The University of Southern Mississippi
The Aquila Digital Community

Master's Theses

Spring 2020

DNA Methylation Analysis for Tissue and Age Determination for Forensic Human Identification

Lauren Satcher

Follow this and additional works at: https://aquila.usm.edu/masters_theses

Recommended Citation

Satcher, Lauren, "DNA Methylation Analysis for Tissue and Age Determination for Forensic Human Identification" (2020). *Master's Theses*. 733.
https://aquila.usm.edu/masters_theses/733

This Masters Thesis is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Master's Theses by an authorized administrator of The Aquila Digital Community. For more information, please contact aquilastaff@usm.edu.

DNA METHYLATION ANALYSIS FOR TISSUE AND AGE DETERMINATION
FOR FORENSIC HUMAN IDENTIFICATION

by

Lauren Satcher

A Thesis
Submitted to the Graduate School,
the College of Arts and Sciences
and the School of Criminal Justice, Forensic Science, and Security
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Master of Science

Approved by:

Dr. Kuppareddi Balamurugan, Committee Chair
Dr. Dean Bertram
Dr. Xuyang He

____________________	____________________	____________________
Dr. Kuppareddi Balamurugan Committee Chair	Dr. Lisa Nored Director of School	Dr. Karen S. Coats Dean of the Graduate School

May 2020

COPYRIGHT BY

Lauren Satcher

2020

Published by the Graduate School

ABSTRACT

Forensic Science is a vital aspect of criminal investigations and new techniques that could be used to advance this field are constantly being investigated. There are current forensic techniques to identify relevant tissues, but tissue specific methylation markers are being examined as a new way for identifying tissue type. In addition, methylation markers are being investigated as a method to determine other relevant information about donors, such as the age of the individual.

In this study, methylation markers were investigated to determine tissue specific markers to differentiate sperm from blood, saliva, and vaginal epithelial cells, and age specific markers to estimate the age group of a donor using semen samples. Samples were collected of the four relevant tissue types and the genomic DNA was extracted, quantitated, and bisulfite converted. The markers were amplified and pyrosequenced to identify the percentage of methylation at the CpG sites. One way ANOVA Tukey's posthoc test by SPSS statistical package was used to determine the level of significance in the difference between tissue types and age groups.

Three age markers and two sperm specific markers were tested. The age markers, cg6304190, SLC22A18AS, and N23, were able to differentiate one age group from others with statistical significance, and the SLC22A18AS marker could be used to predict the approximate age of an individual. The sperm specific markers, ZAP70_01 and Custom_MDFI_2, were able to differentiate semen samples from other tissue types with statistical significance.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Kuppareddi Balamurugan, and my other committee members, Dr. Dean Bertram and Dr. Xuyang He, for all the time and energy they put into making this thesis possible. I would also like thank my fellow graduate student and lab partner, Maggie Staples, for all the support she gave me throughout our long hours in the lab. Thank you to all the other faculty and staff members of the School of Criminal Justice, Forensic Science, and Security for their support during my time at the University of Southern Mississippi.

TABLE OF CONTENTS

ABSTRACT ... ii

ACKNOWLEDGMENTS ... iii

TABLE OF CONTENTS ... iv

LIST OF TABLES ... vi

LIST OF ILLUSTRATIONS ... vii

CHAPTER I – INTRODUCTION ... 1

Current DNA Technology ... 1

Age Determination Methods ... 2

Tissue Identification Methods ... 4

DNA Methylation ... 6

Bisulfite Modification ... 8

Pyrosequencing ... 8

Aims and Objectives ... 9

CHAPTER II – REVIEW OF LITERATURE ... 10

Current Methylation Studies ... 10

Age Methylation Markers ... 10

Age Related Saliva Markers ... 11

Age Related Vaginal Epithelial & Sperm Markers ... 12

Tissue Specific Markers ... 14

CHAPTER III - MATERIALS AND METHODS 16

Sample Collection 16

DNA Extraction 17

DNA Quantitation Using Agarose Gel 18

Bisulfite Conversion 18

CpG marker selection and assay design... 20

PCR Amplification of Target CpG Sites 21

Pyrosequencing 22

Data Analysis 23

CHAPTER IV – RESULTS 24

Age Marker in Sperm cg-6304190 24

Age Marker in Sperm SLC22A18AS 26

Age Marker in Sperm N23 29

Sperm Specific Marker Hs_ZAP70_01 32

Sperm Specific Marker Custom_MDFI 2 33

CHAPTER V – DISCUSSION 35

APPENDIX A - Primer Specifications 40

APPENDIX B –IRB Approval Letter 41

REFERENCES 42

LIST OF TABLES

Table 1 Bisulfite conversion reaction components 19

Table 2 The components and the volume of each for PCR amplification 21

Table 3 Thermal cycling conditions for sample amplification. 22

Table 4. Percent Methylation Averages and Standard Deviations - Marker cg6304190.. 25

Table 5. Levels of significance (p) for 6304190 26

Table 6. Percent Methylation Averages and Standard Deviations Marker SLC22A18AS 28

Table 7. Levels of significance (p) for SLC22A18AS 29

Table 8. Percent Methylation Averages and Standard Deviations for Marker N23 31

Table 9. Levels of significance (p) for N23 31

Table 10. Percent Methylation Averages and Standard Deviations for Hs_ZAP70_01.. 32

Table 11. Percent Methylation Averages and Standard Deviations for Custom_MDFI 2 34

LIST OF ILLUSTRATIONS

Figure 1. Histogram for marker cg-6304190 24

Figure 2. Scatterplot of CpG1 of marker cg6304190 25

Figure 3. Histogram for SLC22A18AS 27

Figure 4. Scatterplot of CpG2 of marker SLC22A18AS 27

Figure 5. Scatterplot of CpG3 of marker SLC22A18AS 28

Figure 6. Histogram for marker N23. 30

Figure 7. Scatterplot of CpG1 of marker N23 30

Figure 8. Histogram for Hs_ZAP70_01.. 32

Figure 9. Histogram for Custom_MDFI 2 marker 33

CHAPTER I – INTRODUCTION

Current DNA Technology

Although forensic DNA analysis is relatively young compared to other branches of forensic science, biology in forensic science has been used for many years before this technology was developed. Forensic biologists prior to 1985 relied on protein markers and blood typing as a means of human identification. ABO blood typing was first developed and used in the early 1900s by Karl Landsteiner and was based off of determining the presence of certain antigens in an individual's blood. The next significant advancement in forensic DNA analysis did not come until the 1980, when Ray White came forward with the discovery of Restriction Fragment Length Polymorphism (RFLP) (Botstein et al., 1980). Geneticist Alec Jeffreys introduced the concept of DNA profiling with the first human identification probe (Gill et al., 1985). Jeffreys was able to look at sequences of repeated DNA and identify individuals based on the unique number of repeating sequences in certain locations of the human genome. These repeated sequences were referred to as Variable Number of Tandem Repeats (VNTR). The use of RFLP and VNTR for human identification became the gold standard in forensic DNA analysis, and it was first put to use in 1986 to prove the innocence of a suspect in a double homicide (Butler, 2010). The true murderer was later identified using RFLP technology, and this method of DNA analysis began to be accepted by the rest of the world for forensic use.

While VNTR analysis had proven to be very effective in forensic DNA analysis, it did have several drawbacks, especially in its effectiveness in forensic cases. While this technology could be used to identify individuals by comparing the unknown DNA found at crime scenes to known DNA from suspects, it relied on the presence of large quantities

(100+ ng) of relativity undegraded DNA (Butler, 2010). This was an issue in that DNA found at crime scenes is often found in very low quantities (<10ng) and is often severely degraded, meaning the samples were not suitable for VNTR analysis. These drawbacks in RFLP technology were overcome by the development of Polymerase Chain Reaction (PCR), by Kary Mullis as well as the identification of Short Tandem Repeat (STR) polymorphisms (Saiki RK, 1985; Edwards et al., 1991). While VNTRs and STRs are very similar in that they are both a set of repeated nucleotides in the DNA that differ between individuals, STRs are a much smaller repeated segment of DNA. Since the target segments of STR markers are much smaller, degraded DNA can be routinely amplified by PCR methods and this overcomes the need for good quality DNA. PCR based STR marker analysis is the preferred method today in all forensic laboratories throughout the world. The quality assurance standards in the US calls for the analysis of 20 STR markers that will provide a higher discriminatory power in criminal cases. The DNA profile generated in a criminal case is compared to the reference samples for a match and if no suitable match is obtained, the profile can be searched for a matching suspect in a DNA database called Combined DNA Index System (CODIS).

Age Determination Methods

Currently, there are several different methods that researchers are looking at to predict the age of the contributor of an unknown DNA sample. Some, such as the examination of teeth and parts of the skeleton, do not rely on DNA technology to make this prediction, but in many cases the parts needed for these tests are not available to the investigators (Zubakov et al., 2016). In cases when only a tissue sample is found at the crime scene, the knowledge of the age of an unknown individual they are looking for

would be immensely helpful to the investigators. Researchers are also looking into methods that can use protein and DNA information from the samples to estimate the age of the sample donor. These methods include the accumulation of D-aspartic acid in proteins, accumulation of advanced glycation end products, shortening of telomeres, accumulation of 4977bp deletion in mitochondrial DNA, amount of decrease in the number of sjTREC molecules, and DNA methylation (Vidaki et al., 2013; Park et al., 2016; Zbieć-Piekarska et al., 2015).

