The Association Between Sperm DNA Methylation and Sperm Mitochondrial DNA Copy Number

The Association Between Sperm DNA Methylation and Sperm Mitochondrial DNA Copy Number

A Thesis Presented

By

EMILY HOULE

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

Master of Science

May 2020

Molecular and Cellular Biology

The Association Between Sperm DNA Methylation and Sperm Mitochondrial DNA Copy Number

A Thesis Presented

By

EMILY HOULE

Approved in style and content by:

_____

Richard Pilsner, Chair

_____

Alexander Suvorov, Member

_____

Kathleen Arcaro, Member

_____

Scott C. Garman, Graduate Program Director

DEDICATION

To my loving parents who have supported me through everything.

ABSTRACT

THE ASSOCIATION BETWEEN SPERM DNA METHYLTION AND SPERM MITOCHONDRIAL DNA COPY NUMBER

MAY 2020

EMILY HOULE, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor J. Richard Pilsner

Background: Infertility has become a growing concern across the world as cases continue to increase each year. Research has now shifted to identifying novel biomarkers to predict male fertility. While mtDNAcn has recently been found to show promising results as potential biomarker, its regulation remains unclear.

Method: Triplex probe-based PCR was used to quantify mtDNA levels, while 850K Array was used to measure methylation levels. A-clustering algorithm followed by generalized estimating equations (GEE) lead to clustering of individual CpG sites, containing a minimum of 2 CpGs within 1000 base pairs of each other. These clusters were used for analysis of the association between mtDNAcn and DNA methylation within sperm. Metascape[1] was used to annotate gene ontology terms.

Result: Generalized estimating equation model analysis produced 6,038 FDR significant ($q<0.05$) DMRs, 2,459 (40.7%) and 3,579 (59.3%) were hyper- and hypo-methylated, respectively. More stringent Bonferroni correction resulted in 1,343 (97.2%) hypermethylated and 39 (2.8%) hypomethylated. Further analysis of gene ontology of genes associated with our DMRs returned pathways functionally relevant in DNA methylation, meiotic cell cycle, and reproduction.

Conclusion: Thus, we show that sperm mtDNAcn is strongly associated with sperm DNA methylation and the associated implicates mtDNAcn as an influence on infertility.

TABLE OF CONTENTS

PAGE

ABSTRACT……iv

LIST OF TABLES……vi

LIST OF FIGURES……vii

CHAPTERS

I. INTRODUCTION……1

II. METHODS……4

A. Study Population and Design……4

B. Semen Sample Collection and Preparation……4

C. DNA Isolation……4

D. Quantification of mtDNAcn……5

E. Illumina 850K/Epic Array Analyses……6

F. Processing 850K and Statistical Analyses……6

G. Bioinformatic Analyses……7

III. RESULTS……8

A. Demographic Analysis and mtDNAcn Outcome……8

B. Association Between Sperm mtDNAcn and DNA Methylation……9

C. Enrichment Analysis……10

D. Gene Ontology Analysis of mtDNAcn-Associated Sperm DMRs……11

IV. DISCUSSION……14

V. CONCLUSION……17

REFERENCES……18

LIST OF TABLES

Table Page

Table 1 Primer and Probe Sequences used in qPCR Analysis……..5

Table 2 Demographic and Biological Data Summary……...8

Table 3 Gene Function and Gene Ontology Terms for Genes Selected for Spearman Correlations…...13

LIST OF FIGURES

Figure Page

Figure 1 All Cluster Analysis…….…...9

Figure 2 Enrichment of FDR Significant DMR…….…..10

Figure 3 Heatmaps for Gene Ontology Analysis……..…….…..11

Figure 4 Spearman Correlation Plots…….…..12

CHAPTER I

INTRODUCTION

Infertility in recent years has become a focus of research as cases rise around the world. The World Health Organization (WHO) reported 1 in 4 couples in developing countries are affected by infertility in 2004.[2] Infertility is defined as "failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse."[3] A majority of research has been focused on female infertility, as females provide the environment for a developing embryo. More recently, research into male-factor infertility has begun to grow. Male-factor infertility is involved in approximately 50% of cases, whether it be the only factor (~30%) or a contributing factor (~20%).[4] Currently, the most common way to evaluate male infertility is the analysis of semen parameters which assess the quality and quantity of sperm. These sperm parameters include sperm concentration, sperm count, sperm motility, and sperm morphology. These parameters are measured and then compared to the standard WHO semen parameters for the general population.[5] Even with the advances in infertility treatments in recent years, 30% of couples still experience unexplained infertility. Unexplained infertility cannot be explained by abnormal sperm quality, deformities/blockages of the reproductive tract, etc.[6] Research has now begun to look into other possible explanations for infertility, and male infertility in particular.

Recently, it has come to light that semen parameters are not an effective way to assess fertility of men. There are many other factors that go into fertilization, including the processes of capacitation and acrosome reaction.[7] These processes prepare the sperm for fertilizing the egg after the sperm enters the female reproductive tract. This shows that sperm parameters do not tell the whole story when it comes to predicting male fertility, as these are measured before these other maturation steps occur. Semen parameters vary over time, so they are not a consistent measure of fertility.[8] Currently, research is being done to find a more reliable marker of male fertility. One new prospect is the circular genome found within the mitochondria.

Mitochondria are important organelles that are responsible for many cellular processes including producing ATP and generating reactive oxygen species (ROS).[9] Mitochondria are located within most eukaryotic cell types, including in the midpiece of sperm. These organelles are essential for proper sperm

development, function, and maturation. Mitochondria within late spermatocytes, spermatids, and sperm are more condensed and efficient.[10] It is believed that in rodents, the mitochondria become functional after epididymal transit and become capable of fertilization.[11] Observations have also shown that mitochondria within human sperm adopt a less condensed conformation during capacitation, which is a secondary maturation step that occurs during transit in the female reproductive tract, and allows for fertilization.[12]

One novel feature of mitochondria is that they have their own circular genome.[13] This genome produces 13 polypeptides, while a majority of proteins essential to mitochondrial function and replication are produced by the nuclear genome.[9] Mitochondrial DNA (mtDNA) can be quantified by finding the ratio of mitochondrial DNA copies to nuclear DNA; this is called mitochondrial DNA copy number (mtDNAcn).[14] This has been shown to be a better predictor of male fertility than normal sperm parameters because it gives a consistent measure of consecutive diagnoses of clinical infertility.[8] Increases in mtDNAcn has already been linked to lower semen parameter and lower fertilization odds.[8,14] While research continues to support mtDNAcn as a potential biomarker for infertility, the regulation of mtDNAcn remain unclear.

