ANALYSIS OF METHYLATION PATTERNS IN EJACULATED SPERM FROM F_1

NELLORE-ANGUS CROSSBRED BULLS

A Thesis

by

MACEE ROBYN PRAUSE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,   Clare A. Gill
Committee Members,   Penny K. Riggs
                        Jason E. Sawyer
Head of Department,   H. Russell Cross

December 2016

Major Subject: Animal Science

ABSTRACT

Methylation of the carbon 5 position of cytosine is a common epigenetic mark in vertebrates, and has been associated with male infertility. Methylation can be affected by environmental changes in utero and our long-term goal is to establish whether maternal nutrition affects the bovine sperm methylome and variation in male fertility. As a first step, our objective here was to establish the general pattern of methylation in ejaculated bovine spermatozoa. We performed whole genome bisulfite sequencing for five F_1 Nellore-Angus crossbred bulls and ~50 Gb of sequence was obtained per sample. Sequences were aligned to a bisulfite-converted version of the UMD3.1 bovine reference assembly.

The average level of CpG methylation was 88.2% and the global methylation patterns for sperm from Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls were similar. Gene bodies were heavily methylated, whereas promoter and CpG island (CGI) tended to be unmethylated. For promoters with unmethylated CGI, there was enrichment for pathways associated with transcription. Some differentially methylated regions were identified between Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls. These data on the bovine sperm methylome also establish a baseline from which future work related to bovine male infertility can proceed.

DEDICATION

To my father.

ACKNOWLEDGEMENTS

NOMENCLATURE

WGBS   Whole-Genome Bisulfite Sequencing

DAZLA   DAZ Like Autosomal

MZ   Monozygotic

MBD   Methyl-CpG-binding Domain

DMR   Differentially Methylated Region

MTR   Methionine Synthase

GEI   Genetic-Environment Interactions

ART   Assisted Reproductive Technology

PGC   Primordial Germ Cell

CGI   CpG Islands

TSS   Transcription Start Site

TF   Transcription Factor

DOHaD   Developmental Origins of Health and Disease

HPG   Hypothalamic-Pituitary-Gonad

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW

In commercial production systems, reproduction is at least five times as important as growth and milk production (Trenkle and Willham, 1977). Improving the reproductive ability of livestock can decrease the cost of production (or input), thereby effectively increasing the overall efficiency of production (Trenkle and Willham, 1977). Therefore, reproductive efficiency and improved fertility in breeding sires can benefit producers. Male fertility is dependent upon sperm quality and normal spermatogenesis (Karimian and Hosseinzadeh Colagar, 2016). Male infertility is an increasing problem in human health care (Cisneros, 2004) and it is also a problem that livestock producers encounter in herds despite animals with normal sperm parameters (Jenkins et al., 2016). Male infertility is the inability to produce a pregnancy over a long span of time, typically at least one year, despite repeated attempts of intercourse without contraception (Cisneros, 2004; Kasturi et al., 2008; Friemel et al., 2014; Li et al., 2016).

Epigenetic mechanisms such as DNA methylation are critical for the proper function of the sperm genome and development of the embryo after fertilization (Ferlin and Foresta, 2014). Improved knowledge of sperm epigenetics in cattle is crucial to understand fertilization and embryo viability, and also the unknown causes that reduce male fertility.

The objective of this experiment is to describe the methylome of bovine ejaculated sperm at nucleotide resolution to establish a baseline from which future work related to bovine male infertility could proceed. These data will also contribute to the

development of a commercial methylation array. Whole-genome bisulfite sequencing (WGBS) will be conducted using DNA from one semen collection from a population of Angus-Nellore crossbred cattle. Global methylation patterns will be characterized. In particular, the extent of differences in methylation patterns between Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls, and the distribution of methylation in CpG islands (CGI) and non-CpG islands will be investigated.

1.1 Spermatogenesis

The hypothalamic-pituitary-gonad (HPG) axis controls reproduction (Cisneros, 2004). Gonadotrophin releasing hormone is produced in short-lived pulses every few hours from specialized neurons in the hypothalamus to stimulate the pulsatile release of follicle stimulating hormone and luteinizing hormone from gonadotroph cells in the anterior lobe of the pituitary gland (Senger, 2012). Luteinizing hormone binds to luteinizing hormone receptors on leydig cells adjacent to the seminiferous tubules within the testes, which stimulates steroidogenic acute regulatory protein-mediated uptake of cholesterol and the production of pregnenolone, which is converted to testosterone. Sertoli cells in the seminiferous tubules have receptors for testosterone and follicle stimulating hormone, which are transduced to promote follicular development of the spermatogonia.

Mature sperm, or spermatozoa, are produced continually in the seminiferous tubules of post-pubertal males due to the hormonal pulses of the HPG axis. The tight junctions between sertoli cells of the seminiferous tubules form a protective environment known as the blood-testis barrier for germ cell development. Spermatogenesis takes 61

days in a bull and has three distinct phases: proliferation, meiotic, and differentiation phases. Spermatogonia are primordial germ cells (PGC) located near the basement membrane of the seminiferous tubules. During the proliferation phase, spermatogonia undergo multiple rounds of mitosis and stem cell renewal, and move towards the lumen. Stem cell renewal allows for continuous spermatozoa production by continually replenishing the source of stem cells from which new spermatogonia can develop. Primary spermatocytes then undergo two rounds of meiosis to become haploid spermatids diffusing from the basal compartment into the protected adluminal compartment of the seminiferous tubules. Recombination during Meiosis I produces genetic diversity among secondary spermatozoa. Finally, spherical spermatids differentiate into the specialized cells of mature spermatozoa (Senger, 2012). The differentiation phase, also referred to as spermiogenesis, is required for proper sperm maturation without further cell divisions. Sperm divide synchronously in cohorts meaning that each sperm cell generation is produced in groups that underwent the spermatogenesis phases at the same time.

A mature sperm consists of a head, containing a highly condensed nucleus, and a flagellum tail driven by the mitochondrial helix midpiece. To be considered normal, the head should be smooth and elongated, not circular. The midpiece, principal piece and endpiece collectively form the tail which normally is slender, connected to the midpiece, and allows proper flagella movement. Abnormal spermatozoa have morphological abnormalities such as head defects, neck and midpiece defects, tail defects and excess residual cytoplasm (World Health Organization, 2010).

1.2 Male Infertility

Male fertility is dependent upon sperm quality and normal spermatogenesis (Karimian and Hosseinzadeh Colagar, 2016). A semen analysis is typically used to categorize normozoospermic males by evaluating the sperm count, morphology, motility, and viability (Aston et al., 2015). Although the main causes of male infertility are abnormalities in spermatogenesis (Karimian and Hosseinzadeh Colagar, 2016), 15-50% of male infertility cases are idiopathic, where no specific cause of the infertility can be diagnosed (Aston et al., 2012; Anawalt, 2013). A few genetic markers correlated with male infertility have been identified, such as Y-chromosome microdeletions (Tiepolo and Zuffardi, 1976), DPY19L mutations (Harbuz et al., 2011), SPATA16 mutations (Dam et al., 2007), and an A2756G-polymorphism in methionine synthase (MTR) (Karimian and Hosseinzadeh Colagar, 2016). However, these genetic markers only explain some male infertility issues, not male infertility associated with idiopathic cases.

1.3 Epigenetics and the Epigenome

Waddington (1942) coined the term epigenetics to describe the field of study that identifies causal mechanisms for the developmental processes that produce specific phenotypes from genotypes (Waddington, 2012). More recently one definition of epigenetics has been refined to mean the study of stable alterations in gene expression that arise during development and cell proliferation (Jaenisch and Bird, 2003). Epigenetic differences are due to heritable alterations in the epigenome, not changes in the genetic DNA sequence (Jaenisch and Bird, 2003; Urdinguio et al., 2015). Despite reprogramming of the epigenome in cells during spermatogenesis and embryogenesis, a

4

fraction of the epigenome from the parental gametes is still inherited and is dependably replicated through cell divisions (Aston et al., 2015). It is now recognized that an epigenome is a set of modifications to the chromatin. Epigenetics is important because alterations in the epigenome, like hypermethylation of genes, at the wrong place or wrong time can lead to diseases and disorders, such as male infertility (Portela and Esteller, 2010).

Epigenetic modifications can be made by: DNA methylation, histone modification, chromatin remodeling and non-coding microRNA (Gotoh, 2015; Kitamura et al., 2015). Histone modifications include methylation, phosphorylation, acetylation, biotinylation, ubiquitination and ADP ribosylation (Portela and Esteller, 2010; Gotoh, 2015). These modifications result from both internal and external stimuli at a particular time and/or intensity which influences gene expression (Gotoh, 2015). This project will focus on DNA methylation.

