The dynamics of the hydroxymethylome and methylome during the progression of Alzheimer's disease

Smith, Michael Allen

http://hdl.handle.net/2144/15072

Boston University
THE DYNAMICS OF THE HYDROXYMETHYLOME AND METHYLOME DURING THE PROGRESSION OF ALZHEIMER’S DISEASE

by

MICHAEL A. SMITH

B.S., University of Minnesota- Twin Cities, 2010

Submitted in partial fulfillment of the requirements for the degree of Master of Science 2014
© 2014 by

MICHAEL ALLEN SMITH

All rights reserved
Approved By

First Reader

Karen Symes, Ph.D.
Professor of Biochemistry
Boston University School of Medicine

Second Reader

Yujiang Geno Shi, Ph.D.
Professor of Biochemistry
Harvard University School of Medicine

Third Reader

Christian Argueta, Ph.D.
Post-Doctoral Fellow
Harvard University School of Medicine
ACKNOWLEDGEMENTS

I would like to sincerely thank all who have helped me during the course of this project. A special thanks goes to Dr. Yujiang Geno Shi for all of his guidance, support, and valuable insights. Thank you to the members of the Shi lab, particularly to Dr. Christian Argueta, Dr. Li Tan, and Di Wu for all of their encouragement, personal help and training. I would also like to thank my advisor, Dr. Karen Symes for all of her effort and support.
THE DYNAMICS OF THE HYDROXYMETHYLOME AND METHYLOME DURING THE PROGRESSION OF ALZHEIMER’S DISEASE

MICHAEL A. SMITH

ABSTRACT

Alzheimer’s disease (AD) is a neurodegenerative condition affecting millions of individuals worldwide and is a major source of mortality in elderly populations. While it is well established that there is a strong genetic basis for the disease, the epigenetic mechanism underlying the disease is largely unknown. The main purpose of this thesis is to understand the alteration of epigenetic modifications associated with the disease and its progression. In particular, we examine how alterations in the cytosine methylation and cytosine hydroxymethylation, two epigenetic modifications that are critically important for the development and function of the brain, are associated with advancing stages of Alzheimer’s disease. Eight progressive AD brain samples were examined for changes in DNA methylation and hydroxymethylation by both dot blot analysis and a new oxidative bisulfite (OXBS) deep sequencing technology. The initial results of dot blot analysis reveal a statistically significant decrease in 5hmC associated with intermediate stage AD among the samples. This data suggests that the alterations in epigenetic modifications is likely associated with the pathophysiology of Alzheimer’s disease, not only shedding new light on our understanding of the epigenetics of the disease, but also providing the
basis for our future investigation on the exact cause and effect relationships of these epigenetic changes and their respective stages in Alzheimer’s.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
<td>i</td>
</tr>
<tr>
<td>COPYRIGHT PAGE</td>
<td>ii</td>
</tr>
<tr>
<td>READER APPROVAL PAGE</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xi</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Introduction to Epigenetic Regulation</td>
<td>1</td>
</tr>
<tr>
<td>Histone Modification</td>
<td>2</td>
</tr>
<tr>
<td>DNA Methylation</td>
<td>3</td>
</tr>
<tr>
<td>DNA Hydroxymethylation</td>
<td>5</td>
</tr>
<tr>
<td>Oxidative Bisulfite Deep Sequencing</td>
<td>7</td>
</tr>
<tr>
<td>Introduction to Alzheimer’s Disease</td>
<td>12</td>
</tr>
<tr>
<td>The Relationship Between Epigenetics and Alzheimer’s Disease</td>
<td>21</td>
</tr>
<tr>
<td>GOALS AND QUESTIONS</td>
<td>26</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>29</td>
</tr>
<tr>
<td>RESULTS</td>
<td>37</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>45</td>
</tr>
<tr>
<td>FUTURE ENDEAVORS</td>
<td>48</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>51</td>
</tr>
</tbody>
</table>
WORKS CITED ................................................................................................................................. 52

CURRICULUM VITAE ....................................................................................................................... 60
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identification of Brain Sample Tissue</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>Concentrations, Volumes, and Total Obtained Genome from Samples</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>Concentrations, Volumes, and Total Obtained Genome after Concentration</td>
<td>38</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pathways of Dynamic Cytosine Methylation and Demethylation</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Pathways of Cytosine Conversion during Reaction with Sodium Bisulfite (Huang et al. 2010b)</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Strategy of Identifying all Variations of Cytosine (Booth et al. 2012)</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Pathways of APP Cleavage (Cole and Vassar 2007)</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Pathways of amyloid recycling via APOE (Hirsch-Reinshagen, L. Burgess, and Wellington 2013)</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Dot blot of 5hmC expression</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Normalized presence of 5hmC</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>Dot blot of 5mC expression</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>Normalized presence of 5mC</td>
<td>43</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