MtDNAcn is dependent on the ability of mtDNA to replicate and the machinery for the replication of mtDNA is located within the nuclear genome. This has led to the examination of epigenetic factors, specifically DNA methylation, as a potential mediator of mtDNAcn. DNA methylation regulates gene expression within the cell by adding or removing methyl groups from certain areas of the gene to suppress or promote transcription. Methylation of genes occur by adding a methyl group to a cytosine at specific sites called cytosine preceding guanine (CpG) sites.[15] Research has mostly focused on three genes related to mtDNA replication, POLG, TFAM, and TWNK. POLG is DNA polymerase gamma which is the main polymerase used in mtDNA replication.[16] TFAM is mitochondrial transcription factor 1 which is a main component of the nucleoid, a DNA-protein complex that act as histones for the mitochondria.[15] While TWNK is the mtDNA helicase used to unravel the circular genome to allow POLG to bind to mtDNA.[16] Any alterations, especially in the methylation pattern of gene, may dysregulate genes and potentially change mtDNAcn levels. Evidence of altered POLG expression affecting mtDNAcn levels has been shown in many different cell types, including oocytes, blastocysts, and embryonic stem cells in mice.[17] The exon 2

region is most sensitive to changes to methylation. A study from Kelly et. al. showed that decreases in methylation in this region was associated with increases in mtDNAcn within different mice tissues. Another study assessed the relation between POLG exon 2 and methylation patterns in human pluripotent/multipotent cells and endometrial tissue. This showed increased methylation in this region which led to lower mtDNAcn.[18]

In this study, we examine if sperm DNA methylation is associated with sperm mtDNAcn levels. Using samples taken from the general population, we assess the association to see if these to factor can be related.

CHAPTER II

METHODS

A. Study Population and Design

This study adopted a cross-sectional design and recruited men who enrolled in the Longitudinal Investigation of Fertility in the Environment (LIFE) study. The inclusion criteria for couples required male participants over 18 years old, in a committed relationship planning a pregnancy within the next 12 months. Female participants must be between ages 18-40 with a menstrual cycle between 21-42 days with no use of contraception for two months and no history of injectable contraceptives within the past year. Participants must have an ability to communicate in English or Spanish.[19] Couples were recruited via hunting/fishing registries in targeted counties in Texas and via commercial marketing database for targeted counties in Michigan.[19]

B. Semen Sample Collection and Preparation

The University of Massachusetts received frozen whole semen samples kept at -80 C until thawed for DNA isolation. Semen parameter and demographic data for each participating couple was also received from the LIFE Study. Whole semen samples were centrifuged to separate sperm from seminal plasma. Seminal plasma was removed from the sample. A 50% gradient spin using Pureception (Cooper Surgical, Connecticut, USA) was performed to further separate mature sperm from any immature sperm cells and somatic cells present in the sample. Sperm were then washed with Quinn's Sperm Wash Medium (Cooper Surgical, Connecticut, USA).

C. DNA Isolation

DNA was then isolated from the processed sperm samples, containing only intact and mature sperm. A previously published method from Wu et. al. was used for isolation.[20] Mature sperm fractions were lysed with 0.2 mm stainless steel beads (Next Advance, New York USA) before using the QiaAMP DNA Mini Kit (Qiagen, Netherlands) for DNA purification. DNA concentration was read using a Fluorometer and used for downstream analyses.

D. Quantification of mtDNAcn

Triplex probe-based Quantitative Polymerase Chain Reaction (qPCR) was used to quantify mitochondrial DNA copy number (mtDNAcn) and mitochondrial DNA deletions (mtDNAdel) based on a previously published method.[21] Briefly, primers were designed to target both minor and major arcs of the mitochondria (Table 1) and amplified using qPCR. The MinorArc, located in the D-Loop of the mitochondrial genome, was targeted for mtDNAcn assessments because it is highly stable with little to no deletion. This led to an accurate measurement of mtDNAcn. Alternatively, the MajorArc region was chosen for mtDNAdel assessments because deletions in this region are common and can lead to accurate measurements of mtDNAdel levels.

Primer Sequences:	
MinorArc	
Forward	5'- CTA AAT AGC CCA CAC GTT CCC-3'
Reverse	5'- AGA GCT CCC GTG AGT GGT TA-3'
MajorArc	
Forward	5'-CTG TTC CCC AAC CTT TTC CT-3'
Reverse	5'-CCA TGA TTG TGA GGG GTA GG-3'
Probe Sequences	
MinorArc:	6FAM-CAT CAC GAT GGA TCA CAG GT-MGBNFQ
MajorArc:	NED- GAC CCC CTA ACA ACC CCC-MGBNFQ

Table 1. Primer and Probe Sequences used in Quantitative Polymerase Chain Reaction (qPCR) Analysis. This is a list of primer sequences and probes used for qPCR analysis for mtDNA quantification. The MinorArc was targeted due to its highly conserved nature. The MinorArc was targeted for mtDNAdel analysis. This region is prone to deletions in the genome, therefore this would give us an accurate reading of mtDNAdel.

Nuclear DNA copy number was determined using an RNAse P assay because it is a single-copy genomic gene with low variability. Therefore, RNAse P was employed to determine the number of copies of nuclear genome due to the consistency of the gene. MtDNAcn was calculated using the ratio of the MinorArc analysis to the nuclear DNA copy number determined by the RNAse P using the following equation:

$$\text{mtDNAcn} = 2(CT_{\text{RNAse P}} - CT_{\text{MinorArc}}).$$

MtDNAdel was calculated using the ratio of MinorArc analysis to the MajorArc analysis:

$$\text{mtDNAdel} = 2(CT_{\text{MinorArc}} - CT_{\text{MajorArc}})$$

All reactions were run in triplicate followed by the calculation of CTs per reaction. Average CTs were then calculated based on the difference of CTs between the triplicate measurements. 100 µM MinorArc Forward Primer Stock and 100 µM MinorArc Reverse Primer Stock were combined to make the MinorArc Primer mix. The primer stock was diluted to 5 µM using TE Buffer. This was repeated using MajorArc Forward and Reverse Primer Stocks to make MajorArc Primer Mix.