1.4 DNA Methylation

DNA methylation describes the covalent addition of a methyl group to cytosine, and is one of the most widely studied epigenetic mechanisms, which is fundamental for cellular processes and regulation (Portela and Esteller, 2010; Urdinguio et al., 2015; Lim et al., 2016). Cytosine methylation is the addition of a methyl group to the fifth position of the pyrimidine ring. In eukaryotes, DNA methylation occurs mainly in CpG dinucleotide sequences accounting for 60 to 90% of methylated cytosines (Bird, 2002; Cisneros, 2004; Portela and Esteller, 2010). Based on nucleotide composition, it is expected there would be 4.41% CpG (International Human Genome Sequencing

5

Consortium, 2001) in the mammalian genome, but instead ~1% of the genome is CpG (Hammoud et al., 2010; Portela and Esteller, 2010; Urdinguio et al., 2015). The underrepresentation of 5-methylcytosine is due to its spontaneous deamination to thymine (Illingworth and Bird, 2009; Cooper et al., 2010).

Because CpG dinucelotides are self-complementary, the patterns of methylated and nonmethylated CpG can be copied when cells divide (Lim et al., 2016). This is regulated by the maintenance of DNA methyltransferase, *DNMT1* (Rajender et al., 2011; Li et al., 2013; Lim et al., 2016). De novo methylation of cytosine is regulated by *DNMT3A* and *DNMT3B* (Ji et al., 2010; Li et al., 2013). DNA methylation plays a role in the proper expression of imprinted genes, biological processes and diseases, and there is evidence of transgenerational inheritance (Lim et al., 2016).

1.5 DNA Methylation in Spermatogenesis

Genome-wide demethylation and remethylation occur in waves during mammalian development: after fertilization of the embryo and during development of the germ cell (Cisneros, 2004; Smallwood and Kelsey, 2012). In mice, within 6 to 8 hours of fertilization the paternal genome is actively demethylated and then the maternal genome is demethylated after several divisions (Mayer et al., 2000). By implantation, most methylation marks in both maternal and paternal genomes are erased in a process known as epigenetic reprogramming. This ensures proper development during the pluripotent cell stage of an organism (Lim et al., 2016). The only methylation to remain at this point is at imprinted genes where DNA methylation patterns are maintained to ensure parent-of-origin monoallelic gene expression (Hammoud et al., 2010).
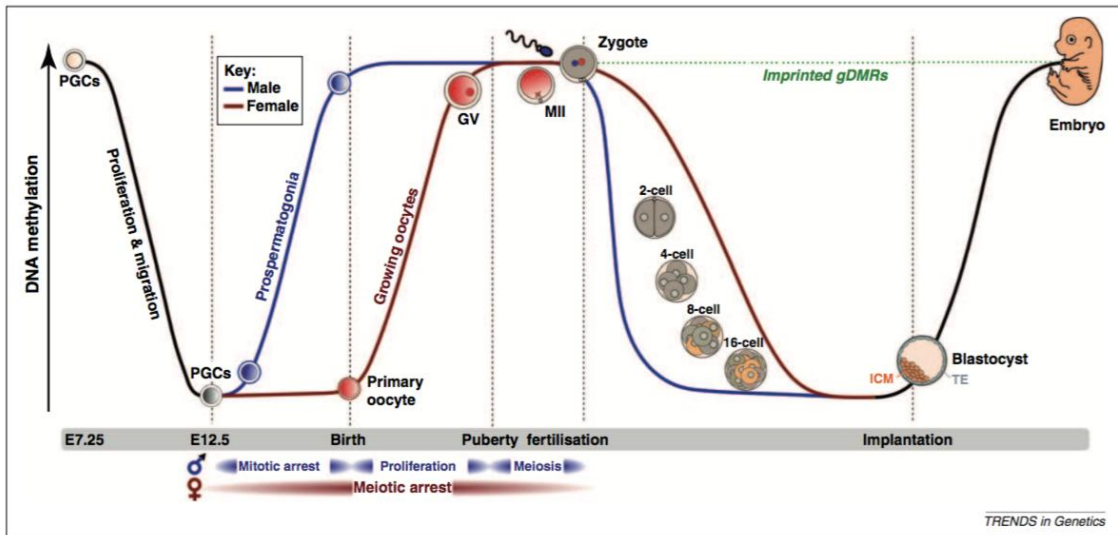
Figure 1. DNA methylation changes during developmental epigenetic reprogramming. PGCs = primordial germ cells. gDMRs = germline differentially methylated regions. From Smallwood and Kelsey (2012); reprinted with permission.

Soon after gastrulation begins, primordial germ cells emerge from the epiblast and they proliferate and colonize the genital ridge, which becomes the gonads. Because PGC originate from a cell type that had already begun to follow a somatic cell fate, during migration to the genital ridge, imprinting marks are erased and reset to ensure totipotency of the next generation (Reik et al., 2001; Smallwood and Kelsey, 2012). After sex determination, male germ cells are mitotically arrested as prospermatogonia. Before meiosis global remethylation occurs in prospermatogonia and the remethylation process is completed before birth (Figure 1). In mice, it has been shown that DNA methyltransferase 3-like (Dnmt3L) is highly expressed in nondividing prospermatogonia in a brief perinatal period associated with global methylation of repetitive DNA (Bourc'his and Bestor, 2004). In Dnmt3L knockout mice, there is hypomethylation of long interspersed nuclear elements and mature sperm are absent, providing evidence that

7

global remethylation is required for correct development of spermatozoa. In addition to prenatal remethylation, there is some postnatal methylation in prospermatogonia that is completed before meiosis (Oakes et al., 2007). It is thought that methylation might suppress gene expression during meiosis (Li, 2002).

In addition to Dnmt3L, other genes affecting methylation have been associated with male infertility. Methionine synthase (MTR) catalyzes the conversion of homocysteine to methionine in folate metabolism (Chen et al., 2001), which is critical for DNA synthesis and methylation reactions (Karimian and Hosseinzadeh Colagar, 2016). There is a common adenine to guanine transition (A2756G) in MTR that can influence enzymatic activity, and the minor allele (G) has been associated with male infertility (Karimian and Hosseinzadeh Colagar, 2016). RHOX gene clusters were found to be hypermethylated in infertile subjects compared to those with normal semen parameters (Richardson et al., 2014). In some infertile men there was aberrant hypermethylation of CCCTC-binding factor (CTCF) binding sites (Wang et al., 2012; Urdinguio et al., 2015).

Errors in the erasure of imprinting marks can also affect fertility. Likewise, errors in global reprogramming could be associated with male infertility (Houshdaran et al., 2007). There is a significant relationship between imprinted loci and aberrant DNA methylation in cases of idiopathic infertility (Hammoud et al., 2010). Kitamura et al. (2015) showed that H19, a paternally imprinted locus, hypomethylated in 16 of the 22 patients with abnormal fertility. Average methylation patterns for H19 were seen in normozoospermic males (Hammoud et al., 2010; Ferlin and Foresta, 2014).

Hypermethylation of MTHFR was also associated with infertility (Ferlin and Foresta, 2014; Urdinguio et al., 2015). Further, it has been observed that with increasing use of artificial reproductive technologies in infertility cases, there has been an increase in the occurrence of imprinting errors compared to mating by natural conception (Hammoud et al., 2011).

1.6 CpG Islands and CpG Island Shores

The distribution of CpG is generally clustered rather than random (Gardiner-Garden and Frommer, 1987; Takai and Jones, 2002). A large portion of methylated CpG are found in repetitive elements to guard chromosomal integrity and avoid instability (Portela and Esteller, 2010). In contrast, unmethylated CpG tend to be clustered in promoter regions of genes and these clusters are known as CpG islands (CGI). About 40% of mammalian gene promoters and exons are associated with CGI (Takai and Jones, 2002). These non-methylated CGI have high G + C content and a high frequency of CpG dinucleotides (Gardiner-Garden and Frommer, 1987; Illingworth and Bird, 2009). Gardiner-Garden and Frommer (1987) defined a CGI as stretches of DNA at least 200bp in length, G + C content of 50% and a CpG frequency of 0.6. To filter CpG that occur in interspersed repeats, such as Alu in the human genome, Takai and Jones (2002) redefined CGI to be stretches of DNA at least 500bp in length, G + C content of 55% and a CpG frequency of 0.65. There was an ~90% reduction in CGI identified by the Takai and Jones (2002) method. Most CGI are hypomethylated, even during global de novo methylation in early development; the mechanism that protects CGI from methylation remains unknown (Illingworth and Bird, 2009).