5caC..............................................................5-carboxycytosine
5fC..............................................................5-formylcytosine
5hmC..........................................................5-hydroxymethylcytosine
5mC............................................................5-methylcytosine
AB......................................................................Alpha-beta
ABCA.......................................................ATP-binding Cassette Transporter
AD......................................................................Alzheimer’s Disease
APP.....................................................................Amyloid Precursor Protein
ADAM..........................................................A Disintegrin and Metalloprotease
AICD..................................................................APP Intracellular Domain
APOE.....................................................................Apolipoprotein E
BIN1..................................................................Bridging Integrator 1
BS-seq............................................................Bisulfite Sequencing
CDK5..................................................................Cyclin-Dependent Kinase 5
CLU......................................................................Clusterin
CpG......................................................................Cytosine phosphate Guanine
ddNTP............................................................Dideoxyribonucleotide Triphosphate
DNA.....................................................................Deoxyribonucleic Acid
DNMT..................................................................DNA Methyltransferase
ECL......................................................................Enhanced Chemiluminescence
EDTA....................................................................Ethylenediaminetetraacetic acid
ERG1..................................................................Early Growth Response 1
EOAD.................................................................Early Onset Alzheimer’s Disease

gDNA..............................................................Genomic Deoxyribonucleic Acid

GSK-3B.............................................................Glycogen Synthase Kinase 3 Beta

GWAS.............................................................Genome Wide Association Studies

KO..............................................................................Knockout

LOAD...............................................................Late Onset Alzheimer’s Disease

MAPT...............................................................Microtubule-Associated Protein Tau

NF-kB.........................................................Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B cells

NGS...............................................................Next Generation Sequencing

PBS............................................................................Phosphate-Buffered Saline

PBST..............................................................Phosphate-Buffered Saline Tween

PCR........................................................................Polymerase Chain Reaction

PICALM.........................................................Phosphatidylinositol-Binding Clathrin Assembly Protein

PSEN.............................................................................Presenilin

OXBS-seq..........................................................Oxidative Bisulfite Sequencing

RAGE..............................................................Receptor for Advanced Glycation Endpoints

SERM..............................................................Selective Estrogen Receptor Modulators

SOP.............................................................................Standard Operating Procedure

SORT...........................................................................Sortilin

TE..............................................................................Tris-EDTA

TET...........................................................................Ten-Eleven Translocase
INTRODUCTION TO EPIGENETIC REGULATION

*Epigenetics* refers to any type of inherited or acquired alteration to gene expression that is not due to changes in the ATCG coding sequence of the nucleic acid. Rather than focus on the largely immutable base-pair sequence-specific effects on genetic phenotype, the study of epigenetics describes how dynamic modifications to a genome affects the activity of genomic information. Chemical modifications to DNA and chromatin by extrachromosomal enzymes, often in response to extracellular signaling, can augment or silence gene expression. (Rosenfeld 2010) These changes can be innate or acquired, and inheritable or non-inheritable. It is by these mechanisms that a wide diversity of differentiated cells found in multicellular organisms can arise from a single totipotent stem cell despite negligible alterations to the genetic code during the development process. (Arney and Fisher 2004) These processes are also involved in cellular homeostasis. The role of these processes in human disease, particularly in cancer, metabolic disorders, and neurological disorders hold a great deal of promise for next-generation treatments. (Martin, Cropley, and Suter 2011)

The need for dynamic and controlled expression of genomic information is paramount in multicellular organisms with many unique and specialized tissues. For instance, despite having the same genetic code as a cardiomyocyte, a heptocyte must produce several enzymes required for functions specific to the liver such as gluconeogenesis or ketogenesis all while specifically avoiding any production of cardiomyocyte-specific
proteins involved with contraction. At the same time, a cardiomyocyte, although it still possesses the genetic information to do so, does not produce enzymes such as ethanol dehydrogenase, cytochrome 450, or pyruvate decarboxylase which are produced at high levels in the liver. A cardiomyocyte will not produce any form of glucagon or insulin such as a cell in the pancreas nor will it adopt the unique shape of a neuron.

The specialized epigenetic modification of the genome is a vital part of embryological development as the single totipotent zygote produced by fertilization gives rise to the fully differentiated cells and specialized tissues of an organism. Undifferentiated embryonic stem cells demonstrate high global levels of genome transcription. (Arney and Fisher 2004) However, as the embryo develops and the embryonic cells progress towards differentiation, several cellular mechanisms reprogram the genome. These mechanisms are promoted by extracellular cytokines, hormones, and growth factors. (Loebel et al. 2003) There are two primary mechanisms by which the genome can be modified in order to alter gene expression—usually in response to messaging from these signaling molecules. Those two mechanisms are histone modification and DNA methylation.

**HISTONE MODIFICATION**

DNA in mammalian cells is tightly packaged in the nucleus using highly evolutionarily conserved histone proteins. This complex of histone protein and DNA is called chromatin. The basic unit of chromatin, which is compromised of an octomer of smaller globular proteins H2A, H2B, H3, and H4, is called a nucleosome. In the nucleosome, the histones are arranged in a fashion that leaves the N terminus, or histone tail, of the
histones unfolded and exposed. These histone tails are subject to heavy short term regulation by cellular enzymes that are capable of methylating, acetylating, phosphorylating, sumoylating, ubiquitinating and glycosylating specific amino acids on those tails. The modifications on those tails chemically control the spacing between each histone octomer and thereby control the overall structure of chromatin. Two general states of chromatin can be found across a packaged genome. Certain regions of chromatin are tightly and densely packaged and allow little transcription because they are not easily accessed by transcription factors. Such “closed” regions are called heterochromatin. In contrast, euchromatin is “open” and allows for high levels of transcription due to a loose packaging of histone proteins. Epigenetic control via histone modifications is a generally short-term mechanism. (Kraushaar and Zhao 2013)