A water control and generated standards were also run in triplicate. Each well contained 8 µL of Super Master Mix, made by combining:

- 0.5 µL FAM labeled MinorArc Probe at 5 µM
- 0.5 µL MinorArc Primer Mix at 5 uM
- 0.5 µL NED labeled MajorArc Probe at 5 µM
- 0.5 µL MajorArc Primer Mix at 5 µM
- 0.5 µL VIC labeled RNAseP Primer/Probe Mix at 20x (Applied Biosystems # A30064)
- 0.5 µL Nuclease-free water
- 5 µL ProAmp 2x Master Mix (Applied Biosystems # A30865)

Then 2 µL of each DNA sample at a concentration of 5 ng/uL was added to the wells. DNA concentration was determined using a NanoDrop2000 UV-Vis Spectophotometer (ThermoFisher Scientific) before the assay was performed. Samples with DNA concentration lower than 5 ng/uL before dilution were excluded from the assay, and thus the study.

E. Illumina 850K/Epic Array Analyses

500 ng of DNA from each sample was aliquoted and sent to Yale University's genomics core where bisulfite conversion was performed on the samples followed by analyses using MethylationEpic beadchip. Illumina Epic Array targets 866,836 methylation sites across the genome. To minimize batch effects, samples were placed on beadchips using a blind test method. Each sample was given a randomly generated barcode, which Yale Genomics Core received, then placed the samples on the plate in a nonbiased manor.

F. Processing 850K and Statistical Analyses

To adjust for background fluorescence, correct for technical variation of background fluorescence, and to adjust for different probe types, the R package minfi was used. Cross-hybridizing probes were removed using DMRcate R package. Lastly, the sva R package ComBat function corrected for batch effect. After processing, 803,063 individual CpG loci were left of downstream analysis.

To generate clusters for analysis, A-clustering algorithm modified for use in 850K analysis was used to assess co-regulated regions. The algorithm was set to generate clusters containing ≥2 correlated CpG sites within 1000 base pairs of each other and these became the unit for analysis. 21,872 clusters were produced that were associated with logmtDNAcn with adjustment for male age, male BMI, cotinine, and race to be further analyzed. General estimating equations (GEE) were used to account for the correlated structure of the CpGs sites within a cluster and to estimate the associations between mtDNAcn and sperm DNA methylation.[22]

Differentially methylated regions (DMRs) associated with logmtDNAcn were identified using m-values to ensure homoscedasticity of the data.[23] Beta-value estimates were used to report methylation since beta-values represent percent methylation (0%-100%). mtDNAcn levels were log transformed to normalize the distribution of the data. A false discovery rate (FDR) of $q<0.05$ was set as the statistically significant threshold for DMRs and p-values were corrected for multiple testing using the Benhamini-Hochberg method. For enrichment analysis of our DMRs relative to total clusters, we created a 2 x 2 table followed by application of Fishers Exact test statistics for determination of significance ($p < 0.05$).

G. Bioinformatic Analyses

Metascape (http://metascape.org)[1] was used to determine functional relevance of genes associated with logmtDNAcn-related DMRs. Enrichment of DMRs were also assessed in the genic (TSS1500, TSS200, 5' UTR, 1st exon, eody, 3' UTR, intergenic regions), CpG features (island, shelves, shores, open sea), and protamine (protamine vs nucleosome) features of cluster using UCSC annotation[24]. Ensembl database was used for non-coding feature analysis[25]. Correlations between methylation sites and logmtDNAcn were assessed for eight genes using Spearman correlations.

CHAPTER III

RESULTS

A. Demographic Analysis and mtDNAcn Outcome

In this study, samples from 379 participants from the general population were used to assess the association between sperm DNA methylation profile and sperm mtDNAcn levels. Summary of demographics and mtDNAcn results can be seen in Table 2. Most male participant were over 30 years of age (56.7%) while the average BMI was 29.86±5.7. The majority of male participants identified as non-Hispanic white (81%). 80 participants (21.1%) were current smokers. The average mtDNAcn was 5.38±4.46 with a range between 0.8 to 35.78.

	Demographic				
	Mean	SD	Median	Minimum	Maximum
Male Age	31.74	4.76	31	19	50
Male BMI	29.86	5.7	28.84	18.11	57.77
Smoking (Cotinine)	n (%)				
Smoking	80 (21.1)				
Non-smoking	295(77.8)				
Race	n (%)				
Non-Hispanic White	307 (81)				
Non-Hispanic Black	15 (4)				
Hispanic	32 (8.4)				
Other	23 (6.1)				
	Biological				
	Mean	SD	Median	Minimum	Maximum
mtDNAcn	5.384	4.46	3.98	0.8	35.78
mtDNAdel	0.313	0.08	0.31	0.11	0.5
Smoking >10 ng/mL cotinine					
Non-smoking ≤10 ng/mL cotinine					

Table 2. Demographic and Biological Data Summary. Summary demographic information and biological data for this study. Most of the study participants were white (81%) and non-smokers (77.8%). The mean age of the population was 31±4.76 years old. This study focuses on mtDNAcn, which this population had an average of 5.384±4.459. Statistical analysis adjusts for male age, male BMI, smoking (cotinine), and race.

B. Association Between Sperm mtDNAcn and DNA Methylation

Cluster analysis used to test the association between logmtDNAcn and the 850k data, generated 21,872 total clusters associated with logmtDNAcn with adjustment for male age, male BMI, smoking, and race. Of these, 6,038 (27.6%) were FDR significant ($q<0.05$) and 1,382 (6.31%) were Bonferroni significant ($q<2.29\times10^{-6}$) (Figure 1A). The FDR significant DMRs and Bonferroni significant DMRs were then split into hypermethylated and hypomethylated DMRs. FDR significant ($q<0.05$) DMRs produced 2,459 (40.7%) hypermethylated and 3,579 (59.3%) hypomethylated DMRs (Figure 1B). While the Bonferroni significant ($q<2.29\times10^{-6}$) DMRs showed 1,343 (97.2%) hypermethylated and 39 (2.8%) hypomethylated DMRs (Figure 1C).