9

Recently, regions within 2kb of a CGI and with lower CpG density have been defined as CGI shores (Ji et al., 2010; Portela and Esteller, 2010). Most tissue-specific or cell-specific DNA methylation occurs at the CGI shores rather than at CGI (Portela and Esteller, 2010). Differential methylation of CGI shores is significantly inversely correlated with gene expression during cellular differentiation (Ji et al., 2010; Portela and Esteller, 2010), and more strongly than for CGI. Hypomethylation of promoters tends to be associated with increased gene expression, whereas hypermethylation of promoters tends to be associated with transcriptional inactivation.

1.7 Differentially Methylated Regions

Several studies have demonstrated that CpG methylation in mammalian sperm is very high. Frommer et al. (1992) found CpG dinucleotides in non-repetitive sequences in human sperm were hypermethylated, whereas CGI were not. Approximately 70% of the DMR in humans are associated with CGI shores (Portela and Esteller, 2010). According to Popp et al. (2010), 80-90% of CpG in mature mouse spermatozoa are methylated, which is the highest level of methylation observed for specific cell types. Although exons are highly methylated, methylation of promoters is considerably lower (35-40%). Promoters associated with genes important in early development tend to be hypomethylated in the sperm genome (Hammoud et al., 2011). Song et al. (2005) identified 14 tissue-specific, differentially methylated CGI that are hypomethylated in mature human sperm, but hypermethylated in somatic cells.

Shojaei Saadi et al. (2014) used a hybridization approach to characterize methylation in bovine sperm. Similar to other species, hypermethylation of bovine

sperm DNA was confirmed when compared to the embryo (Shojaei Saadi et al., 2014).

Greater than 85% of differentially methylated regions (DMR) were to due to

hypermethylation in sperm when compared to blastocysts (Shojaei Saadi et al., 2014).

1.8 Promoter and Gene Body Methylation

Hypermethylation of CGI in the promoter region of a gene is associated with

gene silencing (Cisneros, 2004; Portela and Esteller, 2010). DNA binding proteins are

recruited to target sites in promoters for active transcription and binding is impeded by

methylation (Portela and Esteller, 2010).

Although methylation is associated with silencing in promoters, gene body

methylation is positively correlated with expression. Demethylation of the gene body

can cause transcription to be initiated from incorrect positions, and cause a disease state

(Portela and Esteller, 2010).

1.9 Epigenome and Disease

Increase in risk of assorted diseases has been explained by the Developmental

Origins of Health and Disease (DOHaD) theory, which states that suboptimal growth in

early life can program changes in long-term health (Gotoh, 2015; Kitamura et al., 2015).

Fetal and early life stages are associated with epigenetic variations, but gametes are also

susceptible to environmental factors (Kitamura et al., 2015). Previous studies in both

human and animal models concur that this programming for susceptibility is correlated

to the epigenome (Kaneda et al., 2011; Rajender et al., 2011; Gotoh, 2015; Montjean and

Ravel, 2016). Therefore, differences in the sperm epigenome could also contribute to

health of the next generation. However, a current limitation is that most approaches to

evaluate the sperm epigenome are based on the analysis of a population of sperm cells, rather than an individual sperm (Aston et al., 2012). Results are indicative of the average level of methylation at a specific locus. Nevertheless, Aston et al. (2015) showed that infertile and fertile men have significantly different sperm DNA methylation patterns.

A study of genetically identical monozygotic twins showed that DNA methylation can impact phenotype, including susceptibility to disease (Fraga et al., 2005; Portela and Esteller, 2010). Despite a common genotype, several phenotypic differences between twins were observed that were attributable either to CpG hypermethylation or CpG hypomethylation (Fraga et al., 2005).

1.10 Whole Genome Bisulfite Sequencing

Sodium bisulfite treatment causes cytosine to undergo sulphoration, then hydrolytic deamination, followed by desulphoration, to yield uracil (Saheb et al., 2014). In bisulfite treated DNA that is amplified by PCR, cytosine is detected as thymine. 5-methylcytosine does not react to the bisulfite treatment so methylated positions are unchanged and are detected as cytosine. Most methods developed to characterize methylation patterns in bisulfite-converted DNA rely on enrichment of the methylated fraction of the genome, either by hybridization or immunoprecipitation. Whole-genome shotgun bisulfite sequencing (WGBS) was first reported by Cokus et al. (2008) and is the only method that generates quantitative methylation profiles at single-nucleotide resolution (Johnson et al., 2012; Wu et al., 2015). Libraries of bisulfite converted DNA are sequenced using massively parallel short-read sequencing techniques. Specialized algorithms have been developed to align the bisulfite converted sequences to reference

12

genome sequences. For example, Bismark (Krueger and Andrews, 2011) uses a three-letter alignment method in which all the cytosines in reads and the reference genome sequence are converted to thymine. Reads are mapped and then compared back to the original reads to identify methylated positions. Although this is a relatively simple approach, a known disadvantage of the three-letter method is that decreasing sequence complexity from four nucleotides to three nucleotides causes more reads to be mapped at multiple genomic locations, and therefore results in lower genome coverage for uniquely mapped reads (Li et al., 2013).

2. MATERIALS AND METHODS

Cattle used for the study were produced near McGregor, Texas at the Texas A&M AgriLife Research Center. All procedures involving animals were approved by the Texas A&M University Institutional Animal Care and Use Committee (AUP #2015-013A).

For breeding purposes unrelated to this project, semen samples were collected from two F_1 Angus-Nellore bulls and three F_1 Nellore-Angus bulls by a commercial service provider using standard electroejaculation methods. At the time of collection, semen samples (without extender) were flash frozen in liquid nitrogen and stored at -80°C until they were ready to be processed. Although contamination of semen is reported as 10^3 to 10^7 other cells in humans, there are no reports of contamination levels in cattle (Phillips et al., 1978; Liu, 2008). Sperm DNA was extracted by standard proteinase K digestion followed by extraction with phenol-chloroform. Briefly, for each bull a 150µl aliquot of semen was centrifuged at 1000 x g for 5 min at room temperature to pellet the sperm cells. The pelleted cells were washed three times by resuspension in 1.5 mL TNE (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 8.0), and centrifugation at 1000 x g for 5 min at room temperature. Each pellet was then resuspended in 1000 µl of fresh semen lysis buffer (1% SDS, 1mg/ml proteinase K, 40mM DTT, in TNE, pH 8.0) and incubated with gentle shaking at 37°C overnight. The cell lysate was transferred to a 15 mL conical tube and mixed with 1.5 mL lysis dilution buffer (85% TNE, 1% SDS, 50 mM DTT). Samples were extracted once with an equal

volume of phenol:chloroform:isoamyl alcohol (25:24:1) and twice with an equal volume of chloroform with each extraction followed by centrifugation for 10 min at 1200 x *g*. To precipitate the DNA, 150μl 2M NaCl and 5.5 mL 100% ethanol were added to the aqueous phase, which was mixed by inversion, placed at -80°C for 1 hr, and then centrifuged at 4°C for 30 minutes at 1200g. The DNA pellet was rinsed with 70% ethanol, pelleted again by centrifugation, and then air dried for 15 min. Finally, the DNA was dissolved in 400 μl TE (0.01M Tris pH 7.5, 1 mM EDTA).

DNA integrity was visualized under UV light (302nm) following agarose gel (1% agarose, 1 x TAE, 0.2 μg/mL ethidium bromide) electrophoresis with lambda standards for 15 min at 150V, and DNA concentration was determined using a Nanodrop 1000 spectrophotometer with 0.2 mm path length. A 10 μg aliquot of each DNA sample normalized to 100 ng/μL was shipped to Novogene Ltd.

Bisulfite conversion, library preparation and whole genome bisulfite sequencing were performed by Novogene Ltd. Library construction consisted of: DNA quality control testing, addition of positive control DNA to allow monitoring of the bisulfite conversion, fragmentation of DNA to 200-300 bp using a Covaris S220 focused ultra-sonicator, terminal repair, A-ligation, and ligation of methylated Illumina sequencing adaptors. The final DNA library was bisulfite converted twice using the EZ DNA Methylation Gold Kit (Zymo Research), and finally the bisulfite treated library was size selected and amplified by PCR prior to cluster generation and paired-end 150 bp sequencing with indexing on an Illumina HiSeq 2500. An initial sequencing run combining our libraries with up to 21 other samples in a single lane was performed to

check integrity of the library. Then 4 to 5 additional lanes of data were produced, where our samples were multiplexed with a total of 8 samples per lane. A minimum of 50 Gb of sequence data was generated per animal (i.e. ~18x coverage). Novogene Ltd. performed image deconvolution, quality control calculations, and preparation of FASTQ formatted files. Data files provided by Novogene Ltd. were analyzed using FastQC (Bock, 2012) to trim reads containing adapters, trim reads with > 10% N, and trim reads with low quality.