**DNA METHYLATION**

DNA methylation is a highly established form of long-term epigenetic regulation that can be passed down between a parent and daughter cell. DNA methylation at the 5’ position of cytosine (5mC) is mediated the DNA methyltransferases (DNMTs) family of enzymes, which transfer a methyl group from S-adenosyl-L-homocysteine to their DNA substrate. There are five important DNMTs required for the establishment and maintenance of methylation across the genome. They are DNMT1, DNMT2, DNMT3a, DNMT3b, and DNMT3L, although DNMT2 specifically modifies RNA and DNMT3L does not have any catalytic activity. Due to the abundance of 5mC, it is considered the “fifth base” of DNA. The methylation process often occurs at the promoter regions of DNA containing
repeating clusters of cytosine and guanine dinucleotides called CpG islands. Such dinucleotides are important sites of promoter expression. Hypermethylated CpG regions show reduced expression while hypomethylated regions show increased expression. Presence of a methyl group at a CpG island recruits a methyl CpG Binding protein (MeCP1/2) that blocks the binding of transcriptional enzymes thereby silencing expression at this site. In addition, MeCP2 helps to recruit histone deactylase proteins to the region to promote the formation of heterochromatin. (Nan et al. 1998) Although there will be little emphasis on histone modification in this study, the connection between methylation regulation and histone modification regulation should be stressed.

There are two primary types of methylation: de novo methylation and maintenance methylation. De novo methylation is the process by which undifferentiated cells have their genome specifically methylated in order to designate the cell down a particular route during the process differentiation. (Laurent et al. 2010) This reprogramming process is achieved by DNMT 3a/b.

Maintenance methylation is a dynamic process by which the differentiated cell types both are able to maintain specific methylation patterns as the cells divide and replicate and also to respond to environmental signals. This process is catalyzed by DNMT1. As a cell divides and the genome replicates, the methylated parent strand serves as a template for the formation of an unmethylated daughter strand. This methylated parent and unmethylated daughter strand are called hemimethylated. The DNMT1 then uses the
hemimethylation as a template for full methylation of the daughter strand. (Chen et al. 2014)

DNA HYDROXYMETHYLATION

Though it is a long-term modification, DNA methylation is not permanent nor is it immutable. The cell is able to effectively de-methylate the genome as needed. The 5 hydroxylmethyl cytosine (5hmC) serves as an intermediate for demethylation of the CpG island. (H. Wu and Zhang 2011) This conversion of 5mC into 5hmC is accomplished via the action of the ten-eleven translocase (TET) family of enzymes. These enzymes use an alpha ketogluterate and oxygen to oxidize the methyl group on 5mC to make 5-hydroxymethylcytosine (5hmC). Hydroxymethylcytosine can be further formylated and carboxylated to make 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) respectively. These modifications can serve as an intermediate for the restoration of cytosine and the ultimate reactivation of the promoter region although the exact mechanism by which this orders is still contentious. (L. Tan and Shi 2012) One possible mechanism of demethylation is called passive demethylation. This may occur as the genome replicates. DNMT1 may be unable to recognize 5hmC the same way it would recognize 5mC. Therefore methylation of hemimethylated DNA does not occur and daughter strands of DNA fail to become methylated like the parent strand. More direct mechanisms have been proposed. Although their existence is currently unconfirmed, decarboxylases and deformylases would be capable of recognizing 5caC and 5fC
respectively and removing those modifications to restore unmethylated cytosine. Dynamic regulation of methylation and hydroxymethylation is very important for embryonic development and for maintenance of cellular homoeostasis.

There are three Tet isoforms: TET1, TET2, and TET3. TET proteins are very important regulators of embryological neural development as well as the establishment of pluripotency in early embryological development. (Mayanil 2013) (Costa et al. 2013) TET1 and TET2 are expressed highly in embryonic stem cells and their expression is regulated heavily by OCT4 and Nanog, two transcription factors necessary for
maintaining pluripotency. (Costa et al. 2013) (Y. Wu et al. 2013) Consistently, control of methylation of CpG islands is an important factor in neural development. (Mayanil 2013) TET3 in particular is very important for neural development. (Xu et al. 2012) 5hmC is found in higher levels in the brain than in any other tissue in the body. This high level of hydroxymethylation corresponds with very high levels of gene activity. However, 5hmC in the brain seems to act as a stable epigenetic mark rather than as an intermediate for the eventual demethylation of 5mC. (Hahn et al. 2013) Altogether, these findings suggest that the dynamic regulation of DNA methylation and hydroxymethylation plays a key role the formation and maintenance of the brain. Already, there is a clear link between epigenetic dysfunction and several neurological diseases including Rett’s Syndrome, fragile X syndrome, Prader-Willi syndrome, Angelman syndrome, and Kabuki syndrome as well as with more ubiquitous diseases such as autism spectrum disorder, depression, schizophrenia, and chemical addiction. (Rangasamy, D’Mello, and Narayanan 2013)

In addition to having roles in cellular differentiation, epigenetic mechanisms are highly involved in cellular metabolism and mammalian physiology. The dysregulation of several epigenetic processes are responsible for many forms of human disease.

