer	ATAGAGGC	NEBNext i702 Primer	TCCGGAGA
NEBNext i503 Primer	CCTATCCT	NEBNext i703 Primer	CGCTCATT

NEBNext i504 Primer	GGCTCTGA	NEBNext i704 Primer	GAGATTCC
NEBNext i505 Primer	AGGCGAAG	NEBNext i705 Primer	ATTCAGAA
NEBNext i506 Primer	TAATCTTA	NEBNext i706 Primer	GAATTCGT
NEBNext i507 Primer	CAGGACGT		
NEBNext i508 Primer	GTACTGAC		

Table 15. Pipetting scheme for PCR2 for all 48 pools in DBS run

Reagent	**Volume for one pool (μL)**
10xPCR buffer with MgCl2	5
deoxyribonucleotide triphosphates (dNTPs)	1
Fast Start Taq DNA Polymerase	0.4
DNA template (each pool)	5
Water	36.6
Total	48

Table 16. Touch down PCR thermocycler conditions for amplification of 48 pools in DBS run

Step	**Temperature**	**Time**	
Initial denaturation	95°C	15 minutes	1 cycle
Denaturation	95°C	30 seconds	10 cycles
Primer annealing	68°C to 56°C	30 seconds	
Elongation	72°C	60 seconds	
Final elongation	72°C	10 minutes	1 cycle
	4°C	∞	

All the 48 PCR amplified pools were manually cleaned twice with Agencourt AMPure XP Beads. In short, 50μL PCR product was mixed with 45μL beads by snapping and incubated for 15 minutes at room temperature and then for 5 minutes on the magnet. The clear phase was rejected and beads were washed with 200μL of 75% ethanol. After washing the beads twice, they were dried till little cracks appeared. The mixture was given 40μL of elution buffer, incubated for 2 minutes outside the magnet and then 5 minutes on the magnet. Finally, the clear phase was eluted. The second purification step involved mixing 40μL of eluted DNA with 35.5μL of beads and exactly following the same procedure as described above. The cleaned pools were quantified

using High Sensitivity DNA reagents kit (Agilent, Böblingen, Germany) on 2100 Bioanalyser as per the manufacturer's instructions. Molarity of each pool was taken and pools were diluted to a final concentration of 4000pM using the formula given below.

$$\mu L \text{ of elution buffer} + 3\mu L \text{ pool} = [3 * (\text{molarity from bioanalyser} / 4000)] - 3$$

3µL of each pool with 4000pM concentration was pooled together to one final pool. This final pool was measured on Bioanalyser (Agilent) and the molarity was considered. It should approximately be equal to 4000pM. The pool was stored at -20°C until the day of sequencing.

2.2.9.3 Next Generation Sequencing

The final pool was used for performing deep bisulphite sequencing on the MiSeq platform (Illumina, California, USA) using the Reagent Kit v2 (500 cycles) (Illumina) cartridge following the manufacturer's instructions. The sequencing runs were carried out with 250 basepair paired-end sequencing. After the run, the sequencing reads were processed by Illumina Genome Analyzer and FASTQ files were interpreted using the Amplikyzer2 software [129]. Figure 15 explains a brief overview of the whole protocol.

2.2.9.4 Data processing and analyses

Initially, all analyzed amplicon sequences were aligned to the reference genomic sequence for each gene and allele splitting was done based on single nucleotide polymorphism or genetic variant within the region. Only reads with an overall bisulphite conversion rate of >95% were subsequently analyzed using IBM SPSS version 24. In the case of imprinted genes, the CpGs falling outside the imprinting regions and the CpG sites with SNP in a large number of samples were removed from analysis in all the amplicons. Paternal and maternal allele's methylation levels in fetal cord blood samples were correlated with paternal and maternal factors respectively at each CpG level. Based on the data distribution, either Pearson or Spearman correlations were performed. For comparison between the groups, either T-tests or Mann Whitney U tests were done. Potential confounding factors such as parental age, body mass index and SNP effects were

adjusted. Epimutation rates were determined and analyzed. A p-value <0.05 was considered as statistically significant.

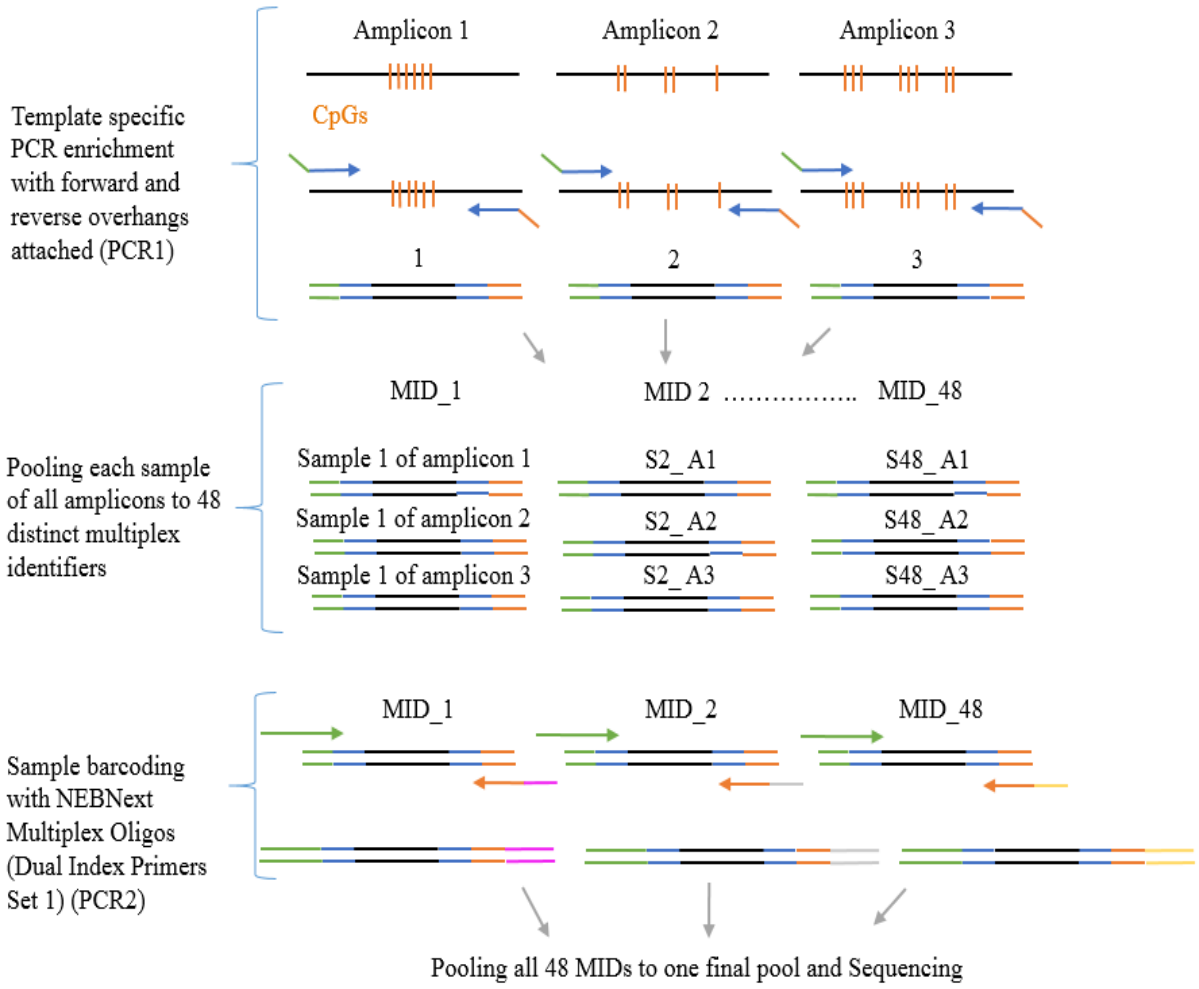


Figure 15. Stepwise representation of the BISPCR2 protocol

Forward and reverse overhangs in PCR1 are represented in green and orange color respectively. Different index primers are labeled with varying colors in PCR2. The figure is taken from [128] and modified.

2.2.10 Luciferase Assay

Dual luciferase reporter assay is used for functional analysis of promoter CpG methylation. The word 'dual' is used because activities of both Renilla and Firefly luciferases are measured simultaneously from a single sample. The DNA region of interest which is known to function as a promoter is cloned into the CpG free vector pCpGL. This vector (Firefly) is co-transfected with an internal control (Renilla) into the desired human cell line. Finally, the assay is carried out in order to determine the promoter methylation influence on the gene expression.

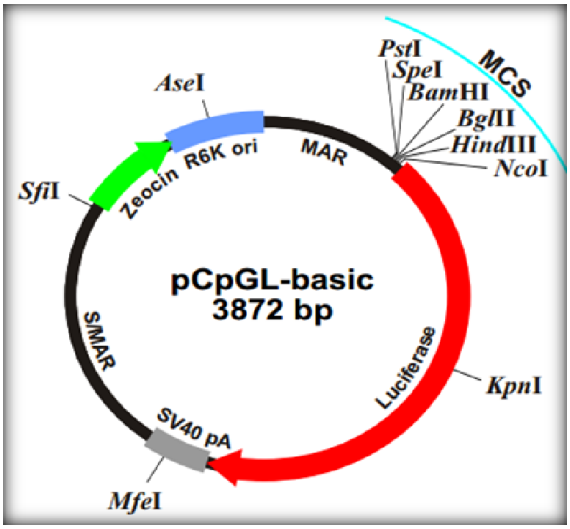


Figure 16. The map of CpG free empty reporter vector pCpGL

This vector is completely free of CpG dinucleotides. MCS denotes multiple cloning sites or poly-linker where the promoter region of interest is cloned. The vector contains an R6K origin of replication and is propagated in bacteria expressing the *PIR* gene. It contains a Zeocin-resistant gene. The figure is taken from [130] and modified.

Protocol

2.2.10.1 Double digestion of vector and the insert

The sequences of chosen restriction enzymes (generally *Bam*H I [5'-GGATCC-3'] and *Hin*d III [5'-AAGCTT-3'] plus 4 more random bases) were added at the end of forward and reverse primers of the desired DNA sequence that includes a transcriptional start site and acts as a

promoter. Polymerase chain reaction (primers are given in table 19) was carried out to amplify the region and subsequently, clean up was performed using QIA quick PCR purification kit (Qiagen, Hilden, Germany). Concentration was determined by NanoDrop 2000c and the DNA was stored at -20ºC until further use. Double digestion of vector and insert was performed as per the below table. Once double digested, vector DNA and insert DNA were cleaned using Nucleospin plasmid kit and Nucleospin PCR clean-up kit (Machery-Nagel, Düren, Germany) respectively.

Table 17. Protocol for double digestion of the vector and the insert

Double digestion of vector	Double digestion of insert
2-3µg of vector	2-3µg of the insert
5µL of 10X cut smart buffer	5µL of 10X cut smart buffer
1µL of restriction enzyme 1 (*Bam*H I)	1µL of restriction enzyme 1 (*Bam*H I)
1µL of restriction enzyme 2 (*Hin*d III)	1µL of restriction enzyme 2 (*Hin*d III)
Make up the volume to 50µL with H_2O	Make up the volume to 50µL with H_2O
Incubate at 37 ºC for one hour	Incubate at 37 ºC for one hour

2.2.10.2 Ligation of vector and the insert

Ligation was done in 1:3 and 1:6 ratios of vector and insert into the multiple cloning sites (MCS) that were upstream of the luciferase gene. The amount of insert to be added to the mix was calculated using ligation calculator (http://www.insilico.uni-duesseldorf.de/Lig_Input.html).

- 2µL of 10X T4 ligase buffer
- Xng (e.g.,50ng or 100ng) of vector DNA (generally 100ng)
- Yng of insert DNA (calculated using ligation calculator)
- 1µL of T4 ligase was finally added and the volume was made to 20µL with H_2O

The control reaction for ligation wherein insert DNA was not added to the mix was also kept simultaneously. All the reactions were incubated at 16 ºC overnight and later stored at -20 ºC till the step of transformation.

2.2.10.3 Preparation of LB medium, agar plates and, transformation procedure

Once prepared, both LB medium and LB agar were autoclaved and allowed to cool down. 75µL of 100mg/mL Zeocin was added in both. Before the LB agar got solidified, plates were loaded and stored at 4°C. PIR one shot competent cells were used for transformation because the vector contains the R6K origin of replication. pCpGL-CMV/EF1 and pCpGL-basic empty reporter (vector without insert) served as positive and negative controls respectively. Transformation protocol was performed as described in the below site following the manufacturer's instructions. (https://tools.thermofisher.com/content/sfs/manuals/oneshotpir_man.pdf).

Table 18. Protocol for the preparation of LB medium and LB agar for plates

Low salt liquid LB medium	**Low salt LB for Agar plates**
1.25g of yeast extract	1.25g of yeast extract
2.5g of tryptone	2.5g of tryptone
1.25g of NaCl	1.25g of NaCl
	3.75g of Agar
Make up the volume to 250mL with H_20	Make up the volume to 250mL with H_20
Adjust pH to 7.5 with 1M NaOH	Adjust pH to 7.5 with 1M NaOH

2.2.10.4 Colony PCR and Sanger Sequencing

These two steps were performed in order to check if the insert was in the right orientation. Two sets of colony PCR primers were designed; First set: Forward primer (FP) was in the vector and reverse primer (RP) was in an insert. Second set: FP in insert and RP in the vector. The positive colonies were detected by checking the appropriate band size in agarose gel, reinoculated in 3-5 mL of LB medium containing Zeocin antibiotic and incubated at 37°C overnight with shaking. Following day, Miniprep was performed using Nucleospin Plasmid kit (Machery-Nagel) as per the manufacturer's recommendations and concentration was measured.

Sanger sequencing was done on the miniprep DNA to check if the insert was complete and was in the right orientation. The sequence was aligned with the original sequence in Clustal W and was compared. If the insert was complete and in the right orientation, Midiprep/Maxiprep was

performed to obtain more amount of DNA. Midiprep/Maxiprep was done using Nucleobond Plasmid kit (Machery-Nagel) by following the manufacturer's instructions. PCR primers for the amplification of human *FOXK1* gene promoter region of interest and colony PCR primers are given in table 19. The procedure and the PCR condition for performing Sanger sequencing are listed in table 20.

Table 19. Primers (PCR and colony PCR) for Luciferase reporter assay for *FOXK1* promoter

Assay	Primer	Sequence (5'-3')	R.S	A.T
PCR	Forward	GGGGGATCCAGGACAGTGGGAGGGTGTT	*Bam*H I	56°C
	Reverse	CCAAGCTTCCAAAATGCACACCTTTCAGC	*Hin*d III	
Colony PCR_1	Forward	AAACCACTGATTTTTGTTTATGTGA		56°C
	Reverse	AGAAAGTGGCTCCAGAGGAA		
Colony PCR_2	Forward	ACCTCAAGGTCTGTTGATCAG		56°C
	Reverse	GACCAGGGCATACCTCTTCA		

A.T stands for annealing temperature for each primer pair. R.S stands for restriction site.

Table 20. Procedure and the PCR conditions for Sanger sequencing

Procedure	PCR conditions	
5µL H_20	96°C for 2 minutes	– 1 cycle
2µL of 5X Sequencing buffer	96°C for 20 seconds	
1µL of miniprep DNA (250ng – 500ng)	56°C for 20 seconds	25 cycles
1µL primer (FP of first set colony PCR)	60°C for 3 minutes	
1µL Big Dye	60°C for 3 minutes	– 1 cycle
	10°C forever	

2.2.10.5 In vitro methylation

pCpGL vector with the desired insert was *in vitro* methylated using *Hha* I, *Hpa* II and *Sss* I methylases. The unmethylated plasmid which served as a control was strictly treated without

adding S adenosyl methionine (SAM) or any other methylases. The final volume was made up to 50µL with H_20. The reactions were incubated at 37°C for 4 hours. 5µL (1600µM) SAM was added only in the methylated plasmid reaction tube after 2 hours. Clean up was done using Nucleospin plasmid kit and concentration of plasmids was measured. The procedure for *in vitro* methylation is given in table 21.

Table 21. Procedure for *in vitro* methylation of plasmid

Methylated Plasmid	Unmethylated plasmid (served as a control)
5µL of 10X NEB buffer 2	5µL of 10X NEB buffer 2
5µg Midi/ Maxiprep DNA	5µg Midi/Maxiprep DNA
3.1µL of *Sss* I (4 U/µL)	
3.1µL of *Hpa* II (4 U/µL)	
1µL of *Hha* I (25 U/µL)	
5µL of SAM (1600µM)	

2.2.10.6 Control digestion

In order to check the efficiency of the *in vitro* methylation step, control digestion was performed with methylation-insensitive *Msp* I and methylation-sensitive *Hpa* II enzymes. The final volume was made up to 25µL with water and the reactions were incubated at 37 °C for an hour. 5µL of 6X gel loading purple dye was added. Agarose gel electrophoresis was performed and visualized. Since *Hpa* II is blocked by methylation, *Hpa* II digested methylated plasmid showed a band with the same size as a completely undigested methylated plasmid.

Table 22. Procedure for control digestion of methylated and unmethylated plasmid

Methylated Plasmid		Unmethylated plasmid	
***Msp* I**	***Hpa* II**	***Msp* I**	***Hpa* II**
500ng DNA	500ng DNA	500ng DNA	500ng DNA
2.5µL 10X buffer	2.5µL 10X buffer	2.5µL 10X buffer	2.5µL 10X buffer
0.25µL (20U/µL) enzyme	0.5µL (10U/µL) enzyme	0.25µL (20U/µL) enzyme	0.5µL (10U/µL) enzyme

2.2.10.7 Cell seeding, Co-transfection and, Dual Luciferase Assay

The day before transfection, ~ 0.5×10^5 cells (e.g. U2OS, HELA etc.) per well were seeded in a 12 well plate. Hemocytometer was used to count the cells. If there were "X" number of cells in 16 squared area of hemocytometer, then the total cells per mL (Y) was calculated using the formula:

$$Y = (X * 1000) / 3.2$$

So, 1000µL (1mL) contained Y number of cells, the required ~ 0.5×10^5 number of cells per well in the plate could be calculated accordingly.

Co-transfection was done using luciferase vector (pCpGL – correlates with the effect of specific experimental condition – firefly) and renilla vector (pRL-TK – provides an internal control system). The final ratio of luciferase and renilla was taken so that it minimizes experimental variability caused by differences in cell viability and transfection efficiency. A ratio greater than 50:1 of luciferase: renilla was considered since it aids in suppressing the occurrence of trans effects between promoter elements which might potentially affect reporter gene expression. Lipofectamine 3000 (Invitrogen) was used as a transfecting reagent. Plasmid DNA transfection was carried out as per the protocol following the manufacturer's recommendations. (http://tools.thermofisher.com/content/sfs/manuals/lipofectamine3000_protocol.pdf).

1000ng DNA per well was used for 12 well plate. pCpGL-CMV served as a positive control whereas empty pCpGL and non-transfected cells as negative controls. All transfections were done in duplicates. To check the efficiency of transfection, some wells were only transfected with green fluorescent protein (GFP) and visualized under fluorescent microscopy. The influence of methylated plasmid, when compared to the unmethylated plasmid was assayed after performing dual luciferase reporter assay (Promega, Wisconsin, USA) following the protocol as per manufacturer's instructions. The final readings were taken from Luminometer (Berthold Technologies, Bad Wildbad, Germany) and the average values of duplicates were considered for analysis.
(https://www.promega.de/~/media/files/resources/protocols/technical%20manuals/0/dual%20luciferase%20reporter%20assay%20system%20protocol.pdf).

3 Results

The effect of male aging on sperm DNA methylation and consequences in the next generation

3.1 Human male aging effects on sperm DNA methylation

3.1.1 Characteristics of study samples: Out of the 294 IVF/ICSI sperm obtained from Fertility Center, Wiesbaden, Germany, 185 samples whose corresponding fetal cord blood were not retrieved were considered as sperm cohort 1, and the remaining 109 samples (respective FCB available) were treated as sperm cohort 2. Additionally, the naturally conceived fertile male sperm samples from Muenster, Germany were also analyzed to examine the effect of male aging on fertile sperm cohort. The range, mean, and median of donor's age (for all the cohorts) are provided in table 23.

Table 23. Range, mean, and median values of the age of the male donor across all three cohorts

Parameter	Sample size	Range	Mean ± SE	Median
Cohort 1_ Age (years)	185	25.71 – 65.82	39.05 ± 0.46	38.75
Cohort 2_ Age (years)	109	28.08 – 52.85	38.85 ± 0.49	38.60
Fertile cohort_Muenster	20	27.00 – 54.00	42.45 ± 1.19	42.00

3.1.2 Methylation of repetitive elements: As repetitive DNA represents a surrogate marker for measuring (epi) genome-wide DNA methylation, they were extensively studied in this thesis to examine the paternal age effect on sperm DNA methylation. Methylation analysis was performed on these regions using bisulphite pyrosequencing. The assays analyzed in humans, PCR and sequencing primers, and the number of CpGs per assay are given in table 6.

3.1.2.1 Ribosomal DNA

The ribosomal DNA methylation in human sperm samples for all the four assays analyzed (promoter, upstream core element [UCE], 18S rDNA, and 28S rDNA) in both cohort_1 and cohort_2 was significantly correlating with the donor's age in a positive direction across all the

individual CpGs. In other words, with an increase in age of the man, the DNA methylation in sperm was significantly increasing in ribosomal DNA regions.

In order to examine if the positive correlation was also observed in naturally conceived males, we considered 20 fertile sperm samples obtained from Muenster. In this cohort as well, the mean and the individual CpG methylation of all the four assays mentioned above was positively correlating with the age of the donor. The promoter and upstream core element assays showed a significant correlation but 18S and 28S rDNA assays did not reach significance probably owing to small sample size (~20). This indicated that even in naturally conceived males, there was a hypermethylation of ribosomal DNA regions in sperm with an increase in age of the donor. The Spearman's rho (correlation coefficient) between the mean methylation of assays and the male age is represented in table 24 and the scatter plots showing the positive correlation for all the assays in the three cohorts are shown in figures 17, 18, and 19.

Table 24. Spearman correlation between donor's age and mean methylation of rDNA assays

Cohort	Assay	No of CpGs	Spearman's rho	p-value
Cohort_1	rDNA promoter	9	+ 0.62	**< 0.0001**
n ~ 185	rDNA UCE	26	+ 0.54	**< 0.0001**
	18S rDNA	8	+ 0.35	**< 0.0001**
	28S rDNA	10	+ 0.44	**< 0.0001**
Cohort_2	rDNA promoter	9	+ 0.57	**< 0.0001**
n ~ 105	rDNA UCE	26	+ 0.58	**< 0.0001**
	18S rDNA	8	+ 0.42	**< 0.0001**
	28S rDNA	10	+ 0.46	**< 0.0001**
Fertile	rDNA promoter	9	+ 0.47	**0.04**
Sperm	rDNA UCE	26	+ 0.45	**0.05**
n ~ 20	18S rDNA	8	+ 0.31	0.22
	28S rDNA	10	+ 0.34	0.16

The technique used was pyrosequencing and the samples analyzed were human sperm. UCE: Upstream core element.

Cohort_1
Human sperm

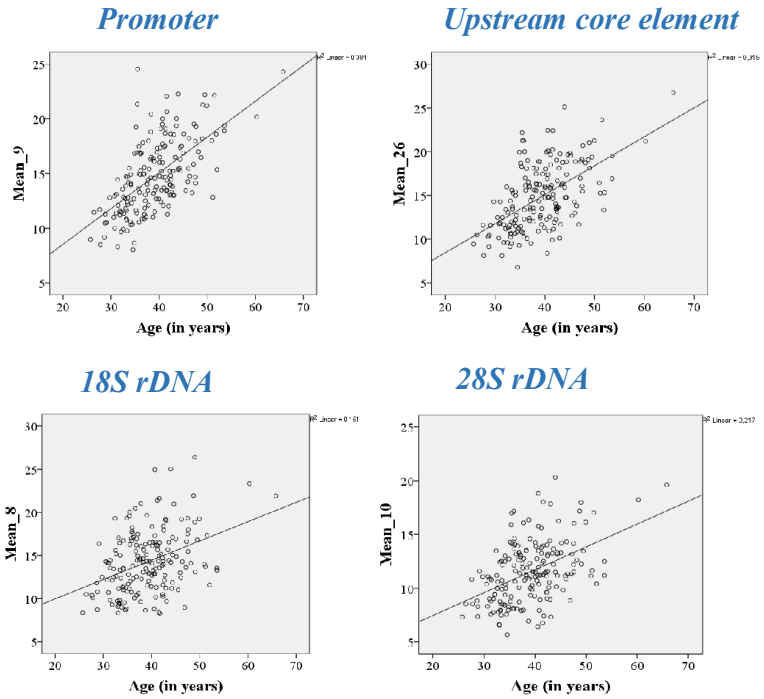


Figure 17. Scatter plots showing a significant positive correlation between donor's age (x-axis) and mean methylation for each rDNA assay (y-axis) in cohort_1 of human sperm samples

Top left plot: promoter, top right: UCE, bottom left: 18S rDNA, and bottom right: 28S rDNA assay.