Sequences were aligned to the cattle reference genome, UMD 3.1, using Bismark v0.16.1 (Krueger and Andrews, 2011). First, for this three-letter alignment method, the bisulfite reads were converted to C-to-T and G-to-A, which were then aligned to C-to-T and G-to-A converted versions of the UMD3.1 bovine reference genome. Bismark uses bowtie2 v2.2.9 for alignments and Phred+33 quality scores to map unique alignments and identify the strand origin of the bisulfite reads. No mismatches were allowed between the converted read and reference sequence. The minimum alignment score was set as a function where, $f(x) = 0 + (-0.6) \cdot L$ where L is the 150 bp read length. Mapping rate was reported by Bismark and calculated as the proportion of reads uniquely mapped to the UMD 3.1 reference genome.

Bismark determines the methylation state of cytosines by comparing the read sequence with the corresponding genomic reference sequence. Bismark methylation extractor was used to extract the methylation status, excluding the first 5 bp of the forward and reverse reads to avoid undesirable sequencing quality. Additionally, Bismark was able to discriminate between cytosines in CpG, CHG, and CHH sequence

16

context. Finally, PCR duplicates were filtered using Bismark_deduplicate. Methylation was classified relative to annotated genome features: CGI, CGI shores, promoters and gene bodies. Annotations based on the UMD3.1.1 assembly for CGI were downloaded from UCSC. The only difference between UMD3.1 and UMD3.1.1 is that NCBI removed some unassigned scaffolds that they determined were not of bovine origin, so annotations for features assigned to chromosomes are identical in these two versions of the bovine reference assembly. CpG island shores were defined as sequences 1000 bp upstream and downstream of each CGI (Doerks et al., 2002; Wu et al., 2015). Annotations for genes from *Bos taurus* Annotation Release 104 were downloaded from the NCBI ftp site and any feature designated "gene" assigned to a chromosome was included in our analysis. Genes in this file included both NCBI Gnomon predictions, tRNAscan-SE predictions, and RefSeq alignments for protein coding genes, pseudogenes, long noncoding RNA, microRNA, and tRNA. Promoters were defined as sequences 2200 bp upstream to 500 bp downstream of the first nucleotide of the "gene" feature (Jin et al., 2014). Gene bodies were defined as beginning 500bp into the "gene" feature and ending at the last nucleotide of the annotation.

To evaluate whether the direction of the F_1 cross (Angus-Nellore vs. Nellore-Angus) affects methylation patterns in bovine sperm, the R packages bsseq and BSmooth were used to identify DMR (Hansen et al., 2012; Wu et al., 2015). Differentially methylated regions were constrained to filter out regions smaller than 200 bp. A 1 Mb sliding window was used for mean counts of methylated CpG sites and to show the methylation patterns for the two different F_1 crosses. Our null hypothesis was

17

there were no differences in methylation patterns due to the treatment (type of cross). A circular plot representing whole genome methylation patterns and DMR was constructed using Circos (Krzywinski et al., 2009).

Next, gene ontology (GO) and KEGG pathway enrichment analyses were performed using DAVID v6.8beta (Huang et al., 2009b; Huang et al., 2009a). Genes within 2 kb upstream or downstream of DMR were input as a list, based on Ensembl gene identifiers, for functional annotation and functional classification with default settings. Within category Benjamini-Hochberg correction was applied to control the false discovery rate.

3. RESULTS AND DISCUSSION

High molecular weight DNA was extracted from the five bovine semen samples

(Figure 2). The yield from one sample (482T) was very low (<1 ng/μL) and the

extraction was repeated using twice as much semen (300 μL). Prior to normalization,

concentrations ranged from 29.5 ng/μL for 482T to 857.6 ng/μL for 473X. To have

sufficient sample for bisulfite sequencing, DNA from 482T was dried down to

concentrate it, whereas there was sufficient material for all the other samples to be

diluted to the required 100 ng/μL concentration. Although the amount of DNA

recovered from 482T was low, he has sired 144 F_2 calves in the McGregor Genomic

population according to herd records. The quality control report from Novogene (Table

1) indicated that the 260/230 ratio for the 482T sample was low, suggestive of carryover

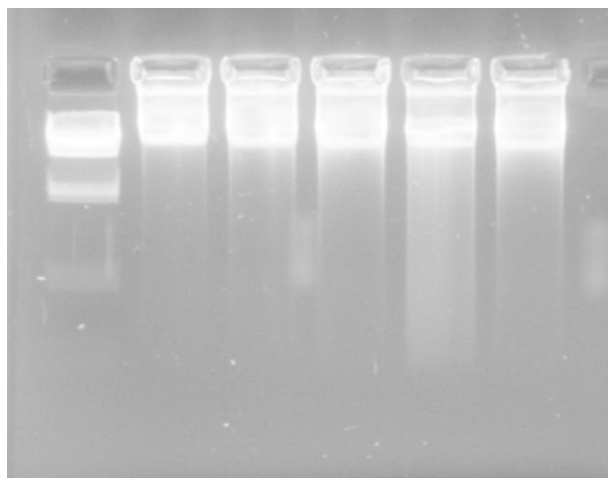phenol or salt contamination (Wilfinger et al., 1997).

Figure 2. Genomic DNA from 5 bovine sperm samples after normalization of
concentration. Lane 1: Hind*III* lambda ladder, 2: 213U, 3: 324T, 4: 473X, and 5: 482T,
and 6: 917A.

Table 1. Summary generated by Novogene of sample quality

Sample Name	F_1 cross	Conc. (ng/µl)	OD260/280	OD260/230
213U	Nellore-Angus	78.1	2.05	1.367
324T	Angus-Nellore	81	1.88	1.741
473X	Nellore-Angus	98.3	1.885	1.885
482T	Nellore-Angus	78.1	1.815	0.71
917A	Angus-Nellore	77.4	1.808	1.679

Raw sequence data generated by Novogene ranged from 49.94 Gb for 213U to 72.32 Gb for 482T (Table 2). Novogene determined that the efficiency of bisulfite conversion was 99.91% for each sample. After quality control filtering, the clean sequence data ranged from 48.06 Gb for 213U to 69.4 Gb for 482T. Mapping rates for the bisulfite-converted sequences ranged from 69.1% for 324T to 74.1% for 917A. The conservative nature of these mapping results is likely because no mismatches in alignments were allowed and because Bismark uses three-letter alignment (Krueger and Andrews, 2011). Using simulated reads, Lee et al. (2015) showed that Bismark mapped reads accurately when no mismatches were allowed and they obtained an 80% mapping rate. Doherty and Couldrey (2014) reported a 61.4% mapping for paired-end 100 bp whole genome bisulfite sequences from sheep aligned with Bismark. As expected, GC content was lower in the bisulfite sequences than in the reference genome because of the conversion of unmethylated cytosines to thymine. GC content for each bisulfite converted sample ranged from 23.83% for 324T to 25.04% for 213U.

Table 2. Summary statistics for bisulfite sequencing

Sample Name	Raw Bases (Gb)	Clean Bases (Gb)	Clean Reads	Mapping Rate	GC Content
213U	49.94	48.06	320,355,580	69.9%	25.04%
324T	58.38	57.06	380,446,588	69.1%	23.83%
473X	58.42	56.82	378,749,946	73.0%	24.38%
482T	72.32	69.4	462,706,440	69.2%	24.00%
917A	57.90	55.38	369,234,782	74.1%	24.64%

Table 3. Annotation of CpG sites in the bovine genome

Feature[1]	CpG Count	No. Features	Mean CpG
CpG Island	2,266,449	36,922	61.38
CpG Island Shore	1,911,051	73,844	25.87
Gene Body	13,205,033	32,314	408.64
Promoter region	2,150,750	32,314	66.56
Other	8,006,634	-	-

[1]Gene body and promoter regions are based on the "gene" feature in bovine annotation release 104 and includes protein coding genes, pseudogenes, long noncoding RNA, microRNA, and tRNA.

In the bovine genome reference sequence there are 27.5 million CpG sites (Table 3) corresponding to 0.96% of the bovine genome (Bovine Genome Sequencing Analysis Consortium, 2009). Gene bodies have the highest number of CpG per feature and CGI-shores the lowest (Table 3). There were 21,115 protein-coding genes in the bovine annotation and 58.2% of the promoters of these genes contained a CGI and 48.8% of gene bodies had at least one CGI.