**OXIDATIVE BISULFITE DEEP SEQUENCING**

This study will utilize a new sequencing technique called oxidative-bisulfite deep sequencing in order to distinguish the hydroxymethylome and methylome of any particular tissue sample. This technique stems from previous methods of determining genomic sequences. In order to properly establish the true methylome and
hydroxymethylome, it is necessary to first establish a next generation sequencing library for that particular genome.

A next generation sequencing library is a full-genome sequence contained in a collection of fragments of genome that have all been ligated to a specific adaptor sequences. These fragments can then be sequenced using Illumina dye sequencing. Illumina dye sequencing operates by a form of shotgun sequencing. The genome must first be fragmented. There are several logistical reasons for this fragmentation. One is to break up the condensed genome into smaller linear fragments so that they can physically be read. The other and more important reason is so that many fragments can be sequenced simultaneously. This allows for much faster sequencing of the whole genome. Once the strand is fragmented, any 3’ overhangs are repaired. Based on the specific size restrictions of Illumina sequencing, only fragments of approximately 350 bp are selected. Then, the fragments are ligated to specific methylated adaptor sequences.

The adaptors are unique to the technique of Illumina sequencing. They are complementary to the annealed sequences at bottom of an Illumina flow cell- a transparent membrane. Fragments with the ligated adaptor sequences are able to bind to the bottom of these flow cells. The adaptors also serve as binding sites for PCR primers that will repeatedly replicate a fragment by several orders of magnitude until it takes up an approximately 1 micron section of the flow cell called a DNA cluster. (Fang et al. 2011) Each cluster has a specific (x,y) coordinate on the blot cell, is made up of several copies of the same unique sequence and can be recognized and analyzed by a computer.
The actual sequencing strategy of shotgun sequencing varies from Sanger sequencing, the older generation form of sequencing that was used to during the human genome project. Sanger sequencing adds a non-reversible color specific phosphorescent ddNTPs to the 3’ end of each genomic fragment and separates the fragments by size in order to establish the sequence. (Morozova and Marra 2008) Shotgun sequencing works as a PCR reaction that can add one reversible florescent ddNTP at a time. The color of an entire DNA cluster will change as each ddNTP is added. As each base is added, the computer will record the color of each cluster and store the sequence of each one, allowing for the quick sequencing of millions of fragments at once. Once the base identity has been determined, the PCR reaction adds another base until it reaches the end of the fragment. The sequence of the entire genome is then reassembled due to overlapping sequences of the fragment.
Bisulfite sequencing (BS-seq) is capable of distinguishing modified cytosine from modified cytosine and thus can shed light on some of the epigenetic modifications on a strand of gDNA. Sodium bisulfite adds as a selective deaminating agent for cytosine. The addition of sodium bisulfite replaces the amine group on cytosine with a sulfite group to form a stable 6-sulfonyluracil intermediate. This sulfite group can then be removed completely in high pH. (Huang et al. 2010a) Upon deamination, cytosine becomes uracil. (Huang et al. 2010b) PCR amplification of that converted strand will replace that uracil
with thymine and will be read as such. However, methylated and hydroxymethylated cytosine resists the deamination effects of the bisulfite and will still read as cytosine during sequencing. (Booth et al. 2013) Specifically, the methyl group of 5mC slows the deamination process by two full orders of magnitude while 5hmC reacts to form cytosine-5-methylenesulfonate which still is read as cytosine. (Huang et al. 2010a) This is why it is particularly important that the adaptors of the NGS library be methylated. If otherwise, the sequence of the adaptors would change upon exposure to the sodium bisulfite and they would not be able to anneal to the blot for later amplification and sequencing.

Of course, to see where the modifications have taken place, this bisulfite-treated sequence must be compared to the sequence of an untreated gDNA strand. This is accomplished
with bioinformatic techniques that are capable of analyzing all the genomic information from the two strands at once.

The major limitation of bisulfite sequencing is that it is incapable of distinguishing methylcytosine from hydroxymethylcytosine. However, one feature of 5hmC can be exploited to distinguish it from 5mC. 5hmC, when exposed to a selective oxidizing reagent potassium perruthenate (KRuO₄), will convert to 5fC. 5fC, unlike 5hmC, is capable of becoming deaminated by the addition of sodium bisulfite and will appear as uracil and then as thymine when the treated strand is sequenced. Therefore, it becomes possible to ascertain the location of hydroxymethylcytosines by comparing this sequence with that obtained by BS-seq and by next generation sequencing. Once is data is subtracted from the BS-seq data, all thymines not accounted for on NGS or BS sequencing correspond to 5hmC. Therefore, combination sequencing of BS-seq and oxBS-seq data will reveal the specific methylation and hydroxymethylation profiles of the genome. (Booth et al. 2013) Of course, the proper identification of the entire hydroxymethylome and methylome is dependent on comparing the sequences obtained by oxidative bisulfite sequencing and bisulfite sequencing to an unconverted next generation sequence.