Cohort_2
Human sperm

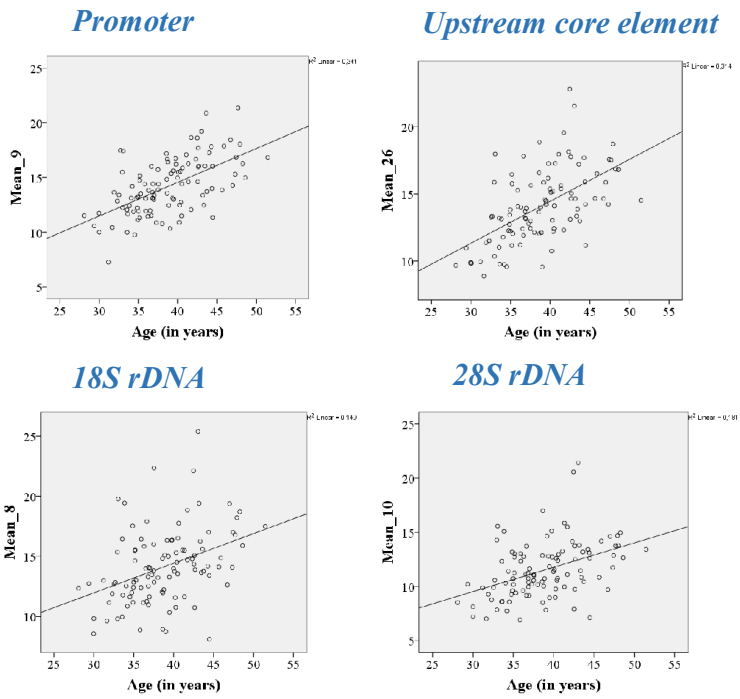


Figure 18. Scatter plots showing a significant positive correlation between donor's age (x-axis) and mean methylation for each rDNA assay (y-axis) in cohort_2 of human sperm samples

Top left plot: promoter, top right: UCE, bottom left: 18S rDNA, and bottom right: 28S rDNA assay.

Fertile human sperm cohort

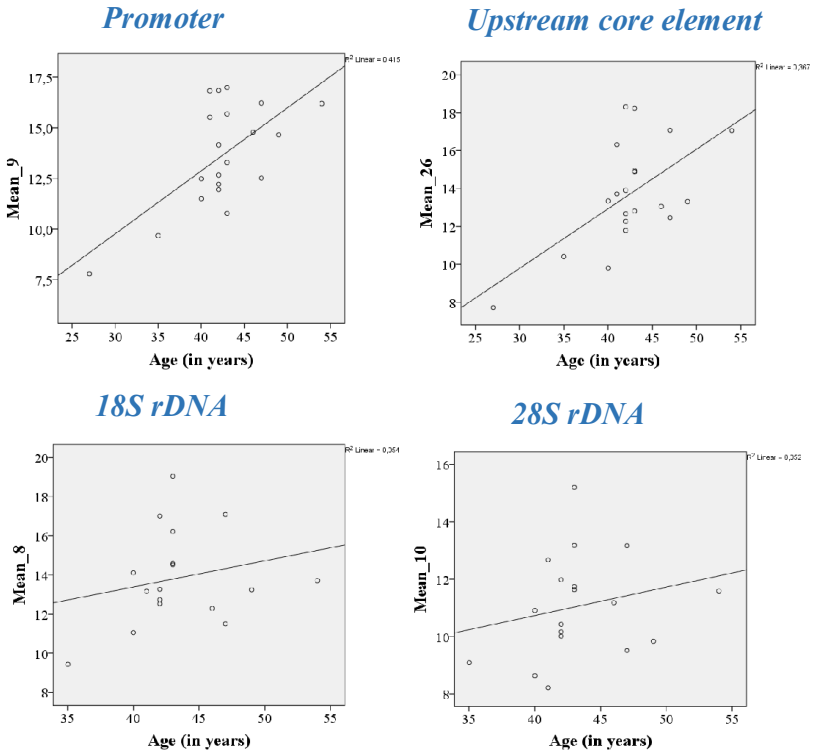


Figure 19. Scatter plots showing a positive correlation between donor's age (x-axis) and mean methylation for each rDNA assay (y-axis) in Muenster cohort of human sperm

Top left plot: promoter, top right: UCE, bottom left: 18S rDNA, and the bottom right: 28S rDNA assay. Promoter and upstream core element assays reached significance but not 18S and 28S rDNA assays.

3.1.2.2 Alpha Satellite, LINE1 and ALU repeats

In addition to rDNA assays, the DNA methylation of repetitive elements Alpha-Satellite, LINE1, and ALU was also analyzed using pyrosequencing technique in human sperm samples.

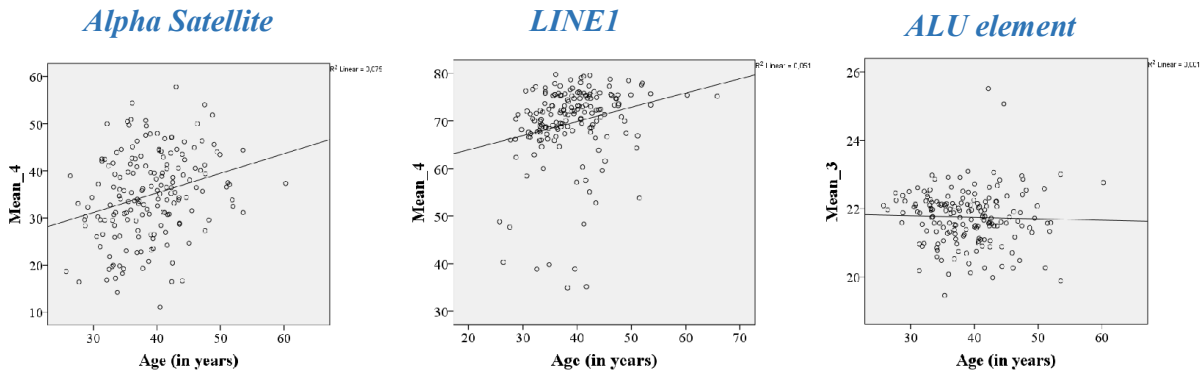
1) The mean DNA methylation levels of Alpha Satellite and LINE1 repeat assays in human sperm were significantly correlating with the age of the donor in cohort_1 and cohort_2. In other words, as the age increased in males, alpha satellite and LINE1 repeats were hyper methylated in human sperm samples.
2) However, ALU elements' methylation in the human sperm did not reveal any correlation with donor's age in any analyzed cohort. The Spearman's rho (correlation coefficient) between the mean methylation for three repeats and the male age is represented in table 25 and the scatter plots showing the positive correlation are shown in figure 20.

Table 25. Spearman correlation between donor's age and mean methylation of assays in alpha satellite, LINE1, and ALU repeats

Cohort	Assay	No of CpGs	Spearman's rho	p-value
Cohort_1	Alpha Satellite	4	+ 0.29	**< 0.0001**
n ~ 185	LINE1	4	+ 0.28	**< 0.0001**
	ALU element	3	- 0.09	0.22
Cohort_2	Alpha Satellite	4	+ 0.21	**0.04**
n ~ 105	LINE1	4	+ 0.24	**0.03**
	ALU element	3	- 0.06	0.59

The technique used was pyrosequencing and the samples analyzed were human sperm.

Cohort_1 Human sperm



Cohort_2 Human sperm

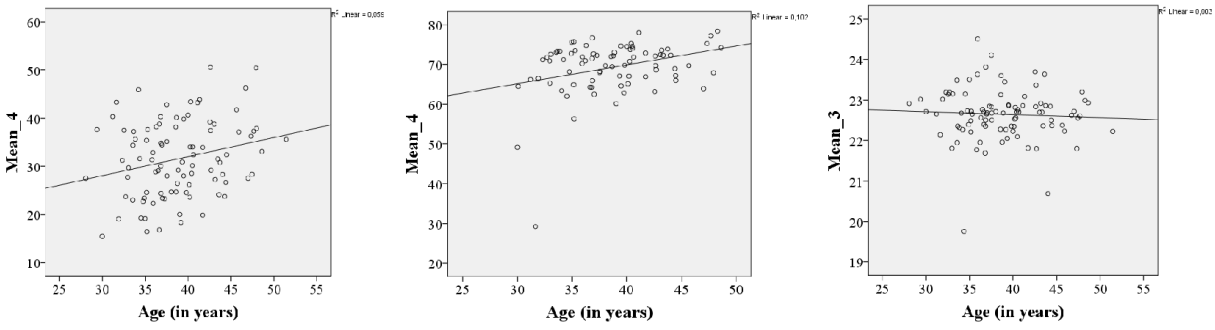


Figure 20. Scatter plots showing a positive correlation between donor's age and mean methylation of α satellite, LINE1, and ALU assays in cohort_1, and cohort_2 of human sperm samples

Left plots: alpha satellite, middle ones: LINE1, and right plots: ALU element. Methylation of alpha satellite and LINE1 showed significant correlation with age in both the cohorts but ALU repeats did not display correlation.

3.1.3 Male age calculator based on sperm rDNA methylation: A prediction tool (or formula) for determining man's age by analyzing his sperm ribosomal DNA regions' methylation was developed using regression analysis. In the statistical models that were implemented, the independent/explanatory variable or predictor (denoted by X) was rDNA methylation in human sperm, and the dependent or outcome variable (denoted by Y) was the age of the donor. Since both rDNA methylation and donor's age were scalar variables, and as there was only one predictor, a simple linear regression model was employed. Four independent regression models were applied for rDNA promoter and UCE regions in both cohort_1 and cohort_2 of human sperm samples. The donor's age prediction was determined when his rDNA promoter/UCE methylation could be 10% in sperm (table 26).

The predicted age of a man with X% of rDNA methylation in sperm can be given by the formula:

$$\mathbf{Y = A + B\,(X)}$$

Y is the outcome/dependent variable (human donor's age),
A is the Y-intercept (the point where the regression line hits the y-axis),
B is the slope of the line (also known as the regression beta coefficient),
X is the explanatory variable or predictor (rDNA methylation in sperm).

Table 26. Regression models showing Y-intercept (A) and beta coefficient (B) when applied on promoter and UCE regions' methylation in cohort_1 and cohort_2 of human sperm samples

Cohort	Assay	Model	Y-intercept (A)	ß coefficient (B)	Y value when X= 10%
Cohort_1	rDNA promoter	Model_1	21.79	1.17	**33.5 years**
n ~ 185	rDNA UCE	Model_3	25.12	0.95	**34.6 years**
Cohort_2	rDNA promoter	Model_2	23.10	1.11	**34.2 years**
n ~ 105	rDNA UCE	Model_4	24.65	1.00	**34.7 years**

Statistical modelling was performed using pyrosequencing data. Promoter and UCE regions harbor 9 and 26 CpGs respectively, and the mean of these CpGs was considered for analysis. rDNA: ribosomal DNA, and UCE: upstream core element.

Inference:

All the four applied models independently predicted that the age of the donor could be 33-34 years when his sperm rDNA methylation in promoter or upstream core element was 10%.

Validation:

The above designed regression models' formulae were applied on naturally conceived fertile men's cohort (from Muenster, n= 20) for the purpose of validation. The age prediction formulae from model_1 and model_2 (X values being the promoter region's methylation levels in Muenster cohort), and from model_3 and model_4 (X values being the UCE region's methylation levels in Muenster cohort) were used to detect the age of the donor in Muenster cohort and was compared with the original age. 75% of the samples (15 out of 20) showed that the predicted age was ≤ ± 6 years when compared to the original age. 10% (2 out of 20) showed that the predicted age was between > + 6 to < +8 years when compared to the original age, and the remaining 15% showed that there was a difference of +13, +11 and +10 years for samples 9, 11, and 17 respectively between the predicted age and the original age of the donor.

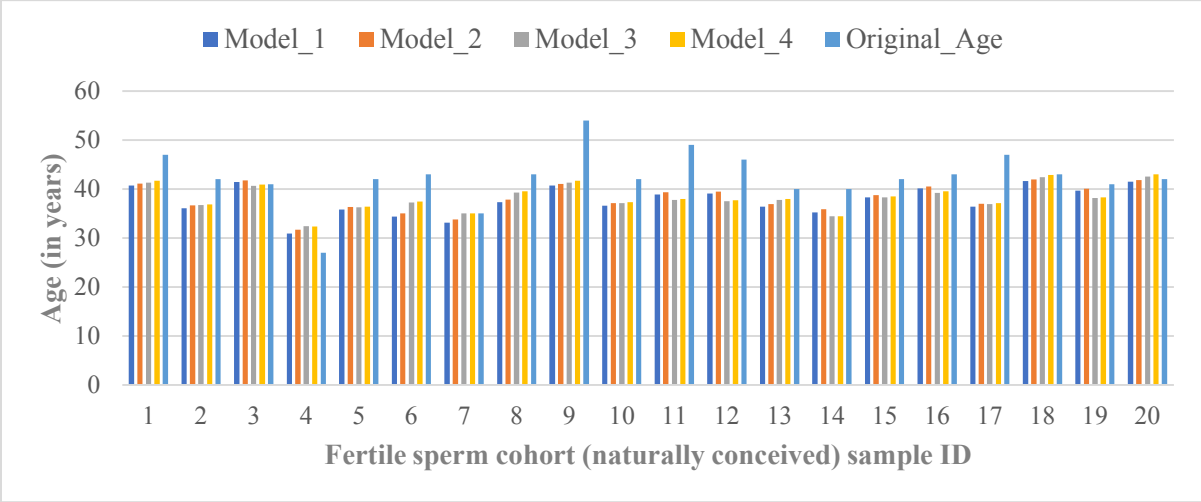


Figure 21. Predicted and original age of donor in Muenster cohort of human sperm samples

Formulae from model_1 and model_2 were used with X values being methylation of the promoter region in Muenster cohort. Formulae from model_3 and model_4 were used with X values being the methylation of UCE region in Muenster cohort. All the four predicted ages were compared with the original age of the donor.

3.1.4 Genetic variant based sperm ribosomal DNA methylation correlation with donor's age: In order to examine if the methylation level of DNA sequence carrying one genetic variant correlated by a larger extent with donor's age than that of the other one, deep bisulphite sequencing on Illumina MiSeq platform was performed. The detailed protocol and the primers used for ribosomal DNA assays are provided under section 2.2.9 and in table 10 respectively. Both assay1 and 2 each contained either genetic variant A (Adenine) or G (Guanine) in their DNA sequences with an allele frequency of ~ 0.34. Human sperm samples were divided into two groups based on the donor's age: young aged (25-35 years) and old aged (>42 years) men each group containing 23 samples.

For both the assays, the age of the sperm donor (in years) significantly correlated with the methylation percentage of ribosomal DNA sequence bearing either A or G underlying genetic variant. However, the amplicon carrying G variant displayed slightly stronger correlation in both the assays than the one with A variant. Table 27 provides the Spearman correlation coefficient between the mean methylation of the two assays across both the variants and the age of the donor, and the scatter plots showing positive correlation are given in figure 22. The difference between the young and old aged group in their sperm ribosomal DNA methylation was distinctly noticeable at individual CpG level as well (figure 23). Heatmaps showing the variation in methylation levels across each CpG between young and old aged sperm donor groups, in both A and G variants of assay 1 and 2 (25 and 38 CpGs respectively) are represented in figure 23. The first 23 rows are the values from young aged males (25-35 years) and the next 23 rows are from old ages males (>42 years) (figure 23).

Figure 24 elaborately depicts the mean methylation values of ribosomal DNA sequences carrying the genetic variants across all the analyzed samples. Interestingly, sequence-dependent methylation was observed in assay 1. Here, the ribosomal DNA sequence bearing A (Adenine) genetic variant displayed more methylation percentage than the one bearing G (Guanine) genetic variant. This was consistently (only with few exceptions) found across all the analyzed 46 samples. The first 23 samples represent sperm from young aged donors (25-35 years) and the samples from 24 to 46 are the sperm from old aged donors (>42 years). Assay 1 and 2 harbors 25 and 38 CpGs respectively.

Table 27. Spearman correlation between donor's age and mean methylation of rDNA assays across both the genetic variants

Assay	Allele	No of CpGs	Spearman's rho	p-value
1	A Allele	25	+ 0.74	**< 0.0001**
	G Allele	25	+ 0.77	**< 0.0001**
2	A Allele	38	+ 0.76	**< 0.0001**
	G Allele	38	+ 0.82	**< 0.0001**

The technique used was deep bisulphite sequencing on MiSeq platform and samples analyzed were human sperm.

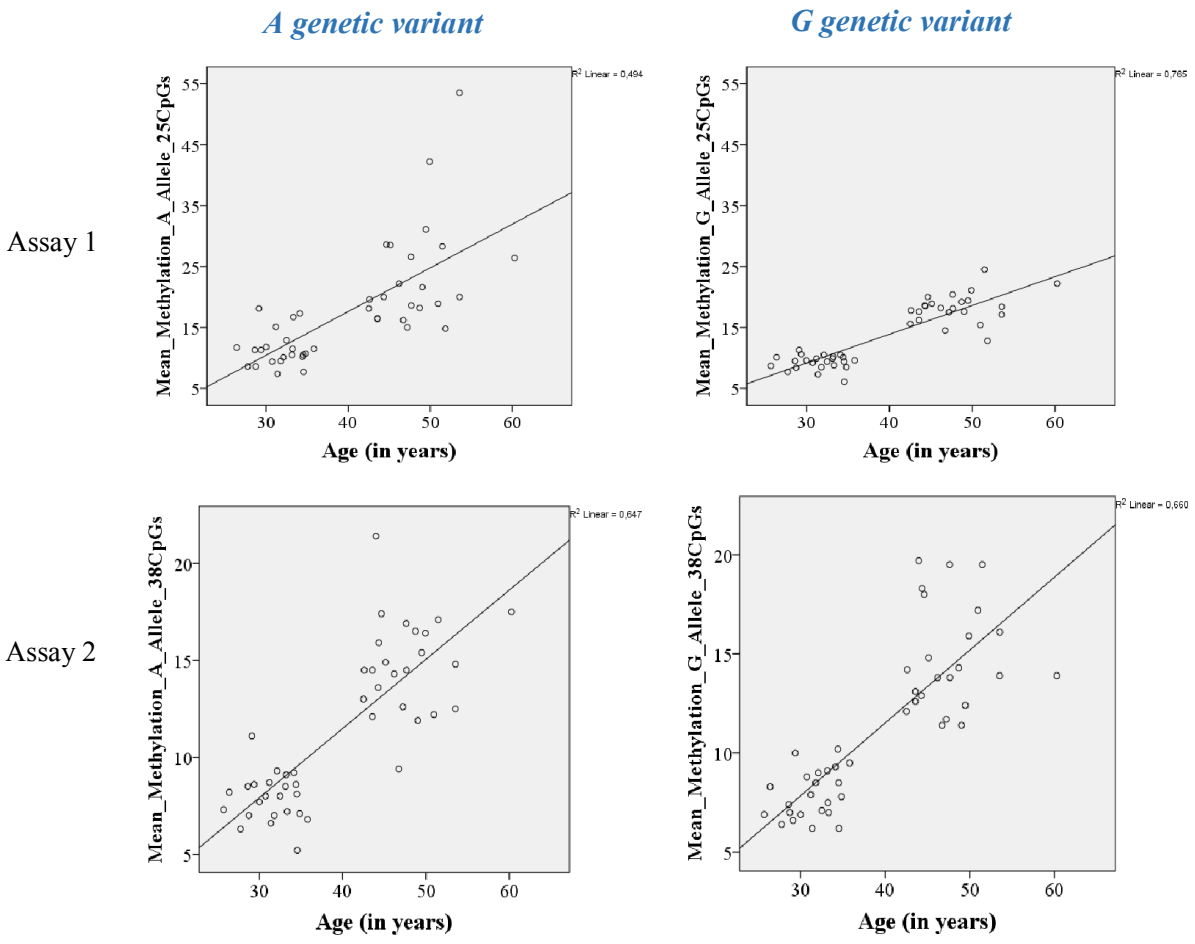


Figure 22. Scatter plots showing a significant positive correlation between donor's age (x-axis) and mean methylation of rDNA assays (y-axis) in variant A, and variant G

The lower and upper 23 dots in each group belong to the sperm sample from young aged males (25-35 years) and old aged males (>42 years) respectively. Assay 1 harbors 25 CpGs, and assay 2 carries 38 CpGs.

Methylation values of rDNA sequence bearing A variant – Assay 1

CpG1	CpG2	CpG3	CpG4	CpG5	CpG6	CpG7	CpG8	CpG9	CpG10	CpG11	CpG12	CpG13	CpG14	CpG15	CpG16	CpG17	CpG18	CpG19	CpG20	CpG21	CpG22	CpG23	CpG24	CpG25
17	5	11	9	4	5	14	7	14	6	10	12	16	13	17	24	31	11	14	14	3	11	7	3	11
16	7	14	16	11	16	28	9	37	17	13	16	22	18	26	35	40	11	13	14	3	26	10	5	28
16	5	8	4	2	3	7	4	10	5	6	8	14	9	12	15	26	8	12	14	3	8	6	3	9
13	5	11	7	6	7	15	4	21	5	8	11	15	14	9	25	28	8	10	11	2	15	6	4	23
11	4	9	6	3	6	14	5	20	10	6	9	15	12	14	23	27	7	11	10	2	13	7	5	18
14	5	8	6	4	4	15	6	17	6	6	9	12	10	12	20	27	10	12	12	2	12	6	5	15
16	6	9	6	4	6	12	5	15	8	8	10	16	14	15	21	29	11	13	14	3	13	7	5	15
13	5	12	11	7	11	27	6	30	14	13	11	12	14	18	25	39	12	18	15	3	16	10	6	30
12	4	9	5	2	3	7	3	9	6	6	7	11	9	10	12	20	5	8	10	3	8	3	2	11
19	8	12	9	6	8	14	7	19	11	11	12	17	13	17	26	31	10	11	15	4	14	7	4	16
19	6	14	14	11	14	27	9	27	16	12	17	23	21	24	31	40	14	16	17	3	21	6	5	25
19	7	12	7	4	7	13	6	19	8	6	10	15	8	15	20	31	10	13	14	3	14	10	6	16
15	5	9	5	3	5	9	5	12	7	8	10	16	15	14	21	28	10	12	14	3	11	6	4	14
14	4	8	6	3	4	9	4	12	9	9	9	11	11	11	13	20	8	9	13	2	7	5	3	8
14	5	9	6	3	5	11	5	15	7	8	10	15	12	15	21	29	10	12	14	3	11	6	4	13
19	4	10	9	8	13	29	9	36	14	12	17	24	20	26	31	38	13	17	12	2	21	7	4	25
17	7	11	8	3	7	12	5	17	9	8	12	16	12	16	21	27	9	13	15	3	13	8	4	15
13	5	8	5	2	5	9	4	13	7	6	9	15	11	13	18	26	8	11	12	2	9	6	3	13
13	4	8	6	4	6	11	5	15	7	7	7	16	13	14	19	28	8	12	13	3	13	7	4	15
13	5	9	8	5	8	10	4	16	8	7	10	12	11	12	17	21	8	10	10	2	11	5	4	12
8	4	7	6	4	5	13	3	12	5	3	3	8	6	10	17	19	5	5	7	1	15	6	3	19
15	6	9	9	7	8	15	5	18	10	10	13	16	14	17	22	30	8	10	12	2	13	7	5	15
23	8	12	13	5	9	23	11	28	17	16	15	27	20	26	35	39	18	20	22	5	21	13	8	20
23	10	15	10	9	15	15	7	21	6	10	14	19	12	25	27	32	12	20	21	5	17	9	7	16
36	13	18	8	5	8	15	10	16	10	15	17	31	18	26	29	52	27	29	31	6	16	14	8	16
31	12	18	13	7	14	20	13	24	18	16	22	33	20	35	36	49	24	27	31	8	23	16	9	23
37	15	22	24	21	22	26	19	35	26	25	32	34	33	41	51	55	34	32	28	11	29	20	11	29
50	25	40	41	34	36	51	28	54	34	36	38	48	35	56	61	73	50	52	47	13	51	33	17	51
25	10	14	15	9	14	25	15	35	15	11	15	27	24	25	35	45	17	24	23	5	24	15	8	30
35	13	18	8	6	8	16	10	19	11	16	20	35	21	29	31	52	27	30	32	6	17	13	8	18
24	10	12	8	6	7	16	10	18	10	10	13	22	14	24	24	37	15	20	19	3	12	13	7	18
28	10	13	9	3	7	10	11	21	14	13	18	26	17	23	26	40	18	22	20	4	15	12	8	20
26	11	15	11	5	11	20	10	25	13	14	19	29	23	27	36	49	22	22	23	5	20	14	9	29
30	13	21	15	11	15	28	11	26	14	15	18	28	24	30	41	54	22	25	25	5	27	16	10	30
29	11	16	7	4	7	12	8	18	12	14	17	28	16	24	27	43	19	24	27	5	13	12	6	14
39	22	30	28	18	27	40	25	51	25	21	24	37	27	31	43	59	31	37	39	10	33	23	17	41
36	14	23	17	11	14	30	16	33	17	21	24	39	28	37	46	63	31	34	36	8	28	18	10	31
41	17	23	14	10	15	27	17	33	20	27	31	44	30	41	45	66	37	38	41	10	24	18	11	27
56	36	51	49	48	45	52	36	71	62	69	68	71	60	74	76	82	53	56	58	11	54	30	20	50
33	13	20	21	18	19	33	19	39	16	24	23	39	26	36	42	57	24	28	27	9	32	17	13	34
30	12	22	21	15	20	38	17	43	22	22	26	38	28	41	53	67	31	30	30	7	33	17	11	38
25	9	13	9	5	8	14	7	20	12	12	15	26	20	22	29	43	18	22	23	4	16	11	8	20
26	11	15	11	6	11	20	9	23	13	15	18	27	19	24	30	44	17	22	24	5	19	13	7	24
28	12	15	9	5	8	15	9	23	14	15	18	31	20	26	31	45	19	25	27	6	18	14	8	21