In bovine ejaculated sperm, CpG sites in gene bodies were strongly methylated (median 94.76%; Figure 3 and Appendix 1). Extensive methylation of gene bodies is common in bovine somatic tissues (Huang et al., 2009b), but has not been previously described for mature sperm. The CpG sites in CGI-shores also were heavily methylated

(median 65.9%), whereas in promoters (median 24.0%) and CGI (median 12.5%) the CpG sites tended to be unmethylated, similar to previous studies (Gardiner-Garden and Frommer, 1987; Irizarry et al., 2009; Portela and Esteller, 2010; Suzuki et al., 2013; Su et al., 2014). However, there was much more variability in the level of methylation of CGI, CGI shores, and promoters than in gene bodies. In particular, 25% of CGI had <2% methylation and 25% of the CGI had >98.8% methylation (Figure 4). Methylation of CGI in promoters is associated with stable silencing of gene expression (Bird, 2002; Illingworth and Bird, 2009). Conversely, active promoters tend to be unmethylated (Cisneros, 2004; Saxonov et al., 2006; Aran et al., 2011). Enrichment analysis using the set of genes near unmethylated CGI showed strong enrichment in pathways associated with transcription (Table 4).

Table 4. Enrichment analysis of unmethylated CpG island nearby genes.

| Term	Count	%	Enrichment *P*-value	Benjamini *P*-value
Membrane-enclosed lumen	202	5.3	1.7e-21	9.1e-19
Intracellular organelle lumen	195	5.1	3.9e-21	1.1e-18
Organelle lumen	195	5.1	4.9e-21	8.9e-19
RNA processing	104	2.7	2.8e-14	8.7e-11
RNA binding	116	3.1	1.2e-11	1.4e-8
Transcription	148	3.9	1.8e-9	1.8e-6
DNA replication	37	1.0	8.5e-7	2.6e-4
ATP binding	243	6.4	5.1e-6	8.0e-4
tRNA metabolic process	40	1.1	9.5e-9	7.3e-6
tRNA aminoacylation	21	0.6	6.0e-6	1.2e-3
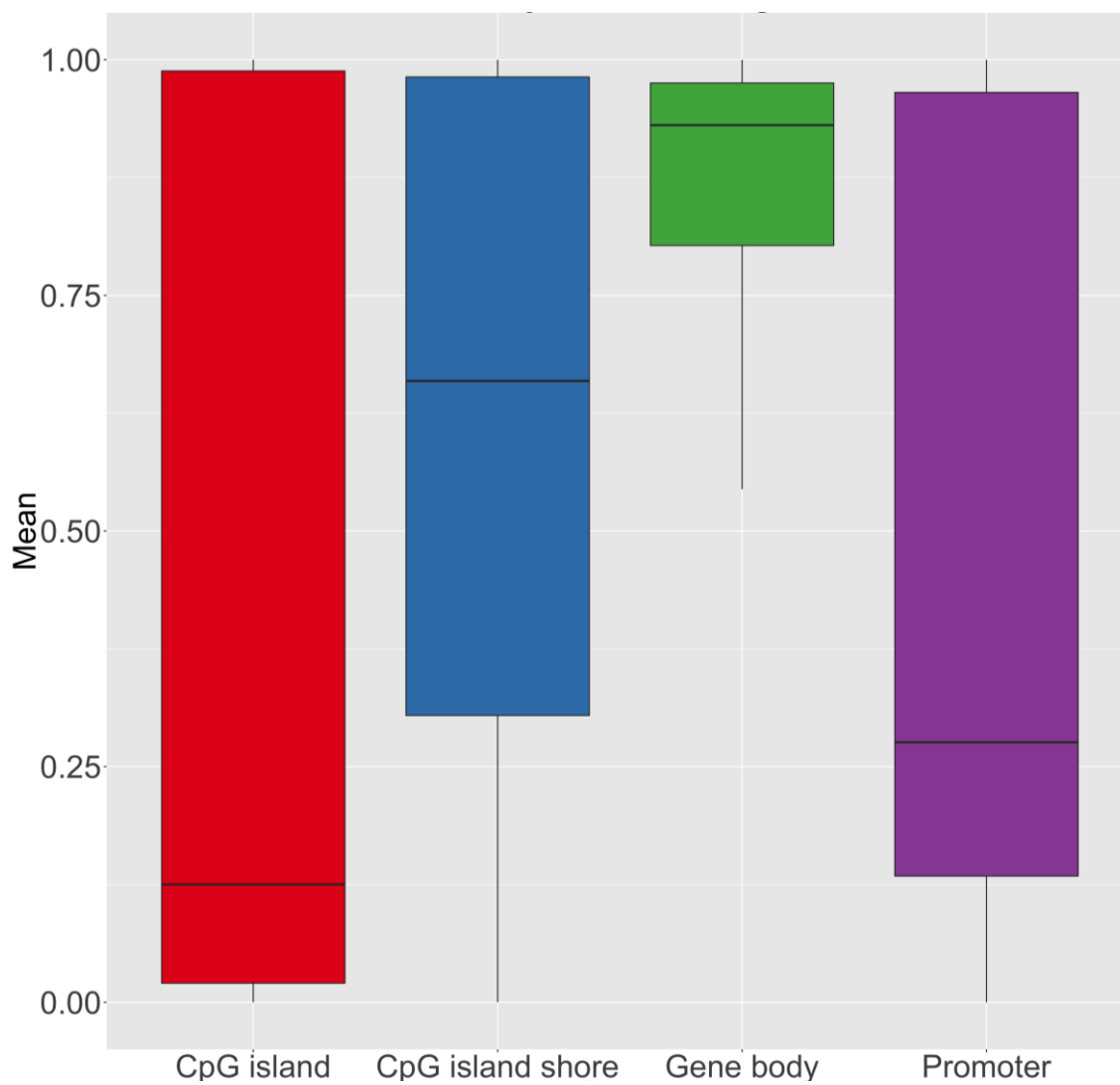
Figure 3. Boxplot of the proportion of methylated CpG sites averaged across 5 samples and the distribution among different genomic features. Promoters and gene bodies are from protein coding genes. Horizontal line is the median, and boxes are the 25th to 75th percentiles.
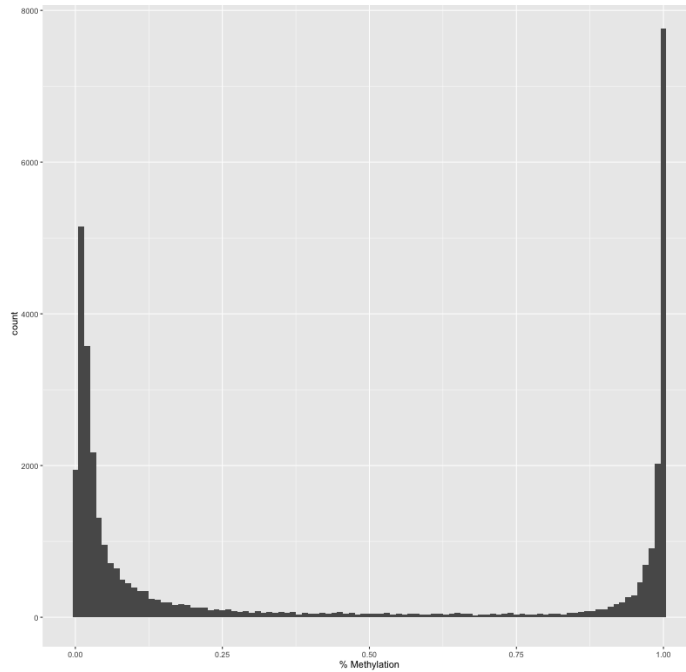
Figure 4. Histogram of CpG island methylation.

Global methylation patterns in F_1 Nellore-Angus and F_1 Angus-Nellore sperm were very similar, and as in other species (Miller et al., 2010; Molaro et al., 2011; Hammoud et al., 2013; Jenkins et al., 2015). The two Angus-Nellore crossbreds had 97% of the methylated sites in common, and the three Nellore-Angus crossbreds were 95% identical. Genome-wide methylation in mature sperm is high (Figure 5). Based on the average methylation in 1 Mb non-overlapping windows across the 5 samples, global methylation was 88.2%. Nevertheless, some DMR were found between the crosses. Few DMR were found on chromosomes 2, 9, 11, and 12, whereas chromosomes 14, 26, and X had a higher number of DMR. Differentially methylated regions had a mean length of 407 bp and were associated with 166 genes within 2 kb upstream and 2 kb