The decrease of telomere length has been highly linked to aging in humans and seemed to be an excellent source of estimating age of individuals, but too often the results from these tests would have technical assay malfunction and variability that was too extreme to be accurate. In addition, there were many other variables that would affect the length of the telomere and were deemed too problematic to be used (Saeed et al., 2012). The same study by Saeed et al., found that there was a difference in telomere length between saliva and blood samples, and studies showed that telomere length was actually longer in the sperm of older men compared to younger men and that it would be unlikely to use this method to determine age in a forensic setting.

The use of the amount of advanced glycation end products that were dependent on age posed the same problem as the use of telomere length. As research continues, DNA methylation is thought to be the most useful and promising method for predicting the age of an unknown donor (Cho et al., 2017). A study by Zubakov et al. (2016) tested the age predictive powers of mRNA, DNA methylation, DNA rearrangement, and telomere length, and found that DNA methylation markers proved to be the most accurate and most forensically relevant method to estimate age of the donor of unknown samples.

Multiple studies have been conducted using DNA methylation at different sites to estimate age of the sample donors (Park et al., 2016; Alghanim et al. 2017; Weidner et al. 2014).

While several different methods have been proposed as viable ways of determining the age of unknown donors, currently, DNA methylation is the most promising one (Zubakov et al., 2016; Alghanim et al., 2017). Many studies have been done to locate and validate sites that are significantly hyper- or hypo- methylated with age. In over 20 studies, multiple markers were selected as potential sites of age-related methylation. One issue that is faced by researchers is the differentiation between biological age and chronological age (Bell et al, 2012). While an individual may only be a certain age chronologically, factors such as disease, environment, or lifestyle may cause their body to age faster than it should, giving them a different biological age than their chronological age. In addition, there may be other factors that contribute to the methylation of certain sites besides age. Diet, environment, lifestyle, and tissue type tested may also contribute to differentiated methylation values between individuals (Zbiec-Piekarska et al., 2015).

Tissue Identification Methods

There are two different groups of tissue identification tests for the forensically relevant tissues, presumptive and confirmatory tests. When a suspected body fluid is found at a crime scene, presumptive tests are used to suggest the identity of the fluid. These tests tend to be more sensitive than confirmatory tests but are not species specific. Confirmatory tests are used in the event that a presumptive test is positive, and these confirm the presence of a certain body fluid. The confirmatory tests, while they may or

may not be species specific, have their own drawbacks. Confirmatory tests are not as sensitive as presumptive tests and run a higher risk of yielding a false negative result.

Each forensically relevant tissue type can be identified by certain presumptive and confirmatory tests that are specific for a particular body fluid. Presumptive blood tests include the Kastle-Meyer test and Hemastix test, and the confirmatory test involves the identification of hemoglobin crystals using a microscope (Gaensslen, R. 1983). Another confirmatory test for blood is the use of the ABA card which relies on the antigen-antibody reaction in blood. Saliva is identified using a test to determine the level of Amylase in the sample, which is an enzyme found in all tissue types, but with a higher level in saliva. Seminal fluid is identified presumptively by testing the presence of Acid Phosphatase in the sample, which is found in high quantities compared to other tissue types. The confirmatory test for seminal fluid involves the identification of Prostate Specific Antigen, or PSA, in the sample, or by staining and identifying spermatozoa under a microscope.

Even though these tests are routinely used in crime labs, they do come with their own limitations. They all use up a portion of the evidence sample that cannot be used for DNA analysis. These tests also run the risk of giving false positives or false negatives, especially when used on diluted samples. In addition, many of the tests are not species specific and often rely on the presence of enzymes found in multiple tissue types. If a test could be developed to accurately identify tissue source of a DNA sample, it would reduce the amount of time needed for analysis and save more sample for further DNA analysis.

DNA Methylation

One new emerging area that has recently been researched as a way to learn more information from an unknown sample is DNA methylation. Methylation is an epigenetic process that regulates gene expression, without altering the genetic code of an organism. Epigenetic processes include nucleosomal remodeling, histone modification, non-coding RNAs, chromatin looping, and DNA methylation (Vidaki, Daniel, & Court, 2013). These processes do not change the DNA sequence, but they alter gene function or cellular phenotype. In DNA methylation, a methyl group (-CH3) is added to the 5' carbon position of cytosine in CpG dinucleotides (Saabeha & Hasnain, 2018). While methylation can change through an individual's lifetime, they are preserved during cellular division (Vidaki, et al., 2013). The only place where it has been observed that there is no methylation at CpG sites is in embryonic stem cells, so these epigenetic changes can be expected in all individuals (Vidaki, et al., 2013).

CpG sites refer to areas in the DNA where a cytosine nucleotide (C) is immediately followed by a guanine nucleotide (G), and they are connected by the phosphate bond (p). In mammals, 60-90% of CpGs are methylated. The sites that are not methylated are usually located in a stretch of DNA that is about 300-3000 base pairs in length and are mostly made up of guanine and cytosine nucleotides (Vidaki et al., 2013). These areas are called CpG islands, and they are usually located around the 5' end of human genes, primarily in the gene promoter regions.

There are multiple factors that can affect the level of methylation in DNA, including diseases, smoking, early life experiences, aging, pollutants, diet, environmental factors, and ethnicity (Alghanim et al., 2017). Different areas of the DNA may become

hyper- or hypo- methylated based on these factors, and these may be useful in determining the lifestyle or other individualizing characteristics of a person (Saabeha & Hasnain. 2019). A study by Naito et.al. (1993) was the first to attempt to use DNA methylation in a relevant manner to forensic genetics. They attempted to use as a method of female sex typing, and it has been proposed to be used in cases involving sex-reversed individuals. Currently, researchers are looking for ways to use methylation levels to determine cause of death, tissue type, health status, and age of unknown samples or individuals (Vidaki et al., 2013). In addition, levels of methylation may be used to authenticate DNA samples, as DNA manufactured *in vitro* produced DNA that was completely unmethylated. Because identical twins have the exact same DNA sequence, it has also been proposed to use DNA methylation as a way to differentiate between identical twins, as studies have shown that the methylation levels will vary between the two, despite the fact that they share a DNA sequence (Vidaki et al., 2013). Studies have shown that identical twins will start off with very similar levels and patterns of methylation, but as they grow and live their different lives, they will be exposed to different environments and diets, along with other factors. Throughout the course of their lives, their individual methylation patterns will begin to differ more and more, becoming more individualized (Vidaki et al., 2013). When studies are done to find forensically useful ways to connect methylation levels to different identifying characteristics of an individual, one of the challenges is to choose genomic areas that are differentially methylated based only on the factor being investigated. For example, if a researcher was trying to conduct a study to prove there was hyper- methylation in a location of the genome if an individual was a smoker, they would have to ensure that the methylation

levels in that location were not also effected by age, gender, diet, tissue type, or any other factor (Vidaki et al., 2013).

Bisulfite Modification

To determine the difference between methylated and unmethylated CpG sites, extra measures must be taken to preserve the methylation status of CpG sites after PCR. The method used to preserve the methylation status is bisulfite modification. During bisulfite modification, the unmethylated cytosines are converted into uracil, which will then be converted into thymine during PCR. However, methylated cytosines are not converted and will remain as cytosine during PCR. After the bisulfite modification, the target segment is amplified during PCR and the samples are run through a pyrosequencer to determine the methylation levels of specific CpG sites.

Pyrosequencing

Pyrosequencing is a sequence by synthesis process that determines the relative % of A/G incorporation at the CpG sites. During the analysis of data, this information is converted as the level of % methylation for each CpG site studied. The pyrosequencer records not only which nucleotide base was used, but also the amount of light that was given off. The unused nucleotides are discarded and the pyrosequncer moves on to the next base in the target sequence. After the entire sequence has been analyzed, the pyrosequencer generates the results in the form of a pyrogram. The pyrogram is a readout that shows the methylation levels of all of the CpG sites found in the target sequence.

This technology has been used successfully in many studies to analyze methylation levels at target CpG sites for age estimation (Zbiec-Piekarska et al., 2015; Huang et al., 2015). A study by Cho et al. successfully used pyrosequencing to analyze

the methylation levels in 5 genes in blood samples to estimate age. Sabeeha & Hasnain have showed a list of advantages of the pyrosequencing technology for methylation analysis such as the ability to run and analyze the samples faster, quantitative methylation analysis, and single base resolution (Sabeeha & Hasnain, 2019).

Aims and Objectives

This study primarily focusses on two objectives, both involving methylation analysis. One is to identify a set of methylation markers in the human genome that can be used to estimate the approximate age of the donor of the DNA sample. For this study, semen samples will be collected from individuals belonging to different age groups from 18-60 years old. The samples will be divided into four groups: 20+ years, 30+ years, 40+ years and 50+ years old.

The second objective is to identify a set of methylation markers in the human genome that can differentiate semen from other body fluids such as blood, vaginal fluid, and saliva. Four different tissue types such as semen, blood, saliva, and vaginal epithelial cells will be used for this study. .

If these markers are validated, they can be combined together or with other markers for forensic identification to develop a multiplex kit that gives investigators more information about an unknown individual in a criminal case.

CHAPTER II – REVIEW OF LITERATURE

Current Methylation Studies

Methylation levels of different tissues in individuals are currently used in forensic studies for information in a number of different areas. Investigators are using methylation data to identify markers that give information about unknown individuals such as hair or eye color, age, and tissue type of origin of a DNA sample (Vidaki, Daniel, & Court 2013; Balamurugan 2014). Because methylation levels can be affected by so many factors, any CpG site that seems to be promising for one aspect have to be tested to ensure that the same CpG site is not affected by other factors. One study found that two CpG sites used as age related markers in saliva studies were not useful as age markers in blood samples (Huang et al., 2015). Another study by Alghanim et al. (2017) showed that the markers they investigated were age related in both blood and saliva. This shows that each tissue must be tested for age related markers because although it would be beneficial to find markers that were age related in every tissue type, there may not be a set of CpG markers that work in every tissue type, rather a set of age related markers specific to each tissue.