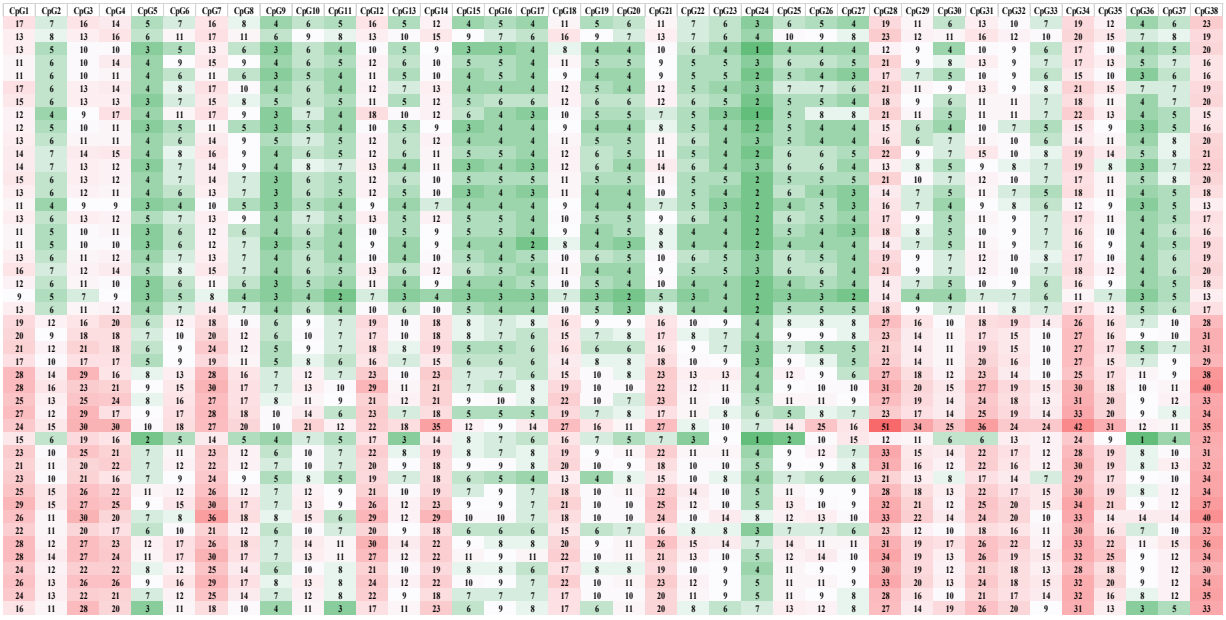
Methylation values of rDNA sequence bearing G variant – Assay 1

CpG1	CpG2	CpG3	CpG4	CpG5	CpG6	CpG7	CpG8	CpG9	CpG10	CpG11	CpG12	CpG13	CpG14	CpG15	CpG16	CpG17	CpG18	CpG19	CpG20	CpG21	CpG22	CpG23	CpG24	CpG25
17	6	9	5	3	5	9	5	12	7	7	10	15	12	14	18	29	9	14	14	3	10	7	4	12
14	6	11	10	7	9	14	5	18	9	8	10	14	12	15	21	25	7	11	11	2	14	7	4	16
16	5	8	4	2	3	6	4	9	6	6	8	13	9	12	15	26	7	12	13	3	7	6	3	8
12	4	9	7	4	6	14	4	16	6	5	8	12	13	12	19	23	6	9	9	2	12	5	4	17
11	4	7	4	2	4	8	3	11	6	6	8	13	13	12	17	22	6	10	10	2	10	5	4	12
15	6	11	6	4	5	14	5	17	6	7	9	13	11	15	20	28	7	12	11	3	12	6	4	15
17	6	9	6	4	6	11	5	14	7	7	10	15	13	15	19	27	8	13	13	3	12	7	4	14
13	5	10	7	4	7	13	4	15	6	6	8	12	12	13	18	24	7	11	10	2	13	6	4	17
12	4	6	4	2	4	6	3	9	5	5	7	12	10	10	14	20	5	9	9	2	8	5	3	9
16	6	10	5	3	5	9	4	11	7	6	9	13	11	13	17	25	7	12	12	3	10	6	4	11
14	5	9	7	4	6	12	5	16	8	7	10	15	13	15	21	26	7	12	12	2	12	6	4	15
20	8	11	5	3	5	8	5	11	6	7	9	14	10	13	16	28	8	13	15	3	10	8	4	10
14	6	8	5	3	5	8	4	12	7	7	10	15	15	13	19	26	8	13	13	3	10	6	4	12
15	6	10	5	3	4	10	4	11	5	6	7	11	9	12	16	25	7	11	11	2	10	5	3	11
11	4	7	4	2	4	6	3	9	6	5	7	12	12	10	15	22	6	10	10	2	8	5	3	10
14	5	8	5	3	5	9	4	12	7	7	9	14	12	13	17	25	7	12	12	2	9	6	4	12
13	5	8	5	3	5	9	4	12	6	6	8	13	12	12	17	22	6	11	11	2	10	6	4	12
15	6	8	6	3	5	9	5	13	7	7	9	15	11	13	18	24	7	12	13	3	11	7	4	11
14	5	8	5	2	5	8	4	12	6	6	8	14	13	13	18	26	7	12	12	2	9	6	4	12
14	5	8	6	3	6	11	5	14	8	7	10	15	14	14	19	26	8	12	13	2	11	6	4	13
14	6	8	5	3	4	7	4	10	6	6	8	13	10	11	15	23	6	11	12	3	10	6	4	11
9	3	6	4	2	3	7	2	9	5	4	5	9	7	8	13	14	3	7	8	1	9	4	2	10
13	6	9	6	4	6	11	4	14	7	7	9	14	12	14	18	23	6	10	11	2	11	6	4	13
24	10	13	9	5	8	13	8	20	12	13	16	27	19	23	29	35	11	20	21	4	16	11	6	18
27	12	17	10	8	9	15	9	20	13	14	18	28	19	25	29	39	15	24	24	5	17	13	8	18
30	13	16	7	4	7	11	8	16	10	12	15	25	15	23	25	38	13	22	23	5	13	12	6	13
31	13	17	9	5	9	14	10	19	13	13	17	28	19	26	29	41	15	25	27	7	17	14	7	16
33	14	18	10	6	10	16	12	18	13	14	19	28	21	26	29	49	20	28	29	7	16	13	8	17
33	13	19	11	7	11	20	12	25	14	17	22	34	21	32	36	49	20	30	31	7	19	16	9	21
28	12	17	13	9	11	20	11	25	13	14	18	27	20	27	32	40	14	22	23	5	19	12	8	22
32	13	17	9	5	8	15	10	20	12	15	19	31	18	27	31	45	17	28	29	6	15	13	8	17
25	9	13	5	3	5	9	6	15	9	10	13	22	12	19	22	29	9	19	20	4	12	11	5	12
26	11	14	8	4	7	11	7	17	10	11	14	24	17	21	25	33	11	21	21	5	13	12	6	15
28	11	15	10	5	10	17	9	21	12	13	17	27	22	25	31	43	15	24	24	6	18	13	8	23
26	11	16	11	7	11	18	10	23	14	13	17	27	23	26	32	41	15	23	24	5	20	13	8	22
30	13	16	7	4	6	12	8	17	11	14	17	27	15	24	26	40	14	25	26	6	14	13	6	13
32	15	19	12	7	12	18	12	22	14	15	19	28	19	27	31	43	16	26	28	8	19	16	9	21
33	13	19	12	7	11	21	12	25	14	16	20	31	22	29	34	46	17	27	28	7	21	15	9	23
38	17	22	14	9	13	22	15	26	17	21	25	37	26	35	39	60	25	36	36	10	21	17	10	22
29	13	16	9	5	7	13	9	19	12	14	17	28	18	25	28	40	15	25	26	6	16	14	7	17
34	16	21	15	10	13	22	13	27	15	17	21	32	22	31	36	48	19	30	30	8	22	17	10	25
27	11	18	14	10	13	24	11	28	14	15	18	27	22	28	35	45	16	23	23	6	22	13	9	27
26	10	15	9	6	10	18	9	22	12	12	15	26	22	25	31	43	16	25	24	5	18	12	8	21
27	12	16	14	8	13	23	11	26	14	15	18	27	20	26	33	42	14	23	24	6	21	14	8	26
29	12	16	9	5	9	15	9	22	13	14	18	29	21	26	31	41	14	25	26	6	18	14	8	21
32	15	19	9	5	8	15	10	20	12	15	19	30	18	27	31	43	17	27	29	7	18	14	7	17

Graded Color Scale → Increase

The first 23 rows are young aged male sperm samples and the next 23 rows are old aged male sperm samples

Methylation values of rDNA sequence bearing A variant – Assay 2



Methylation values of rDNA sequence bearing G variant – Assay 2

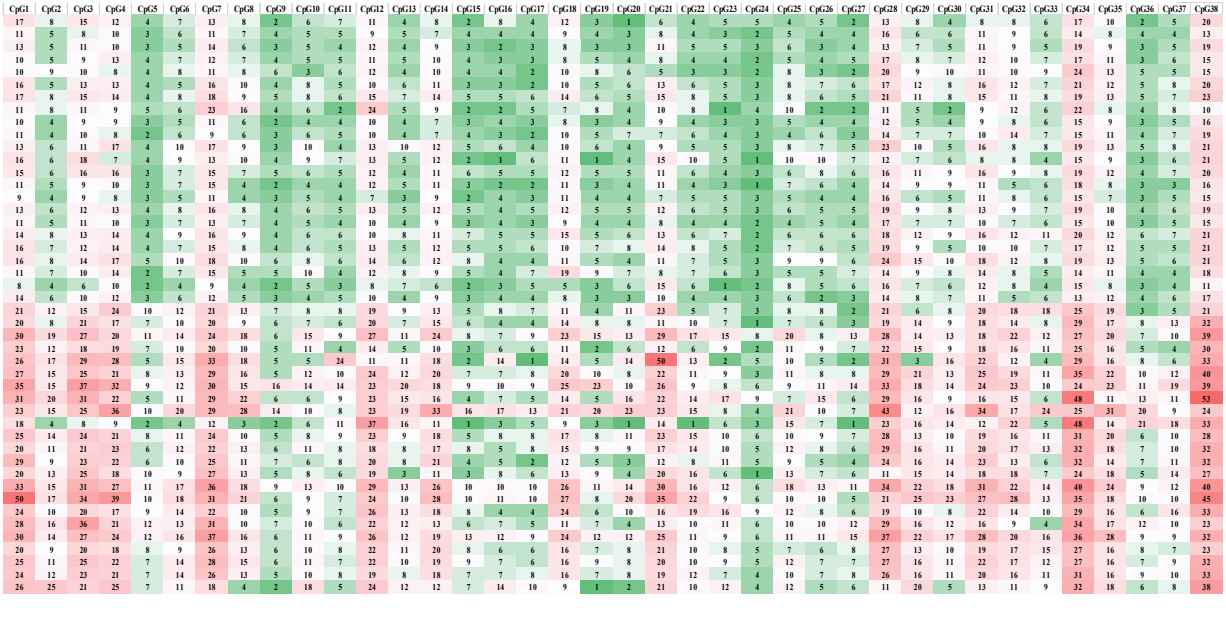


Figure 23. Heatmaps showing variation in methylation levels across individual 25 and 38 CpGs of assay 1 and 2 (each column represents one CpG) in ribosomal DNA region

The first 23 and the next 23 rows are methylation values in sperm from young aged and old aged human male donors respectively. An increase in methylation values across each CpG can be noticed between the two groups.

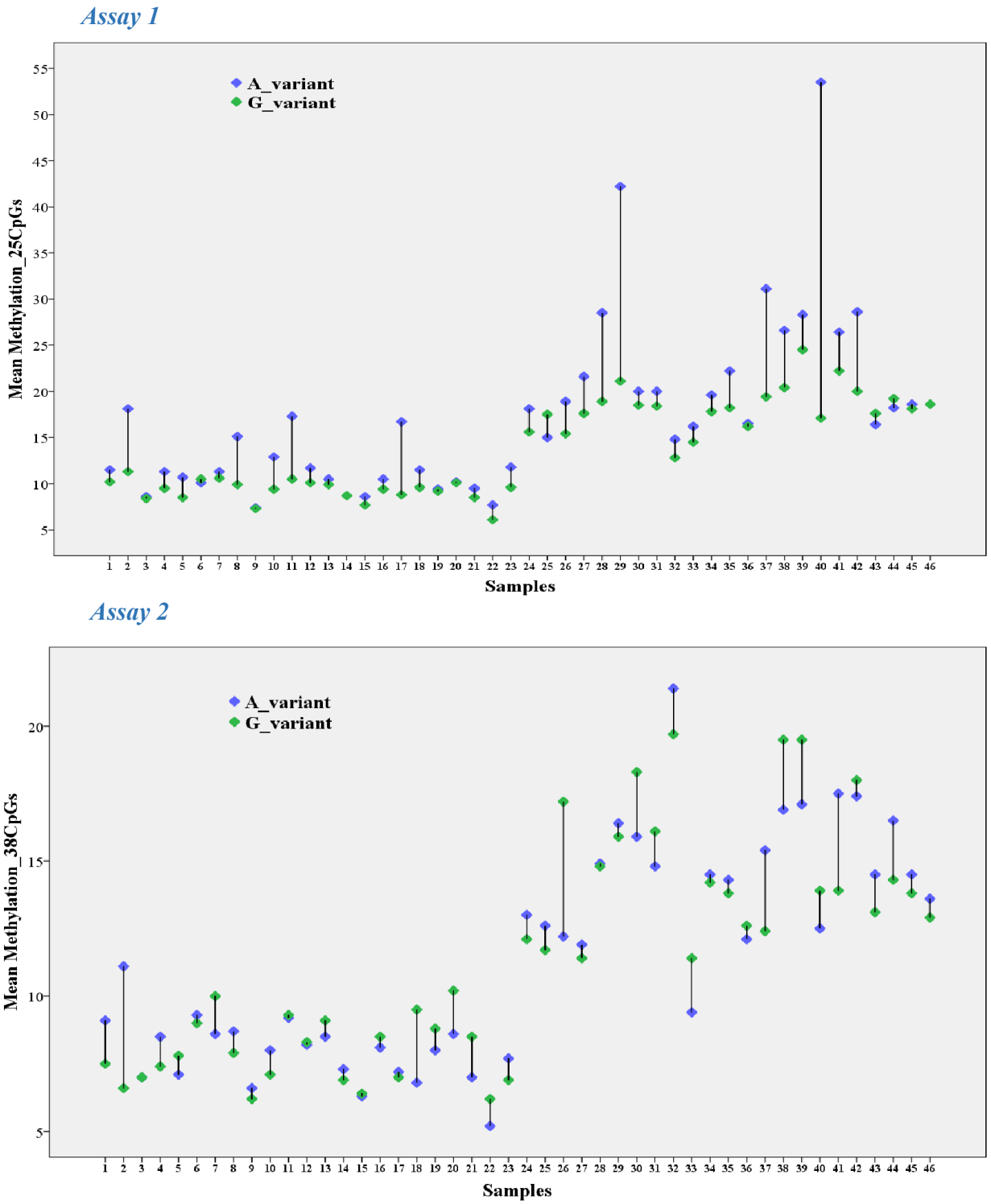


Figure 24. Mean methylation of rDNA assays with both genetic variants across all samples

The first 23 samples are from young males (25-35 years) and the next 23 samples are from old males (>42 years). Sequence-dependent methylation is observed in assay 1 wherein the DNA sequence carrying A (blue) showed more methylation than the one with G (green). The technique used was deep bisulphite sequencing on MiSeq platform.

3.1.5 Epimutation rate: Epimutations, which are defined as alleles exhibiting >50% abnormally (de) methylated CpGs, were calculated and subsequently epimutation rate (ER) was determined. Since the methylation map of ribosomal DNA repeats in sperm is expected to be close to 0%, all alleles/molecules showing >50% methylation values were considered as epimutations. The epimutation rate for sperm from old men was significantly higher than the sperm from young men in both the genetic variants (A and G) of two assays. Figure 25 depicts the box plots showing a significant difference in epimutation rate between the sperm from young and old men in both the genetic variants of the assays. Figure 26 displays an increased variation of epimutation rate across individual samples in assay 1 and 2.

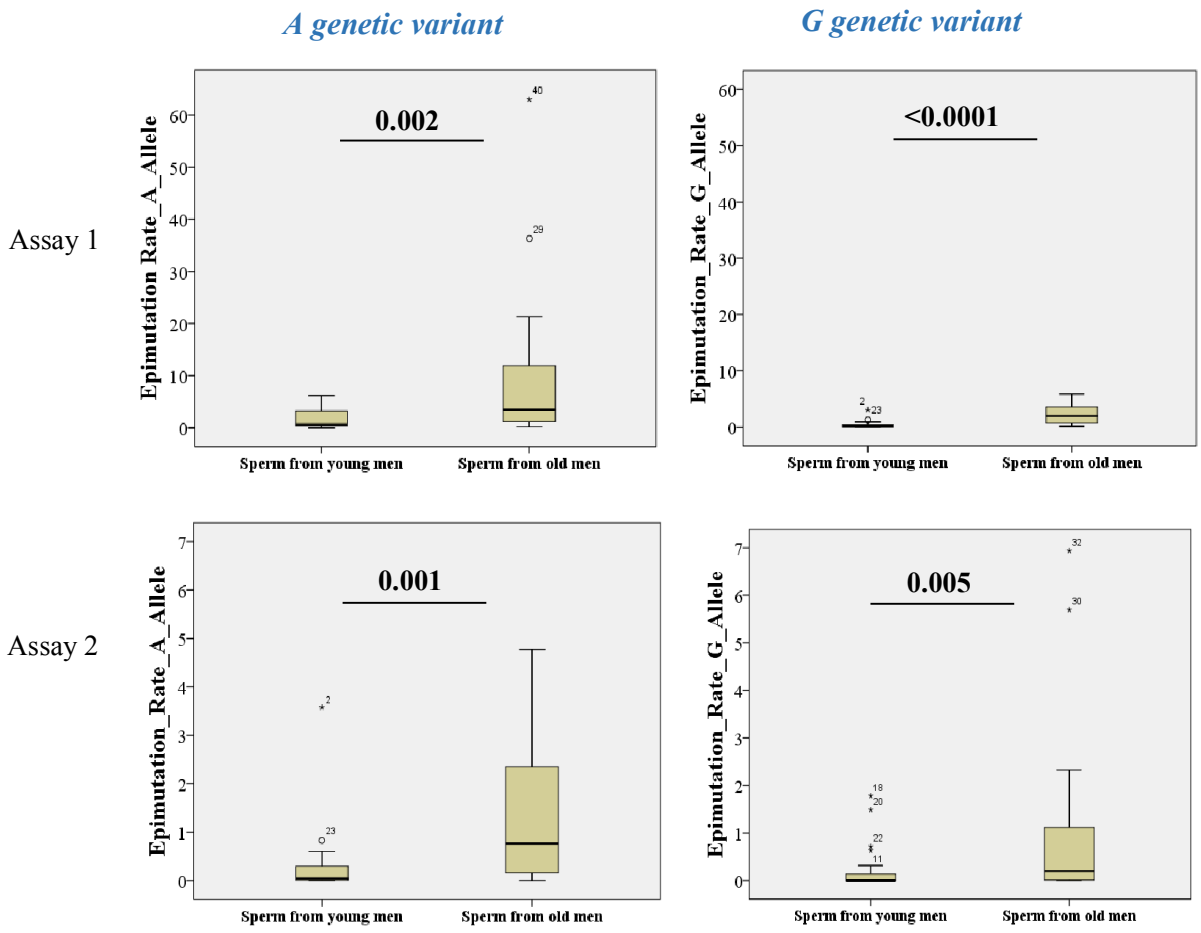
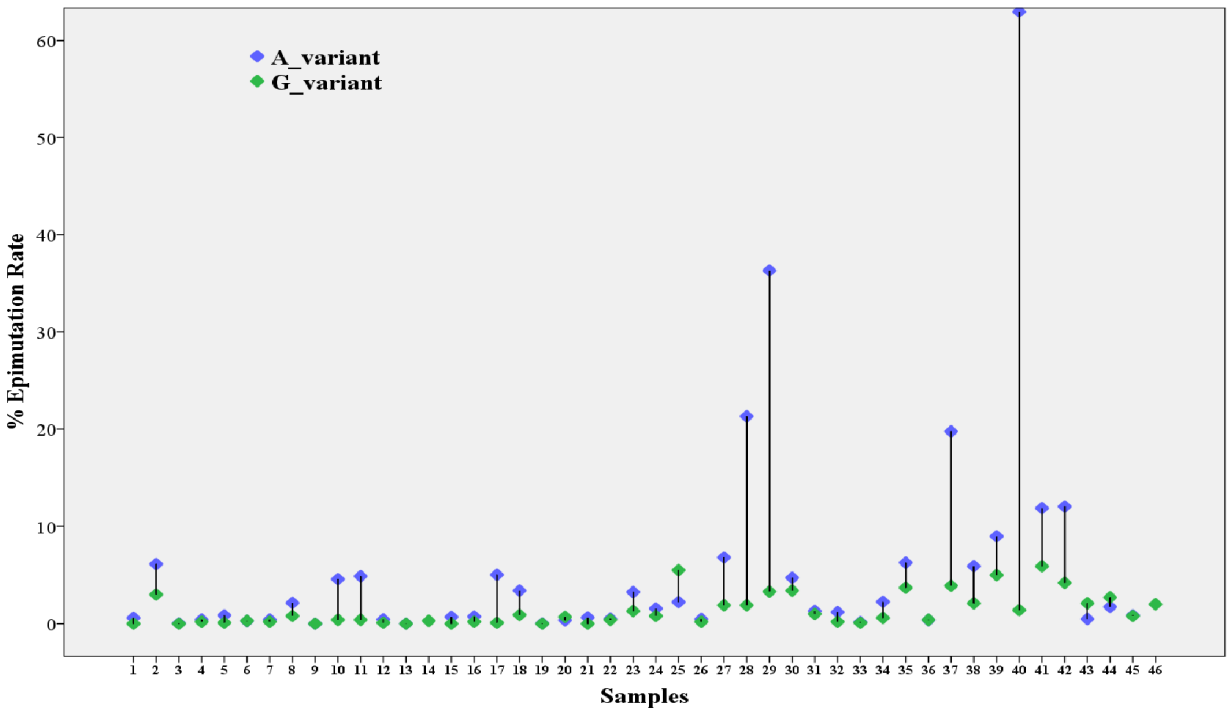


Figure 25. Box plots depicting the significant difference in epimutation rate (%) between the sperm from young and old men in both variants across two assays

The alleles displaying >50% methylation levels were considered as epimutations, and subsequently epimutation rate was determined.