INTRODUCTION TO ALZHEIMER’S DISEASE

The focus of this thesis will be to explore the epigenetic components that become deregulated and contribute to the progression of Alzheimer’s disease (AD). We intend to identify epigenetic components that can be manipulated for the eventual treatment of AD.
AD is a neurodegenerative disease affecting approximately 5.2 million Americans and approximately 50 million individuals worldwide. It is particularly prevalent among geriatrics, affecting 11% of the population above 65 years of age and 32% of the population above 85 years of age. (Ferri et al. 2005) (Mayeux and Yaakov 2012) Although much still has to be learned about the specific pathology and etiology of the disease, it is a well established hypothesis that neurodegeneration occurs as a result of increased neural apoptosis and necrosis relating to the accumulation of extracellular β-amyloid fibrils and intracellular hyperphosphorylated tau protein into visible senile plaques. (Masters and Selkoe 2012) The degree to which these plaques have accumulated across different parts of the brain is measured as a particular value (I-VI) called the Braak scale. (Braak and Braak 1991)

The precise mechanism by which the plaques lead to neurodegeneration remains controversial. Several pathways are associated with the progression of the disease. Defects in endocytosis and endocytosis may lead to increased extracellular deposition. Changes in lipid metabolism and clearance may be result in decreased clearance of the extracellular plaques. Changes in intracellular phosphatase and kinase activity promote the hyperphosphorylation of tau. The accumulation to plaques may lead to apoptosis via interference with proper neuronal calcium signaling (LaFerla 2002), mitochondrial oxidation and/or by capase activation via accumulation of AICD protein, a product of amyloid cleavage, in the endoplasmic reticulum (Kögel et al. 2012) (Takahashi et al. 2009). There is also evidence to suggest that accumulations of amyloid fibrils can activate microglial cells via the Receptor for Advanced Glycation End Products (RAGE).
Activation of this receptor leads to a NF-kb-mediated inflammation response which may contribute to neuron death. (Y. Wang et al. 2011) (Yan et al. 2012) Still other studies suggest that the accumulation of amyloid plaques are secondary effects of the disease and are not directly involved with the neurodegeneration. Instead, the same genetic defects and aberrant protein expression may lead to impaired protein and lipid transport that interfere with proper synaptic function. Such hypothesis are supported by studies in which patients present with amyloid plaques but without symptoms associated with Alzheimer’s disease.

The exact etiology of the disease is likely due to an accumulation of several interconnected cellular pathways. Both amyloid and non-amyloid pathways have a genetic and epigenetic basis.

Familiar early onset Alzheimer’s Disease (EOAD), which is characterized as AD onset at <60 years of age, can be traced to genetic polymorphisms of the amyloid precursor peptide (APP), Presenilin 1 (PSEN1), Presenilin 2(PSEN2) and Apolipoprotein E (APOE). The particularly pathogenic ε4 allele of the APOE is currently the strongest genetic risk factor for developing the disease. Individuals with a homozygous ε4 genotype have a 90% chance of developing the disease.(Bertram and Tanzi 2012)(De-Paula et al. 2012)(Tanzi 2012) These alleles are inherited in a Mendelian fashion. However, autosomal dominant mutations in the APP, PSEN1/2, and APOE gene regions accord only to cases of EOAD—a small fraction of total Alzheimer’s cases. (Lunnon and Mill 2013). Each are briefly discussed below.
**Amyloid Precursor Protein**

Although the exact function of APP is unknown, it is theorized to participate in the formation of new synapses. (Priller et al. 2006) APP knockout (KO) mice show declined cognitive function and increased gliosis. (Dawson et al. 1999) APP is an integral membrane protein. Mutations in APP may be associated with altered transport of the protein to the membrane and/or altered cleavage of the protein. Multiple lipid transport pathways regulate both the transport and the removal of APP from the cell membrane. Clathrin-coated pits have been shown to remove APP from the cell membrane via endocytosis.

Although expression of APP increases with age, its expression has been shown to be lower in an advanced AD brain. (Barger et al. 2008) Strangely, the amount of APP expression in neurons decreases with increasing proximity to the neuron to a neurofibillary Aβ tangle in patients with advanced Alzheimer’s disease. (Matsuyama et al. 2007)

Although expression of APP increases with age, its expression has been shown to be lower in an advanced AD brain.
Alpha Secretase

The Alpha secretase family of proteases is described as a family of a disintegrin and metalloproteases (ADAMs). Overexpression of a particular ADAM (ADAM10) has been demonstrated to improve the formation of new synaptic junctions. (Bell et al. 2008) Alpha secretase makes the first cleave in APP during non-amyloid-producing pathway.

Beta Secretase

The sequential cleavage of Beta secretase by alpha and then gamma secretase produces the non-pathogenic P3 peptide. In contrast, sequential cleavage of APP by beta secretase followed by gamma secretase produces a protein that can be further cleaved by epsilon secretase to produce AICD and the pathogenic Aβ_{42} polypeptide. (Cole and Vassar 2007)
PSEN1/2

Presenilin 1 and 2 (PSEN 1/2) are integral membrane proteins and components of the four protein presenilin complex. The presenilin complex is the catalytic domain of the gamma secretase. Mutations in PSEN1 are more associated with the development of AD than are mutations in PSEN2. (Suárez-Calvet et al. 2013) Mutations in the presenilins may be associated with altered cleavage of the APP protein or increased affinity for the APP β peptide- the product of beta secretase on the APP protein.