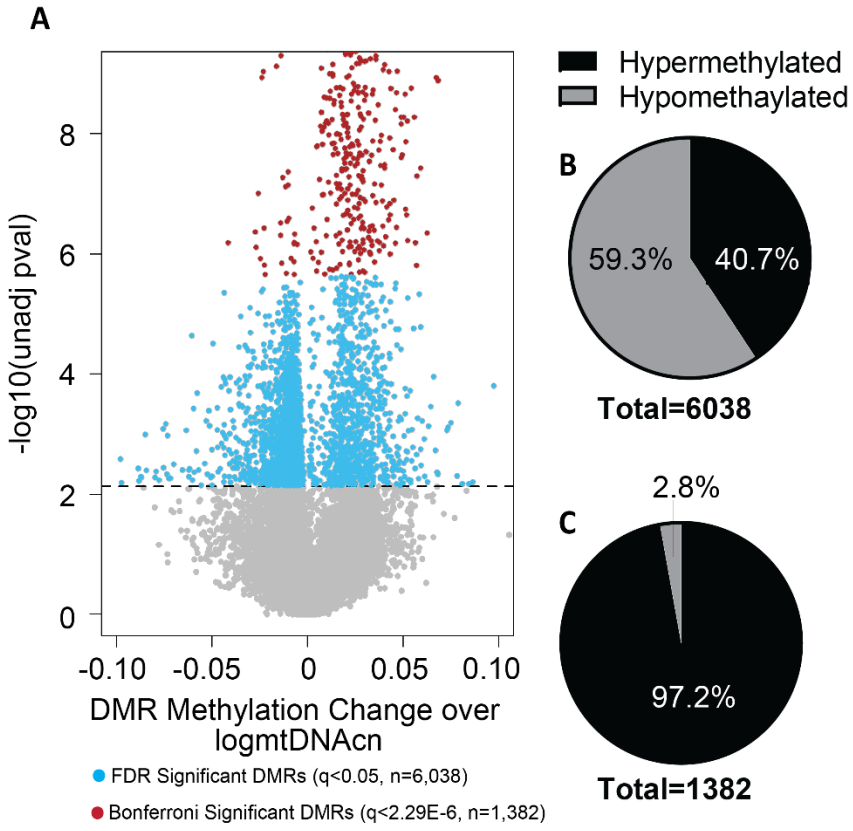

Figure 1: All Cluster Analysis. (A) Volcano plot of association between sperm mtDNAcn and methylation at regional level where grey, blue, and red-colored data points represent non-significant (n = 15,834), FDR significant (n = 6,038), and Bonferroni significant (n = 1,382), respectively. (B) Percentage of hypermethylated and hypomethylated DMRs found as FDR significant ($q<0.05$). (C) Percentage of hypermethylated and hypomethylated DMRs found as Bonferroni significant ($q<2.26\times10$^-16).

C. Enrichment Analysis

Enrichment analysis was then used to classify the FDR significant (q<0.05) DMRs associated with logmtDNAcn. Analysis of the CpG features revealed that only the open sea was significantly more enriched with our DMRs compared to the 21,872 total clusters generated (66.5% vs 55.9%, p-value=2.20×10^{-16}) (Figure 2A). All other CpG features displayed less enrichment compared to all clusters; islands (19.3% vs 22.4%), shores (18% vs 26.3%), and shelves (3.4% vs 5.2%). Analysis also showed that the nucleosome was not significantly enriched.

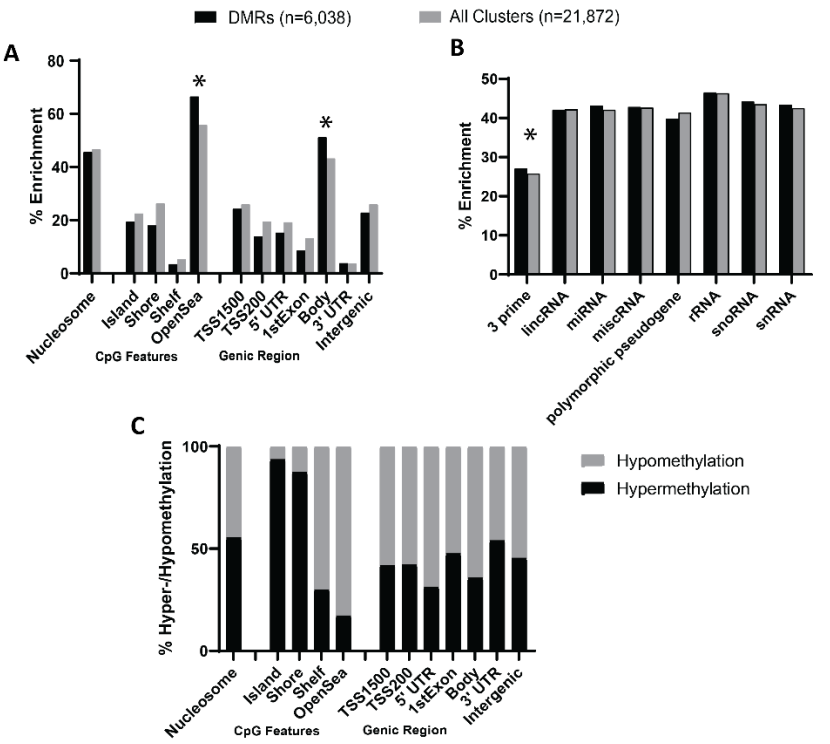

Figure 2: Enrichment Analysis of FDR Significant DMRs (q<0.05, n=6,038). (A) Genic region enrichment analysis. Results showed that the CpG feature open sea was significant (66.5% vs 55.9%, p-value=2.20×10^-16) and the body genic region (51.2% vs 43.2%, p-value= 2.20×10^-16). (B) Non-coding region analysis produced a borderline significant result for the 3 prime (3' UTR) (27% vs 25.8%, p-value=0.046). (C) Results of the methylation enrichment analysis. Many DMRs located in CpG island (93.5%) and shores (87.2%) were hypermethylated, whereas the DMRs located in CpG shelves (70.2%) and open sea (83%) were mostly hypomethylated.