downstream of that gene. Although DMR were found between the crosses, there was no
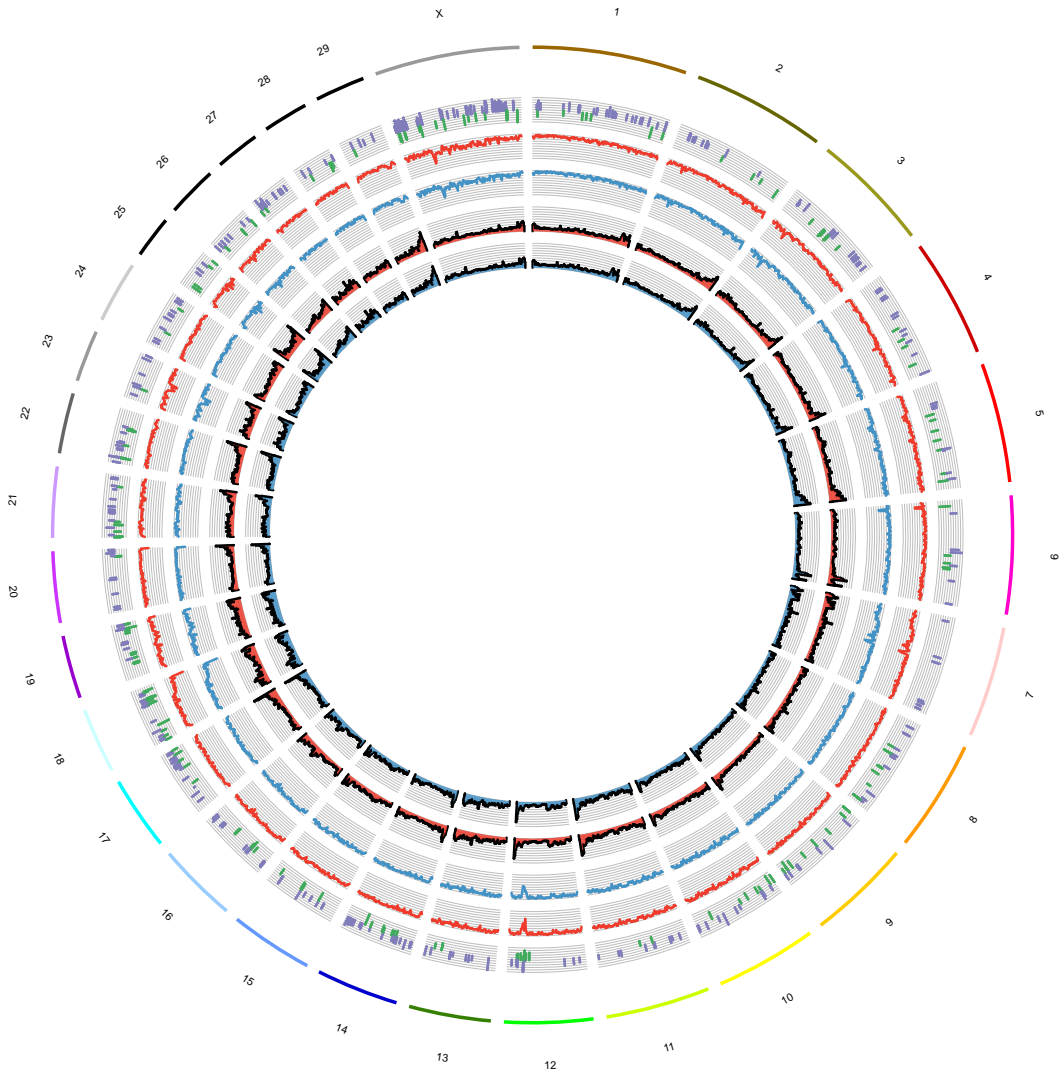
enrichment in specific pathways (Table 5).

Figure 5. Circos plot of the distribution of CpG sites, whole genome methylation and differentially methylated regions. The two inner histogram tracks show the total CpG sites within 1 Mb non-overlapping windows across the whole genome for the Nellore-Angus cross (red) and the Angus-Nellore cross (blue). The two line tracks show proportion of methylated CpG sites in 1 Mb non-overlapping windows for the whole genome. The track below the ideogram shows differentially methylated regions. Regions hypermethylated in sperm from the Nellore-Angus F_1 bulls compared to the Angus-Nellore F_1 bulls are in purple and hypomethylated regions are in green.

Table 5. Enrichment analysis of differentially methylated regions

Category	Term	Gene Number	%	p-value	Benjamin
KEGG Pathway	Renin secretion	4	3.3	6.1e-3	4.9e-1
KEGG Pathway	cGMP-PKG signaling pathway	5	4.2	1.3e-2	5.3e-1
GO Biological	Positive regulation of canonical Wnt signaling pathway	2	1.7	5.6e-2	9.9e-1

There was one region of chromosome 12 with substantially lower methylation than the global methylation levels, where there was also a cluster of DMR. One of the DMR coincided with the promoter of LOC100337006, a computationally predicted gene similar to ATP-binding cassette, subfamily C (ABCC4), member 4, also known as multidrug resistance-associated protein 4 (MRP4). MRP4 regulates cAMP-dependent signaling pathways and controls proliferation of smooth muscle cells (Sassi et al., 2008). The CpG sites in the promoter region of the sperm from Nellore-Angus bulls were mostly unmethylated, whereas in the sperm from Angus-Nellore bulls they were methylated (Figure 6).
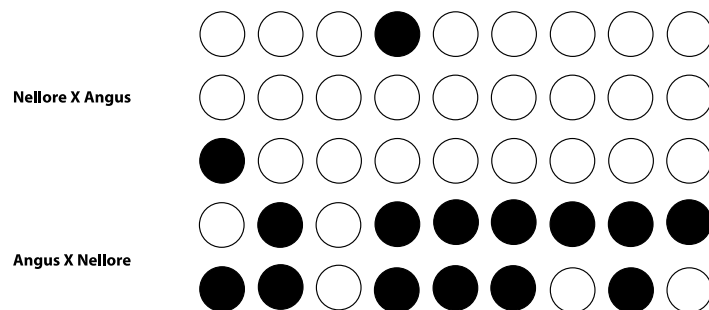

Figure 6. CpG sites in the promoter region for LOC100337006 are differentially methylated in sperm from F_1 bulls. Black circles indicate that the CpG site is methylated and white circles indicate that the CpG site is unmethylated.

Future examination of all the DMR in mature sperm may be useful in identifying the cause of reciprocal differences in birth weights in *Bos indicus* x *Bos taurus* crosses (Amen et al., 2007).

Although it was beyond the scope of this study, DNA methylation can occur in CHG and CHH nucleotides, where H = A, C, or T (Li et al., 2013). In embryonic stem cells, ~17.3% of 5-methylcytosine occurred in CHG regions (Cooper et al., 2010). Lister et al. (2009) found that stem cells showed CHG and CHH methylation enrichment in gene bodies and depletion in protein binding sites. This CHG and CHH methylation accounted for 25% of the 5-methylcytosines identified (Lister et al., 2009). CHH and CHG methylation levels were inversely related to the amount of *DNMT3A* and *DNMT3B* (Lister et al., 2009). There are no reports on the extent of CHH and CHG methylation in sperm and the data produced in this study could be used for future work on non-CpG methylation.

4. SUMMARY

The data in this experiment represent the first evaluation of the bovine sperm methylome at nucleotide resolution based on whole-genome bisulfite sequencing. Mature sperm were collected from Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls. Across the genome, the average level of CpG methylation was 88.2% and the global methylation patterns for sperm from Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls were very similar. Gene bodies were heavily methylated, whereas promoter and CGI tended to be unmethylated. For promoters with unmethylated CGI, there was enrichment for pathways associated with transcription. Some differentially methylated regions were identified between Angus-Nellore F_1 bulls and Nellore-Angus F_1 bulls and future work will focus on the role of these genes in growth and development. These data on the bovine sperm methylome also establish a baseline from which future work related to bovine male infertility can proceed.

LITERATURE CITED

Amen, T. S., A. D. Herring, J. O. Sanders, and C. A. Gill. 2007. Evaluation of reciprocal

differences in *Bos indicus* x *Bos taurus* backcross calves produced through

embryo transfer: I. Birth and weaning traits. J. Anim. Sci. 85: 365-372.

Anawalt, B. D. 2013. Approach to male infertility and induction of spermatogenesis. J.

Clin. Endocrinol. Metab. 98: 3532-3542.

Aran, D., G. Toperoff, M. Rosenberg, and A. Hellman. 2011. Replication timing-related

and gene body-specific methylation of active human genes. Hum. Mol. Genet.

20: 670-680.

Aston, K. I., V. Punj, L. Liu, and D. T. Carrell. 2012. Genome-wide sperm

deoxyribonucleic acid methylation is altered in some men with abnormal

chromatin packaging or poor in vitro fertilization embryogenesis. Fertil. Steril.

97: 285-292.

Aston, K. I., P. J. Uren, T. G. Jenkins, A. Horsager, B. R. Cairns, A. D. Smith, and D. T.

Carrell. 2015. Aberrant sperm DNA methylation predicts male fertility status and

embryo quality. Fertil. Steril. 104: 1388-1397.

Bird, A. 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16: 6-21.

Bock, C. 2012. Analysing and interpreting DNA methylation data. Nat. Rev. Genet. 13:

705-719.

Bourc'his, D., and T. H. Bestor. 2004. Meiotic catastrophe and retrotransposon

reactivation in male germ cells lacking Dnmt3L. Nature 431: 96-99.