Age Methylation Markers

One of the most commonly tested tissue types used for age estimation in methylation studies is blood. Studies done by different investigators have identified different CpG sites with methylation levels that are significantly different between age groups. Most studies use Illumina platforms such as Illumina HumanMethylation 450K BeadChip to select CpG sites that showed promising methylation patterns and furthered investigated a number of these sites (Freire-Aradas et al., 2016). In many of these studies, pyrosequencing was used to collect the methylation data used in the age estimation,

showing that it is a readily available and reliable method for collecting methylation data (Zbieć-Piekarska et al., 2015). Yi et. al. (2014) were able to identify two CpG sites in blood samples from 105 donors between the ages of 10-72 years old that were significantly differentially methylated between age groups. Many other studies used this method to study blood samples from different age groups and comparing the methylation levels between the different age groups. A study by Jenkins et al. (2014) studied blood samples from a group of individuals, collected at two different time points that differed by 10 years and compared the difference in methylation levels from these two separate sets of samples. Other investigators performed studies on old bloodstains and found they were still able to collect useful methylation data to estimate age, though there was some decline in successful PCR amplification (Huang et al., 2015; Zbiec-Piekarska et al., 2014).

While there have been numerous studies done on blood samples to determine the difference in methylation levels between age groups, often the studies are done primarily on individuals from one ethnicity. A study by Park et al., (2016) points out that some methylation patterns can be population specific, and studies should be done to ensure that these markers are age related in all populations.

Age Related Saliva Markers

While most age-related methylation studies use blood samples, there are a number of studies that used saliva samples as well. Hong et al. were able to achieve a high age predictability in saliva form 226 individuals using 7 CpG markers. One of these markers, found in the KLF14 gene, also shows promise as a valuable marker to predict age in blood, leading investigators to believe that it may be possible to find markers that predict

age accurately in all tissues, although more studies are needed to identify such markers (Hong et al., 2017)

In another study, investigators found 23 markers that may be linked to age in blood samples, and then attempted to use those same markers as age markers in saliva samples (Vidaki et al., 2016). The markers gave accurate age predictions when tested in saliva samples, but researchers express concerns with the validity of these results due to the manner of saliva collection. These samples were collected using Oragene DNA collection kits, which are commonly used for DNA extraction from white blood cells rather than buccal cells. This would lead to skewed results which would tend to lean towards the same results as the blood samples (Vidaki et al., 2016). For this reason, the manner of sample collection must be taken into account in the future when using saliva samples to test for age markers. Studies may also be needed to test for differences in methylation results between saliva samples and buccal swabs.

Age Related Vaginal Epithelial & Sperm Markers

At this time, very few studies have been published regarding the use of methylation markers to determine the age of an individual using vaginal secretions and methylation markers. In a study by Alsaleh et al. 3 vaginal secretion samples were used in a multi-tissue age related methylation study. In an attempt to eliminate the need to identify the tissue type before estimating the age of an individual, these researchers used twenty-four samples from four tissue types, one of which being vaginal secretions, to estimate age (Alsaleh et al., 2017). While the mean absolute deviation (MAD) for the vaginal samples was 6.9 years, this data was collected only from 3 samples. Further

studies need to be done using more samples to validate the usefulness of the ten CpG sites used in this study.

One of the other most forensically relevant tissue types commonly found at crime scenes is semen, although there are few studies done on age related markers in this tissue type as well. However, results of the few studies that have been done give promising results. The first study by Lee et al. (2015) identified 24 specific CpG sites thought to be related to age in sperm. Three of these CpG sites were chosen due to their high R value and validated in additional samples and were able to predict the chronological age of individuals with a mean absolute deviation of 5 years. While the results were promising, only 68 semen samples were used, and this small sample size means additional studies on these markers are needed to validate them as age-related sperm markers (Lee et al., 2015).

These same three CpG sites were investigated further in 2018 by Lee et al. Twelve more semen samples were used to test the three markers as age-related markers in semen, and the results showed a promising trend with age. Researchers noted, however, that the accuracy of the prediction decreased with the age of the individual, with the MAD from the chronological age for samples of individuals in their 20s was 2.9 years, with a MAD from the chronological age for samples of individuals in their 50s was 7.2 years. These markers were also tested using forensic casework samples from 19 different cases. The deviation from chronological age increased with these actual casework samples, but still showed a promising correlation to age (Lee et al., 2018). However, there were no studies done with mixed tissue samples, and this may play a factor in the usefulness of these markers, and this warrants additional studies.

Tissue Specific Markers

As DNA methylation studies become more popular in the field of forensic science, more and more studies are undertaken to find tissue specific methylation markers in sperm, vaginal epithelial, saliva, and blood. One of the first studies done in this area was by Frumkin et al. They were able to discriminate between different tissue types using restriction enzymes that were methylation sensitive and capillary electrophoresis. Hypermethylated samples gave high electropherogram signals while hypomethylated samples gave low electropherogram signals (Frumkin et al., 2011). These signals were compared and used to determine the methylation status of the CpG sites. Studies by Fu et al. found four regions in different genes that showed a significant difference in methylation between venous blood, menstrual blood, saliva, sperm, and vaginal fluid. These researchers used specially designed primers for PCR and pyrosequencing to collect methylation data, and found they were able to distinguish between all 5 tissue types when evaluated in 40 body samples (Fu et al., 2015). In further testing, the researchers were also able to show that the methylation levels at these CpG sites were not influenced by sex.

Lee et al. conducted a study in 2015 using markers from data obtained through the Illumina Infinium HumanMethylation450K BeadChip array. The researchers were able to test 1-3 CpG sites each for blood, saliva, semen, vaginal fluid, and menstrual blood for a total of 8 sites. These were then analyzed using a methylation SNaPshot reaction and were able to successfully discriminate between different tissues, even with mixed samples and aged samples (Lee et al., 2015).

Methylation studies performed by Balamurugan et. al. (2014) were able to differentiate Sperm cells form four other tissues such as blood, vaginal secretions, buccal, and skin cells using multiple methylation markers such as β_SPTB_03, INSL6_03, DACT1, USP 49, and ZC3H12D. In another study, Park et. al. identified two novel markers each for blood, saliva, vaginal secretions, and sperm. Similar to the study done by Fu et. al., these researchers used pyrosequencing to obtain methylation data to determine if the markers were significant (Park et al., 2014). In using pyrosequencing, researchers were able to evaluate not only the target CpG site, but also the neighboring sites adjacent to the target site.

CHAPTER III - MATERIALS AND METHODS

Sample Collection

Biological fluids were collected from normal healthy volunteers after informed consent and Institutional Review Board (IRB) approval. Approximately ten samples were collected from volunteers for the following body fluid: blood, saliva, semen and vaginal epithelial cells.

Blood samples were collected from healthy volunteers by sterilizing the volunteer's fingertip with 70% ethanol, followed by applying an autolet lancet device to the fingertip and sample collected on sterile cotton swabs, air-dried, placed in labeled paper envelopes, and stored frozen until analysis.

Buccal cells were collected from volunteers by providing the volunteers with sterile cotton swabs and asking the volunteers to firmly rub the swab inside of the cheek for several seconds. The swabs were air-dried, returned to labeled paper envelopes and stored frozen until analysis.

Volunteers who choose to give a vaginal epithelial sample were provided with sterile cotton swabs and instructions which were given to collect the sample themselves and then provide the sample to the investigator. Volunteers were instructed to wait at least three to four days after sexual activity before collecting samples. The swabs were air-dried and placed in labeled paper envelopes and stored frozen until analysis

Semen samples were collected by volunteers in a way convenient to them. A specimen cup was provided to the volunteer to collect the sample in a manner that is private and convenient. The participant brought the sample to the investigator and samples were stored frozen until analysis.

When sufficient numbers of samples were not available, samples were purchased through commercial sources. All samples were assigned a unique identifier number and the information was entered into the password protected spreadsheet.

DNA Extraction

DNA was extracted from all samples using standard organic extraction protocols (Budowle et. al., 2000). For the samples on the swab, the cotton portion of the swab was separated from the stick and placed in a 1.5mL tube. Four hundred microliters of stain extraction buffer and 10µL of Proteinase K was added to the tube and were incubated at 56°C overnight.

Once the incubation period was completed, the swab was transferred to a spin basket column and centrifuged for three minutes to remove the substrate. To the extracted DNA, 500µL of phenol/chloroform/isoamyl alcohol was added and vortexed to obtain a milky emulsion. The tubes were spun for 10 minutes at 12,000rpm, separating the liquid in the tubes into two layers. The top aqueous layer containing the DNA was removed from the tubes using a disposable pipette and placed into a centrifugal filter unit (Amicon Corportion) and centrifuged at 5,000 rpm for 10 minutes. Following centrifugation, the bottom reservoir was removed from the unit and the liquid disposed of. An additional 500µl of TE was added to the filter unit and centrifuged and this process was repeated at least three times to clean up the DNA. After 3-4 washes the filter unit was inverted into a new 2.0ml tube and spun at approximately 3000rpm for 3 minutes to collect the DNA. After purification and concentration, the DNA was stored at -20°C.