APOE

Apolipoprotien E is a lipid transporter unique to the brain. It is produced and secreted by astrocytes and microglial cells in the brain. Once secreted, it is the primary apolipoprotein of HDL in the brain. The brain is the most cholesterol-rich part of the body as it contains approximately 25% of the cholesterol in the body despite only comprising 2% of the total body mass. (Dietschy and Turley 2001) Much of that cholesterol is necessary in for the construction of complex lipids in the central nervous system such as the myelin sheath of neurons. Since cholesterol cannot pass the blood brain barrier, it must be produced in situ and removed via conversion to 24S-hydroxycholesterol. (Hirsch-Reinshagen, L. Burgess, and Wellington 2013) For that reason, precisely controlled lipid transport is essential in the brain.
APOE also transports β-Amyloid plaques. There are two pathways by which amyloid is removed from the brain. Receptor-mediated endocytosis through microglial cells and astrocytes is one path. Transport through the blood brain barrier is the other. Defects or underexpression in APOE can lead to Alzheimer’s symptoms both due to decreased Aβ removal and due to impaired cholesterol transport. Interestingly, APOE expression is increased by progesterones. (Fan et al. 2013) However, APOE is shown to be expressed at higher levels in the AD brain. It remains to be seen if this is a compensatory action or if it is directly involved in the pathogenesis. (Martínez-Morillo et al. 2014)

Other genetic contributions to the progression of the disease are being discovered. The genome wide association studies (GWAS) compiled genomic information from several
Alzheimer’s cases and identified several other genes that are associated with increased risk of the disease. Among them, mutations of the genes *ABCA7*, *BIN1*, *CD2AP*, *CD33*, *CLU*, *CR1*, *EPHA1*, *MS4A*, and *PICALM* and *URPG* have been shown to be associated with the disease. (Waring SC and Rosenberg RN 2008)

**CLU**

Clusterin, also known as Apolipoprotien J, is another apolipoprotein associated with lipid metabolism in the brain. It has two forms: a pro-apoptotic cytoplasmic form and secreted from the cell. Clusterin is expressed in high quantities in response to cerebral trauma or neuroinflammation. (Loison et al. 2006) Its effects are similar to those of APOE. It has been shown to bind with amyloid, prevent the formation of amyloid fibrils and assist with its removal from the brain. In addition, it has an important role in regulating neuroinflammation, cell death, and lipid metabolism. (Desikan RS et al. 2013)

**PICALM**

Phosphatidylinositol-binding clathrin assembly protein (PICALM) is essential in the formation of clatherin-coated pits that are needed for the removal of APP from the lipid membrane of the neuron. (Ando et al. 2013)

**ABCA7**
ATP-binding Cassette transporter 7 is a lipid transporter for immune cells. Expression of ABCA7 is increased by the presence of LDL and decreased by the presence of HDL. High levels of expression are associated with decreased AD risk, but expression also increases as the disease progresses, indicating that increased expression is a compensatory measure.

**BIN1**

Bridging integrator 1 (BIN1) is second only to APOE in its association with Alzheimer’s disease. It is present in all tissue by is highly expressed in brain and muscle tissue. It has several known functions, several of which may be associated with AD pathology. It is involved with the amyloid exocytosis pathway, calcium signaling. (M.-S. Tan, Yu, and Tan 2013) It is a myc-associated binding protein involved with apoptosis and also has interactions with tau. It has been shown to be highly upregulated in AD brains. (Chapuis et al. 2013)

**EGR1**

Early Growth Response 1 (ERG1) is a zinc finger transcription factor that promotes the production of Cdk5. Active Cdk5 phosphorylates Tau protein. AD brain samples have shown that ERG1 is upregulated as compared to healthy controls. This increased phosphorylation leads to decreased stability of neuronal microtubules (Lu et al. 2011) It is upregulated in response to a number of physiological factors including immune response, growth factor stimulation, and brain trauma. (Beckmann and Wilce 1997) (McMahon and Monroe 1996)

**THE RELATIONSHIP BETWEEN EPIGENETICS AND ALZHEIMERS**
**DISEASE**

Although there is a strong genetic basis for the pathology of Alzheimer’s Disease, inheritance of genetic polymorphisms account for less than 10% of AD cases. (Lunnion and Mill 2013) It is clear that classical genetics offers only partial explanations about the onset of disease.