Furthermore, enrichment analysis of the genic coding region showed that only the gene body was significantly more enriched with our DMRs compared to all clusters (51.2% vs 43.2%, p-value= 2.20×10^{-16}) (Figure 2A). Most other coding region features were significantly less enriched when compared to all clusters; TSS1500 (24.2% vs 25.9%), TSS200 (13.9% vs 19.5%), 5' UTR (15.3% vs 19.2%), 1st exon (8.5% vs 13.1%), and intergenic region (22.8% vs 25.9%). The 3' UTR had a p-value greater than 0.05, thus was not significantly enriched.

The non-coding region produced only the 3' UTR regions as significantly enriched compare to all clusters (27% vs 25.8%, p-value=0.046) (Figure 2B). While the polymorphic pseudogenes were

significantly less enriched when compared to all clusters (39.8% vs 41.4%). LincRNAs, miRNAs, miscRNAs, rRNAs, snoRNAs, and snRNAs were not significantly enriched.

Analysis of the methylation patterns of the CpG features revealed that both the islands and the shores were mostly hypermethylated, 93.5% and 87.2%, respectively (Figure 2C). While the shelves and open sea were mostly hypomethylated, 70.2% and 83%, respectively.

D. Gene Ontology Analysis of mtDNAcn-Associated Sperm DMRs

Gene ontology (GO) analysis was performed on DMRs associated with logmtDNAcn and divided into four groups based on significance and methylation. The first group was FDR significant (q<0.05) DMRs that were hypermethylated and contained 3,579 DMRs. This showed most DMRs in the group were linked to DNA methylation involved with gamete generation, followed by those belonging to neuronal systems, and those associated with cellular process involved in reproduction in multicellular organism (Figure 3A). The second group were the DMRs that were FDR significant (q<0.05) and were

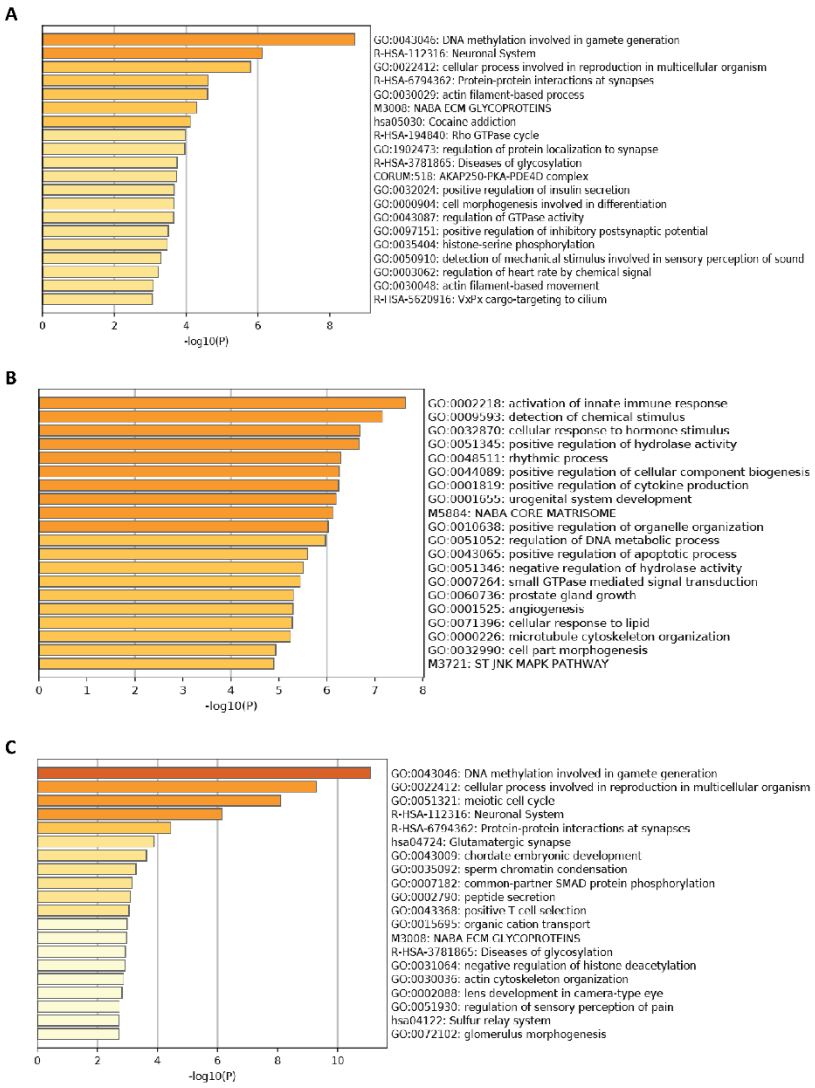

Figure 3: Heatmaps for Gene Ontology Analysis. (A) Heatmap for the gene pathways FDR significant (q<0.05), hypermethylated DMRs are involved with (n=2,459). (B) Heatmap for the gene pathways FDR significant (q<0.05), hypomethylated DMRs are involved with (n=3,579). (C) Heatmap for the gene pathways Bonferroni significant (q<2.20×10^-16), hypermethylated DMRs are involved with (n=1,382).

hypomethylated, which contained 2,489 DMRs. This group showed many significant pathway associations. The most significant GO terms were activation of innate immune response, detection of chemical stimulus, and positive regulation of hydrolase activity (Figure 3B).