For the semen sample DNA extraction, 150µL of TNE, 50µL of 20% Sarkosyl, 40µL of 0.39M DTT, 150µL of diH2O, and 10 µL of Proteinase K per sample was added to a 1.5ml tubes containing 25µl of samples. The mixture was vortexed briefly followed by quick spin in a microcentrifuge and incubated overnight at 37°C. After incubation the samples were cleaned and concentrated as discussed for the blood samples above. After purification and concentration, the DNA was stored at -20°C.

DNA Quantitation Using Agarose Gel

A 1% agarose gel was used to electrophorese all of the extracted DNA samples to determine the quality and quantity of the extracted DNA. Appropriate molecular weight markers were used along with samples. A mixture containing one µL of sample and two µL of bromophenol blue (BPB) loading dye was loaded into the wells of the agarose gel and was electrophoresed in 1x TAE (Tris/Acetic Acid/EDTA) buffer for 20 minutes at 120V. The gel was stained in an ethidium bromide solution and visualized using a UV Transilluminator. The gel images were photographed and stored in a computer.

Bisulfite Conversion

Approximately 200-300ng of DNA samples were used for bisulfite conversion using the Qiagen Epitect Bisulfite kit following manufacturer's recommendations (Qiagen Inc).

The following steps describe the details of the bisulfite conversion process:

1. The components required for bisulfite conversion were added to 0.2mL PCR tubes. The different components for conversion and the volume of each reagent for conversion are listed in Table 1.

Component	**Volume per reaction (μL)**
DNA solution (1 ng-2 μg)	Variable* (maximum 20 μL)
Deionized water	Variable*
Bisulfite Mix solution	85
DNA Protect Buffer	35
Total volume	**140**

Table 1 Bisulfite conversion reaction components

*The combined volume of DNA and RNase-free water is 20μL.

2. After a brief vortex and quick spin, the samples were placed in an ABI 9700 thermal cycler.
3. All samples underwent a three-stage cycling process as below:
    a. Denaturation at 95°C for 5 minutes, followed by 60°C for 25 minutes.
    b. Denaturation at 95°C for 5 minutes, followed by 60°C for 85 minutes.
    c. Denaturation at 95°C for 5 minutes followed by 60°C for 175 minutes
    d. Hold at 20°C until they are removed from the thermal cycler.

Cleanup of the converted DNA:

1. The converted samples were removed from the thermal cycler and the samples were transferred to a clean 1.5mL tubes.
2. 560μL of loading buffer (Buffer BL) containing 10μL of carrier RNA was added to each tube, and samples were vortexed and centrifuged briefly at 12,000 rpm.
3. Samples were transferred to *EpiTect*® DNA spin columns with collection tubes and centrifuged for 1 minute at 12,000 rpm.

4. The flow-through discarded, and 500μL of wash buffer (Buffer BW) was added to the spin columns.
5. Samples were centrifuged at 12,000 rpm for 1 minute, and the flow-through discarded.
6. 500μL desulphonation buffer (Buffer BD) was added to each column and was incubated at room temperature for 15 minutes.
7. Samples were spun at 12,000rpm for 1 minute the flow-through was discarded.
8. 500μL of wash buffer (BW) was added, followed by spinning the samples at 12,000 rpm for 1 minute. The resulting flow through was discarded.
9. Repeated step 8 one more time.
10. The columns were transferred to new 2mL tubes, centrifuged for another minute, followed by incubation at 60°C for 5 minutes with open lids to promote ethanol evaporation.
11. To elute the converted DNA from the spin columns, the columns were placed in clean 1.5mL tubes, and 20µL of eluting buffer was added directly onto the column membrane.
12. Allowed the tubes to incubate for 1 minute at room temperature and then eluted the DNA by centrifuging the samples at 12,000rpm.
13. Stored converted DNA -20°C until ready for use.

CpG marker selection and assay design

The potential markers for age determination and tissue identification were chosen from available literature (Lee et al., 2015). Several markers were selected for this study

and promising markers were studied in detail for age and tissue source determination. The CpG marker information, such as the location and position of the marker in the genomic DNA was used to download the DNA sequence information using the University of California Santa Cruz (UCSC) Genome Browser using the human genome assembly build 37 (GRCh37/hg19). After locating the sites within the human genome, a sequence of approximately 250 bases on both the 5' and, 3'sides were downloaded. This downloaded sequence was used to design primers for marker specific PCR amplification and sequencing using the Pyromark assay design software (Qiagen Inc.).

PCR Amplification of Target CpG Sites

Bisulfite-converted DNA was used for site specific amplification and approximately 10 samples from each tissue were used for tissue source determination. For age determination study, approximately 10 different semen samples were used for each of the following age groups: age 20+years, 30+ years, 40+years, and 50+ years. Table 2 describes the components of the PCR reaction in detail.

Bisulfite-converted DNA	2 μL
10x PCR Primer set	2 μL
Coral load solution	2 μL
Q-solution	4 μL
2x PyromarkPCR Master mix (Qiagen)	10 μL
Reaction Volume	20 μL

Table 2 The components and the volume of each for PCR amplification

Negative controls were included to check for contamination in reagents. Annealing temperatures were determined by subtracting 5°C from the melting temperatures (Tm) of the forward and reverse primers. The PCR cycling conditions for the amplification are described in Table 3.

Initial Incubation Step	Denaturation	Annealing	Extension	Final Extension	Final Soak
HOLD	**CYCLE (45 cycles)**			**HOLD**	**HOLD**
95ºC 15 min	94ºC 30 sec	Tm-5ºC 30 sec	72ºC 30 sec	72ºC 10 min	4ºC ∞

Table 3 Thermal cycling conditions for sample amplification.

To check the robustness of amplification, a 2% agarose gel quantitation was utilized. Two micro liters of amplicons along with loading dye was used for quantitation of the products. Appropriate molecular weight markers were used alongside the samples to determine the size of the PCR products.

Pyrosequencing

Pyrosequencing assays were created using the PyroMarkQ24 assay design software. Amplified samples were arranged in three rows of eight per the assay design. Eighteen µl of PCR products were aliquoted into each well of a 24 well format PCR tray containing 62µl of pyrosequencing cocktail (2µl streptavidin beads (GE health care), 40µl binding buffer, and 20µl diH_2O). The tray was covered with 8 well strip caps and shaken on a microplate shaker at approximately 1000 rpm for 10 minutes. After the agitation, the caps were removed, and the PCR products bound to the streptavidin beads

were lifted using a vacuum pump and processed using the PyroMark Q24 workstation. The streptavidin beads and bound PCR products were released onto a pyrosequencing plate containing 25µl of 1x sequencing primer in each well. The plate was incubated on a hot plate at 80°C for 2 minutes, then allowed to cool down at room temperature for 10 minutes.

The pyrosequencing cartridge was prepared with enzyme, substrate, and dNTPs in amounts determined by the PyroMark Q24 software. The samples were sequenced using the PyroMark Q24 pyrosequencer as per manufacturer's protocol. Once sequencing is completed, the Pyromark software was used to analyze the methylation data for each CpG site.

Data Analysis

Once the pyrosequencing data collection was complete, the results of the raw methylation data at each CpG site for each sample were determined using the PyroMark Q24 software (Qiagen). Once the methylation data was determined, the data was entered into an Excel spread sheet, grouping each tissue separately. Mean percent methylation and standard deviation values were calculated by averaging the methylation values at each CpG site for each group of tissues tested and each age group tested. Mean methylation values of the different tissues were compared using a one-way ANOVA and Tukey's Post Hoc pair-wise comparisons to determine if there were any statistically significant differences exist between the methylation values of different tissues studied and different age groups studied. Methylation differences were considered statistically significant when p- values are less than 0.05 ($p<0.05$). SPSS software package version 22 (IBM) was used for statistical analysis (Balamurugan et al., 2014).

CHAPTER IV – RESULTS

Locus specific primers were used to amplify the CpG sites of interest and prior to pyrosequencing, the efficiency of amplification for each marker was determined using a 2% agarose gel.

Age Marker in Sperm cg-6304190

For epigenetic marker cg-6304190, there were 13 sperm samples pyrosequenced for the 20 year old age group, 12 sperm samples for the 30 year old age group, 9 sperm samples for the 40 year old age group, and 10 samples for the 50+ year old age group. Figure 1 shows the histogram of average methylation levels and standard deviations for each age group at each CpG site for age marker cg-6304190.

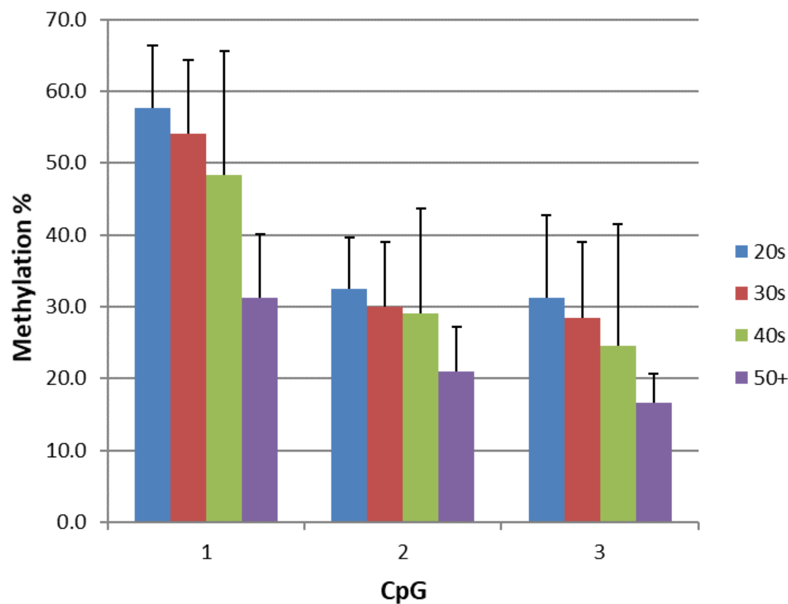

Figure 1. Histogram for marker cg-6304190

Figure 2 shows a scatter plot of CpG 1 of marker 6304190, where the x axis is the age of the individuals and the y axis is the methylation percentage.