Bovine Genome Sequencing Analysis Consortium. 2009. The genome sequence of
taurine cattle: a window to ruminant biology and evolution. Science 324: 522-
528.

Chen, Z., A. C. Karaplis, S. L. Ackerman, I. P. Pogribny, S. Melnyk, S. Lussier-Cacan,
M. F. Chen, A. Pai, S. W. M. John, R. S. Smith, T. Bottiglieri, P. Bagley, J.
Selhub, M. A. Rudnicki, S. J. James, and R. Rozen. 2001. Mice deficient in
methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and
decreased methylation capacity, with neuropathology and aortic lipid deposition.
Hum. Mol. Genet. 10: 433-443.

Cisneros, F. J. 2004. DNA methylation and male infertility. Front. Biosci. 9: 1189-1200.

Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan,
S. F. Nelson, M. Pellegrini, and S. E. Jacobsen. 2008. Shotgun bisulphite
sequencing of the *Arabidopsis* genome reveals DNA methylation patterning.
Nature 452: 215-219.

Cooper, D. N., M. Mort, P. D. Stenson, E. V. Ball, and N. A. Chuzhanova. 2010.
Methylation-mediated deamination of 5-methylcytosine appears to give rise to
mutations causing human inherited disease in CpNpG trinucleotides, as well as in
CpG dinucleotides. Human Genomics 4: 406-410.

Dam, A. H., I. Koscinski, J. A. Kremer, C. Moutou, A. S. Jaeger, A. R. Oudakker, H.
Tournaye, N. Charlet, C. Lagier-Tourenne, H. van Bokhoven, and S. Viville.
2007. Homozygous mutation in SPATA16 is associated with male infertility in
human globozoospermia. Am. J. Hum. Genet. 81: 813-820.

Doerks, T., R. R. Copley, J. Schultz, C. P. Ponting, and P. Bork. 2002. Systematic
identification of novel protein domain families associated with nuclear functions.
Genome Res. 12: 47-56.

Doherty, R., and C. Couldrey. 2014. Exploring genome wide bisulfite sequencing for
DNA methylation analysis in livestock: a technical assessment. Frontiers in
Genetics 5: 1-8.

Ferlin, A., and C. Foresta. 2014. New genetic markers for male infertility. Curr. Opin.
Obstet. Gynecol. 26: 193-198.

Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-
Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-
Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector,
Y. Z. Wu, C. Plass, and M. Esteller. 2005. Epigenetic differences arise during the
lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. 102: 10604-10609.

Friemel, C., O. Ammerpohl, J. Gutwein, A. G. Schmutzler, A. Caliebe, M. Kautza, S.
von Otte, R. Siebert, and S. Bens. 2014. Array-based DNA methylation profiling
in male infertility reveals allele-specific DNA methylation in PIWIL1 and
PIWIL2. Fertil. Steril. 101: 1097-1103.

Frommer, M., L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L.
Molloy, and C. L. Paul. 1992. A genomic sequencing protocol that yields a
positive display of 5-methylcytosine residues in individual DNA strands. Proc.
Natl. Acad. Sci. U. S. A. 89: 1827-1831.

Gardiner-Garden, M., and M. Frommer. 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196: 261-282.

Gotoh, T. 2015. Potential of the application of epigenetics in animal production. Animal Production Science 55: 145-158.

Hammoud, S. S., B. R. Cairns, and D. T. Carrell. 2013. Analysis of gene-specific and genome-wide sperm DNA methylation. Methods Mol. Biol. 927: 451-458.

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

Hammoud, S. S., J. Purwar, C. Pflueger, B. R. Cairns, and D. T. Carrell. 2010. Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. Fertil. Steril. 94: 1728-1733.

Hansen, K. D., B. Langmead, and R. A. Irizarry. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 13.

Harbuz, R., R. Zouari, V. Pierre, M. Ben Khelifa, M. Kharouf, C. Coutton, G. Merdassi, F. Abada, J. Escoffier, Y. Nikas, F. Vialard, I. Koscinski, C. Triki, N. Sermondade, T. Schweitzer, A. Zhioua, F. Zhioua, H. Latrous, L. Halouani, M. Ouafi, M. Makni, P. S. Jouk, B. Sele, S. Hennebicq, V. Satre, S. Viville, C. Arnoult, J. Lunardi, and P. F. Ray. 2011. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. Am. J. Hum. Genet. 88: 351-361.

Houshdaran, S., V. K. Cortessis, K. Siegmund, A. Yang, P. W. Laird, and R. Z. Sokol. 2007. Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. PLoS One 2: e1289.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37: 1-13.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44-57.

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

International Human Genome Sequencing Consortium. 2001. Initial sequence and analysis of the human genome. Nature 409: 860-921.

Irizarry, R. A., C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan, and A. P. Feinberg. 2009. Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41: 178-186.

Jaenisch, R., and A. Bird. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33: 245-254.

Jenkins, T. G., K. I. Aston, T. D. Meyer, J. M. Hotaling, M. B. Shamsi, E. B. Johnstone, K. J. Cox, J. B. Stanford, C. A. Porucznik, and D. T. Carrell. 2016. Decreased fecundity and sperm DNA methylation patterns. Fertil. Steril. 105: 51-57.

Jenkins, T. G., K. I. Aston, C. Trost, J. Farley, J. M. Hotaling, and D. T. Carrell. 2015. Intra-sample heterogeneity of sperm DNA methylation. Mol. Hum. Reprod. 21: 313-319.

Ji, H., L. I. R. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M. J. Aryee, R. A. Irizarry, K. Kim, D. J. Rossi, M. A. Inlay, T. Serwold, H. Karsunky, L. Ho, G. Q. Daley, I. L. Weissman, and A. P. Feinberg. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 467: 338-342.

Jin, L., Z. Jiang, Y. Xia, P. e. Lou, L. Chen, H. Wang, L. Bai, Y. Xie, Y. Liu, W. Li, B. Zhong, J. Shen, A. a. Jiang, L. Zhu, J. Wang, X. Li, and M. Li. 2014. Genome-wide DNA methylation changes in skeletal muscle between young and middle-aged pigs. BMC Genomics 15: 1-12.

Johnson, M. D., M. Mueller, L. Game, and T. J. Aitman. 2012. Single nucleotide analysis of cytosine methylation by whole-genome shotgun bisulfite sequencing. Curr. Protoc. Mol. Biol. 99:21.23: 21.23.21-21.23.28.

Kaneda, M., S. Akagi, S. Watanabe, and T. Nagai. 2011. Comparison of DNA methylation levels of repetitive loci during bovine development. BMC Proc. 5(Suppl 4): 1-5.

Karimian, M., and A. Hosseinzadeh Colagar. 2016. Methionine synthase A2756G transition might be a risk factor for male infertility: Evidences from seven case-control studies. Mol. Cell. Endocrinol. 425: 1-10.

Kasturi, S. S., J. Tannir, and R. E. Brannigan. 2008. The metabolic syndrome and male infertility. J. Androl. 29: 251-259.

Kitamura, A., N. Miyauchi, H. Hamada, H. Hiura, H. Chiba, H. Okae, A. Sato, R. M. John, and T. Arima. 2015. Epigenetic alterations in sperm associated with male infertility. Congenital Anomalies 55: 133-144.

Krueger, F., and S. R. Andrews. 2011. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics 27: 1571-1572.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19: 1639-1645.

Lee, J. H., S. J. Park, and N. Kenta. 2015. An integrative approach for efficient analysis of whole genome bisulfite sequencing data. BMC Genomics 16(Suppl 12): 1-11.

Li, E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. 3: 662-673.

Li, P., F. Demirci, G. Mahalingam, C. Demirci, M. Nakano, and B. C. Meyers. 2013. An integrated workflow for DNA methylation analysis. Journal of Genetics and Genomics 40: 249-260.

Li, X. P., C. L. Hao, Q. Wang, X. M. Yi, and Z. S. Jiang. 2016. H19 gene methylation status is associated with male infertility. Exp. Ther. Med. 12: 451-456.

Lim, C. Y., B. B. Knowles, D. Solter, and D. M. Messerschmidt. 2016. Epigenetic control of early mouse development. Curr. Top. Dev. Biol. 120: 311-360.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315-322.

Liu, J. 2008. Recovery and evaluation of somatic cells from ovine and bovine semen for use in nuclear transfer. Dissertation, Texas A&M University.

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

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

Molaro, A., E. Hodges, F. Fang, Q. Song, W. R. McCombie, G. J. Hannon, and A. D. Smith. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell 146: 1029-1041.

Montjean, D., and C. Ravel. 2016. Male fertility: Genome integrity and sperm DNA methylation. Stem Cell Epigenetics 3: 1-6.

Oakes, C. C., S. La Salle, D. J. Smiraglia, B. Robaire, and J. M. Trasler. 2007. Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. Dev. Biol. 307: 368-379.