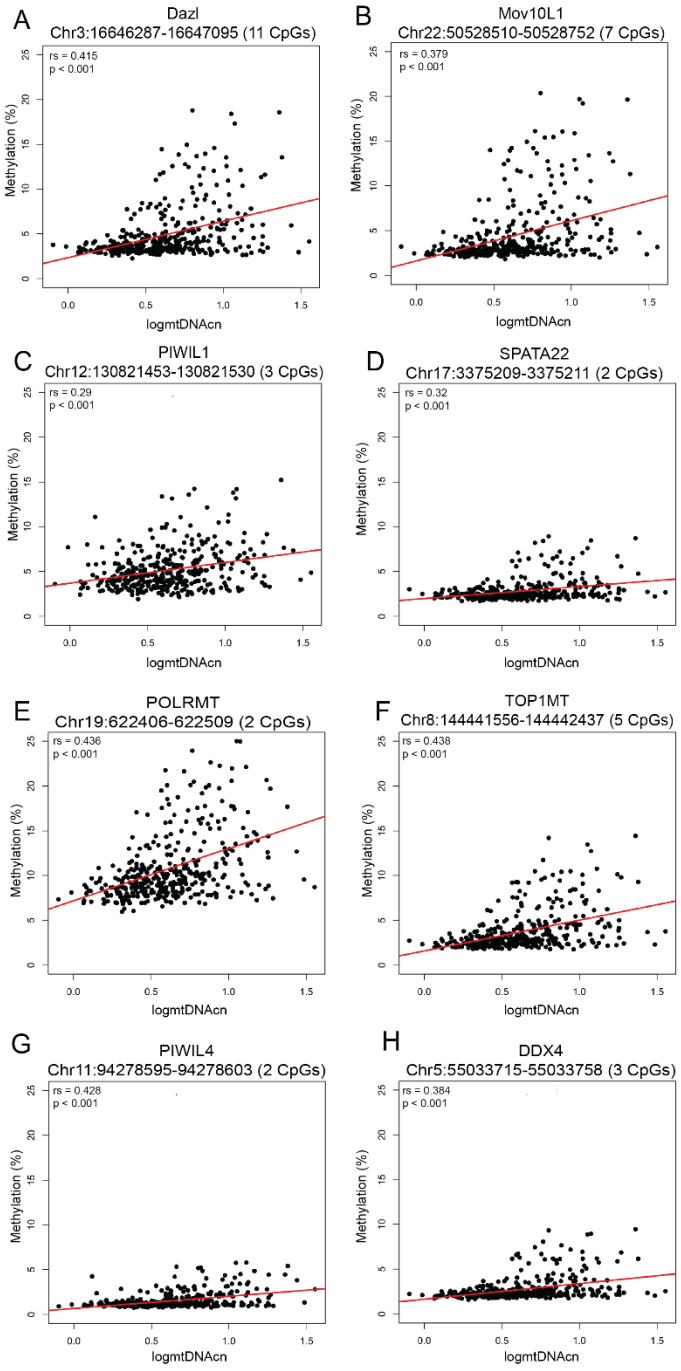

Figure 4: Spearman Correlation Plots. Correlation plots between logmtDNAcn and Percent Methylation for Selected Genes Related to Meiosis, DNA Methylation in Gametes, and mtDNA Replication. These plots show that the selected genes have a positive correlation between mtDNAcn and DNA methylation.

The next group was Bonferroni significant DMRs that were hypermethylated (n = 1,343). This showed most DMRs in the group were linked to biological functions such as DNA methylation involved with gamete generation, those associated with cellular process involved in reproduction in multicellular organism, and those belonging to the meiotic cell cycle (Figure 3C). The last group contains 39 Bonferroni significant hypomethylated DMRs. This group showed low amounts of significant GO terms. The most significant GO term was within the cellular response to nutrient pathway (data not shown).

Correlation plots were created for eight genes (DAZL, MOV1L1, PIWIL1, SPATA22, POLRMT, TOP1MT, PIWIL4, DDX4) within the Bonferroni significant,

hypermethylated DMRs (Figure 4). These correlation plots display the results of a Spearman correlation between percent methylation and logmtDNAcn. All genes showed positive correlations between percent methylation and logmtDNAcn. That is, as percent methylation increases so does logmtDNAcn with a p-value<0.001.

Gene	Function	GO Term
DAZL	deleted in azoospermia like	GO:0007147 female meiosis II GO:0070935 3'-UTR-mediated mRNA stabilization GO:0001556 oocyte maturation
DDX4	DEAD-box helicase 4	GO:1990511 piRNA biosynthetic process GO:0043046 DNA methylation involved in gamete generation GO:0007141 male meiosis I
MOV10L1	Mov10 like RISC complex RNA helicase 1	GO:0043046 DNA methylation involved in gamete generation GO:0007141 male meiosis I GO:0034587 piRNA metabolic process
PIWIL1	piwi like RNA-mediated gene silencing 1	GO:0035093 spermatogenesis, exchange of chromosomal proteins GO:0035092 sperm chromatin condensation GO:0007289 spermatid nucleus differentiation
PIWIL4	piwi like RNA-mediated gene silencing 4	GO:0043046 DNA methylation involved in gamete generation GO:0010529 negative regulation of transposition GO:0010669 epithelial structure maintenance
POLRMT	RNA polymerase mitochondrial	GO:0006391 transcription initiation from mitochondrial promoter GO:0006390 mitochondrial transcription GO:0000959 mitochondrial RNA metabolic process
SPATA22	spermatogenesis associated 22	GO:0000711 meiotic DNA repair synthesis GO:0007129 synapsis GO:0045143 homologous chromosome segregation
TOP1MT	DNA topoisomerase I mitochondrial	GO:0006265 DNA topological change GO:0071103 DNA conformation change GO:0006260 DNA replication

Table 3. Gene Function and Gene Ontology Terms for Genes Selected for Spearman Correlation. List of genes selected for correlation analysis. This table contains the gene, the gene function, and the GO terms associated with the gene.

CHAPTER IV

DISCUSSION

How could mtDNAcn be this highly associated with DNA methylation? Research into cancer cells, glioblastoma, has shown there is a bidirectional control of mtDNAcn and DNA methylation[26]. Using cells with decreasing levels of mtDNAcn and blockages of normal methylation machinery, this study was able to show that methylation is modulated by mtDNAcn levels. This enabled the cell to return to normal mtDNAcn levels to promote tumorigenesis by regulating expression of genes.