Assay 1



Assay 2

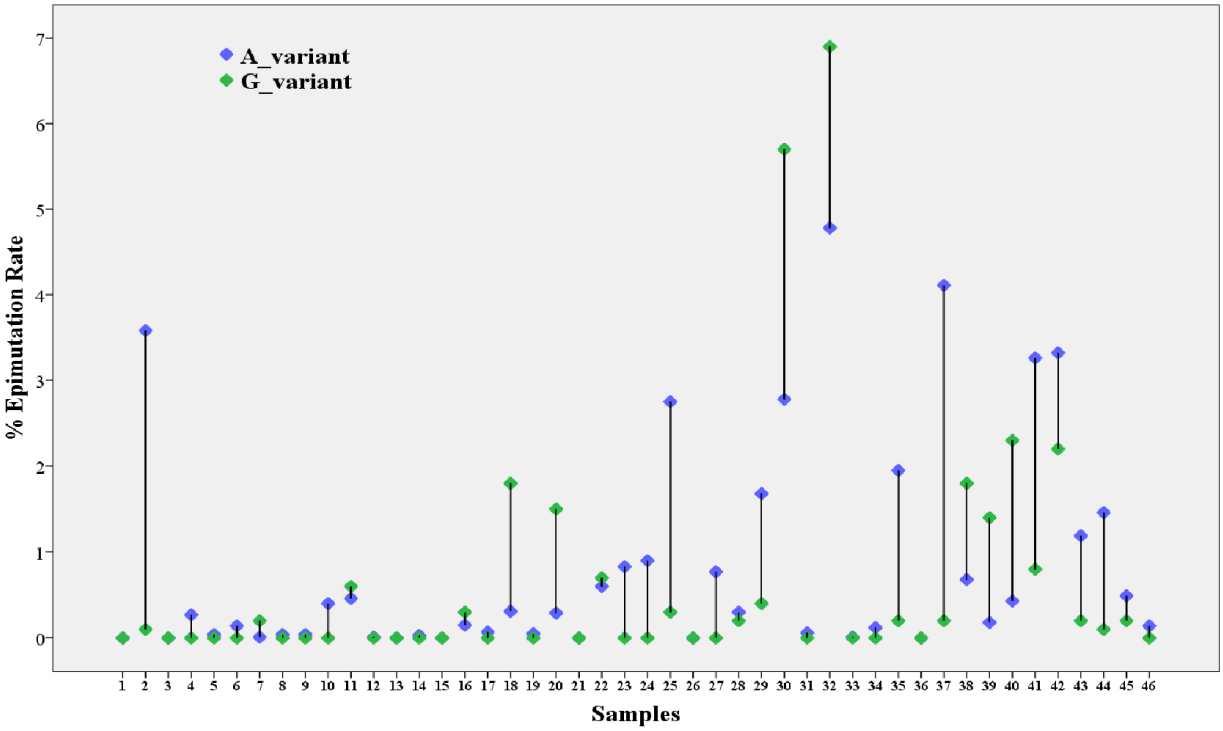


Figure 26. Increased variation of epimutation rate in individual samples of both rDNA assays

The first 23 samples are from young males (25-35 years) and the next 23 samples are from old males (>42 years).

3.2 The effect of promoter CpG methylation on gene expression

Functional analysis using Dual-Luciferase Reporter Assay

It was shown in our lab that the methylation levels of the human *FOXK1* gene promoter region in sperm DNA from 162 male donors, and in 191 fetal cord blood samples from resulting children significantly correlated with donor/paternal age [131]. In order to examine the effect of *FOXK1* amplicon's promoter methylation on the expression of the gene, dual luciferase reporter assay was performed. Detailed methodology and primers for *FOXK1* amplicon are provided in section 2.2.10 and in table 19 respectively.

The human *FOXK1* differentially methylation region (DMR) of 711 bp was cloned into pCpGL vector. This vector is fully devoid of CpG dinucleotides in its backbone that might eventually influence the expression of the downstream luciferase gene. This human *FOXK1* desired region was known to function as a promoter (ENSR00000156049). The firefly experimental vector with either un-methylated or methylated insert was co-transfected with an internal renilla control in three different cell lines: SH SY5Y (human neuroblastoma), U2OS (human osteosarcoma), and HELA (human cervical carcinoma).

After normalizing firefly activity against renilla control, luciferase expression of the methylated pCpGL vector containing methylated *FOXK1* promoter was significantly lower than that of the unmethylated insert in all the cell lines ($p = 0.002$ in SH SY5Y, $p = 0.005$ in U2OS and $p = 0.001$ in HELA). In other words, the methylation at the promoter CpG region in the human *FOXK1* amplicon was suppressing the expression of the downstream luciferase gene. The experiment was carried out in duplicates; empty pCpGL vector (only plasmid without any insert) and non-transfected cells were used as negative controls; CMV (cyto-megalo virus) served as positive control. The human *FOXK1* amplicon is a single copy gene whose promoter methylation in sperm and corresponding fetal cord blood samples significantly correlated with donor/paternal age, and the gene expression was influenced by promoter methylation [131]. Similar to this, attempts will be made to examine the influence of human ribosomal DNA repeat promoter methylation on the expression of the downstream gene using dual luciferase reporter assay.

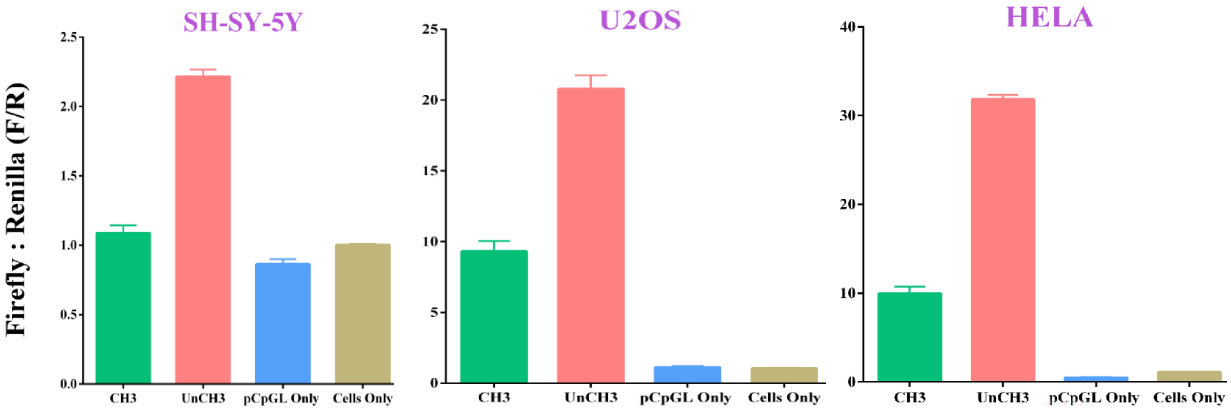


Figure 27. Luciferase activity of pCpGL vector containing methylated and un-methylated *FOXK1* promoter, normalized to the activity of an internal renilla vector

Firefly and renilla vector were co-transfected in a 500:1 ratio in all the above-mentioned cell lines. Green bars indicate the methylated vector and the red bars indicate the unmethylated vector. Blue bars indicate empty pCpGL vector without any insert and the brown bars indicate non-transfected cells. Methylated vector (green) clearly displayed low gene expression levels when compared to the unmethylated vector (red) in all the analyzed cell lines.

3.3 Paternal aging effects on embryo grading quality of next generation

3.3.1 Data collection: After *in vitro* fertilization of sperm samples (obtained from Fertility Center, Wiesbaden, Germany), the resulting embryos were graded before implanting into their mothers. We were successful in obtaining the respective grading information for each embryo.

3.3.2 Blastocyst stage embryos: Though embryos were graded at different stages of development (zygote, 2-celled stage, 4-celled stage, 8-celled/cleavage stage, 16-celled stage/morula, and blastocyst), we considered only blastocyst stage (figure 28) quality grading for analysis. A morphological scoring system was used by the clinicians to grade the embryos at the blastocyst stage. It was performed based on blastocoel, inner cell mass, and trophectoderm grades. Criteria such as expansion and hatching status were considered for blastocoel grading. On the other hand, the number and the symmetrical arrangement of cells were taken into account for inner cell mass and trophectoderm quality. Depending upon the overall grading criteria at the blastocyst stage, embryos were categorized into three groups, namely good, moderate, and poor graded ones. Figure 28 depicts blastocyst stage embryo representing different regions.

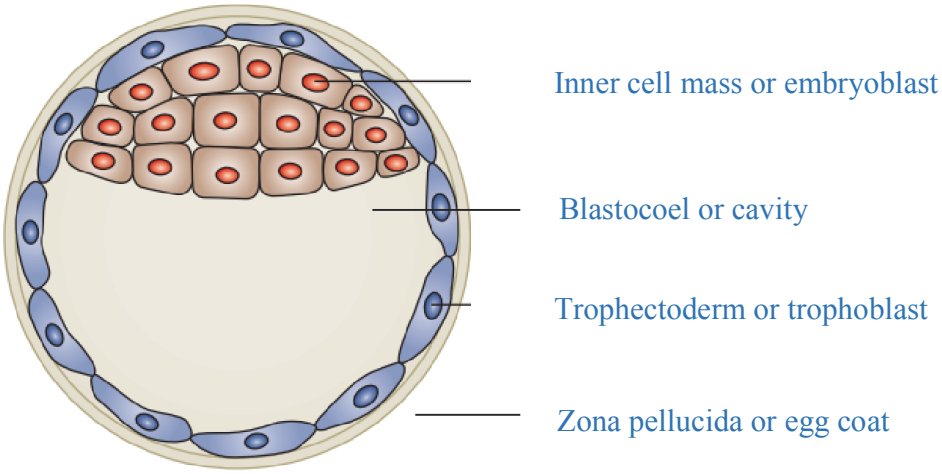


Figure 28. Blastocyst stage embryo

Embryos with an expanded cavity and thinning of egg coat (blastocoel grade), many tightly packed cells (inner cell mass quality), and many cells forming a cohesive layer (trophectoderm grade) are generally considered as good ones. The figure is taken from [132] and modified.

3.3.3 Analysis: To examine how grading of the next generation blastocyst stage embryos varied with paternal age, and with the methylation in sperm of the father in rDNA promoter and UCE regions, analysis was performed on the obtained grading information. The data, in this case, was highly skewed since there were more of good graded embryos (n = 276) than moderate (n = 88) and poorly graded (n = 8) embryos that were implanted into their corresponding mothers.

a) When paternal age was plotted against embryo (offspring) grading information, there was a significant increase in father's age from good to poorly graded embryos (figure 29). Good quality embryos were obtained after IVF/ICSI from the sperm of younger fathers than that of the older ones. In other words, with the sperm from older males, the chance for the occurrence of poorly graded embryo increased.

b) When embryo grading information was plotted against the father's sperm rDNA promoter and upstream core element regions' DNA methylation, there was a tendency for poorly graded embryos to occur with an increase in methylation of rDNA promoter and UCE amplicons. However, the increase in methylation of these regions between good and poorly graded embryos did not reach significance (figure 30).

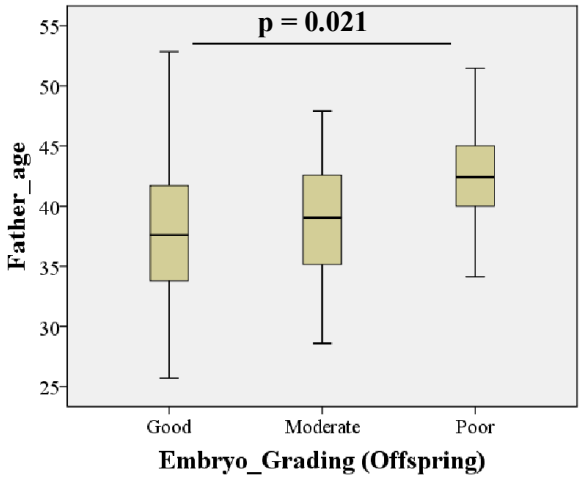


Figure 29. Box plots representing embryo grading (good, moderate and poor) information on the x-axis and the paternal age (in years) on the y-axis

There is a significant tendency for poorly graded embryos to occur from the sperm of older males.

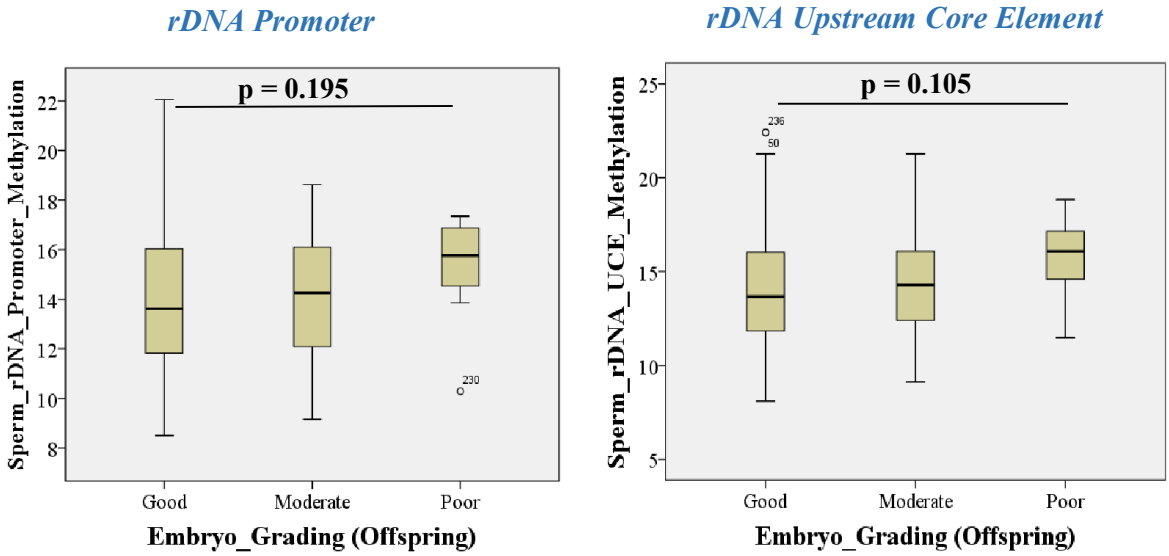


Figure 30. Box plots representing embryo grading (good, moderate and poor) information on x-axis and methylation of rDNA promoter and UCE regions in sperm of father on y-axis

Poorly graded embryos tend to occur from father's sperm that had high methylation in rDNA promoter and UCE regions.

3.4 Aging effects on sperm DNA methylation in bull

3.4.1 Characteristics of study samples: Out of 36 semen samples acquired from 15 different bulls from Friedrich Loeffler Institute, Mariensee, Germany, 9 bulls had their samples collected at three different ages and for 3 bulls at two different ages. Overall, semen samples from 12 bulls at 2 or 3 different ages were acquired. The age of each bull at which the semen sample was collected is given in table 28. The age was categorized into three different groups which are summarized in table 29.

Table 28. The names of the bulls and the age at which semen samples were collected

Name of the Bull	Age (in years) at which semen sample was collected		
Blomdahl	1.1	5.0	6.6
Bowers	1.1	3.1	5.4
Mowambo	1.2	5.3	7.3
Puki	1.1	3.8	7.0
X-man	1.3	2.8	4.9
Armor Red	1.5	3.5	4.2
Bahrain	1.3	5.1	5.8
Elwood	2.8	6.8	7.8
Mergim	0.7	9.4	10.9
Benstrup	1.0		7.0
Bengal	1.0	5.3	
Emidio		4.9	7.3
Truman			7.7
Mavid			9.8
Stylist			12.3

Table 29. The sample size for each category of bull based on their age

Age of the bull	Category	Sample size	Color code
<3 years	Young	12	Green
3 – 6 years	Middle aged	12	Blue
>6 years	Old	12	Red

3.4.2 Methylation of repetitive elements: To examine the age effect among the three categories and across individual bull, methylation analysis of repetitive elements was performed on bull sperm samples using bisulphite pyrosequencing. The assays analyzed in bulls, the number of CpGs per assay, and the PCR and pyrosequencing primers are given in table 7.

3.4.2.1 Ribosomal DNA

The promoter sequence of ribosomal DNA region in the bull is not annotated. On the other hand, the primer sequences (for PCR and pyro) in 18S and 28S rDNA regions are conserved between humans and bulls. We confirmed it by designing primers (for Sanger sequencing) at the flanking regions of the 18S and 28S human rDNA assay amplicons and performed Sanger sequencing on bovine genomic DNA (data not shown). Therefore, we considered human 18S and 28S rDNA PCR and pyro primers for methylation analysis in bull as well (table 7). When all the samples were treated as individual ones, the mean methylation of 18S rDNA amplicon in sperm exhibited a significant correlation, and the mean methylation of 28S rDNA showed a trend with the age of the bull (table 30 and figure 31). When the mean methylation was analyzed across the three categories, there was a significant increase in methylation between young vs old bulls in 28S rDNA, and 18S rDNA displayed a trend (box plots in figure 32). Mean methylation in sperm of individual bull was also clearly increasing as its age increased (figure 33).

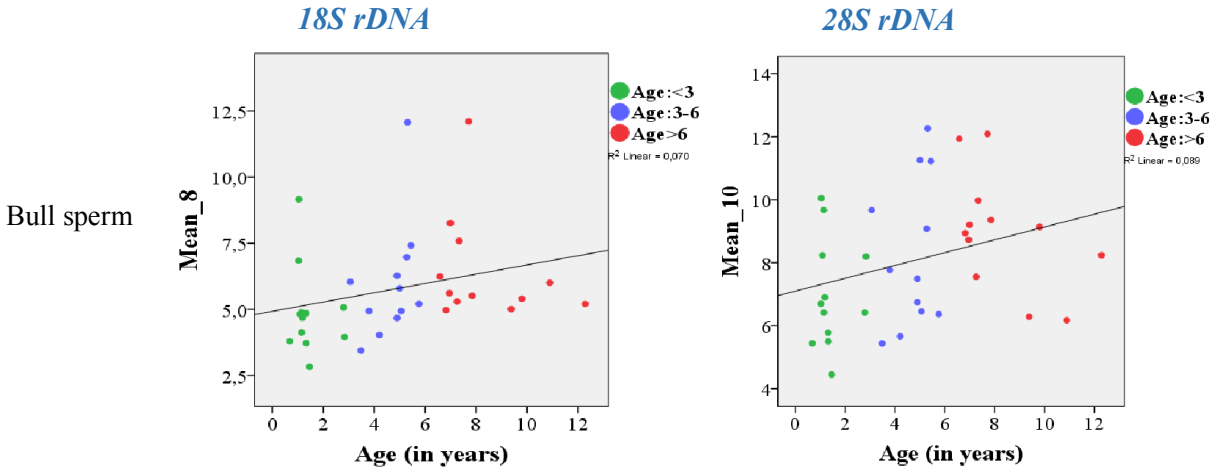


Figure 31. Scatter plots showing the positive correlation between bulls' age (x-axis) and mean methylation of rDNA assays (y-axis)

The technique used was pyrosequencing and the samples analyzed were bull sperm.

Table 30. Spearman correlation between bulls' age and mean methylation of rDNA assays

Cohort	Assay	No of CpGs	Spearman's rho	p-value
Bull sperm	18S rDNA	8	+ 0.45	**0.006**
n ~ 36	28S rDNA	10	+ 0.32	**0.058**

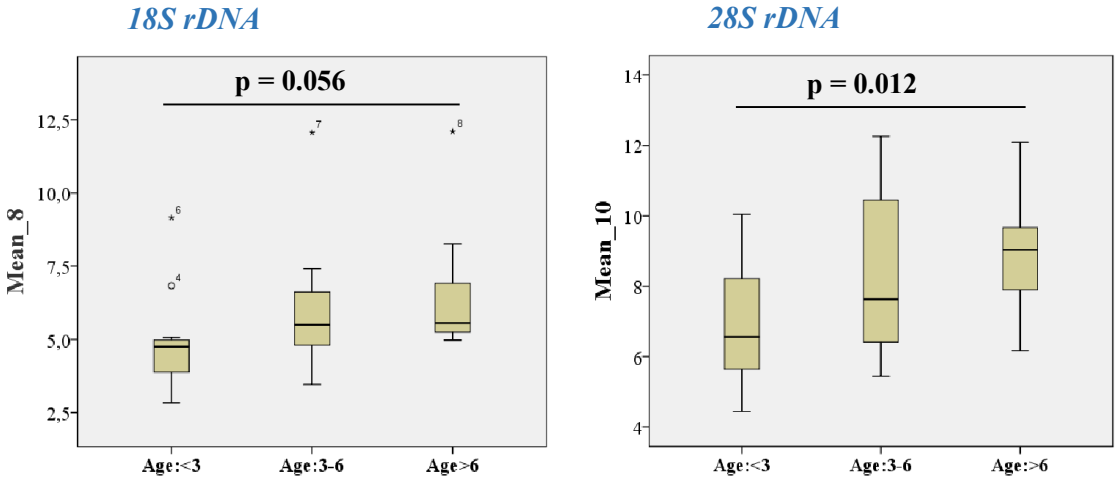


Figure 32. Box plots representing the distribution of methylation values in sperm samples of bulls categorized into three different groups based on their age

A significant difference in methylation was observed in 28S rDNA region between young vs old bulls.

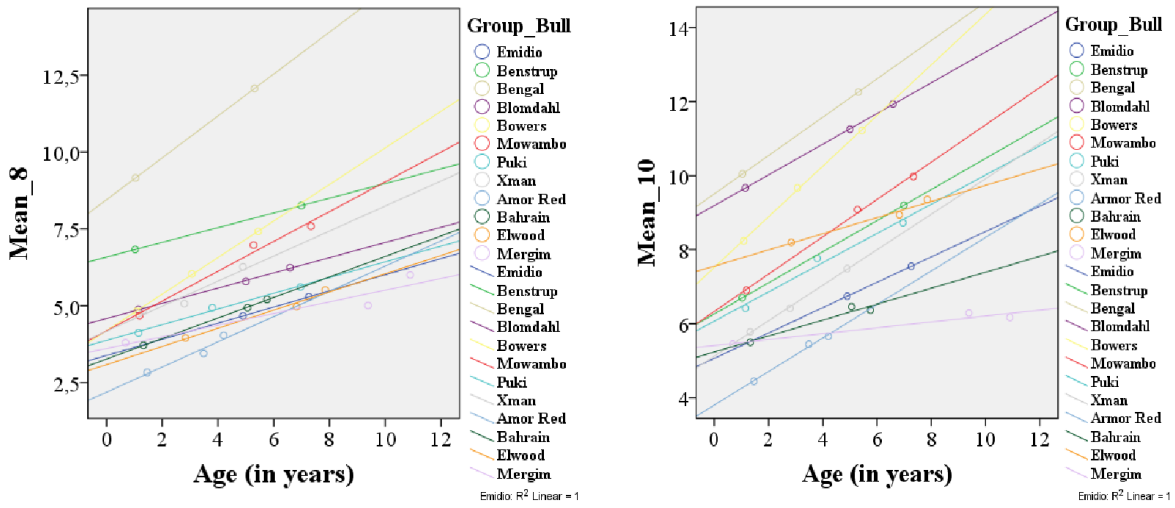


Figure 33. Scatter plots showing an increase in the methylation (in sperm sample) of 18S and 28S rDNA regions across individual bulls at different ages in their lives

The 12 bulls for which we could retrieve the semen sample at 2 or 3 different ages were plotted in the above figures.

3.4.2.2 Bovine Alpha satellite and Testis satellite repeats

Next, the aging effect on other repetitive elements in all the sperm samples of bulls was examined. Methylation analysis was performed using bisulphite pyrosequencing on bovine alpha satellite and testis satellite repeats. When all the 36 samples were analyzed individually, the mean methylation of both repeats displayed a significant correlation with age of the bulls (table 31 and figure 34). When the distribution of mean methylation of sperm was considered across the three age categories (box plots in figure 35), there was also a significant increase in methylation in both the repeats (between young vs middle-aged, middle-aged vs old, and young vs old bulls). Finally, the mean methylation in sperm of each individual bull in both the repetitive elements was also apparently increasing as the age of the corresponding bull increased (figure 36).

Table 31. Spearman correlation between bulls' age and mean methylation of repeat assays

Cohort	Assay	No of CpGs	Spearman's rho	p-value
Bull sperm	Alpha satellite	12	+ 0.62	**< 0.0001**
n ~ 36	Testis satellite	9	+ 0.76	**< 0.0001**

The technique used was pyrosequencing and the samples analyzed were bull sperm.