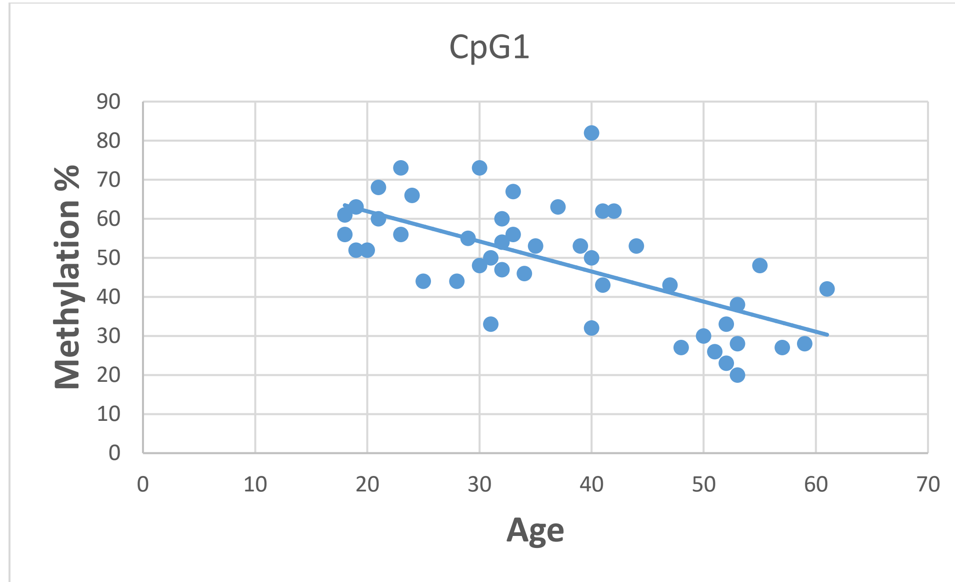

Figure 2. Scatterplot of CpG1 of marker cg6304190

Table 4 shows the averages of the methylation levels and standard deviations for each age group at each CpG site.

		CpG1	CpG2	CpG3
20s	Average +/- sd	57.7 ± 8.7	32.5 ±7.2	31.3 ± 11.4
30s	Average +/- sd	54.1 ± 10.2	30.0 ± 9	28.4 ± 10.6
40s	Average +/- sd	48.4 ± 17.1	29.1 ± 14.6	24.6 ± 16.9
50s	Average +/- sd	31.3 ± 8.9	21.0 ± 6.2	16.6 ± 4.1

Table 4. Percent Methylation Averages and Standard Deviations - Marker cg6304190

Statistical analysis was performed on the average methylation percentages for each age group compared to the others using one-way ANOVA and Tukey's HSD test. Tables 5 shows the significance for the 50s age group compared to the other age groups for marker 6304190.

Cg_6304190			
	CpG1	**CpG2**	**CpG3**
20s	0.000	0.030	0.010
30s	0.000	0.121	0.074
40s	0.009	0.254	0.406

Table 5. Levels of significance (p) for 6304190 (For 50+ compared to other age groups)

The p values for CpG 1 of this marker were only found to be below 0.05 between the 50+ age group compared to the other age groups. At CpG 2 and CpG 3, the p values were only below 0.05 when comparing the 20s age group to the 50+ age group. There is significant difference in methylation levels between age groups only where the p value is below 0.05. This shows that this marker was only able to differentiate the 50+ age group from the other age groups at CpG 1 and differentiate the 20s age group from the 50+ age group at CpG 2 and CpG3.

Age Marker in Sperm SLC22A18AS

For epigenetic marker SLC22A18AS, there were 13 sperm samples pyrosequenced for the 20 year old age group, 12 sperm samples for the 30 year old age group, 9 sperm samples for the 40 year old age group, and 10 samples for the 50+ year old age group. Figure 3 shows the histogram with the average methylation levels and standard deviations for each age group at each CpG site for age marker SLC22A18AS.

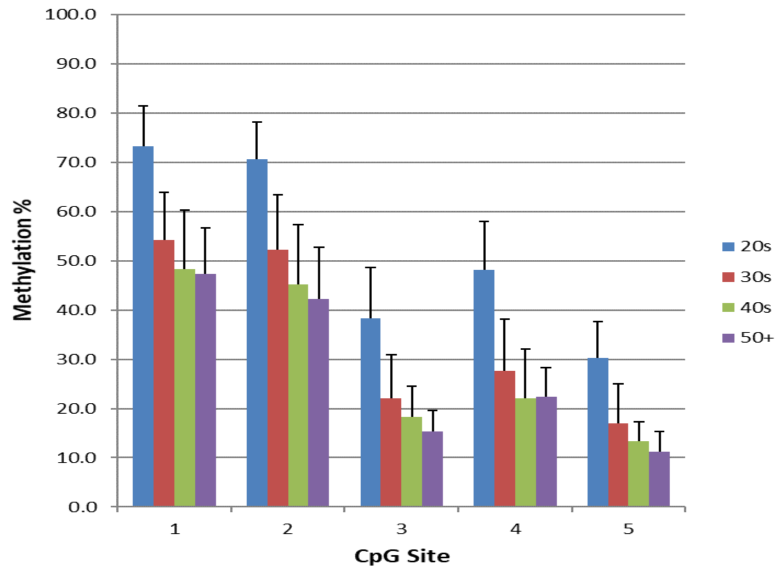

Figure 3. Histogram for SLC22A18AS

Figures 4 and 5 show scatter plots of CpG 2 and CpG3 of marker SLC22A18AS, with the x axis as the age of the individuals and the y axis as the methylation percentage.

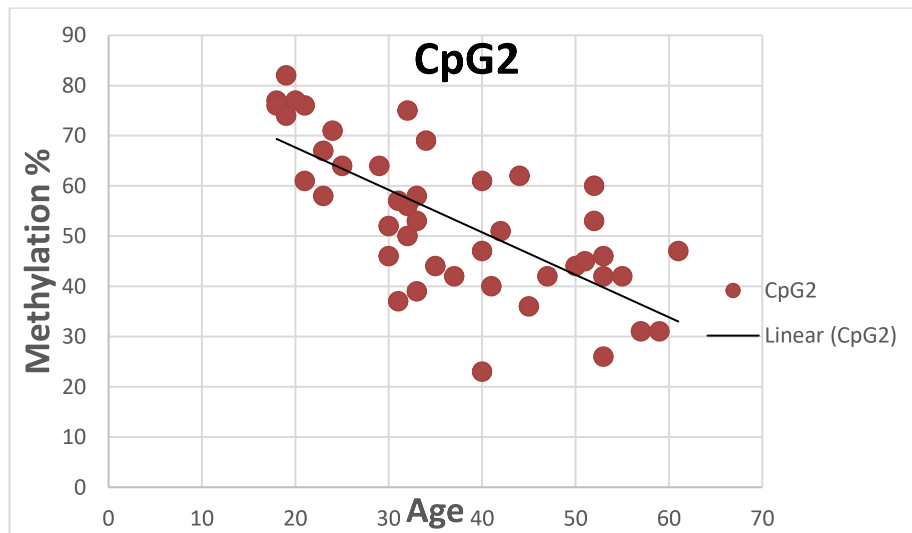

Figure 4. Scatterplot of CpG2 of marker SLC22A18AS

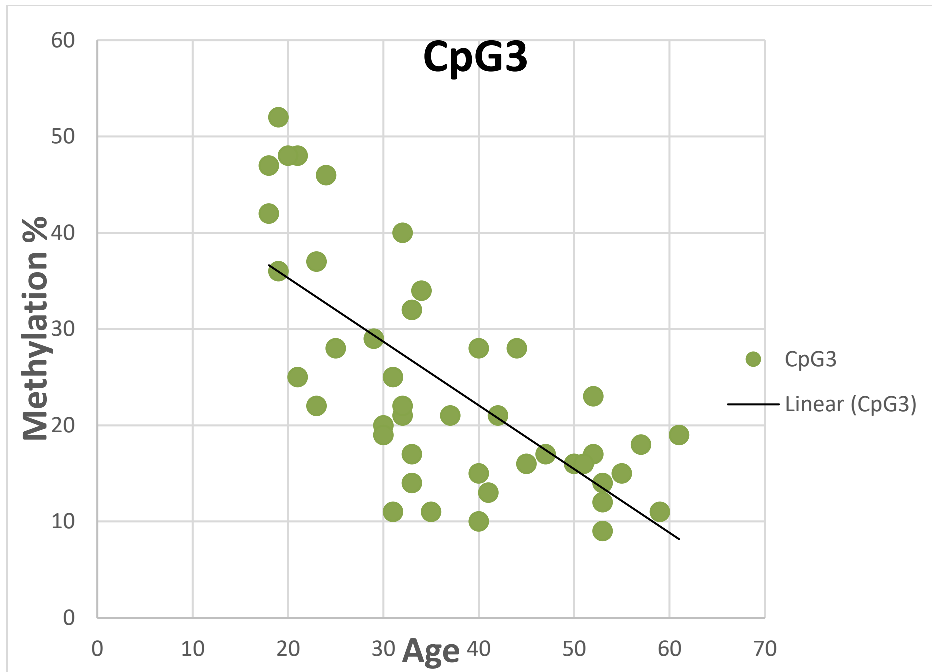

Figure 5. Scatterplot of CpG3 of marker SLC22A18AS

Table 6 shows the averages of the methylation levels and standard deviations for each age group at each CpG site.