Twin studies of monozygotic twins have shown that the development of the disease and the age of onset can vary considerably, suggesting that other factors may play a large role in the progression of the disease. (Räihä et al. 1997)

While Mendelian inheritance can explain some of the progression of the disease, epigenetic considerations may hold a promise of further understanding the disease and better treating it in the foreseeable future. Support for an epigenetic routes of disease progression emerged after it was observed that individuals with trisomy-21, also known as Down Syndrome, have drastically increased rates of Early Onset Alzheimer’s Disease. It has been hypothesized that the EOAD phenotype results from increased APP protein production due to the presence of the third copy of the APP gene, which is located on chromosome 21. (Wilcock and Griffin 2013)

The primary risk factor for developing AD is age. Even with individuals who carry risky alleles for the disease, the probability of developing the disease at a young or middle age is staggeringly low. There are several physiological changes in the body associated with advancing age including decreased sex steroid signaling, increased inflammation, altered glucose metabolism that are tightly involved with regulation of the epigenome.

Regulation of the DNA methylome has been shown to change with advanced age.
In the brain, age is positively correlated with decreases in methylation and increases in hydroxymethylation. (S.-C. Wang, Oelze, and Schumacher 2008a) (Wilson et al. 1987) The presence of amyloid plaques in particular seems to exacerbate dysregulation of the methyome in the brain. (Taher et al. 2013) In addition, altered methylation patterns in aging brains have been linked to dysregulation of APP and tau. (Iwata et al. 2013)

Newer developments have indicated several genes in addition to those with known disease-prone alleles that are differentially expressed in the case of Alzheimer’s disease compared to a healthy cohort. Several studies show that expression levels change both overtime during Alzheimer’s and temporally with different sections of the brain responding differently. (Winnie S. Liang et al. 2010) (W. S. Liang et al. 2008) Some of the involved genes are SORT1, ABCA1, MAPT, CDK5, and GSK-3B.

**SORT1**

Sortilin is a receptor on neurons that may play a role in the endocytosis of APOE and β-amyloid complexes. In vivo studies have shown that down-regulation of SORT1 leads to exacerbated Aβ disposition. (Rogaeva et al. 2007)

**ABCA1**

ATP-binding Cassette Transporter A1 is an important lipid transporter responsible for lipating APOE and the CNS HDL lipid transporters. Defects in ABCA1 lead to an 80%
loss of functional APOE. Overexpression of ABCA1 in mice leads to decreased amyloid deposition. (Hirsch-Reinshagen, L. Burgess, and Wellington 2013) (Wahrle et al. 2008)

**MAPT**

Microtubule-associated protein tau is a protein needed to stabilize axonal microtubules necessary for transport of neurotransmitters from the cell body to the synaptic junction. In addition, these microtubules are necessary for the formation of new synaptic junctions. Unphosphorylated tau proteins are necessary for building up the microtubules. The mechanism by which the microtubules are recycled and broken down involves phosphorylation of tau. Excess phosphorylation of tau leads to the formation of insoluble intraneuronal deposits of phosphorylated tau called neurofibrillary tangles- the main pathological hallmark of AD progression. MAPT is shown to be upregulated in an AD brain. There are two possible explanations for the link between increased MAPT expression and the formation of the tangles. One is that an originally increased MAPT expression invariably leads to more tau protein, phosphorylated or otherwise. Another possibility is that an original increase in the activity of tau kinases leads to a decrease in unphosphorylated tau and the increase in MAPT activity represents an epigenetic compensatory response on the part of the neuron. Tau is largely phosphorylated by CDK5 and GSK-3B. (Jayapalan and Natarajan 2013)

**CDK5**

Cyclin-dependent kinase 5 (Cdk5) is a proline-directed serine–threonine kinase that phosphorylates tau on S202, T212, S396, and S404. Increased activity of Cdk5 is
associated with the progression of Alzheimer’s diseases. (Shukla, Skuntz, and Pant 2012) The process by which Cdk5 activity is increased is well established.

**GSK-3B**

Glycogen synthase kinase 3 beta tightly regulates the activity of several enzymes involved with glucose metabolism including glycogen synthase. It has also been shown to phosphorylate several other enzymes. Glycogen synthase kinase-3B is upregulated in an AD brain. The mechanism by which it is upregulated may be that the extracellular accumulation of amyloid blocks Wing Integrase (wnt) signaling and thereby prevents inhibition of GSK-3B. (Hernández et al. 2010) Once active, GSK-3B is a kinase for both tau and presenilin. (Hernández et al. 2010) (Takashima 2006) It has also been shown that its activity is pro-apoptotic.

Though Alzheimer’s epigenetics is still in its infancy, several viable epigenetic mechanisms have been proposed. One major epigenetic process for which there is a strong association with Alzheimer’s disease is DNA methylation. DNA methylation of promoter regions is a key point of genetic regulation.

The genomic methylation profiles of Alzheimer’s patients vary from those of healthy individuals of the same age group. (S.-C. Wang, Oelze, and Schumacher 2008b) (Johansson, Enroth, and Gyllensten 2013) (Rao et al. 2012) (Bradley-Whitman and Lovell 2013) Some degree of differential DNA methylation is associated with aging (Bakulski et al. 2012) According to a hypothesis involving DNA-methylation, genetic factors promoting the formation of β-amyloid deposits as well as those promoting tau
production and phosphorylated are being increasingly overexpressed in Alzheimer’s brain, either by hypomethylation or by excessive hydroxymethylation, and/or genetic factors disrupting β-amyloid deposition and tau phosphorylation are down-regulated, likely because they are being inappropriately hypermethylated due to a loss of proper epigenetic homoeostatic controls including TET. For this report, we would like to examine the relationship between epigenetic modification of 5mC and 5hmC and the progression of the disease.