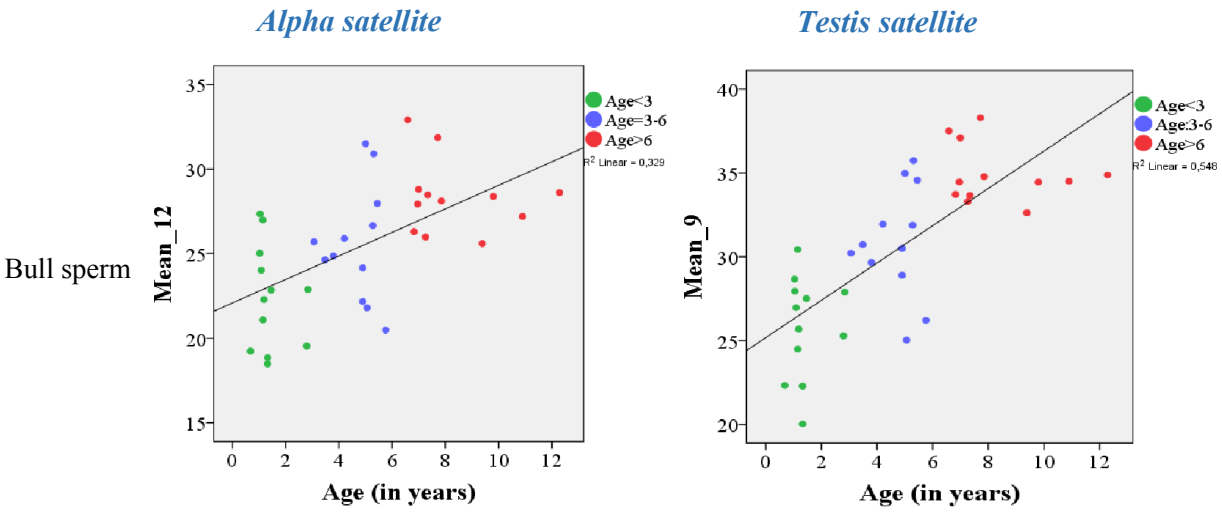


Figure 34. Scatter plots showing the significant positive correlation between bulls' age (x-axis) and mean methylation of alpha satellite and testis satellite assays (y-axis)

The left plot: alpha satellite and the right plot: testis satellite assay.

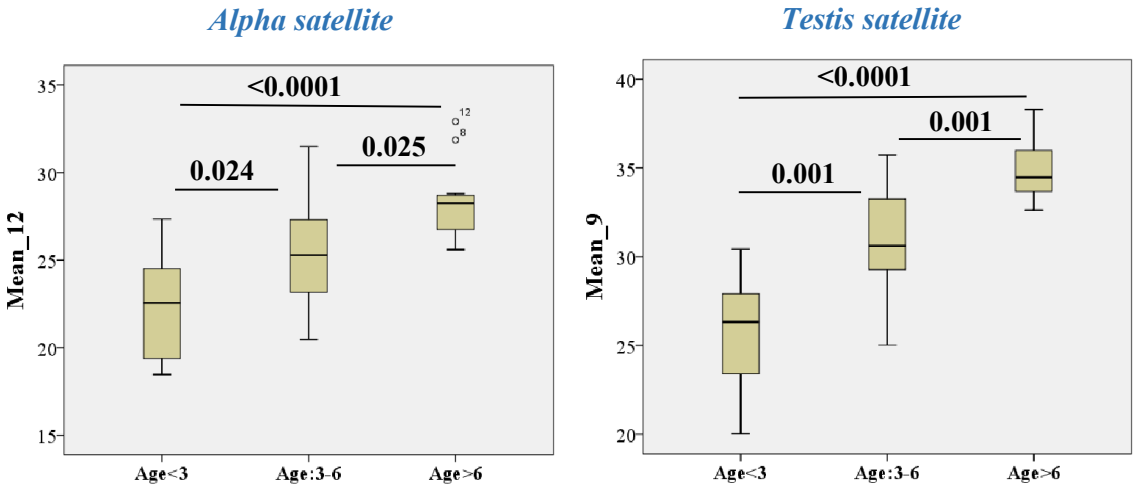


Figure 35. Box plots representing the distribution of methylation values in sperm samples of bulls categorized into three different groups based on their age

A significant difference in methylation was observed in both the repeats. The left plot: alpha satellite and the right plot: testis satellite repeat.

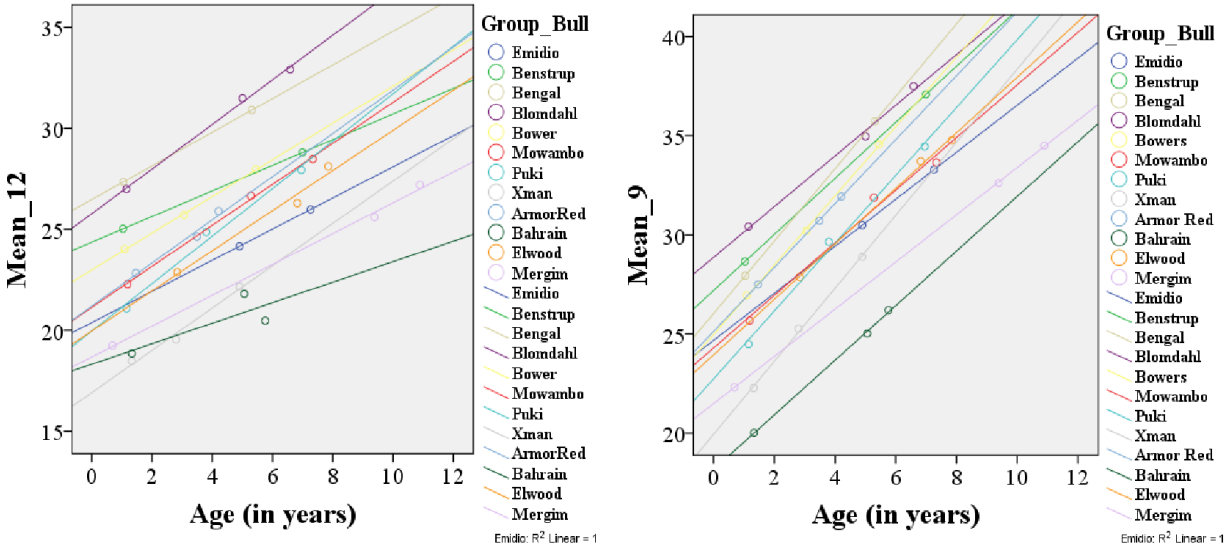


Figure 36. Scatter plots showing an increase in the methylation (in sperm sample) of bovine alpha satellite and testis satellite repeats across individual bulls at different ages in their lives

The 12 bulls for which we could retrieve the semen sample at 2 or 3 different ages were plotted in the above figures. The left plot: alpha satellite and the right plot: testis satellite assay. The technique used was pyrosequencing and the samples analyzed were bull sperm.

The effect of male body mass index on sperm and next generation cord blood DNA methylation

3.5 Human male body mass index (BMI) effects on sperm DNA methylation

***3.5.1 Characteristics of study samples*:** For the study of BMI effects, all the 294 IVF/ICSI sperm samples from Fertility Center, Wiesbaden, Germany were considered as a single cohort. BMI information of the participating male donor was calculated from measured weight (in kilograms) and height (in centimeters). Based on the WHO guidelines of BMI categorization for adults, the percentage of samples falling in different groups is given in table 32. Range, mean and median values for BMI, age and sperm concentration for all the 294 sperm samples are given in table 33.

Table 32. Percentage of samples in different categories based on WHO criteria of BMI classification

Group	WHO classification	Number of samples	Percentage
Underweight	<18.5	1	0.34%
Normal weight	18.5-24.9	144	48.97%
Pre-obese/Obese	>25.0	149	50.68%

Table 33. Range, mean, and median values of BMI, age, and sperm concentration of male donor

Parameter	Range	Mean ± SE	Median
BMI (kg/m^2)	17.30 – 40.30	25.74 ± 0.18	25.00
Age (in years)	25.71 – 65.82	38.94 ± 0.34	38.63
Sperm concentration (million/mL)	0.2 – 210	31.80 ± 2.14	19.00

***3.5.2 Methylation signatures in imprinted genes and in HIF3A amplicon*:** Since imprinted genes are well known to be involved in growth and development and also they escape the genome-wide epigenetic reprogramming mechanism, they were extensively studied in this thesis to detect the male donor's BMI effect on sperm and next generation fetal cord blood DNA methylation. In-depth methylation analysis using pyrosequencing at the regulatory regions of five paternally expressed genes (*MEST*/*PEG1*, *PW1*/*PEG3*, *SNRPN*/*PEG4*, *NNAT*/*PEG5*, and *SGCE*/*PEG10*), three maternally expressed genes (*H19*, *IGF2* DMR0, and *GTL2*/*MEG3*-IG

DMR), and one obesity-related non-imprinted gene *HIF3A* was performed. The number of CpGs quantified for each amplicon, PCR and sequencing primers for bisulphite pyrosequencing of 294 IVF/ICSI sperm samples are provided in table 8.

The methylation measurements of paternally expressed imprinted genes in sperm (haploid cell) are usually close to zero percent (un-methylated), and that of the maternally expressed imprinted genes are close to hundred percent (fully methylated). Samples showing abnormal methylation values (>20% for PEG genes and <70% for MEG genes) were excluded from analysis because they might have been contaminated with somatic cells. Potential confounding factors such as donor's age, sperm concentration, and male fertility status (successful and non-successful live birth after IVF/ICSI) were adjusted by regression analysis using beta models.

Average of all the CpG sites for each amplicon was considered for analysis. *GTL2* (*MEG3*-IG DMR) sperm DNA methylation showed a significant positive correlation with donor's BMI and obesity-related gene *HIF3A* showed trend towards a positive correlation. This indicated that with an increase in male BMI, the methylation levels of *GTL2* (*MEG3*-IG DMR) and *HIF3A* genes in sperm DNA were also increasing. The remaining genes did not reveal a significant correlation between donor's BMI and sperm DNA methylation. Table 34 shows the regression models when sperm DNA methylation of the corresponding amplicon was correlated with BMI of the donor.

Table 34. Regression analysis showing the correlation of sperm DNA methylation and donor's BMI

Amplicon	Estimate	Standard error	p-value
MEST	-0.004	0.009	0.661
PEG3	0.001	0.010	0.917
SNRPN	-0.001	0.009	0.939
NNAT	-0.010	0.010	0.324
PEG10	-0.001	0.011	0.936
H19	0.012	0.009	0.191
***MEG3*-IG DMR**	0.014	0.007	**0.043**
IGF2 DMR0	0.003	0.005	0.632
HIF3A	0.008	0.004	**0.080**

The model was adjusted for sperm concentration, age, and fertility status of the man.

3.6 Paternal BMI effects on next generation fetal cord blood DNA methylation

3.6.1 Characteristics of fetal cord blood (FCB) samples: Among the 113 FCBs that were retrieved from offspring born after IVF/ICSI of above-mentioned sperm, 54 samples (44.62%) were female and 59 samples (48.76%) were male. Range, mean, and median values for birth weight, gestational week of the child, age and BMI of father and mother are represented in table 35. Birth weight positively correlated with the gestational week of the child (rho = 0.67, $p < 0.0001$). Father's age significantly correlated with mother's age (rho = 0.56, $p < 0.0001$) and to a lesser extent, father's BMI also correlated with mother's BMI (rho = 0.21, p = 0.02).

Table 35. The range, mean, and median values of FCB characteristics

Parameter	Range	Mean ± SE	Median
Birth weight: Child (grams)	1510.00 – 4950.00	3154.17 ± 57.14	3130.00
Gestational week: Child	34 - 42	39.54 ± 0.16	40.00
BMI: Father (kg/m^2)	17.30 – 40.30	25.71 ± 0.30	25.00
BMI: Mother (kg/m^2)	17.00 – 39.80	23.04 ± 0.36	22.20
Age: Father (in years)	28.08 – 52.85	38.72 ± 0.46	38.54
Age: Mother (in years)	22.08 – 42.68	34.62 ± 0.35	34.53

3.6.2 Transmission of methylation marks: To investigate the probable transmission of paternal BMI induced sperm DNA methylation signatures in FCB samples of next generation, bisulphite pyrosequencing was performed on 113 samples which were retrieved after successful live birth. Methylation analysis was carried out on the nine above-mentioned genes (section 3.5.2). The methylation levels of imprinted genes in fetal cord blood (diploid cells) are generally close to fifty percent since there is one allele in the FCB sample from each parent. To adjust for potential confounding factors such as birth weight and the gestational week of the child, father's age, mother's age and BMI, regression analysis using beta models was performed.

Interestingly, the influence of father's body mass index on the fetal cord blood DNA methylation varied depending on the gender of the child. Here, *GTL2* (*MEG3*-IG DMR) and *HIF3A* cord blood DNA methylation of male but not female children displayed a significant positive

correlation with the BMI of the father. This was in line with the positive trend observed in these two genes between sperm DNA methylation and donor's BMI (section 3.5.2 and table 34). Additionally, *IGF2* DMR0 displayed a significant negative correlation with paternal BMI in female FCBs whereas *NNAT*/*PEG5* exhibited a negative trend towards significance in male FCB samples. It is noteworthy to mention that the methylation values of none of the nine analyzed genes correlated with maternal BMI. Table 36 and 37 shows the regression beta models when FCB DNA methylation of the corresponding amplicons was correlated with paternal and maternal BMI respectively. Figure 37 shows scatter plots of the four significant genes when paternal BMI was correlated with FCB DNA methylation in male and female children separately.

Table 36. Regression analysis showing the correlation of FCB methylation and father's BMI

Amplicon	FCB Gender	Estimate	Standard error	p-value
MEST	Female	0.003	0.002	0.279
	Male	0.002	0.003	0.505
PEG3	Female	0.001	0.001	0.354
	Male	0.000	0.001	0.821
SNRPN	Female	-0.001	0.001	0.278
	Male	-0.001	0.001	0.313
NNAT	Female	0.003	0.002	0.106
	Male	-0.003	0.002	**0.085**
PEG10	Female	-0.001	0.001	0.299
	Male	0.001	0.001	0.285
H19	Female	-0.004	0.003	0.165
	Male	0.002	0.003	0.524
***MEG3*-IG DMR**	Female	0.000	0.001	0.564
	Male	0.002	0.001	**0.016**
***IGF2* DMR0**	**Female**	-0.003	0.001	**0.025**
	Male	0.001	0.002	0.601
HIF3A	Female	-0.001	0.001	0.546
	Male	0.004	0.002	**0.033**

The model was adjusted for birth weight and the gestational week of the child, father's age, mother's age and mother's body mass index.

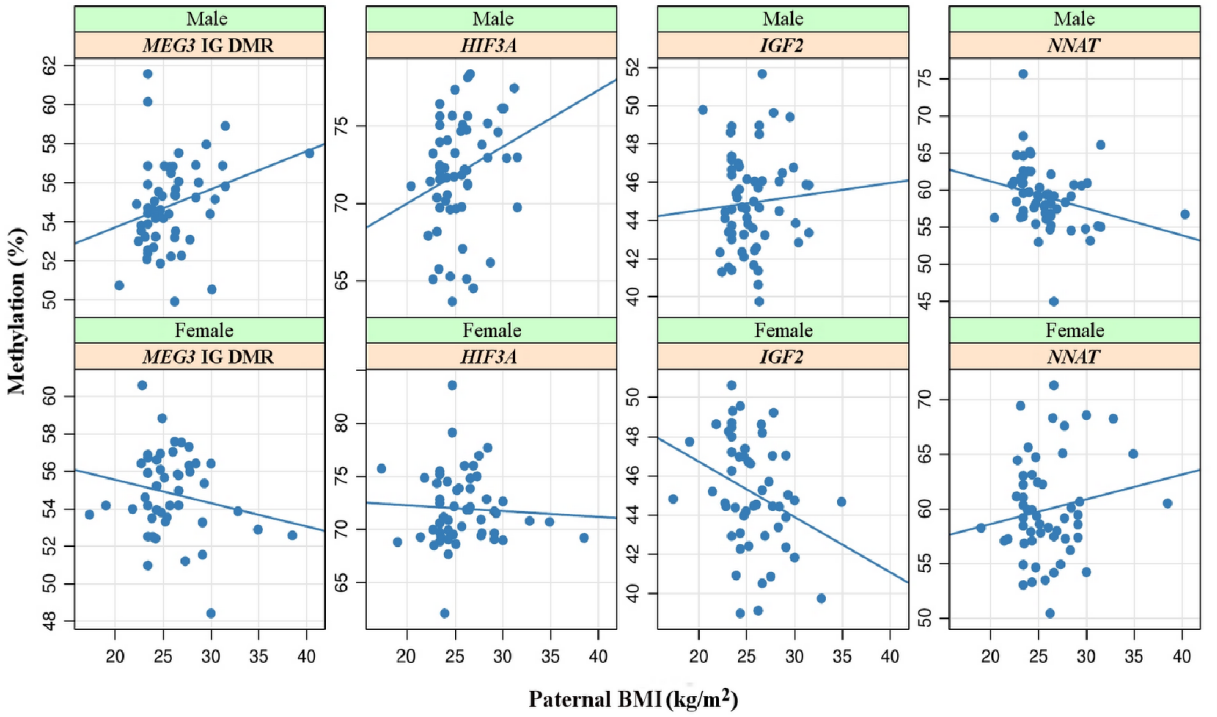


Figure 37. Scatter plots showing the correlation between *MEG3*-IG DMR, *HIF3A*, *IGF2* DMR0, and *NNAT* methylation and paternal BMI in male (upper) and female (lower) FCB samples

Each dot in the plot represents average methylation of several targeted CpGs in an individual cord blood sample after measurement by bisulphite pyrosequencing. Regression line suggests the direction of correlation with paternal BMI.

Table 37. Regression analysis showing the correlation of FCB methylation and mother's BMI

Amplicon	FCB Gender	Estimate	Standard error	p-value
MEST	Female	0.002	0.002	0.296
	Male	-0.001	0.002	0.580
PEG3	Female	0.000	0.000	0.694
	Male	-0.001	0.001	0.486
SNRPN	Female	0.000	0.001	0.927
	Male	-0.001	0.001	0.430
NNAT	Female	0.000	0.002	0.854
	Male	-0.001	0.002	0.640
PEG10	Female	0.001	0.001	0.545
	Male	0.001	0.001	0.411
H19	Female	0.000	0.002	0.962
	Male	-0.002	0.003	0.411
MEG3-IG DMR	Female	0.000	0.001	0.605
	Male	0.000	0.001	0.657
IGF2 DMR0	Female	-0.001	0.001	0.442
	Male	0.001	0.001	0.344
HIF3A	Female	0.000	0.001	0.955
	Male	0.000	0.002	0.795

None of the genes exhibited significant correlation with maternal BMI in any gender.

3.7 Allele-specific methylation association of imprinted genes in FCBs with parental factors

3.7.1 Genotyping of FCB samples: All the available 113 fetal cord blood samples were genotyped for eight imprinted genes to pre-select the heterozygous samples for deep bisulphite sequencing (DBS) on the Illumina MiSeq platform. The covariates for FCB samples and the primers for genotyping assays on bisulphite converted DNA are provided in table 35 and 9 respectively. The protocols for genotyping and DBS are given under sections 2.2.7 and 2.2.9. Table 38 shows the number of heterozygous FCB samples selected after genotyping of each amplicon along with the single nucleotide polymorphism considered, and its corresponding allele frequency in the European population (as DBS run on MiSeq was designed for 48 IDs for each amplicon, we could only select a maximum of 46 samples and other two IDs were for controls).

Table 38. Genotyping results showing no. of obtained heterozygous FCB samples for each gene along with the SNP considered and its allele frequency in the European population

Gene	Single nucleotide polymorphism	Allele frequency in European population	Heterozygous FCBs for DBS run (n)
MEST (*PEG1*)	rs 3778859	0.44 A / 0.56 G	46
PW1 (*PEG3*)	rs 2302376	0.29 A / 0.71 G	46
SNRPN (*PEG4*)	rs220030	0.34 G / 0.66 A	46
NNAT (*PEG5*)	rs 6019062	0.27 A / 0.73 G	39
SGCE (*PEG10*)	rs3814105	0.15 G / 0.85 A	20
H19	rs 11042167	0.49 G / 0.51 A	46
IGF2	rs 3741210	0.30 G / 0.70 A	46
GTL2 (*MEG3*)	rs 12437020	0.29 A / 0.71 G	44

3.7.2 Parental allele methylation correlation with parental factors: The heterozygous fetal cord blood samples were deep bisulphite sequenced on Illumina MiSeq platform for six imprinted genes excluding *PEG4* and *PEG10* amplicon. The primers for the next generation sequencing run for imprinted genes are provided in table 11. Using single nucleotide polymorphism, the parental alleles were differentiated based on the parental origin of imprinted genes as following: For *PEG1*, *PEG3*, and *PEG5* genes, the unmethylated and methylated alleles were considered as paternal and maternal alleles respectively since these are paternally expressed and maternally

imprinted (methylated) genes. For *H19*, *IGF2* DMR0, *MEG3*-IG DMR amplicons, the unmethylated and methylated alleles were treated as maternal and paternal alleles respectively as they are maternally expressed and paternally methylated regions in the genome. Regression analysis using beta models was performed which included factors like parental age, BMI, FCB gender, and the genetic variant (SNP) used to separate the alleles. Parental allele's DNA methylation was correlated with the corresponding parental factors like body mass index and age.

3.7.2.1 BMI effects

The paternally unmethylated allele's DNA methylation in FCB samples for *PEG1*/*MEST* gene significantly correlated in a positive direction with father's BMI. This was particularly detected in female offspring (p = 0.006; n = 20). Similar to pyro results (section 3.6 and table 37), the maternal allele's DNA methylation of none of the studied genes displayed a significant correlation with the mother's BMI. This might probably imply that the mother's body mass index is not influencing methylation of the studied imprinted genes in fetal cord blood DNA. Table 39 summarizes the father's BMI effects on the corresponding *PEG1* paternal allele's DNA methylation in fetal cord blood samples of next generation. Maternal allele's methylation correlation with mother's BMI is also provided.

Table 39. Paternal body mass index effects on the corresponding paternal *PEG1* allele's DNA methylation in fetal cord blood samples of next generation

Gene	Allele in FCB	Correlation	Estimate	St. Error	p-value
PEG1	Unmethylated / **Paternal**	**Father's BMI**	0.004	0.002	**0.005**
	Methylated / Maternal	Mother's BMI	-0.001	0.001	0.439

The technique used was DBS on Illumina MiSeq platform and the samples were heterozygous FCB samples.

3.7.2.2. Age effects

The maternally methylated allele's DNA methylation in FCB samples for *PEG3* amplicon significantly correlated with mother's age in a positive direction. This was especially evident in male offspring (p = 0.028; n = 22). On the other hand, the paternally methylated allele's DNA methylation in FCB samples for the *MEG3*-IG DMR significantly correlated with father's age in

a negative direction. The unmethylated allele's DNA methylation of these two genes did not exhibit any correlation with their corresponding parental age. Table 40 summarizes the parental age effects of these two genes and figure 38 depicts the scatter plots.

Table 40. Parental allele-specific methylation correlation in FCB samples with parental age

Gene	Allele in FCB	Correlation	Estimate	St. Error	p-value
PEG3	Unmethylated / Paternal	Father's age	0.000	0.000	0.349
	Methylated / **Maternal**	**Mother's age**	0.001	0.000	**0.024**
MEG3	Unmethylated / Maternal	Mother's age	-0.003	0.002	0.114
	Methylated / **Paternal**	**Father's age**	-0.001	0.001	**0.055**

Regression analysis using beta models showing a significant correlation of the methylated maternal allele of *PEG3* and methylated paternal allele of *MEG3* in FCBs with mother's and father's age respectively. Technique: DBS.

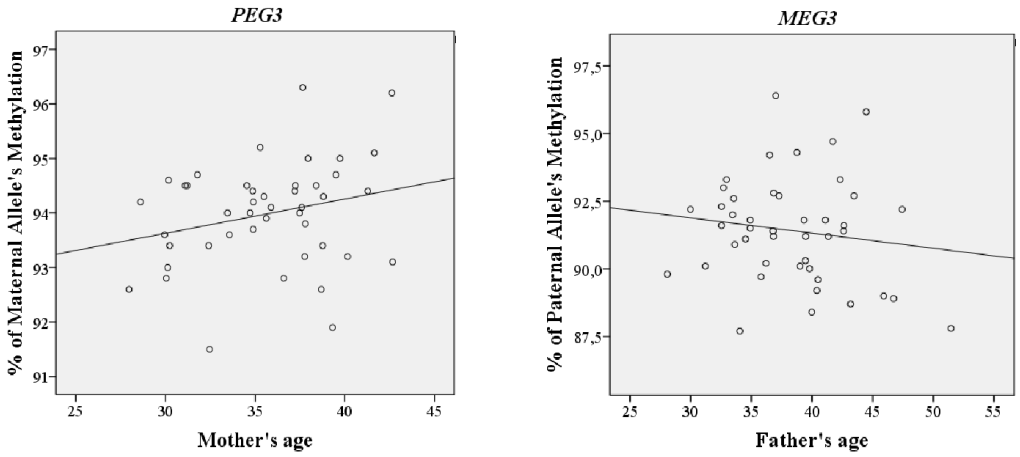


Figure 38. Scatter plots of *PEG3* (left) and *MEG3* (right) showing a correlation of their parental allele's methylation with corresponding parental age

3.7.3 Allele-specific methylation map was depended on the underlying genetic variant: A single nucleotide polymorphism was used to distinguish the methylated allele from the unmethylated one in each sample across all the six amplicons. The rs numbers for all the SNPs are provided in table 38. Interestingly, deep bisulphite sequencing of fetal cord blood DNA regions containing the mentioned SNPs revealed that the levels of the methylation map depended on the underlying genetic variant in certain amplicons of the genome. This effect was the strongest associated

confounding factor in our deep bisulphite sequencing analysis. This was particularly detected in the case of *PEG1*, *PEG5*, *H19* and *IGF2* amplicons and to a certain extent in *PEG3* (table 41).