		CpG1	CpG2	CpG3	CpG4	CpG5
20s	Avg +/-sd	76.0 ± 8.3	73.1 ± 7.6	42.1 ±10.3	51.4 ±9.8	32.6 ±7.3
30s	Avg +/-sd	54.4 ± 9.6	52.1 ±11.2	22.5 ±8.8	28.8 ±10.5	16.0 ± 8.1
40s	Avg +/-sd	48.2 ±12.0	45.1 ±12.2	18.2 ± 6.3	22.0 ±10.0	13.3 ±3.9
50s	Avg +/-sd	47.3 ± 9.4	42.3 ±10.5	15.4 ± 4.1	22.4 ±5.9	11.3 ± 4.0

Table 6. Percent Methylation Averages and Standard Deviations Marker SLC22A18AS

Statistical analysis was performed on the average methylation percentages for each age group compared to the others. Table 7 shows the significance for the 20s age group compared to other age groups for marker SLC22A18AS.

SLC22A18AS					
	CpG1	**CpG2**	**CpG3**	**CpG4**	**CpG5**
30s	0.001	0.001	0.001	0.001	0.000
40s	0.000	0.000	0.000	0.000	0.000
50+	0.000	0.000	0.000	0.000	0.000

Table 7. Levels of significance (p) for SLC22A18AS (For 20s compared to other age groups)

The p values for all CpG sites for this marker were only found to be below 0.05 between the 20s age group compared to the all other age groups. This shows that this marker was only able to differentiate the 20s age group from the other age groups at all CpG sites. There was no significant difference between any other age groups, as all other p values are above 0.05.

Age Marker in Sperm N23

For epigenetic marker N23, there were 13 semen samples pyrosequenced for 20-year-old age group, 13 semen samples for the 30-year-old age group, 9 semen samples for the 40 year old age group, and 10 semen samples for the 50+ age group. Figure 3 shows the histogram of the average methylation levels and standard deviations for each age group at each CpG site for age marker N23.

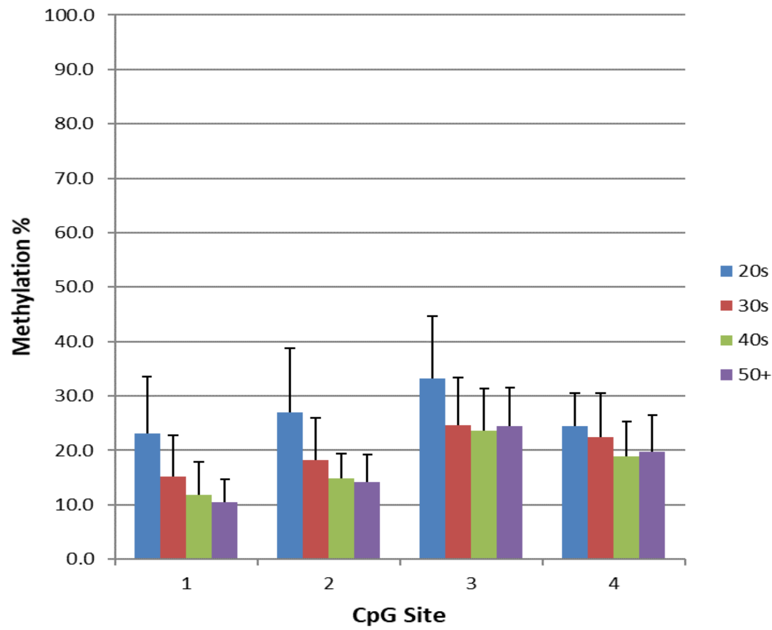

Figure 6. Histogram for marker N23.

Figure 7 shows a scatter plot of CpG 1 of marker N23, where the x axis is the age of the individuals and the y axis is the methylation percentage.

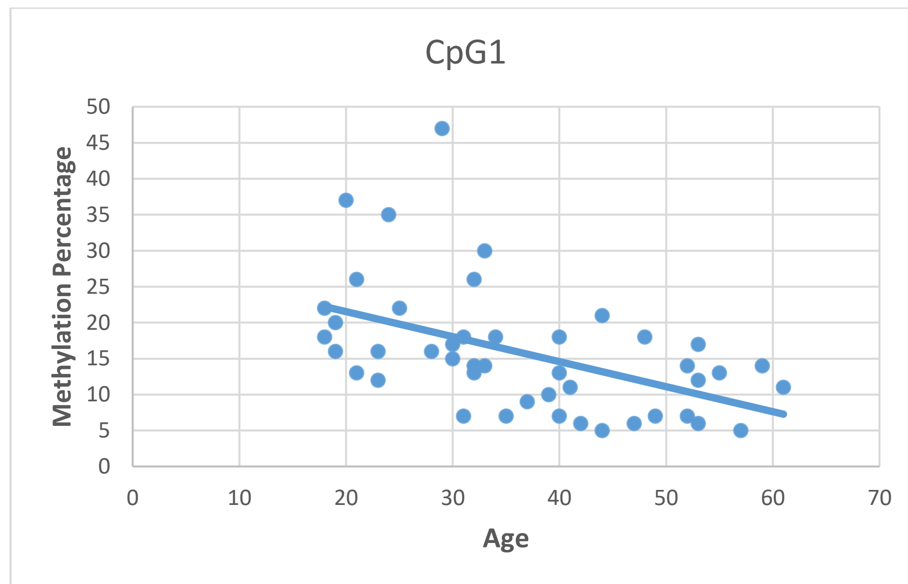

Figure 7. Scatterplot of CpG1 of marker N23

Table 8 shows the averages of the methylation levels and standard deviations for each age group at each CpG site.

		CpG1	**CpG2**	**CpG3**	**CpG4**
20s	Average (+/- s.d.)	23.1 ±10.5	26.9 ±11.9	33.2 ±11.4	24.4 ±6.0
30s	Average (+/- s.d.)	15.2 ±7.4	18.2 ±7.8	24.6 ±8.7	22.4 ±8.2
40s	Average (+/- s.d.)	11.8 ±6.0	14.8 ±4.7	23.7 ±7.7	18.9 ±6.5
50s	Average (+/- s.d.)	10.5 ±4.2	14.1 ±5.0	24.4 ± 7.0	19.7 ± 6.7

Table 8. Percent Methylation Averages and Standard Deviations for Marker N23

Statistical analysis was performed on the average methylation percentages for each age group compared to the others. Table 9 shows the significance for the 20s age group compared to other age groups for marker N23.

N23				
	CpG1	**CpG2**	**CpG3**	**CpG4**
30s	0.054	0.045	3.48451	2.62842
40s	0.007	0.007	3.85227	2.90583
50+	0.002	0.003	3.73672	2.81867

Table 9. Levels of significance (p) for N23 (For 20s compared to other age groups)

The p values for marker N23 were only below 0.05 when comparing the 20s age group to all others at CpG 1 and CpG 2, displaying a significant difference in methylation levels only between the 20s age group when compared to all others at these two CpG sites.

Sperm Specific Marker Hs_ZAP70_01

For the sperm specific marker, 11 sperm samples, 10 buccal samples, 10 vaginal epithelial samples, and 15 blood samples were used for pyrosequencing. Figure 8 shows the histogram with the average methylation levels and standard deviations for each tissue type at each CpG site for sperm marker Hs_ZAP70_01.

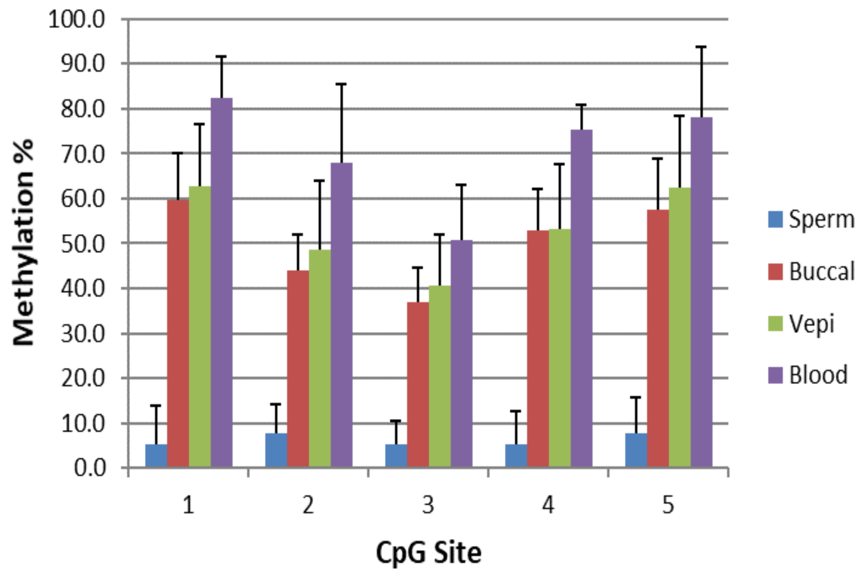

Figure 8. Histogram for Hs_ZAP70_01.

Table 10 shows the averages of the methylation levels and standard deviations for each age group at each CpG site.