Phillips, S. G., D. M. Phillips, E. A. Kabat, and O. J. Miller. 1978. Human semen as a
source of epithelial cells for culture. *In Vitro* 14: 639-650.

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

Portela, A., and M. Esteller. 2010. Epigenetic modifications and human disease. Nat.
Biotechnol. 28: 1057-1068.

Rajender, S., K. Avery, and A. Agarwal. 2011. Epigenetics, spermatogenesis and male
infertility. Mutat. Res. 727: 62-71.

Reik, W., W. Dean, and J. Walter. 2001. Epigenetic reprogramming in mammalian
development. Science 293: 1089-1093.

Richardson, M. E., A. Bleiziffer, F. Tuttelmann, J. Gromoll, and M. F. Wilkinson. 2014.
Epigenetic regulation of the RHOX homeobox gene cluster and its association
with human male infertility. Hum. Mol. Genet. 23: 12-23.

Saheb, A., S. Patterson, and M. Josowicz. 2014. Probing for DNA methylation with a
voltammetric DNA detector. Analyst 139: 786-792.

Sassi, Y., L. Lipskaia, G. Vandecasteele, V. O. Nikolaev, S. N. Hatem, F. Cohen Aubart,
F. G. Russel, N. Mougenot, C. Vrignaud, P. Lechat, A. M. Lompre, and J. S.
Hulot. 2008. Multidrug resistance-associated protein 4 regulates cAMP-
dependent signaling pathways and controls human and rat SMC proliferation. J.
Clin. Invest. 118: 2747-2757.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. 103: 1412-1417.

Senger, P. L. 2012. Pathways to pregnancy & parturition. 3 ed. Current Conceptions, Pullman, WA.

Shojaei Saadi, H. A., A. M. O'Doherty, D. Gagne, E. Fournier, J. R. Grant, M. A. Sirard, and C. Robert. 2014. An integrated platform for bovine DNA methylome analysis suitable for small samples. BMC Genomics 15: 1-18.

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

Song, F., J. F. Smith, M. T. Kimura, A. D. Morrow, T. Matsuyama, H. Nagase, and W. A. Held. 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc. Natl. Acad. Sci. U. S. A. 102: 3336-3341.

Su, J., Y. Wang, X. Xing, J. Liu, and Y. Zhang. 2014. Genome-wide analysis of DNA methylation in bovine placentas. BMC Genomics 15: 1-11.

Suzuki, M. M., A. Yoshinari, M. Obara, S. Takuno, S. Shigenobu, Y. Sasakura, A. R. Kerr, S. Webb, A. Bird, and A. Nakayama. 2013. Identical sets of methylated and nonmethylated genes in *Ciona intestinalis* sperm and muscle cells. Epigenetics & Chromatin 6: 1-12.

Takai, D., and P. A. Jones. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. U. S. A. 99: 3740-3745.

Tiepolo, L., and O. Zuffardi. 1976. Localization of factors controlling spermatogenesis
in the nonfluorescent portion of the human Y chromosome long arm. Hum.
Genet. 34: 119-124.

Trenkle, A., and R. L. Willham. 1977. Beef production efficiency. Science 198: 1009-
1015.

Urdinguio, R. G., G. F. Bayon, M. Dmitrijeva, E. G. Torano, C. Bravo, M. F. Fraga, L.
Bassas, S. Larriba, and A. F. Fernandez. 2015. Aberrant DNA methylation
patterns of spermatozoa in men with unexplained infertility. Hum. Reprod. 30:
1014-1028.

Waddington, C. H. 2012. The epigenotype. 1942. Int. J. Epidemiol. 41: 10-13.

Wang, H., M. T. Maurano, H. Qu, K. E. Varley, J. Gertz, F. Pauli, K. Lee, T. Canfield,
M. Weaver, R. Sandstrom, R. E. Thurman, R. Kaul, R. M. Myers, and J. A.
Stamatoyannopoulos. 2012. Widespread plasticity in CTCF occupancy linked to
DNA methylation. Genome Res. 22: 1680-1688.

Wilfinger, W. W., K. Mackey, and P. Chomczynski. 1997. Effect of pH and ionic
strength on the spectro-photometric assessment of nucleic acid purity.
BioTechniques 22: 474-481.

World Health Organization. 2010. WHO laboratory manual for the examination and
processing of human semen. 5 ed. World Health Organization, Switzerland.

Wu, H., T. Xu, H. Feng, L. Chen, B. Li, B. Yao, Z. Qin, P. Jin, and K. N. Conneely.
2015. Detection of differentially methylated regions from whole-genome
bisulfite sequencing data without replicates. Nucleic Acids Res. 43: 1-9.
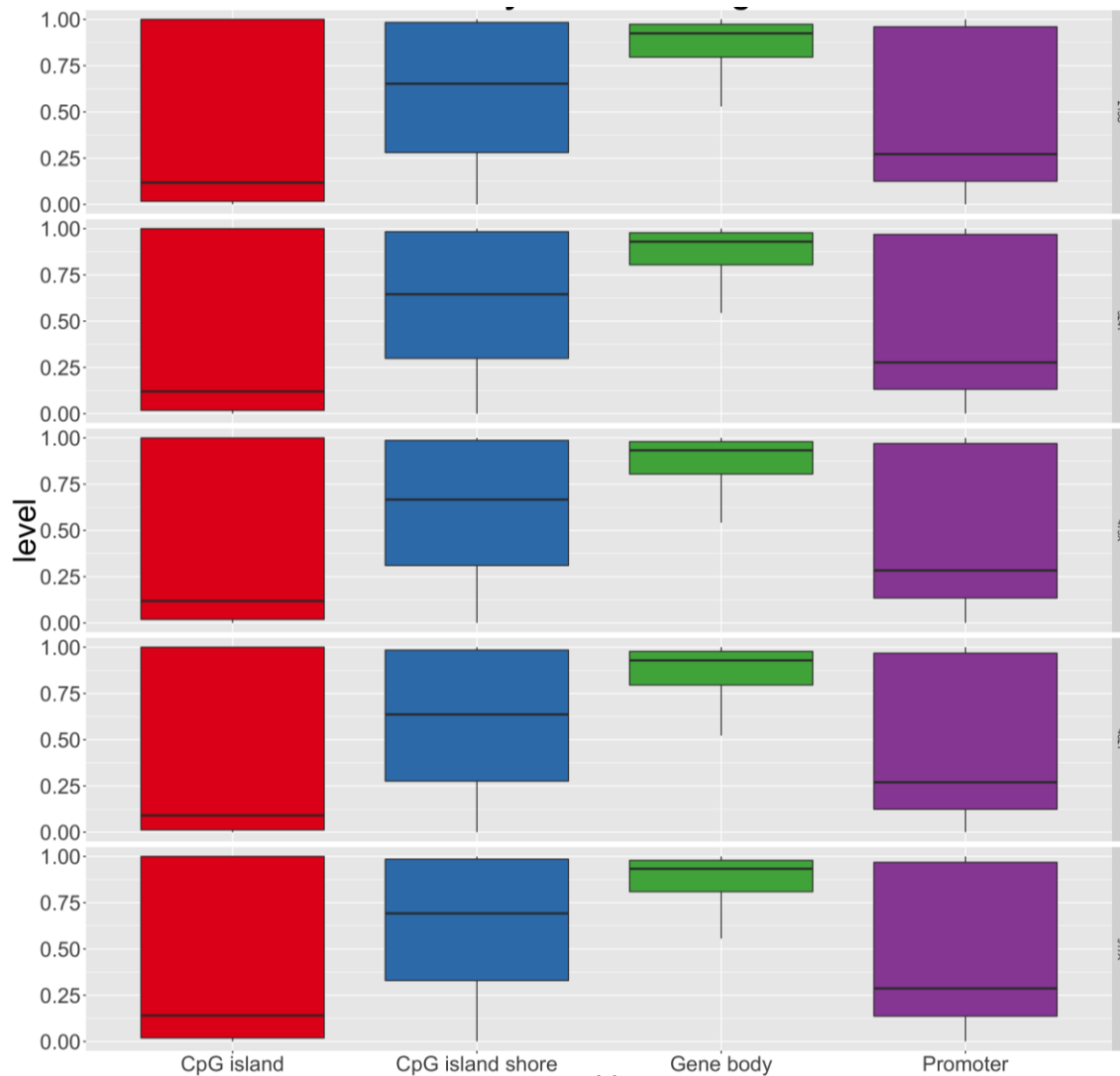
Figure A1. Boxplot of the proportion of methylated CpG sites across 5 samples and the distribution among different genomic features. Horizontal line is the median, and boxes are the 25th to 75th percentiles. Outliers (below 10th percentile and above 90th percentile) were omitted.