In the unmethylated alleles of *PEG1*, *PEG3*, and *H19* amplicons, the alleles carrying G (guanine) variant exhibited more methylation values than the ones bearing A (adenine) variant. On the other hand, in the methylated molecules, alleles with A variant showed more methylation than the ones with G variant. Opposite trend was identified in *PEG5*/*NNAT* gene. However, both in unmethylated and in methylated alleles of *IGF2* gene, DNA molecules carrying the G variant displayed significant lower methylation than the ones bearing A variant. *MEG3*-IG DMR was the only region that did not reveal a significant SNP effect. Figure 39 demonstrates methylation marks between alleles with A and G genetic variants, across 32 and 8 individual CpGs of *PEG1* and *MEG3*-IG DMR amplicons in both methylated and unmethylated molecules.

Table 41. Methylation correlation of alleles carrying G genetic variant over ones bearing A variant

Gene	Allele in FCB	Sample Size	Methylation correlation of G over A		
			Estimate	St. Error	p-value
PEG1	Unmethylated / Paternal	46	0.113	0.013	**0.000**
	Methylated / Maternal	46	-0.034	0.009	**0.001**
PEG3	Unmethylated / Paternal	46	0.004	0.002	**0.059**
	Methylated / Maternal	46	-0.006	0.003	**0.035**
PEG5	Unmethylated / Paternal	39	-0.102	0.018	**0.000**
	Methylated / Maternal	39	0.032	0.006	**0.000**
H19	Unmethylated / Maternal	46	0.018	0.005	**0.001**
	Methylated / Paternal	46	-0.033	0.013	**0.013**
IGF2	Unmethylated / Maternal	46	-0.045	0.010	**0.000**
	Methylated / Paternal	46	-0.023	0.009	**0.023**

MEG3	Unmethylated / Maternal	44	-0.009	0.014	0.494
	Methylated / Paternal	44	0.002	0.007	0.732

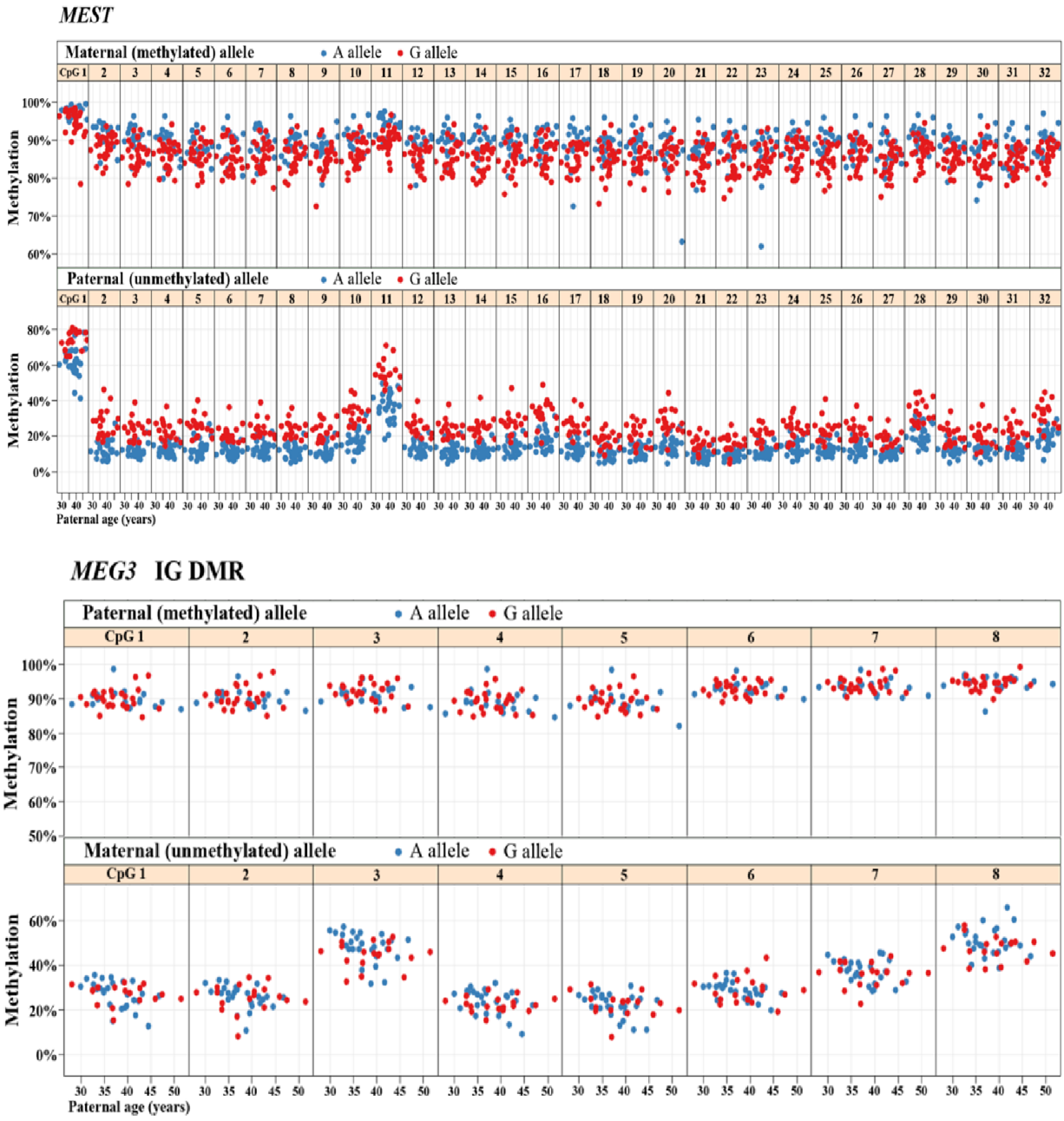


Figure 39. The difference in the percentage of methylation values based on the underlying genetic variants (A or G) in the analyzed DNA sequences of *PEG1* and *MEG3* amplicons

DBS of these genes harbour 32 and 8 CpGs respectively. *MEG3*-IG DMR was the only amplicon that did not exhibit a significant SNP effect (lower panel).

3.7.4 Allele-specific epimutation rates: Epimutations are defined as alleles displaying >50% aberrantly (de) methylated CpGs. For an unmethylated allele (paternal allele for paternally expressed genes and maternal allele for maternally expressed genes), CpGs exhibiting >50% methylation values were considered as epimutations. For a methylated allele (maternal allele for *PEG* genes and paternal allele for *MEG* genes), CpGs with <50% methylation levels were taken as epimutations. Epimutation rate was subsequently calculated by the formula 100 * (epimutations/total number of reads).

It was previously showed in our lab that the epimutation rates (ER) of the unmethylated allele of the materally expressed *MEG3*-IG DMR and paternally expressed *PEG1*/*MEST* genes were significantly higher than those on the corresponding methylated allele [133]. Haertle *et al*. (2017) named the phenomenon as the hypermethylation of the non-imprinted allele (HNA) [133]. This was confirmed in the present DBS run for both *MEG3*-IG DMR and *PEG1*/*MEST* amplicons and was additionally detected in the paternally expressed *PEG5*/*NNAT* region. Despite the methylated alleles of these genes exhibiting a mean methylation value of >85% (approximately close to the theoretical 100%), their unmethylated alleles which were assumed to have a methylation map of 0%, displayed a noticeable hypermethylation (e.g., 35.58% and 32.56% in *PEG5* and *MEG3* genes respectively). Table 42 and figure 40 depict the mean methylation of methylated and unmethylated alleles for all the three genes along with their epimutation rates. Figure 41 consistently demonstrates that the epimutation rate for each individual sample on its unmethylated allele was higher than that of the methylated allele in oppositely imprinted *PEG5* and *MEG3* amplicons.

In contrast to the above phenomenon, for the maternally expressed *H19*-IG DMR and *IGF2* DMR0 amplicons, the methylated paternal allele revealed a higher epimutation rate than that of the unmethylated maternal allele. The difference between both the alleles was much higher for the primary *H19*-IG DMR ($p < 0.0001$) than that of the secondary *IGF2* DMR0 ($p = 0.05$). This increase in epimutation rate for the methylated alleles compared to the unmethylated ones was a novel observation in these two amplicons. This was opposite to the effect of hypermethylation of non-imprinted allele (HNA). On the other hand, unlike all the other analyzed imprinted genes, *PEG3* amplicon exhibited very low epimutation rate on both the parental alleles.

Table 42. Mean methylation and epimutation rate of both alleles in oppositely imprinted genes

Gene	Allele in FCB	Sample Size	% of Mean Methylation	Epimutation Rate %	p-value for ER
PEG1	Unmethylated / Paternal	46	19.33 ± 1.00	14.75 ± 0.99	
	Methylated / Maternal	46	87.25 ± 0.45	10.89 ± 0.50	**0.008**
PEG5	Unmethylated / Paternal	39	35.58 ± 1.35	33.49 ± 1.59	
	Methylated / Maternal	39	85.15 ± 0.38	8.3 ± 0.34	**<0.0001**
MEG3	Unmethylated / Maternal	44	32.56 ± 0.63	17.42 ± 0.74	
	Methylated / Paternal	44	91.42 ± 0.29	5.66 ± 0.29	**<0.0001**

Note that the significant hypermethylation of non-imprinted allele (HNA) was observed in all the above mentioned three genes. Unmethylated allele displays more epimutation rate than the methylated allele (p-value for ER).

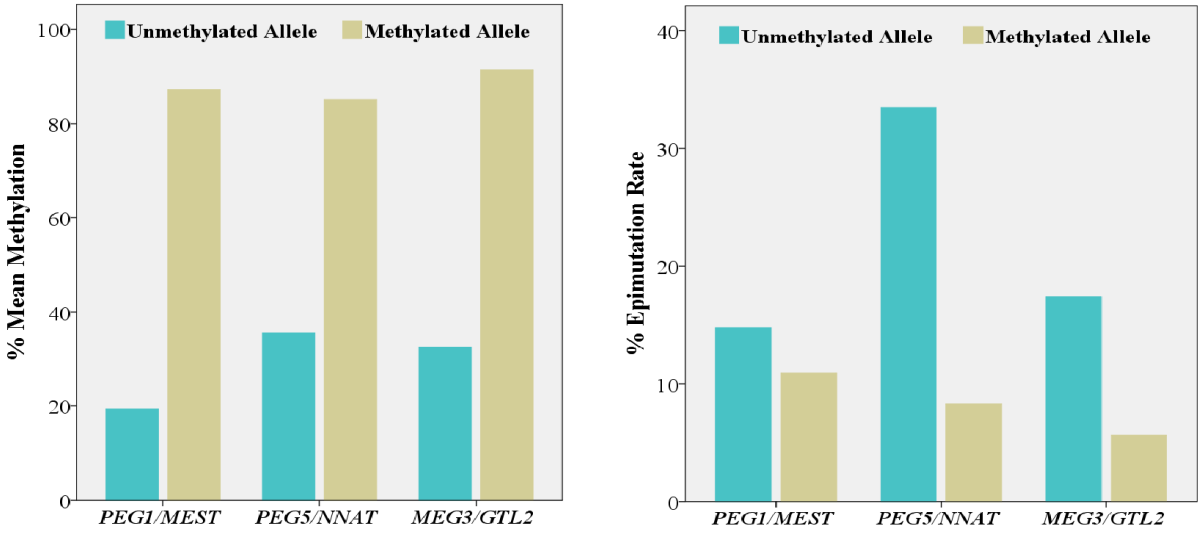


Figure 40. Clustered bar charts showing mean methylation (left) and epimutation rate (right) across the genes PEG1, PEG5, and MEG3 between the unmethylated and methylated alleles

Unmethylated alleles (light blue bars) display more methylation than the theoretical 0%, in all the genes (left diagram). The epimutation rate for the unmethylated allele (light blue bars) is significantly higher than methylated allele (right diagram). The technique used was deep bisulphite sequencing on Illumina MiSeq platform and the samples analyzed were heterozygous fetal cord blood samples.

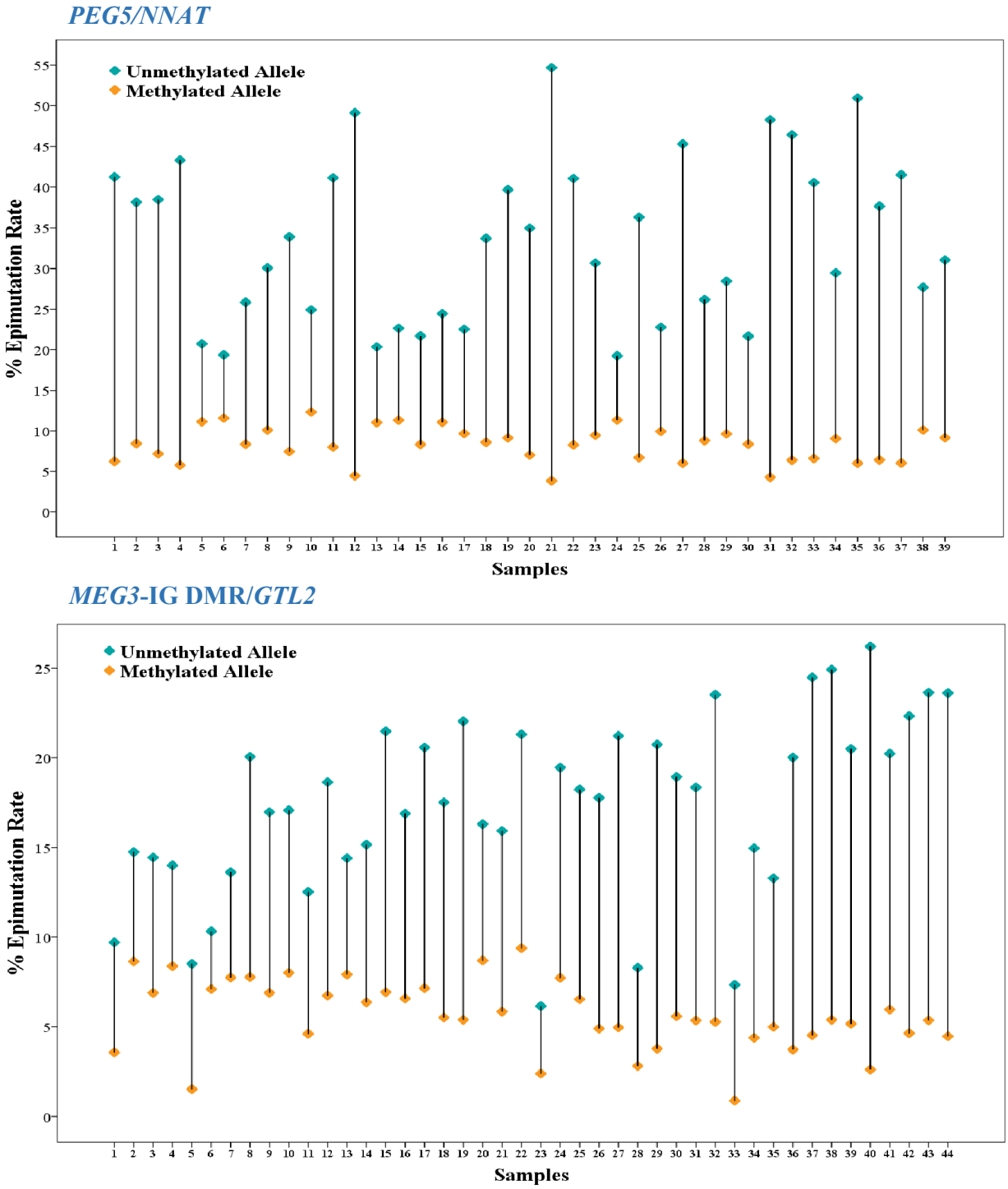


Figure 41. Percentage of epimutation rate between methylated (orange) and unmethylated alleles (blue) of oppositely imprinted genes PEG5 and MEG3 across individual FCB sample

Note that the epimutation rate of unmethylated (paternal allele in *PEG5* and maternal allele in *MEG3*) allele is constantly higher than the methylated allele in both the amplicons.

4 Discussion

4.1 Hypermethylation of repetitive DNA in sperm with aging and impact on next generation

Repetitive sequences approximately comprise 52% of all CpG sites in the human genome [134]. Since these regions are highly methylated [37], repeat DNA is generally used as markers for alterations in global DNA methylation [88]. The methylation status of repeats in adult cells is associated with maintenance of genome stability [135]. However, studies showed that sperm cells exhibit lower global DNA methylation levels compared to somatic cells [136]. DNA methylation signatures differ between various types of repeats in response to a wide range of processes, namely genome-wide demethylation after fertilization [17], gametogenesis [137], and knock-down of epigenetic modifiers [138]. In line with this, we have observed different levels of methylation values in repetitive DNA analyzed in sperm samples ranging from ~5-25% in rDNA, ~10-60% in alpha satellite, ~20-25% in ALU, and ~30-80% in LINE1 (figure 17,18,19, and 20). This variability among repeats was proposed to contribute to phenotypic plasticity [139].

Changes in sperm rDNA region's methylation with aging in humans and in bulls

This work demonstrated a significant positive correlation between ribosomal DNA (rDNA) methylation in sperm and male aging. In humans, it was detected at the promoter, upstream core element, 18S, and 28S regions of rDNA. In bovine sperm cohort, the methylation of 18S and 28S regions of rDNA correlated with aging within individual and across different age-grouped bulls.

During the aging process, alterations in the biology of nucleolus and ribosomes have been proposed [140, 141]. This dysfunctional regulation influences protein synthesis and thus has implications in aging. Previous research revealed that age-dependent methylation changes in ribosomal genes occurred in somatic tissues of mice which resulted in a loss of rDNA function [142]. Methylation of rDNA promoter was associated with unexpressed rRNA [143] depicting a negative influence of methylation on gene expression. Furthermore, hypermethylation of rDNA with aging was reported in sperm samples and in the liver of male rats [144]. Loss of rDNA function with aging was shown in human somatic tissues as well [145].

The methylation analysis of human rDNA region in this thesis was carried out both by pyrosequencing and deep bisulphite sequencing (DBS) techniques. While the former method detects mean methylation of all the molecules in a given sample at a particular site (CpG), the latter reveals an in-depth methylation analysis at individual molecule/allele (sperm) level. DBS additionally showed a significant group difference in rDNA methylation levels (figure 22 and 23) and epimutation rates (figure 25) in sperm samples between young males (25-35 years) and old males (>42 years) representing variation at single sperm level as well. Irrespective of the underlying genetic variant in the sequence (at least in the analyzed assays), male age correlated with sperm rDNA methylation (figure 22) in DBS analysis.

Male age calculator based on human sperm rDNA methylation

Several age estimators based on the levels of DNA methylation in somatic tissues were proposed by researchers. Giuliani *et al*. (2015) reported that the DNA methylation signatures of *ELOVL2*, *FHL2*, and *PENK* genomic regions correlated with age in human teeth, and constructed a mathematical model to predict the chronological age of the individuals [146]. Based on DNA methylation values, molecular age estimators termed as 'epigenetic clock(s)' were proposed. The most predominant ones are Horvath's clock [147] and Hannum's clock [148]. In spite of the significant difference in the epigenetic profiles amongst various somatic tissues, Horvath's clock is sufficiently consistent in a wide-range tissues/organ such as whole blood, peripheral blood mononuclear cells, cerebellar samples, occipital cortex, buccal epithelium, colon, adipose, liver, lung, saliva, and uterine cervix [147]. On the other hand, several chronological age prediction models based on the DNA methylation signatures in whole blood were reported [148-150].

For the first time, Lee *et al*. (2015) designed an age-predictive tool based on the DNA methylation signatures in semen sample [151]. Recently, Jenkins *et al*. (2017) also proposed a model to predict the chronological age of the donor using the methylation map in sperm [152]. The above two mentioned studies reported the models based on Illumina Human Methylation 450K Bead Chip array data. In contrast, the model proposed in the present work was based on methylation data using pyrosequencing of ribosomal DNA repeat (table 26) and could be highly cost-effective. Presently, we are making attempts to better fit the proposed model(s) by

considering the highly correlated CpGs (amongst others) using regression models. A possible application of such an age prediction tool based on sperm DNA methylation map might serve as an important tool in legal medicine. It can assist as a crucial investigative lead to narrow down the search range for unknown suspects in the case of crime incidents such as rape [153]. Finally, compared to the array-based epigenome-wide platform which needs complicated bioinformatics analysis and high-quality DNA, the age prediction models that are constructed using DNA methylation measurements from pyrosequencing seem less challenging and more promising to be easily incorporated into forensic procedures [153].

Alterations in methylation of other repeats in human and bull sperm with aging

In addition to rDNA regions, the methylation levels of alpha satellite and LINE1 repeats in human sperm significantly correlated with male aging in a positive direction. On the other hand, bovine alpha satellite and bovine testis satellite repeats were also hyper methylated in sperm with aging, within the individual and across different age-grouped bulls.

Since low methylation map of centromeric DNA was proposed as an epigenetic signature to differentiate germ cells from that of the somatic cell lineages [154], hypermethylation of satellite repeats in aging sperm can be hypothesized as a causative factor for loss of germ cell identity. Research showed that *Dnmt3l*-overexpressing cells exhibited hypermethylation in major and minor satellite repeat sequences in mouse [138]. As it was known that the DNMT3B and DNMT3L work together in the *de novo* methylation mechanism of centromeric satellite sequences [155], the ectopic expression of DNMT3L might have assisted DNMT3B to assess the satellite repeats for methylation. Whatsoever, hypermethylation of human and bovine alpha satellite repeat DNA in sperm of older males might lead to a more condensed state of chromatin compared to the sperm from young males, which might eventually have a considerable impact in the offspring. On the other hand, a previous study showed an increase in methylation of human sperm LINE1 repeats between young (<45 years) and old (>45 years) males [156] similar to the significant positive correlation seen in this work (table 25 and figure 20).

Potential mechanisms for methylation changes with aging

a) It is possible to assume that repeat DNA attracts *de novo* methylation which was detected in the case of retroviral elements [157] and transgenes [158]. The repetitive nature of these regions might make them preferable potential targets for *de novo* methylation when compared to the non-repetitive single copy genes.
b) The elevated methylation levels noticed in sperm of older men suggest that there might probably be inadequate time or availability of appropriate factors to assure complete epigenetic reprogramming mechanism.
c) Hypermethylation of repeats in sperm DNA during male aging could arise as a consequence of over and/or under activity of 'methylating' and 'demethylating' factors respectively. Research showed that the DNA methyltransferase levels increased in the age-related disease like cancer [159]. A similar analogy could possibly be applied to aging germ cells also.
d) In case of rDNA repeats, the upstream non-transcribed spacers were shown to be highly methylated [160] and might facilitate 5' to 3' spreading of methylation events which was also shown in age-dependent disease such as cancer [161].

Methylation levels of none of the studied repetitive elements in human sperm correlated with donor's body mass index. But, there seemed to be a minor influence of sperm concentration. As of now, the results presented in this project were not adjusted for confounding factors. Nevertheless, we believe that the significant correlation of sperm repeat DNA methylation with donor's age holds true since it was consistently detected in pyrosequencing analysis of human and bovine sperm cohorts, and also in NGS (deep bisulphite sequencing) data using human sperm. We are presently taking attempts to adjust for potential confounding factors (e.g., sperm concentration) using regression models.

Impact of rDNA hypermethylation in aging sperm on next generation embryos

Since rDNA codes for ribosomal RNA molecules which form an integral part of ribosomes, it can be hypothesized that the hypermethylation pattern observed in this region with aging sperm might have a convincing detrimental effect on the developing embryo of the next generation. Upon

plotting fathers' sperm rDNA methylation levels against human blastocyst stage embryo grading information, a trend for highly methylated sperm resulting in 'poorly' graded embryo was noticed (figure 30). Though the data distribution was extremely skewed with many 'good' compared to 'poorly' graded embryos, these preliminary results in humans provided an initial hint that hypermethylation in rDNA regions in sperm with aging could have a negative impact on the next generation embryos. Because of the ethical concerns involved in research with human embryos, we plan to carry out further work using a bovine model.

We were successful in acquiring the semen samples from 15 different bulls, out of which 12 individual bulls had their semen samples collected at 2 or 3 different ages within their lifetime (table 28). Bovine sperm cohort is advantageous for analyzing aging effects on sperm (within the same bull), given that the underlying inter-individual differences at the genetic and/or epigenetic level could be nullified. This is especially important when performing functional studies on F1 embryos obtained via *in vitro* fertilization of oocytes from one female with F0 sperm samples from a particular bull at different ages. This is indeed possible with the presently available bovine sperm cohort, and we are currently making attempts to carry out experimentation in this direction.