		CpG1	**CpG2**	**CpG3**	**CpG4**	**CpG5**
Sperm	Avg ±s.d.	5.3 ±8.6	7.8 ±6.5	5.4 ±5.2	5.3 ±7.5	7.6 ±8.2
Buccal	Avg ±s.d.	59.6 ±10.5	44.0 ±8.0	36.9 ±7.8	52.9 ±9.2	57.4±11.4
V. Epi	Avg ±s.d.	62.7 ±14.0	48.7 ±15.3	40.7 ±11.2	53.3±14.3	62.4±16.0
Blood	Avg ±s.d.	82.5 ±9.2	68.0 ±17.4	50.7 ±12.4	75.4 ±5.5	78.0 ±15.8

Table 10. Percent Methylation Averages and Standard Deviations for Hs_ZAP70_01

Statistical analysis was performed on the average methylation percentages for sperm samples. It was found that the methylation data of sperm samples were statistically significant compared to other tissues ($p<0.05$).

The marker Hs_ZAP70_01 was found to differentiate sperm from blood, buccal, and vaginal epithelial cells.

Sperm Specific Marker Custom_MDFI 2

For epigenetic marker Custom_MDFI 2, there were 11 sperm samples, 13 buccal samples, 11 vaginal epithelial samples, and 13 blood samples used for pyrosequencing. Figure 9 shows the histogram with the average methylation levels and standard deviations for each tissue type at each CpG site for sperm marker Custom_MDFI 2

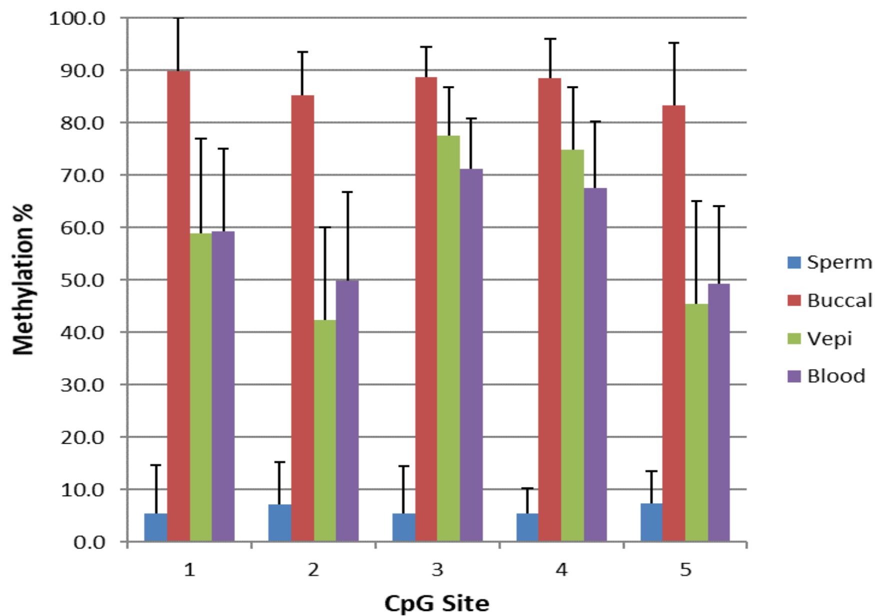

Figure 9. Histogram for Custom_MDFI 2 marker

Table 11 shows the averages of the methylation levels and standard deviations for each age group at each CpG site.

		CpG1	CpG2	CpG3	CpG4	CpG5
Sperm	Avg ± s.d.	5.5 ±9.1	7.2 ±8.1	5.5 ±8.9	5.5 ±4.8	7.4 ±6.2
Buccal	Avg ± s.d.	89.9 ±10.2	85.3 ±8.3	88.8 ±5.8	88.5 ±7.4	83.3 ±11.9
V. Epi	Avg ± s.d.	58.9 ±18.1	42.4 ±17.8	77.5 ±9.3	74.8 ±11.9	45.5 ±19.6
Blood	Avg ± s.d.	59.2 ±15.8	49.9 ±17.0	71.2 ±9.6	67.5 ±12.8	49.2 ±14.8

Table 11. Percent Methylation Averages and Standard Deviations for Custom_MDFI 2

Statistical analysis was performed on the average methylation percentages for each tissue type compared to the others. It was found that the methylation data of sperm samples were statistically significant compared to other tissues ($p<0.05$).

CHAPTER V – DISCUSSION

The ability to estimate the age of an unknown donor or identify tissue type of a biological sample can be of great importance in criminal cases. While tissue identification is already part of criminal investigations, it is a vital part of human identification and DNA profiling that can lead to the convictions or link an individual to a crime scene or piece of evidence. The type of tissue found can also assist investigators in determining what kind of crime occurred, and how violent that crime was. For example, a casual handshake may transfer some skin cells from person to person, while the presence of blood or semen indicates a violent act. There are already some tests in place to identify tissue type, but a method of testing that uses small amount of DNA would be extremely beneficial in cases with a very limited amount of biological material. Currently there is no method being practiced to estimate the age of an unknown donor of a biological sample. This would be extremely beneficial in criminal investigations to give investigators some sort of useful information when they have no suspects or no one to match a DNA sample to. Giving investigators the age range of an unknown individual whose sample was collected during an investigation would be highly beneficial in helping them narrow their search range, especially if it can be combined with other physical characteristics gleaned from future methylation studies.

The purpose of this study was to look at certain methylation markers in sperm to determine if an age range could be estimated based on the methylation percentages found in different age groups, and to determine if certain methylation markers could be used to distinguish sperm from blood, vaginal epithelial cells, and buccal cells based on the difference in methylation levels. The potential age markers were taken from a study in

which a single CpG site was reported as age specific, and this was modified to include several other CpG sites located upstream and/or downstream of that site in the testing area (Lee et al., 2015). The sperm specific tissue markers were chosen from preliminary data available from Balamurugan research group and adapted to determine their usefulness as tissue specific sperm markers, to see if they would be useful in identifying sperm from other forensically relevant tissues. In this study two sperm identification markers and three potential age predicting loci were identified and one age predicting locus is very promising.

Epigenetics is the process in which chemical changes occur to alter the gene expression without changing the DNA sequence. One of the epigenetic changes that occurs naturally in an organism is methylation (Vidaki et al., 2013). Methylation is the addition of a methyl group (CH3) to the 5-carbon of a cytosine that is followed by a guanine and this usually occurs in the promoter region of gene as a method to regulate gene expression. (Saabeha & Hasnain, 2018). This epigenetic change has been observed in all individuals and is preserved during cell division (Vidaki et al., 2013). The importance of predicting the age of an individual in a criminal investigation cannot be overemphasized and this is a rapidly developing area in forensic science.

While there were several potential age-related markers selected from literature, the most promising markers identified were cg-6304190, SLC22A18AS, and N23. The most promising markers for sperm identification were Hs_ZAP70_01 and Custom_MDFI.

For age marker cg-6304190, the methylation data shows that the biggest difference in methylation levels was the 50+ age group compared to other age groups.

Methylation levels in the 50+ age group were hypomethylated when compared to other age groups. For age marker SLC22A18AS, the methylation data shows that while there is a pattern of hypomethylation as age increases, there is only a significant difference in methylation at this marker when comparing the 20s age group to the rest of the age groups. The third age marker, N23, shows that there was only a significant level of difference in methylation levels in the first two CpG sites, and only between the 20s age group when compared to the other age groups.

Of the two tissue specific markers, Hs_ZAP70_01 showed that sperm samples were significantly hypomethylated ($p<0.05$) compared to blood, buccal, and vaginal epithelial samples. For the custom-MDFI marker, the methylation levels for sperm were significantly hypermethylated ($p<0.05$) compared to other tissues. These data provide sufficient evidence that the two markers listed above could be used to differentiate sperm samples from other biological samples.

As forensic DNA analysis continues to evolve, epigenetic modifications such as DNA methylation have received increasing attention by researchers in recent years as a potential avenue for forensic advantages. Studies have been done by researchers in recent years to determine the usefulness of DNA methylation as a tissue specific marker to a reasonable degree of success (Balamurugan et al., 2014, Park et al., 2014, Fu et al., 2015, Lee et al., 2015). Advances have also been made in using DNA methylation analysis as a method to estimate the age of an individual in studies done by other researchers using various tissues (Alsaleh et al., 2017, Freire-Aradas et al., 2016, Vidaki et al., 2016).

The potential advantages of methylation markers for age estimation are novel in that there are currently no reliable methods of determining the age of an unknown donor

of a biological sample other than anthropological studies. If a method, such as DNA methylation analysis, could be used to give investigators information about an unknown donor, this could lead the investigators in narrowing down their search parameters when attempting to locate a person of interest in a crime. If this could be added to other methylation studies being done to determine other physical characteristics such as hair color, eye color, ethnicity, or obesity, investigators could use this to get a clear idea of what individuals they should be searching for. The other advantage with methylation analysis a separate DNA extraction is not necessary since the DNA is already available for case work analysis. This could save time, sample, and money.

Even though conventional serological tests are used routinely in crime labs, the methylation analysis offers additional advantages. Conventional serological tests may offer false positive or false negative results especially when the protein in the sample is degraded while DNA is much more stable than proteins and works better on degraded DNA as well. Multiplexing in DNA analysis has become a routine practice in forensics and these methylation assays can be combined in a kit format that can save time and samples.

Once a marker has been identified and validated, it is necessary to make sure that the methylation data obtained is not influenced by external factors. More research is needed to study the effect of such external factors such as environment, ethnicity, diet, and smoking on DNA methylation. Certain diseases may also influence methylation levels, so further studies would need to be completed to ensure these do not affect the methylation pattern. (Vidaki et al., 2013).

In conclusion the available experimental evidence proves that the two markers, custom-MDFI and Hs_ZAP70_01 can be used successfully to differentiate semen samples from other bodily fluids. Also, the age specific methylation differences in the SLC22A18AS marker can be used to predict the approximate age of an individual.