Our study found a strong association between mtDNAcn and methylation in sperm. Of the total DMRs found to be associated with logmtDNAcn, nearly half displayed hypermethylation. This implies that half of the DMRs had the potential of down regulating gene expression depending on the location of the increased methylation, with a few exceptions. On the other hand, 59.3% of these DMRs were hypomethylated, which means these DMRs had the potential to increase gene expression. The more stringent analysis produced 97.2% Bonferroni significant hypermethylated DMRs. This means that the highly significant DMRs could have a large impact on gene expression, more specifically, by decreasing expression.[27]

Enrichment analysis of the 6,038 DMRs showed little significance in terms of the coding region and non-coding region. CpG feature enrichment analysis showed that only the open sea was significantly enriched compared to all clusters (21,872) (Figure 2A). This is significant in because the open sea contains long-range enhancers[28] and repetitive elements[27] that are controlled by methylation. Loss of regulation of these elements can lead to different diseases such as cancer.[27] Analysis of the genic region showed the intergenic region was significantly enriched compared to all clusters (Figure 2A). The function of the intergenic region is not fully understood. This is the area between genes, which is believed to have few functions. Though research suggests that new genes can be found in this region.[29] This study suggests a new gene discovered in mice could influence sperm motility and testis weight.

Results from the methylation analysis of the CpG features were interesting. The analysis showed that both the CpG islands (93.5%) and shores (87.3%) were hypermethylated while the shelves (70.2%) and

open sea (83%) were hypomethylated (Figure 2C). This is interesting because the CpG islands and shores are where most transcription factor binding sites, as well as other regulatory elements, are located. Increased methylation of these regions could lead to downregulation of the genes in these regions. The shelves and open sea being demethylated is also significant because these regions are normally hypermethylated to stability in the genome by preventing repetitive elements from moving throughout it.[30] These findings provided important information leading into GO analysis.

Hypermethylated clusters from the FDR significant DMRs showed that they were mostly involved in DNA methylation involved with gamete generation, followed by those belonging to neuronal systems, and those associated with cellular process involved in reproduction in multicellular organism (Figure 3A). This means that DMRs that are associated with alterations of mtDNAcn levels are influencing both methylation patterns and reproductive processes. This relationship is interesting because it has been suggested that malformations in the midpiece of the sperm could indicate alterations in mtDNAcn level stemming from errors in spermatogenesis.[31,32] Hypomethylated clusters from the FDR significant DMRs were related to genes involved in the activation of innate immune response, detection of chemical stimulus, and positive regulation of hydrolase activity (Figure 3B).

When analyzing the Bonferroni significant DMRs, it was found that almost all of these genes were hypermethylated (97.2%, Figure 1C) and are associated with processes in DNA methylation in gamete generation, male meiosis, and cellular processes involved with reproduction (Figure 2C). The hypomethylated DMRs from the Bonferroni significant data showed involvement with the cellular response to nutrient pathway (data not shown). Further analysis of the Bonferroni significant, hypermethylated data using Spearman correlation plots of selected genes displayed a positive correlation between mtDNAcn which produced results with a p-value<0.001 (Figure 4A-G). This means that with increasing DNA methylation, mtDNAcn also increased.

The eight genes selected were chosen based on their relation to the gene ontology analysis and their associated function. These selected genes included DAZL, DDX4, MOV10L1, PIWIL1, SPATA22, POLRMT, TOP1MT, and PIWIL4 (Table 3). DDX4, MOV10L1, and PIWIL4 are related to DNA methylation in gamete generation. While DAZL, MOV10L1, and SPATA22 are involved in meiotic

pathways. PIWIL4 is involved in spermatogenesis. Lastly, POLRMT and TOP1MT are related to the replication of mtDNA. All these genes showed increases in methylation meaning that they could be dysregulated due to alterations in the methylation patterns in these genes and lead to possible changes in mtDNAcn replication.

Originally, this research sought to look at the association of sperm DNA methylation and mtDNAcn. Our evidence suggest that there is a strong association between DNA methylation within sperm and sperm mtDNAcn. But, we cannot conclusively state that DNA methylation is a potential regulator of mtDNAcn. There were few genes located in the hypermethylation data related to replication of mtDNAcn; POLRMT a mitochondrial RNA polymerase and TOP1MT is DNA topoisomerase 1 mitochondrial. Further testing of methylation of the located mitochondrial genes is needed to see how the change in methylation of these genes effect the expression. There was no relation to the genes, POLG, TWNK, or TFAM, which are highly researched regulators of mtDNAcn and are sensitive to methylation changes. Though these data have not conclusively supported the idea that sperm DNA methylation plays a role in the expression of mtDNAcn, our data display the high association between genome-wide methylation within sperm and sperm mtDNAcn. This study suggest that mtDNAcn may control DNA methylation, which could be a potential explanation for our high association in sperm.[14]

CHAPTER V

CONCLUSION

There is a strong association between sperm mtDNAcn and sperm DNA methylation. In the future, we will look for enrichment transcription factors and methylation patterns within imprinted genes related to significant DMRs to see how they are associated with mtDNAcn.

REFERENCES

1 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;**10**:1523. https://doi.org/10.1038/s41467-019-09234-6.

2 *WHO | Global prevalence of infertility, infecundity and childlessness*. WHO. World Health Organization; n.d. URL: https://www.who.int/reproductivehealth/topics/infertility/burden/en/ (Accessed 18 April 2020).

3 *WHO | Infertility definitions and terminology*. WHO. World Health Organization; n.d. URL: http://www.who.int/reproductivehealth/topics/infertility/definitions/en/ (Accessed 18 April 2020).

4 Quaas A, Dokras A. Diagnosis and Treatment of Unexplained Infertility. *Rev Obstet Gynecol* 2008;**1**:69–76.

5 Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HWG, Behre HM, *et al.* World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010;**16**:231–45. https://doi.org/10.1093/humupd/dmp048.

6 Effectiveness and treatment for unexplained infertility. *Fertility and Sterility* 2006;**86**:S111–4. https://doi.org/10.1016/j.fertnstert.2006.07.1475.