4.2 Male BMI correlation with sperm and next generation cord blood DNA methylation

Past work demonstrated that imprinted genes play a crucial role in a broad array of biological processes and diseases like obesity and diabetes mellitus [162]. The methylation marks at these regions are stably inherited through the germline ensuing mono-allelic expression in all somatic tissues [29]. In this regard, one of the projects of the present thesis focused on studying male body mass index (BMI) effects on sperm and next generation cord blood DNA methylation levels at eight imprinted differentially methylated regions (DMRs). In addition, *HIF3A* amplicon which was shown to be strongly associated with obesity [71] was also studied.

It is very vital to differentiate the methylation errors at single CpG sites from that of the entire cis-regulatory region. The presence of a single or a few methylated CpGs within an entire unmethylated promoter region most likely represents either a stochastic biological phenomenon

without any functional outcome or a technical fault during the bisulphite conversion step [163]. Even though single CpG methylation at some regions of the genome, i.e. in transcription factor binding site, might regulate the gene expression [164], it is generally the density of CpG methylation rather than individual CpGs in a cis-regulatory region that turns a gene "off" or "on". In this context, we have systematically analyzed the methylation state over several CpG sites for each amplicon and considered their (many CpG sites) mean for analysis.

MEG3 intergenic differentially methylated region (IG DMR)

After adjusting for possible confounding factors like male age and sperm concentration using regression beta models, the analysis showed that the methylation levels of paternally imprinted *MEG3*-IG DMR in sperm significantly correlated with donor's BMI (table 34). This amplicon displayed the same directional (as that of the sperm) significant correlation between the next generation male fetal cord blood (FCB) DNA methylation and paternal BMI (table 36), supporting the hypothesis of potential epigenetic inheritance at imprinted regions. *MEG3*-IG DMR is positioned at the intergenic region between the paternally expressed *PEG9*/*DLK1* and maternally expressed *MEG3*/*GTL2* (figure 4). This DMR was shown to serve as the imprinting control region (ICR) for the *DLK1*/*GTL2* cluster and governs imprinting and gene expression during development [165]. Hypermethylation of *MEG3* DMR in pancreatic islets of Type 2 diabetes mellitus patients was strongly associated with dramatic downregulation of microRNAs in this gene cluster [117]. Functional analysis of this gene knockdown promoted adipogenic differentiation [166].

HIF3A

This study identified that the methylation signatures of obesity-related *HIF3A* gene in sperm displayed a trend towards positive significance with donor's BMI (table 34) and also a similar significant positive association between methylation of male FCB DNA and paternal BMI (table 36). Previously, a large epigenome-wide association study of DNA methylation in whole blood and adipose tissue identified a significant correlation between BMI and methylation at three CpG sites in the first intron of *HIF3A* [71]. Several studies replicated the aforementioned association at

the *HIF3A* locus across different cohorts including umbilical cord DNA of newborns [167-169]. Results from one study supported the idea that HIF (hypoxia-inducible factor) pathways play a crucial part in the development of adipose tissue dysfunction in obesity [170]. Altogether, these findings along with our result indicate that *HIF3A* plays an important role in the etiology of obesity, and possible epigenetic transmission might occur to the next generation at this locus.

Methylation of IGF2 and NNAT in FCB samples

After adjusting for potential confounding factors such as birth weight, gestational week of the child, father's age, mother's age and BMI, we identified a significant negative correlation between female FCB DNA methylation and paternal BMI in *IGF2* amplicon (table 36). Same directional significant association between *IGF2* hypomethylation and paternal obesity was previously detected in neonates of Newborn Epigenetics Study cohort [171]. Additionally, an association between *IGF2* methylation and lipid profile in obese children was previously reported [172]. Circulating *IGF2* levels were used to predict weight gain and obesity development in humans [173]. This gene's methylation was shown to be a modulator of newborn's development and growth [174]. On the other hand, the methylation values of *PEG5*/*NNAT* gene in male FCB samples exhibited a trend towards significant correlation with paternal BMI. One study demonstrated an association between SNPs in human *PEG5* gene and severe childhood and adult obesity [111], while another work showed a similar correlation between paternal pre-conceptional obesity and *PEG5* gene's methylation in umbilical cord blood leukocytes of newborn children [175]. Nevertheless, the observed paternal BMI effects on FCB DNA methylation of these genes were of small effect size similar to the previous studies [171, 175].

Gender-specific effects in next generation FCB samples; Correlation with maternal BMI

Almost all the obtained significant genes displayed a correlation between paternal BMI and FCB DNA methylation in a gender-specific manner (figure 37). Such sex-dependent effects are in line with several previously published studies. Paternal BMI was shown to be associated with the growth of the male but not female offspring [78]. Here, authors displayed a correlation between father's body mass index and birth parameters such as birth weight, head and abdominal

circumference, pectoral and abdominal diameter, and cortisol levels in newborn males. Huypens *et al*. (2016) reported sex-dependent effects on insulin levels and body-weight gains in mice fathered by males fed high-fat diet [176]. In rodents, Ng SF *et al*. (2010) showed that chronic high-fat diet exposure in F0 father's generation programmed β cell dysfunction in F1 female rat offsprings [177].

Fetal cord blood DNA methylation signatures in none of the studied imprinted (and non-imprinted *HIF3A*) genes displayed a significant correlation with maternal BMI values both in pyro (table 37) and deep bisulphite sequencing (section 3.7.2.1) experiments. This might potentially indicate that paternal obesity is affecting methylation of the studied genes in cord blood DNA by a greater extent when compared to maternal BMI.

Limitations

Handling somatic contamination in sperm samples

The obtained swim-up sperm fraction for each sample was purified using PureSperm two-phase density gradient protocol to get rid of contamination from bacteria or somatic cells. Despite theoretically considering 'pure sperm fraction' for experimentation, there were some samples which displayed abnormal (e.g., higher values for paternally expressed genes in sperm) methylation levels across all/many studied imprinted genes. All those samples, which might have still probably contained a tiny fraction of somatic cells, were eliminated prior to the analysis step. Therefore, the attained correlations between donor's BMI and human sperm DNA methylation were the results from the analysis of pure sperm samples.

Adjusting for confounding factors like sperm concentration and male aging

Methylation aberrations at imprinted genes in sperm were frequently associated with male infertility and low-quality spermiogram parameters [178, 179]. This is one of the limitations related to this work because the considered human sperm samples (cohort_1 and cohort_2) for methylation analysis of imprinted genes were collected at a fertility center. Male age and sperm

concentration were the two strong confounding factors in our analysis of methylation data in imprinted genes. However, we have corrected for these traits using multivariate beta regression models and studied methylation maps in a considerable number of samples (combining both cohort_1 and cohort_2; in total 294 sperm samples). This allowed us to identify minor methylation changes in imprinted genes associated with obesity in sperm.

Small FCB sample number in NGS run

The measurements by pyrosequencing of fetal cord blood DNA represent the average methylation values from both maternal and paternal alleles. Therefore, in order to separately delineate paternal versus maternal alleles' DNA methylation in somatic cord blood cells of the offspring, next generation sequencing technology (NGS) on Illumina MiSeq Platform was performed. Heterozygous FCB samples were preselected based on genotyping of the corresponding imprinted gene, and paternal allele's methylation values in FCB were correlated with paternal body mass index.

Here, the paternal allele's DNA methylation levels of *PEG1*/*MEST* gene in FCB significantly correlated with paternal BMI (table 39). This gene was shown to be a negative regulator of human adipogenesis [180] where the authors displayed that the knockdown of *PEG1* enhanced adipocyte differentiation in humans while overexpression impaired adipogenesis. *MEST* is also known to be a primary candidate gene for developmental programming of metabolic phenotypes. On the other hand, the allele-specific methylation analysis of other genes did not reach significance in detecting paternal BMI effects on FCB samples, probably owing to the small sample size per amplicon in NGS run. Increasing the number of heterozygous FCB samples, which in turn depends on the minor allele frequency (MAF) of the available SNP in the region of interest, might help us evaluate if paternal BMI influences paternal allele's methylation in *MEG3*-IG DMR, *HIF3A*, *IGF2*, and *NNAT* amplicons as well.

4.3 SNP haplotypes influencing methylation signatures

Deep bisulphite sequencing (DBS) of amplicons containing a single nucleotide polymorphism or genetic variant enables us to differentiate between the methylation patterns of parental alleles in diploid somatic tissues of the next generation offspring. In case of imprinted genes, 5 out of 6 regions displayed significant cis-regulatory sequence polymorphism effects (table 41). The remarkable effect of SNP rs3778859 in the *MEST*/*PEG1* promoter (figure 39 upper panel) showed that the A (adenine) genetic variant of SNP haplotype was significantly correlated with an increased (>3% points; table 41 estimate) methylation map of imprinted maternal and a decreased (>10% points) methylation values of the non-imprinted paternal allele. For this amplicon, the average methylation difference between parental AA alleles was approximately 15% higher than GG alleles. All the studied imprinted genes except for the *MEG3* amplicon (figure 39 lower panel) exhibited this strong SNP effect. This effect was not only observed in case of imprinted genes but also in assay 1 of DBS run in ribosomal DNA repeat. Here, DNA sequences with the A genetic variant exhibited consistently more methylation values than that of the ones bearing G (guanine) genetic variant across all the analyzed samples. (figure 24). Thus, it is important to consider the effects of cis-regulatory SNP haplotypes when measuring methylation levels.

Previous studies demonstrated that allele-specific methylation (ASM) was largely determined by cis-acting polymorphisms [181]. Kerkel *et al*. (2008) showed that genotype-dependent ASM was associated with allele-specific expression at different loci in the human genome [182]. Genome-wide analysis revealed that ASM was a widespread phenomenon across the genome, and a spectrum of it exhibited between-individual heterogeneity [183, 184]. A subtle skewing of DNA methylation between alleles for many non-imprinted genes was also reported [185]. Overall, our deep bisulphite sequencing data also supports allele-specific methylation phenomenon which could be widely explained by the underlying genetic variation.

4.4 Allele-specific epimutations

Because DBS generates allele-specific methylation information on thousands of individual DNA molecules, it enables one to determine allele-specific epimutation rates. Imprinted genes in the diploid somatic tissue of offspring contain one methylated (imprinted) and one unmethylated (non-imprinted) parental allele. Epimutations are alleles exhibiting >50% aberrantly (de) methylated CpG sites.

Hypermethylation of non-imprinted allele (HNA)

A new phenomenon named hypermethylation of the non-imprinted allele (HNA) was recently reported in our lab in *PEG1*/*MEST* and *MEG3*-IG DMR amplicons [133]. It demonstrated that the unmethylated non-imprinted allele was more prone to epimutations than the methylated imprinted allele irrespective of the parent of origin of the imprinted gene. The analysis using DBS in this present work confirmed the previously published results and additionally detected the similar phenomenon in the *NNAT*/*PEG5* amplicon (table 42, figure 40 and 41). In all the studied samples of two oppositely imprinted genes (figure 41), the methylation levels of the non-imprinted unmethylated allele, which was theoretically expected to be 0%, displayed a considerable hypermethylation (e.g., 36% for *PEG5* and 33% for *MEG3*; table 42).

Since imprinted genes are functionally haploid and they resist the genome-wide epigenetic reprogramming after fertilization, they might particularly be prone to adverse environmental conditions during early development. The non-imprinted allele should be protected effectively against the *de novo* methylation mechanism so that the functionally important parent-specific imprinting patterns can be maintained during several cell divisions in the developing embryo. Previous studies showed that the differential methylation was retained by binding of zinc finger proteins to non-imprinted allele [186, 187]. Histone proteins such as H3K4me2 and H3K4me3 are enriched on the non-imprinted allele which might also help prevent the action of DNA methyltransferases. Ten-eleven translocation (TET) proteins are involved in enzymatically removing the aberrantly captured methylation signatures [188]. Overall, an interplay of various mechanisms is needed for proper maintenance of the unmethylated state at the non-imprinted

allele. During somatic development, stochastic errors in this complex process could lead to hypermethylation of non-imprinted unmethylated allele. Research demonstrated that during post-natal growth cessation, the network of imprinted genes that regulates mammalian somatic growth, becomes transcriptionally down-regulated in many organs [189]. HNA which leads to biallelic methylation (somatic mosaicism) might be part of a physiological process for limiting the growth and development and contributing to the complexity of phenotypic diversity in mammals [133].

Loss of imprinting at IGF2-H19 locus (Hypomethylation of the imprinted allele)

In contrast to the previously described HNA effect, for *H19*-IG DMR and *IGF2* DMR0 regions, the imprinted methylated allele displayed more epimutations than that of the non-imprinted unmethylated allele. In other words, the methylation levels of the imprinted methylated allele, which was theoretically expected to be 100%, exhibited a considerable hypomethylation at both the DMRs. The difference between allele-specific epimutation rates was much higher for the primary *H19*-IG DMR than for the secondary *IGF2* DMR0.

One previous example of loss of imprinting (LOI) at the transcriptional levels was the biallelic expression of *IGF2* in blood samples from normal individuals [190, 191]. Colorectal cancer patients showed several times more frequent biallelic *IGF2* expression in colonic mucosa and blood when compared to healthy individuals [192-194], supporting the hypothesis that LOI predisposes to cancer. Hypomethylation of DMR0 was observed in Wilms tumor as well [195]. Previous research showed that the *H19*-IG DMR influences methylation at the *IGF2* DMR0 in cis [196]. Increased susceptibility of the imprinted *H19-IGF2* allele to epimutations may be the mechanism underlying biallelic IGF2 expression in 10-20% of blood samples from normal healthy individuals [190, 191].

4.5 Future perspectives

In spite of male germ cells being in a continuous state of renewal, results from this thesis evidently demonstrated that DNA methylation in spermatozoa was vulnerable to adverse BMI and aging-related aberrations. In this regard, some of the future perspectives of this work are as follows:

1) Future research is highly encouraged to decode if the observed alterations in mature spermatozoa also exist in the spermatogonial stem cell population and/or if a change in the testicular environment is accountable for these modifications in sperm DNA methylation.

2) The copy number of repetitive DNA sequences is highly variable among individuals. Therefore, we are making efforts to check if sperm DNA methylation during aging correlates with repeat copy number. We hypothesize that there might not be a major change in the number of copies of the repeats with the aging of an individual. However, aging could have induced DNA methylation on many more copies in the repeats.

3) To check if the methylation changes in repetitive elements in sperm during aging would impact the next generation embryos at a functional level, we plan to carry out functional studies on bovine embryos obtained via ICSI using sperm from young vs old bulls. Silver staining of nucleolar organized regions (in case of rDNA), RT PCR and RNA sequencing procedures will be used to compare embryos fathered from sperm with low and high repeat DNA methylation.

4) Finally, longitudinal follow-up research studies are to be carried out in order to elucidate if the effects on the next generation offspring can influence the possibility of developing aberrant phenotype during adulthood.

5 References

1. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
2. Gudbjartsson, D.F., et al., *Many sequence variants affecting diversity of adult human height.* Nat Genet, 2008. **40**(5): p. 609-15.
3. Weedon, M.N., et al., *Genome-wide association analysis identifies 20 loci that influence adult height.* Nat Genet, 2008. **40**(5): p. 575-83.
4. Lettre, G., et al., *Identification of ten loci associated with height highlights new biological pathways in human growth.* Nat Genet, 2008. **40**(5): p. 584-91.
5. Simeone, P. and S. Alberti, *Epigenetic heredity of human height.* Physiol Rep, 2014. **2**(6).
6. Trerotola, M., et al., *Epigenetic inheritance and the missing heritability.* Hum Genomics, 2015. **9**: p. 17.
7. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
8. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
9. Jones, P.A. and D. Takai, *The role of DNA methylation in mammalian epigenetics.* Science, 2001. **293**(5532): p. 1068-70.
10. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
11. Goll, M.G. and T.H. Bestor, *Eukaryotic cytosine methyltransferases.* Annu Rev Biochem, 2005. **74**: p. 481-514.
12. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mech Dev, 2002. **117**(1-2): p. 15-23.
13. Messerschmidt, D.M., B.B. Knowles, and D. Solter, *DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos.* Genes Dev, 2014. **28**(8): p. 812-28.
14. Sasaki, H. and Y. Matsui, *Epigenetic events in mammalian germ-cell development: reprogramming and beyond.* Nat Rev Genet, 2008. **9**(2): p. 129-40.
15. Smallwood, S.A. and G. Kelsey, *De novo DNA methylation: a germ cell perspective.* Trends Genet, 2012. **28**(1): p. 33-42.
16. Mayer, W., et al., *Demethylation of the zygotic paternal genome.* Nature, 2000. **403**(6769): p. 501-2.
17. Lane, N., et al., *Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse.* Genesis, 2003. **35**(2): p. 88-93.
18. Fresard, L., et al., *Epigenetics and phenotypic variability: some interesting insights from birds.* Genet Sel Evol, 2013. **45**: p. 16.
19. Gillman, M.W., *Developmental origins of health and disease.* N Engl J Med, 2005. **353**(17): p. 1848-50.
20. Barker, D.J., *The origins of the developmental origins theory.* J Intern Med, 2007. **261**(5): p. 412-7.
21. Painter, R.C., et al., *Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life.* BJOG, 2008. **115**(10): p. 1243-9.

22. Stein, A.D. and L.H. Lumey, *The relationship between maternal and offspring birth weights after maternal prenatal famine exposure: the Dutch Famine Birth Cohort Study.* Hum Biol, 2000. **72**(4): p. 641-54.
23. Schieve, L.A., et al., *Low and very low birth weight in infants conceived with use of assisted reproductive technology.* N Engl J Med, 2002. **346**(10): p. 731-7.
24. El Hajj, N., et al., *Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus.* Diabetes, 2013. **62**(4): p. 1320-8.
25. Wei, Y., H. Schatten, and Q.Y. Sun, *Environmental epigenetic inheritance through gametes and implications for human reproduction.* Hum Reprod Update, 2015. **21**(2): p. 194-208.
26. Blake, G.E. and E.D. Watson, *Unravelling the complex mechanisms of transgenerational epigenetic inheritance.* Curr Opin Chem Biol, 2016. **33**: p. 101-7.
27. Hammoud, S.S., et al., *Distinctive chromatin in human sperm packages genes for embryo development.* Nature, 2009. **460**(7254): p. 473-8.
28. Wilkins, J.F., F. Ubeda, and J. Van Cleve, *The evolving landscape of imprinted genes in humans and mice: Conflict among alleles, genes, tissues, and kin.* Bioessays, 2016. **38**(5): p. 482-9.
29. Reik, W. and J. Walter, *Genomic imprinting: parental influence on the genome.* Nat Rev Genet, 2001. **2**(1): p. 21-32.
30. Smith, F.M., A.S. Garfield, and A. Ward, *Regulation of growth and metabolism by imprinted genes.* Cytogenet Genome Res, 2006. **113**(1-4): p. 279-91.
31. Bartolomei, M.S. and A.C. Ferguson-Smith, *Mammalian genomic imprinting.* Cold Spring Harb Perspect Biol, 2011. **3**(7).
32. Moore, T. and D. Haig, *Genomic imprinting in mammalian development: a parental tug-of-war.* Trends Genet, 1991. **7**(2): p. 45-9.
33. Ideraabdullah, F.Y., S. Vigneau, and M.S. Bartolomei, *Genomic imprinting mechanisms in mammals.* Mutat Res, 2008. **647**(1-2): p. 77-85.
34. Kameswaran, V. and K.H. Kaestner, *The Missing lnc(RNA) between the pancreatic beta-cell and diabetes.* Front Genet, 2014. **5**: p. 200.
35. Singer, M.F., *SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes.* Cell, 1982. **28**(3): p. 433-4.
36. Rubin, C.M., et al., *Alu repeated DNAs are differentially methylated in primate germ cells.* Nucleic Acids Res, 1994. **22**(23): p. 5121-7.
37. Bestor, T.H., *The host defence function of genomic methylation patterns.* Novartis Found Symp, 1998. **214**: p. 187-95; discussion 195-9, 228-32.
38. Marques, A.H., et al., *The influence of maternal prenatal and early childhood nutrition and maternal prenatal stress on offspring immune system development and neurodevelopmental disorders.* Front Neurosci, 2013. **7**: p. 120.
39. Geraghty, A.A., et al., *Nutrition During Pregnancy Impacts Offspring's Epigenetic Status-Evidence from Human and Animal Studies.* Nutr Metab Insights, 2015. **8**(Suppl 1): p. 41-7.
40. Champagne, F.A., *Epigenetic mechanisms and the transgenerational effects of maternal care.* Front Neuroendocrinol, 2008. **29**(3): p. 386-97.
41. Jenkins, T.G. and D.T. Carrell, *The sperm epigenome and potential implications for the developing embryo.* Reproduction, 2012. **143**(6): p. 727-34.

42. Tanphaichitr, N., et al., *Basic nuclear proteins in testicular cells and ejaculated spermatozoa in man.* Exp Cell Res, 1978. **117**(2): p. 347-56.
43. Chen, Q., W. Yan, and E. Duan, *Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications.* Nat Rev Genet, 2016. **17**(12): p. 733-743.
44. Nugent, D. and A.H. Balen, *The effects of female age on fecundity and pregnancy outcome.* Hum Fertil (Camb), 2001. **4**(1): p. 43-8.
45. de la Rochebrochard, E. and P. Thonneau, *Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study.* Hum Reprod, 2002. **17**(6): p. 1649-56.
46. Frans, E.M., et al., *Advancing paternal age and bipolar disorder.* Arch Gen Psychiatry, 2008. **65**(9): p. 1034-40.
47. Miller, B., et al., *Advanced paternal age and parental history of schizophrenia.* Schizophr Res, 2011. **133**(1-3): p. 125-32.
48. Smith, R.G., et al., *Advancing paternal age is associated with deficits in social and exploratory behaviors in the offspring: a mouse model.* PLoS One, 2009. **4**(12): p. e8456.
49. Auroux, M., *Decrease of learning capacity in offspring with increasing paternal age in the rat.* Teratology, 1983. **27**(2): p. 141-8.
50. Wilson, V.L. and P.A. Jones, *DNA methylation decreases in aging but not in immortal cells.* Science, 1983. **220**(4601): p. 1055-7.
51. Maegawa, S., et al., *Widespread and tissue specific age-related DNA methylation changes in mice.* Genome Res, 2010. **20**(3): p. 332-40.
52. Thompson, R.F., et al., *Tissue-specific dysregulation of DNA methylation in aging.* Aging Cell, 2010. **9**(4): p. 506-18.
53. Crow, J.F., *The origins, patterns and implications of human spontaneous mutation.* Nat Rev Genet, 2000. **1**(1): p. 40-7.
54. Dada, R., et al., *Epigenetics and its role in male infertility.* J Assist Reprod Genet, 2012. **29**(3): p. 213-23.
55. Goriely, A. and A.O. Wilkie, *Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease.* Am J Hum Genet, 2012. **90**(2): p. 175-200.
56. Goriely, A., et al., *"Selfish spermatogonial selection": a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders.* Am J Psychiatry, 2013. **170**(6): p. 599-608.
57. Risch, N., et al., *Spontaneous mutation and parental age in humans.* Am J Hum Genet, 1987. **41**(2): p. 218-48.
58. Hultman, C.M., et al., *Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies.* Mol Psychiatry, 2011. **16**(12): p. 1203-12.
59. Jones, M.J., S.J. Goodman, and M.S. Kobor, *DNA methylation and healthy human aging.* Aging Cell, 2015. **14**(6): p. 924-32.
60. Milekic, M.H., et al., *Age-related sperm DNA methylation changes are transmitted to offspring and associated with abnormal behavior and dysregulated gene expression.* Mol Psychiatry, 2015. **20**(8): p. 995-1001.
61. Smith, R.G., et al., *Advanced paternal age is associated with altered DNA methylation at brain-expressed imprinted loci in inbred mice: implications for neuropsychiatric disease.* Mol Psychiatry, 2013. **18**(6): p. 635-6.