APPENDIX A - Primer Specifications

Marker	Primer Sequence	PCR Product Size
6304190	Forward 5' TGTTAGGGGTGTTTTAGATGG Reverse 5' CAAAACAAAAAACCCCACCAATA Sequencing 5' GGTGTTTTAGATGGTTGA	115
N23	Forward 5' TGTGAGATAGGGATTTGTTAGTGGT Reverse 5' AAAACATCAACACAACCTAAAACTA Sequencing 5' GGATTTGTTAGTGGTGT	104
SLC22A18AS	Forward 5' TGAGAGAGTAGTAGAGGAAAGGTATTA Reverse 5' ATCTACAACCTCCATCTAACCAAC Sequencing 5' AGTTATTTTGTATATAAGTGGG	222
Custom_MDFI_CpGI20	Forward 5' GGGGGTTAGGAGAGTTTAAGA Reverse 5' ACACCAAACCACCTTTTCT Sequencing 5' AGTTTAAGAAGTGGGG	111
Hs_ZAP70_01	Qiagen premade-sequence not available	90

APPENDIX B –IRB Approval Letter

Office of Research Integrity

118 COLLEGE DRIVE #5125 • HATTIESBURG, MS | 601.266.6576 | USM.EDU/ORI

NOTICE OF INSTITUTIONAL REVIEW BOARD ACTION

The project below has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services regulations (45 CFR Part 46), and University Policy to ensure:

- The risks to subjects are minimized and reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered involving risks to subjects must be reported immediately. Problems should be reported to ORI via the Incident template on Cayuse IRB.
- The period of approval is twelve months. An application for renewal must be submitted for projects exceeding twelve months.

PROTOCOL NUMBER: IRB-18-134
PROJECT TITLE: DNA Methylation Analysis for Tissue and Age Determination for Forensic Human Identification
SCHOOL/PROGRAM: School of CJFS, Criminal Justice, Forensic Sci
RESEARCHER(S): Lauren Satcher, Kuppareddi Balamurugan

IRB COMMITTEE ACTION: Approved
CATEGORY: Expedited

2b. Collection of blood samples by finger stick, heel stick, ear stick, or venipuncture from other adults and children, considering the age, weight, and health of the subjects, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these subjects, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.

3. Prospective collection of biological specimens for research purposes by noninvasive means.

PERIOD OF APPROVAL: February 11, 2019 to February 11, 2020

Donald Sacco

Donald Sacco, Ph.D.
Institutional Review Board Chairperson

REFERENCES

Alghanim, H., Antunes, J., Silva, D. S., Alho, C. S., Balamurugan, K., & Mccord, B. (2017). Detection and evaluation of DNA methylation markers found at SCGN and KLF14 loci to estimate human age. Forensic Science International: Genetics, 31, 81-88. doi:10.1016/j.fsigen.2017.07.011

Alsaleh, H., McCallum, N. A., Halligan, D. L., & Haddrill, P. R. (2017). A multi-tissue age prediction model based on DNA methylation analysis. Forensic Science International: Genetics Supplement Series, 6, e62-e64.

Balamurugan, K., Bombardi, R., Duncan, G., & McCord, B. (2014). Identification of spermatozoa by tissue-specific differential DNA methylation using bisulfite modification and pyrosequencing. Electrophoresis, 35(21-22), 3079-3086.

Bell, J. T., Tsai, P., Yang, T., Pidsley, R., Nisbet, J., Glass, D., . . . Deloukas, P. (2012). Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. PLoS Genetics, 8(4). doi:10.1371/journal.pgen.1002629

Botstein, D., White, R., Skolnick, M., & Davis, R. (1980). Construction of a Genetic Linkage Map in Man Using Restriction. American Journal of Human Genetics, 32, 314-331.

Budowle, B., Smith, J., Moretti, T., & DiZinno, J. (2000). DNA Typing Protocols: Molecular Biology and Forensic Analysis. Natick, MA: Eaton Publishing.

Butler, J. M. 2010. Fundamentals of Forensic DNA Typing. London: Academic Press

Cho, S., Jung, S., Hong, S. R., Lee, E. H., Lee, J. H., Lee, S. D., & Lee, H. Y. (2017).

Independent validation of DNA-based approaches for age prediction in blood. Forensic Science International: Genetics, 29, 250-256. doi:10.1016/j.fsigen.2017.04.020

Edwards, A., Civitello, A., Hammond, H., & Caskey, C. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics, 49(4), 746-756.

Freire-Aradas, A., Phillips, C., Mosquera-Miguel, A., Girón-Santamaría, L., Gómez-Tato, A., Cal, M. C., . . . Lareu, M. (2016). Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. Forensic Science International: Genetics, 24, 65-74. doi:10.1016/j.fsigen.2016.06.005

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. Forensic Science International: Genetics, 5(5), 517-524.

Fu, X. D., Wu, J., Wang, J., Huang, Y., Hou, Y. P., & Yan, J. (2015). Identification of body fluid using tissue-specific DNA methylation markers. Forensic Science International: Genetics Supplement Series, 5, e151-e153.

Gaensslen, R. (1983). Sourcebook in forensic serology, immunology, and biochemistry. National Institute of Justice, US Department of Justice.

Gill, P., Jeffreys, A., & Werrett, D. (1985). Forensic application of DNA 'fingerprints'. Nature, 318(6046), 577-579

Hasnain, S. E. (2019). Forensic Epigenetic Analysis: The Path Ahead. Medical Principles and Practice.

Hong, S. R., Jung, S. E., Lee, E. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2017). DNA methylation-based age prediction from saliva: high age predictability by combination of 7 CpG markers. Forensic Science International: Genetics, 29, 118-125.

Huang, Y., Yan, J., Hou, J., Fu, X., Li, L., & Hou, Y. (2015). Developing a DNA methylation assay for human age prediction in blood and bloodstain. Forensic Science International: Genetics, 17, 129-136. doi:10.1016/j.fsigen.2015.05.007

Jenkins, T. G., Aston, K. I., Pflueger, C., Cairns, B. R., & Carrell, D. T. (2014). Age-Associated Sperm DNA Methylation Alterations: Possible Implications in Offspring Disease Susceptibility. PLoS Genetics, 10(7). doi:10.1371/journal.pgen.1004458

Lee, H. Y., An, J. H., Jung, S. E., Oh, Y. N., Lee, E. Y., Choi, A., ... & Shin, K. J. (2015). Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. Forensic Science International: Genetics, 17, 17-24.

Lee, H. Y., Jung, S., Oh, Y. N., Choi, A., Yang, W. I., & Shin, K. (2015). Epigenetic age signatures in the forensically relevant body fluid of semen: A preliminary study. Forensic Science International: Genetics, 19, 28-34. doi:10.1016/j.fsigen.2015.05.014

Lee, J. W., Choung, C. M., Jung, J. Y., Lee, H. Y., & Lim, S. K. (2018). A validation study of DNA methylation-based age prediction using semen in forensic casework samples. Legal Medicine, 31, 74-77.

Naito, E., Dewa, K., Yamanouchi, H., Takagi, S., & Kominami, R. (1993). Sex

determination using the hypomethylation of a human macro-satellite DXZ4 in female cells. Nucleic acids research, 21(10), 2533.

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., ... & Kim, Y. S. (2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. Forensic Science International: Genetics, 13, 147-153.

Park, J., Kim, J. H., Seo, E., Bae, D. H., Kim, S., Lee, H., . . . Kim, Y. S. (2016). Identification and evaluation of age-correlated DNA methylation markers for forensic use. Forensic Science International: Genetics, 23, 64-70. doi:10.1016/j.fsigen.2016.03.005

Saeed, M., Berlin, R. M., & Cruz, T. D. (2012). Exploring the utility of genetic markers for predicting biological age. Legal Medicine, 14(6), 279-285. doi:10.1016/j.legalmed.2012.05.003

Saiki RK, et al. (1985). Enzymatic Amplification of β-globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. Science, 230, 1350-1354.

Vidaki, A., Daniel, B., & Court, D. S. (2013). Forensic DNA methylation profiling—Potential opportunities and challenges. Forensic Science International: Genetics, 7(5), 499-507. doi:10.1016/j.fsigen.2013.05.004

Vidaki, A., Ballard, D., Aliferi, A., Miller, T. H., Barron, L. P., & Court, D. S. (2016). DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. Forensic Science International: Genetics, 28, 225-236. doi:10.1016/j.fsigen.2017.02.009

Weidner, C., Lin, Q., Koch, C., Eisele, L., Beier, F., Ziegler, P., . . . Wagner, W. (2014).

Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biology,15(2). doi:10.1186/gb-2014-15-2-r24

Yi, S. H., Xu, L. C., Mei, K., Yang, R. Z., & Huang, D. X. (2014). Isolation and identification of age-related DNA methylation markers for forensic age-prediction. Forensic Science International: Genetics, 11, 117-125. doi:10.1016/j.fsigen.2014.03.006

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż, Pałeczka, A., . . . Branicki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. Forensic Science International: Genetics, 17, 173-179. doi:10.1016/j.fsigen.2015.05.001

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Makowska, Ż, Spas, A., Parys-Proszek, A., Kucharczyk, K., Ploski, R., Branicki, W. (2015). Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic

Zubakov, D., Liu, F., Kokmeijer, I., Choi, Y., Meurs, J. B., Ijcken, W. F., Uitterlinden, A., Horman, A., Broer, L., Duijin, C., Lewin, J., Kayser, M. (2016). Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. Forensic Science International: Genetics, 24, 33-43. doi:10.1016/j.fsigen.2016.05.014.