7 Austin CR. Membrane fusion events in fertilization. *J Reprod Fertil* 1975;**44**:155–66. https://doi.org/10.1530/jrf.0.0440155.

8 Wu H, Huffman AM, Whitcomb BW, Josyula S, Labrie S, Tougias E, *et al.* Sperm mitochondrial DNA measures and semen parameters among men undergoing fertility treatment. *Reproductive BioMedicine Online* 2019;**38**:66–75. https://doi.org/10.1016/j.rbmo.2018.10.004.

9 Demain L a. M, Conway GS, Newman WG. Genetics of mitochondrial dysfunction and infertility. *Clinical Genetics* 2017;**91**:199–207. https://doi.org/10.1111/cge.12896.

10 Ramalho-Santos J, Varum S, Amaral S, Mota PC, Sousa AP, Amaral A. Mitochondrial functionality in reproduction: from gonads and gametes to embryos and embryonic stem cells. *Hum Reprod Update* 2009;**15**:553–72. https://doi.org/10.1093/humupd/dmp016.

11 Aitken RJ, Nixon B, Lin M, Koppers AJ, Lee YH, Baker MA. Proteomic changes in mammalian spermatozoa during epididymal maturation. *Asian J Androl* 2007;**9**:554–64. https://doi.org/10.1111/j.1745-7262.2007.00280.x.

12 Vorup-Jensen T, Hjort T, Abraham-Peskir JV, Guttmann P, Jensenius JC, Uggerhøj E, *et al.* X-ray microscopy of human spermatozoa shows change of mitochondrial morphology after capacitation. *Hum Reprod* 1999;**14**:880–4. https://doi.org/10.1093/humrep/14.4.880.

13 Amaral A, Lourenço B, Marques M, Ramalho-Santos J. Mitochondria functionality and sperm quality. *Reproduction* 2013;**146**:R163–74. https://doi.org/10.1530/REP-13-0178.

14 Wu H, Whitcomb BW, Huffman A, Brandon N, Labrie S, Tougias E, *et al.* Associations of sperm mitochondrial DNA copy number and deletion rate with fertilization and embryo development in a clinical setting. *Hum Reprod* 2019;**34**:163–70. https://doi.org/10.1093/humrep/dey330.

15 Mposhi A, Van der Wijst MG, Faber KN, Rots MG. Regulation of mitochondrial gene expression, the epigenetic enigma. *Front Biosci (Landmark Ed)* 2017;**22**:1099–113. https://doi.org/10.2741/4535.

16 Gustafsson CM, Falkenberg M, Larsson N-G. Maintenance and Expression of Mammalian Mitochondrial DNA. *Annu Rev Biochem* 2016;**85**:133–60. https://doi.org/10.1146/annurev-biochem-060815-014402.

17 Kelly RDW, Mahmud A, McKenzie M, Trounce IA, St John JC. Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A. *Nucleic Acids Res* 2012;**40**:10124–38. https://doi.org/10.1093/nar/gks770.

18 Lee W, Johnson J, Gough DJ, Donoghue J, Cagnone GLM, Vaghjiani V, *et al.* Mitochondrial DNA copy number is regulated by DNA methylation and demethylation of POLGA in stem and cancer cells and their differentiated progeny. *Cell Death Dis* 2015;**6**:e1664. https://doi.org/10.1038/cddis.2015.34.

19 Buck Louis GM, Sundaram R, Schisterman EF, Sweeney A, Lynch CD, Kim S, *et al.* Semen quality and time to pregnancy: the Longitudinal Investigation of Fertility and the Environment Study. *Fertility and Sterility* 2014;**101**:453–62. https://doi.org/10.1016/j.fertnstert.2013.10.022.

20 Wu H, de Gannes MK, Luchetti G, Pilsner JR. Rapid method for the isolation of mammalian sperm DNA. *BioTechniques* 2015;**58**:293–300. https://doi.org/10.2144/000114280.

21 Phillips NR, Sprouse ML, Roby RK. Simultaneous quantification of mitochondrial DNA copy number and deletion ratio: a multiplex real-time PCR assay. *Sci Rep* 2014;**4**:3887. https://doi.org/10.1038/srep03887.
22 Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;**44**:1049–60.
23 Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;**11**:587. https://doi.org/10.1186/1471-2105-11-587.
24 *The Human Genome Browser at UCSC*. n.d. URL: https://genome.cshlp.org/content/12/6/996.abstract (Accessed 3 June 2020).
25 Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, *et al.* Ensembl variation resources. *Database (Oxford)* 2018;**2018**:. https://doi.org/10.1093/database/bay119.
26 Sun X, St John JC. Modulation of mitochondrial DNA copy number in a model of glioblastoma induces changes to DNA methylation and gene expression of the nuclear genome in tumours. *Epigenetics & Chromatin* 2018;**11**:53. https://doi.org/10.1186/s13072-018-0223-z.
27 Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, *et al.* Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Molecular and Cellular Biology* 2002;**22**:480–91. https://doi.org/10.1128/MCB.22.2.480-491.2002.
28 Skvortsova K, Stirzaker C, Taberlay P. The DNA methylation landscape in cancer. *Essays Biochem* 2019;**63**:797–811. https://doi.org/10.1042/EBC20190037.
29 Heinen TJAJ, Staubach F, Häming D, Tautz D. Emergence of a New Gene from an Intergenic Region. *Current Biology* 2009;**19**:1527–31. https://doi.org/10.1016/j.cub.2009.07.049.
30 Zheng Y, Joyce BT, Liu L, Zhang Z, Kibbe WA, Zhang W, *et al.* Prediction of genome-wide DNA methylation in repetitive elements. *Nucleic Acids Res* 2017;**45**:8697–711. https://doi.org/10.1093/nar/gkx587.
31 Song GJ, Lewis V. Mitochondrial DNA integrity and copy number in sperm from infertile men. *Fertility and Sterility* 2008;**90**:2238–44. https://doi.org/10.1016/j.fertnstert.2007.10.059.
32 Andreu AL, Martinez R, Marti R, García-Arumí E. Quantification of mitochondrial DNA copy number: Pre-analytical factors. *Mitochondrion* 2009;**9**:242–6. https://doi.org/10.1016/j.mito.2009.02.006.