62. *Obesity: preventing and managing the global epidemic. Report of a WHO consultation.* World Health Organ Tech Rep Ser, 2000. **894**: p. i-xii, 1-253.
63. van Vliet-Ostaptchouk, J.V., H. Snieder, and V. Lagou, *Gene-Lifestyle Interactions in Obesity.* Curr Nutr Rep, 2012. **1**: p. 184-196.
64. Drong, A.W., C.M. Lindgren, and M.I. McCarthy, *The genetic and epigenetic basis of type 2 diabetes and obesity.* Clin Pharmacol Ther, 2012. **92**(6): p. 707-15.
65. Albuquerque, D., et al., *The contribution of genetics and environment to obesity.* Br Med Bull, 2017. **123**(1): p. 159-173.
66. Herrera, B.M., S. Keildson, and C.M. Lindgren, *Genetics and epigenetics of obesity.* Maturitas, 2011. **69**(1): p. 41-9.
67. Elks, C.E., et al., *Variability in the heritability of body mass index: a systematic review and meta-regression.* Front Endocrinol (Lausanne), 2012. **3**: p. 29.
68. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
69. Ohta, T., et al., *Imprinting-mutation mechanisms in Prader-Willi syndrome.* Am J Hum Genet, 1999. **64**(2): p. 397-413.
70. Butler, M.G., *Prader-Willi Syndrome: Obesity due to Genomic Imprinting.* Curr Genomics, 2011. **12**(3): p. 204-15.
71. Dick, K.J., et al., *DNA methylation and body-mass index: a genome-wide analysis.* Lancet, 2014. **383**(9933): p. 1990-8.
72. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
73. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
74. Barres, R., et al., *Weight loss after gastric bypass surgery in human obesity remodels promoter methylation.* Cell Rep, 2013. **3**(4): p. 1020-7.
75. Barres, R., et al., *Acute exercise remodels promoter methylation in human skeletal muscle.* Cell Metab, 2012. **15**(3): p. 405-11.
76. Ronn, T., et al., *A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue.* PLoS Genet, 2013. **9**(6): p. e1003572.
77. van Dijk, S.J., et al., *Recent developments on the role of epigenetics in obesity and metabolic disease.* Clin Epigenetics, 2015. **7**: p. 66.
78. Chen, Y.P., et al., *Paternal body mass index (BMI) is associated with offspring intrauterine growth in a gender dependent manner.* PLoS One, 2012. **7**(5): p. e36329.
79. Freeman, E., et al., *Preventing and treating childhood obesity: time to target fathers.* Int J Obes (Lond), 2012. **36**(1): p. 12-5.
80. Kaati, G., L.O. Bygren, and S. Edvinsson, *Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period.* Eur J Hum Genet, 2002. **10**(11): p. 682-8.
81. Schagdarsurengin, U. and K. Steger, *Epigenetics in male reproduction: effect of paternal diet on sperm quality and offspring health.* Nat Rev Urol, 2016. **13**(10): p. 584-95.
82. Donkin, I., et al., *Obesity and Bariatric Surgery Drive Epigenetic Variation of Spermatozoa in Humans.* Cell Metab, 2016. **23**(2): p. 369-78.
83. de Castro Barbosa, T., et al., *High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring.* Mol Metab, 2016. **5**(3): p. 184-97.

84. Carone, B.R., et al., *Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals.* Cell, 2010. **143**(7): p. 1084-96.
85. Palmer, N.O., et al., *Impact of obesity on male fertility, sperm function and molecular composition.* Spermatogenesis, 2012. **2**(4): p. 253-263.
86. Soubry, A., et al., *Obesity-related DNA methylation at imprinted genes in human sperm: Results from the TIEGER study.* Clin Epigenetics, 2016. **8**: p. 51.
87. de Koning, A.P., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS Genet, 2011. **7**(12): p. e1002384.
88. Yang, A.S., et al., *A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements.* Nucleic Acids Res, 2004. **32**(3): p. e38.
89. Ostertag, E.M. and H.H. Kazazian, Jr., *Biology of mammalian L1 retrotransposons.* Annu Rev Genet, 2001. **35**: p. 501-38.
90. Eichler, D.C. and N. Craig, *Processing of eukaryotic ribosomal RNA.* Prog Nucleic Acid Res Mol Biol, 1994. **49**: p. 197-239.
91. Srivastava, A.K. and D. Schlessinger, *Structure and organization of ribosomal DNA.* Biochimie, 1991. **73**(6): p. 631-8.
92. Karahan, G., et al., *Relative expression of rRNA transcripts and 45S rDNA promoter methylation status are dysregulated in tumors in comparison with matched-normal tissues in breast cancer.* Oncol Rep, 2015. **33**(6): p. 3131-45.
93. Lee, C., et al., *Human centromeric DNAs.* Hum Genet, 1997. **100**(3-4): p. 291-304.
94. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
95. Schueler, M.G. and B.A. Sullivan, *Structural and functional dynamics of human centromeric chromatin.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 301-13.
96. Garavis, M., et al., *Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs.* Chemistry, 2015. **21**(27): p. 9816-24.
97. Kazazian, H.H., Jr. and J.L. Goodier, *LINE drive. retrotransposition and genome instability.* Cell, 2002. **110**(3): p. 277-80.
98. Batzer, M.A. and P.L. Deininger, *Alu repeats and human genomic diversity.* Nat Rev Genet, 2002. **3**(5): p. 370-9.
99. Hasler, J. and K. Strub, *Alu elements as regulators of gene expression.* Nucleic Acids Res, 2006. **34**(19): p. 5491-7.
100. Nishita, Y., et al., *Genomic imprinting and chromosomal localization of the human MEST gene.* Genomics, 1996. **36**(3): p. 539-42.
101. Kosaki, K., et al., *Isoform-specific imprinting of the human PEG1/MEST gene.* Am J Hum Genet, 2000. **66**(1): p. 309-12.
102. Takahashi, M., Y. Kamei, and O. Ezaki, *Mest/Peg1 imprinted gene enlarges adipocytes and is a marker of adipocyte size.* Am J Physiol Endocrinol Metab, 2005. **288**(1): p. E117-24.
103. Lefebvre, L., et al., *Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest.* Nat Genet, 1998. **20**(2): p. 163-9.
104. Kim, J., et al., *The human homolog of a mouse-imprinted gene, Peg3, maps to a zinc finger gene-rich region of human chromosome 19q13.4.* Genome Res, 1997. **7**(5): p. 532-40.
105. Curley, J.P., et al., *Increased body fat in mice with a targeted mutation of the paternally expressed imprinted gene Peg3.* FASEB J, 2005. **19**(10): p. 1302-4.

106. Kim, J., et al., *Peg3 mutational effects on reproduction and placenta-specific gene families.* PLoS One, 2013. **8**(12): p. e83359.
107. Horsthemke, B. and K. Buiting, *Imprinting defects on human chromosome 15.* Cytogenet Genome Res, 2006. **113**(1-4): p. 292-9.
108. Albuquerque, D., et al., *Polymorphisms in the SNRPN gene are associated with obesity susceptibility in a Spanish population.* J Gene Med, 2017. **19**(5).
109. Joe, M.K., et al., *Crucial roles of neuronatin in insulin secretion and high glucose-induced apoptosis in pancreatic beta-cells.* Cell Signal, 2008. **20**(5): p. 907-15.
110. Higuchi, H., T. Niki, and T. Shiiya, *Feeding behavior and gene expression of appetite-related neuropeptides in mice lacking for neuropeptide Y Y5 receptor subclass.* World J Gastroenterol, 2008. **14**(41): p. 6312-7.
111. Vrang, N., et al., *The imprinted gene neuronatin is regulated by metabolic status and associated with obesity.* Obesity (Silver Spring), 2010. **18**(7): p. 1289-96.
112. Ono, R., et al., *A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21.* Genomics, 2001. **73**(2): p. 232-7.
113. Ono, R., et al., *Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality.* Nat Genet, 2006. **38**(1): p. 101-6.
114. Hishida, T., et al., *peg10, an imprinted gene, plays a crucial role in adipocyte differentiation.* FEBS Lett, 2007. **581**(22): p. 4272-8.
115. Zhang, X., et al., *Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions.* Endocrinology, 2010. **151**(3): p. 939-47.
116. Gejman, R., et al., *Selective loss of MEG3 expression and intergenic differentially methylated region hypermethylation in the MEG3/DLK1 locus in human clinically nonfunctioning pituitary adenomas.* J Clin Endocrinol Metab, 2008. **93**(10): p. 4119-25.
117. Kameswaran, V., et al., *Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets.* Cell Metab, 2014. **19**(1): p. 135-45.
118. Leighton, P.A., et al., *Disruption of imprinting caused by deletion of the H19 gene region in mice.* Nature, 1995. **375**(6526): p. 34-9.
119. Reik, W., et al., *Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by altered imprinting pattern in the IGF2-H19 domain.* Hum Mol Genet, 1995. **4**(12): p. 2379-85.
120. Zeschnigk, M., et al., *IGF2/H19 hypomethylation in Silver-Russell syndrome and isolated hemihypoplasia.* Eur J Hum Genet, 2008. **16**(3): p. 328-34.
121. Hernandez-Valero, M.A., et al., *Interplay between polymorphisms and methylation in the H19/IGF2 gene region may contribute to obesity in Mexican-American children.* J Dev Orig Health Dis, 2013. **4**(6): p. 499-506.
122. Chao, W. and P.A. D'Amore, *IGF2: epigenetic regulation and role in development and disease.* Cytokine Growth Factor Rev, 2008. **19**(2): p. 111-20.
123. Casellas, A., et al., *Insulin-like Growth Factor 2 Overexpression Induces beta-Cell Dysfunction and Increases Beta-cell Susceptibility to Damage.* J Biol Chem, 2015. **290**(27): p. 16772-85.
124. Bai, B., et al., *Different epigenetic alterations are associated with abnormal IGF2/Igf2 upregulation in neural tube defects.* PLoS One, 2014. **9**(11): p. e113308.
125. Drevytska, T., et al., *HIF-3alpha mRNA expression changes in different tissues and their role in adaptation to intermittent hypoxia and physical exercise.* Pathophysiology, 2012. **19**(3): p. 205-14.

126. Hatanaka, M., et al., *Hypoxia-inducible factor-3alpha functions as an accelerator of 3T3-L1 adipose differentiation.* Biol Pharm Bull, 2009. **32**(7): p. 1166-72.
127. Ronaghi, M., *Pyrosequencing sheds light on DNA sequencing.* Genome Res, 2001. **11**(1): p. 3-11.
128. Bernstein, D.L., et al., *The BisPCR(2) method for targeted bisulfite sequencing.* Epigenetics Chromatin, 2015. **8**: p. 27.
129. Rahmann, S., et al., *Amplikyzer: Automated methylation analysis of amplicons from bisulfite flowgram sequencing.* PeerJ PrePrints, 2013. **1**: p. e122v2.
130. Klug, M. and M. Rehli, *Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector.* Epigenetics, 2006. **1**(3): p. 127-30.
131. Atsem, S., et al., *Paternal age effects on sperm FOXK1 and KCNA7 methylation and transmission into the next generation.* Hum Mol Genet, 2016. **25**(22): p. 4996-5005.
132. Arnold, S.J. and E.J. Robertson, *Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo.* Nat Rev Mol Cell Biol, 2009. **10**(2): p. 91-103.
133. Haertle, L., et al., *Hypermethylation of the non-imprinted maternal MEG3 and paternal MEST alleles is highly variable among normal individuals.* PLoS One, 2017. **12**(8): p. e0184030.
134. Su, J., et al., *Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts.* Genomics, 2012. **99**(1): p. 10-7.
135. Putiri, E.L. and K.D. Robertson, *Epigenetic mechanisms and genome stability.* Clin Epigenetics, 2011. **2**(2): p. 299-314.
136. Monk, M., M. Boubelik, and S. Lehnert, *Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development.* Development, 1987. **99**(3): p. 371-82.
137. Sanford, J.P., et al., *Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse.* Genes Dev, 1987. **1**(10): p. 1039-46.
138. Takashima, S., et al., *Abnormal DNA methyltransferase expression in mouse germline stem cells results in spermatogenic defects.* Biol Reprod, 2009. **81**(1): p. 155-64.
139. Sharif, J., Y. Shinkai, and H. Koseki, *Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs?* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1609): p. 20110340.
140. Rattan, S.I., *Synthesis, modifications, and turnover of proteins during aging.* Exp Gerontol, 1996. **31**(1-2): p. 33-47.
141. Comai, L., *The nucleolus: a paradigm for cell proliferation and aging.* Braz J Med Biol Res, 1999. **32**(12): p. 1473-8.
142. Swisshelm, K., et al., *Age-related increase in methylation of ribosomal genes and inactivation of chromosome-specific rRNA gene clusters in mouse.* Mutat Res, 1990. **237**(3-4): p. 131-46.
143. Santoro, R. and I. Grummt, *Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription.* Mol Cell, 2001. **8**(3): p. 719-25.
144. Oakes, C.C., et al., *Aging results in hypermethylation of ribosomal DNA in sperm and liver of male rats.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1775-80.
145. Thomas, S. and A.B. Mukherjee, *A longitudinal study of human age-related ribosomal RNA gene activity as detected by silver-stained NORs.* Mech Ageing Dev, 1996. **92**(2-3): p. 101-9.

146. Giuliani, C., et al., *Inferring chronological age from DNA methylation patterns of human teeth.* Am J Phys Anthropol, 2016. **159**(4): p. 585-95.
147. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.
148. Hannum, G., et al., *Genome-wide methylation profiles reveal quantitative views of human aging rates.* Mol Cell, 2013. **49**(2): p. 359-367.
149. Naue, J., et al., *Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression.* Forensic Sci Int Genet, 2017. **31**: p. 19-28.
150. Vidaki, A., et al., *DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing.* Forensic Sci Int Genet, 2017. **28**: p. 225-236.
151. Lee, H.Y., et al., *Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study.* Forensic Sci Int Genet, 2015. **19**: p. 28-34.
152. Jenkins, T.G., et al., *Paternal germ line aging: DNA methylation age prediction from human sperm.* bioRxiv, 2018.
153. Sang-Eun, J., et al., *DNA methylation-based age prediction from various tissues and body fluids.* BMB Reports, 2017. **50**(11): p. 546-553.
154. Yamagata, K., et al., *Centromeric DNA hypomethylation as an epigenetic signature discriminates between germ and somatic cell lineages.* Dev Biol, 2007. **312**(1): p. 419-26.
155. Kato, Y., et al., *Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse.* Hum Mol Genet, 2007. **16**(19): p. 2272-80.
156. Jenkins, T.G., et al., *Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility.* PLoS Genet, 2014. **10**(7): p. e1004458.
157. Arnaud, P., et al., *SINE retroposons can be used in vivo as nucleation centers for de novo methylation.* Mol Cell Biol, 2000. **20**(10): p. 3434-41.
158. McBurney, M.W., et al., *Reexpression of a cluster of silenced transgenes is associated with their rearrangement.* Genes Chromosomes Cancer, 2001. **32**(4): p. 311-23.
159. el-Deiry, W.S., et al., *High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer.* Proc Natl Acad Sci U S A, 1991. **88**(8): p. 3470-4.
160. Brock, G.J. and A. Bird, *Mosaic methylation of the repeat unit of the human ribosomal RNA genes.* Hum Mol Genet, 1997. **6**(3): p. 451-6.
161. Turker, M.S., *Gene silencing in mammalian cells and the spread of DNA methylation.* Oncogene, 2002. **21**(35): p. 5388-93.
162. Peters, J., *The role of genomic imprinting in biology and disease: an expanding view.* Nat Rev Genet, 2014. **15**(8): p. 517-30.
163. Hansmann, T., et al., *Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer.* Hum Mol Genet, 2012. **21**(21): p. 4669-79.
164. Griswold, M.D. and J.S. Kim, *Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription.* Biol Reprod, 2001. **64**(2): p. 602-10.
165. Takada, S., et al., *Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19.* Hum Mol Genet, 2002. **11**(1): p. 77-86.

166. Li, Z., et al., *Long non-coding RNA MEG3 inhibits adipogenesis and promotes osteogenesis of human adipose-derived mesenchymal stem cells via miR-140-5p.* Mol Cell Biochem, 2017. **433**(1-2): p. 51-60.
167. Pan, H., et al., *HIF3A association with adiposity: the story begins before birth.* Epigenomics, 2015. **7**(6): p. 937-50.
168. Demerath, E.W., et al., *Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci.* Hum Mol Genet, 2015. **24**(15): p. 4464-79.
169. Ronn, T., et al., *Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood.* Hum Mol Genet, 2015. **24**(13): p. 3792-813.
170. Pfeiffer, S., et al., *Hypoxia-inducible factor 3A gene expression and methylation in adipose tissue is related to adipose tissue dysfunction.* Sci Rep, 2016. **6**: p. 27969.
171. Soubry, A., et al., *Paternal obesity is associated with IGF2 hypomethylation in newborns: results from a Newborn Epigenetics Study (NEST) cohort.* BMC Med, 2013. **11**: p. 29.
172. Deodati, A., et al., *IGF2 methylation is associated with lipid profile in obese children.* Horm Res Paediatr, 2013. **79**(6): p. 361-7.
173. Sandhu, M.S., et al., *Low circulating IGF-II concentrations predict weight gain and obesity in humans.* Diabetes, 2003. **52**(6): p. 1403-8.
174. St-Pierre, J., et al., *IGF2 DNA methylation is a modulator of newborn's fetal growth and development.* Epigenetics, 2012. **7**(10): p. 1125-32.
175. Soubry, A., et al., *Newborns of obese parents have altered DNA methylation patterns at imprinted genes.* Int J Obes (Lond), 2015. **39**(4): p. 650-7.
176. Huypens, P., et al., *Epigenetic germline inheritance of diet-induced obesity and insulin resistance.* Nat Genet, 2016. **48**(5): p. 497-9.
177. Ng, S.F., et al., *Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring.* Nature, 2010. **467**(7318): p. 963-6.
178. El Hajj, N., et al., *Methylation status of imprinted genes and repetitive elements in sperm DNA from infertile males.* Sex Dev, 2011. **5**(2): p. 60-9.
179. Poplinski, A., et al., *Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1.* Int J Androl, 2010. **33**(4): p. 642-9.
180. Karbiener, M., et al., *Mesoderm-specific transcript (MEST) is a negative regulator of human adipocyte differentiation.* Int J Obes (Lond), 2015. **39**(12): p. 1733-41.
181. Schilling, E., C. El Chartouni, and M. Rehli, *Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences.* Genome Res, 2009. **19**(11): p. 2028-35.
182. Kerkel, K., et al., *Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation.* Nat Genet, 2008. **40**(7): p. 904-8.
183. Schalkwyk, L.C., et al., *Allelic skewing of DNA methylation is widespread across the genome.* Am J Hum Genet, 2010. **86**(2): p. 196-212.
184. Meaburn, E.L., L.C. Schalkwyk, and J. Mill, *Allele-specific methylation in the human genome: implications for genetic studies of complex disease.* Epigenetics, 2010. **5**(7): p. 578-82.
185. Cheung, W.A., et al., *Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome.* Genome Biol, 2017. **18**(1): p. 50.

186. Schoenherr, C.J., J.M. Levorse, and S.M. Tilghman, *CTCF maintains differential methylation at the Igf2/H19 locus.* Nat Genet, 2003. **33**(1): p. 66-9.
187. Engel, N., J.L. Thorvaldsen, and M.S. Bartolomei, *CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus.* Hum Mol Genet, 2006. **15**(19): p. 2945-54.
188. Kelsey, G. and R. Feil, *New insights into establishment and maintenance of DNA methylation imprints in mammals.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1609): p. 20110336.
189. Lui, J.C., et al., *An imprinted gene network that controls mammalian somatic growth is down-regulated during postnatal growth deceleration in multiple organs.* Am J Physiol Regul Integr Comp Physiol, 2008. **295**(1): p. R189-96.
190. Sakatani, T., et al., *Epigenetic heterogeneity at imprinted loci in normal populations.* Biochem Biophys Res Commun, 2001. **283**(5): p. 1124-30.
191. Rancourt, R.C., et al., *The prevalence of loss of imprinting of H19 and IGF2 at birth.* FASEB J, 2013. **27**(8): p. 3335-43.
192. Cui, H., et al., *Loss of IGF2 imprinting: a potential marker of colorectal cancer risk.* Science, 2003. **299**(5613): p. 1753-5.
193. Cui, H., et al., *Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability.* Nat Med, 1998. **4**(11): p. 1276-80.
194. Cui, H., et al., *Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2.* Cancer Res, 2002. **62**(22): p. 6442-6.
195. Sullivan, M.J., et al., *Relaxation of IGF2 imprinting in Wilms tumours associated with specific changes in IGF2 methylation.* Oncogene, 1999. **18**(52): p. 7527-34.
196. Murrell, A., et al., *Distinct methylation changes at the IGF2-H19 locus in congenital growth disorders and cancer.* PLoS One, 2008. **3**(3): p. e1849.

6 Annexure

6.1 List of acronyms, abbreviations and units

A-N		From P	
ATP	Adenosine triphosphate	p arm	Short arm of chromosome
CMV	Cytomegalovirus	piRNA	Piwi-interacting RNA
d H_2O	Distilled water	pmol	Picomole
DLK1	Delta like 1 homolog	p-value	Probability value
DNA	Deoxyribonucleic acid	q arm	Long arm of chromosome
EDTA	Ethylenediaminetetraacetic acid	rho	Spearman's rank correlation coefficient
F 1	Filial 1	RNA	Ribonucleic acid
g	Gravity	rpm	Revolutions per minute
g	Gram	rs	Reference SNP
HCl	Hydrochloric acid	RT	Room temperature
HeLa	Henrietta Lacks	RT PCR	Reverse transcription PCR
H3K27me3	Histone 3 Lysine 27 Methylation	S	Sedimentation rate
kb	Kilo base pair	SAM	S-adenosine methionine
kg	Kilogram	SDS	Sodium dodecyl sulphate
m^2	Square meter	SE	Standard error
M	Molar	siRNA	Small interfering RNA
mg	Milligram	sncRNA	Small noncoding RNA
$MgCl_2$	Magnesium Chloride	U	Units
miRNA	MicroRNA	vs	Versus
mL	Milliliter	α	Alpha
mm	Millimeter	β	Beta
mM	Millimolar	^oC	Degree celsius
mRNA	Messenger RNA	%	Percentage
n	Number (sample size)	µg	Microgram
NaCl	Sodium chloride	µL	Microliter
NaOH	Sodium hydroxide	µM	Micromolar
ng	Nanogram		

6.2 List of figures

6.3 List of tables

7 Publications list

1. *Allele-specific methylation analysis of imprinted genes reveals parental age and BMI effects in cord bloods*.

   Ramya Potabattula, Marcus Dittrich, Martin Schorsch, Thomas Hahn, Larissa Haertle, Tobias Müller, Nady El Hajj and Thomas Haaf. (First author paper submitted).

2. *Male obesity effects on sperm and next generation cord blood DNA methylation.*

   Ramya Potabattula, Martin Schorsch, Thomas Hahn, Marcus Dittrich, Thomas Haaf, Nady El Hajj. (First author paper submitted).

3. *Hypermethylation of the non-imprinted maternal MEG3 and paternal MEST alleles is highly variable among normal individuals.*

   Larissa Haertle, Anna Maierhofer, Julia Böck, Harald Lehnen, Yvonne Böttcher, Matthias Blüher, Martin Schorsch, **Ramya Potabattula**, Nady El Hajj, Silke Appenzeller, Thomas Haaf. *PLoS ONE* 2017, *doi: 10.1371/journal.pone.0184030*; (co-author).

4. *Gene expression and epigenetic aberrations in F1-placentas fathered by obese males.*

   Megan Mitchell, Reiner Strick, Pamela L. Strissel, Ralf Dittrich, Nicole O. McPherson, Michelle Lane, Galyna Pliushch, **Ramya Potabattula**, Thomas Haaf, Nady El Hajj. *Molecular Reproduction and Development* 2017, *doi: 10.1002/mrd.22784*; (co-author).

5. *Paternal age effects on sperm FOXK1 and KCNA7 methylation and transmission into the next generation.*

   Stefanie Atsem*, Juliane Reichenbach*, **Ramya Potabattula***, Marcus Dittrich, Caroline Nava, Christel Depienne, Lena Böhm, Simone Rost, Thomas Hahn, Martin Schorsch, Thomas Haaf☯ and Nady El Hajj☯. *Human Molecular Genetics* 2016, *doi: 10.1093/hmg/ddw328*; (shared first author).

Conferences and Presentations

1. *Aging effects on sperm DNA methylation*
   Eureka International Symposium (2017)
   Julius Maximilians University, Germany — Oral presentation

2. *Paternal body mass index effects on sperm DNA methylation and impact on next generation*
   28th Annual Meeting of the German Society of Human Genetics (2017)
   Ruhr University, Bochum, Germany — Oral presentation

3. *Paternal age effects on sperm DNA methylation and transmission into the next generation*
   Latsis Symposium: Transgenerational Epigenetic Inheritance (2017)
   Swiss Federal Institute of Technology, Zurich, Switzerland — Poster

4. *The effects of paternal body mass index on sperm DNA methylation*
   27th Annual Meeting of the German Society of Human Genetics (2016)
   Lübeck University, Lübeck, Germany — Poster

5. *Paternal age effects on sperm DNA methylation and transmission into the next generation*
   Eureka International Symposium (2016)
   Julius Maximilians University, Germany — Poster