This examination will be achieved using a new sequencing technique called oxidative bisulfite deep sequencing which is capable of showing the entire hydroxymethylome and methylome. This study will be performed using eight brain samples at distinct stages in the disease progression.
GOALS AND QUESTIONS

A. Establish trends in methylation and hydroxymethylation in the genome as the disease progresses.

How are global DNA modifications altered during disease progression? We wish to establish a firm link between the progression of the disease and changes in global DNA modification during that progression. Changes in global trends of hydroxymethylation and methylation, though it can tell us very little about the specific expression patterns of many of the genes associated with AD, can provide two useful pieces of information. If a significant trend can be established, this trend may be able help predict or track AD progression in a clinical setting. More importantly for a laboratory setting, global changes in these methylation and hydroxymethylation can help establish which DNA-modifying enzymes are most involved with the progression of the disease.

This can fortunately be accomplished with simpler and cheaper techniques than sequencing. We will attempt to answer this question using a dot blot. Previous studies have shown that, compared to healthy individuals of the same age, hydroxymethylation and methylation levels are both lower in the AD brain. (Mastroeni et al. 2010) However, we hope to see at which points during the progression of the disease that these changes are most drastic.

B. Establish an epigenetic signature of the AD brain
The epigenetic signature of the AD brain refers to the unique epigenetic modifications that are present in the genome an AD brain that can be contrasted with those on a healthy genome. This goes beyond the general trends in global methylation and hydroxymethylation as mentioned the first goal. Specifically, we hope to see which unique modifications in which sections of the genome are most associated with the progression of the disease and how these modifications can serve as a basis for predicting the onset and progression of the disease.

C. Establish which genes are the most differentially expressed during the progression of the disease.

Which genes associated with the progression of the disease are activated or suppressed at which stages in the disease? If we have the sequences of the methylome and hydroxymethylome at different stages in the progression of the disease for specific genes, the data could provide more insights into the pathophysiology of the disease. For instance, we could distinguish which genes are differently regulated as consequences or as compensatory reactions to the AD progression and which ones are originally altered to manifest in AD symptoms.

D. Establish physiological mechanisms promoting the progression of the disease.

Should this study find a clear relationship between changes in DNA methylation of promoter regions of key Alzheimer’s genes and the progression of the disease, there will be fertile new research areas for Alzheimer’s study and treatment. Specifically, this will help establish an outline for the specific role that DNMTs, TETs, and other epigenetic
enzymes have in the disease and what extracellular and intracellular signals and pathways help localize these enzymes to the target sequences.

For instance, we have previously demonstrated that TET expression is altered by estrogen signaling in a breast cancer cell model. Increased estrogen signaling can increase the expression of TET2. However, this has not yet been demonstrated in a neuronal cell line though our group has evidence to suspect that some of this effect is preserved in neurons.

The risk of developing AD increases significantly after menopause. There is evidence to suggest that estrogen hormone replacement therapy (HRT) and the agonistic effects of Selective Estrogen Receptor Modulators (SERMs) can decrease the risk of developing Alzheimer’s disease. (Henderson 2009) Tamoxifen, an important SERM used to treat breast cancer, acts as an estrogen antagonist that can downregulate estrogen receptor in breast tissue. However, it has agonistic effects in other tissues including in bone and brain. Deciphering how the TETs are involved during the dynamic alteration of the hydroxymethylome and methylome and how these alter gene expression during the progression of Alzheimer’s disease presents a potentially fruitful avenue to better understand and treat the disease.
MATERIALS AND METHODS

Preparation of Brain Tissue

Brain samples were obtained through the Alzheimer’s Disease Center at Boston University School of Medicine. A total of eight samples were obtained. Two were disease-free control samples. Two were early-stage brain samples (Braak I/II), two were intermediate stage (Braak III/IV), and two were late stage (Braak V/VI) samples all obtained from the pre-frontal cortex. The pre-frontal cortex was chosen due to its relatively large size compared to other parts of the brain associated with AD such as the hippocampus. Samples were delivered on dry ice and stored at -20 degrees C. Once ready for preparation, brain samples were wrapped in aluminum foil and immersed in liquid nitrogen for one minute. Then, using a mortar and pestle, the brain samples were ground into a course powder that could be massed with greater ease. The brain samples did not readily stay in a solid state, resulting in the samples quickly thawing within seconds into a gelatinous and cohesive stage. This made massing and storing the sample quite difficult. This was remedied by constantly re-submerging ground samples in liquid nitrogen and massing while on dry ice. Once the ground samples were obtained, 25 mg were massed and placed in a 1.7 ml centrifuge tube on dry ice.

Genomic DNA Extraction

The extraction of genomic DNA was performed using a Qiagen DNeasy Tissue Kit. The samples were mixed with 180 ul of proteinase K digestion buffer with 20 ul of proteinase K and placed in 56 degrees C water bath and left to digest over night in order to fully
disintegrate the tissue sample. Then, 200 ul of a lysis buffer was added to break up the cells. 200 ul of ethanol was added in order to make the DNA less soluble. The solution was then passed through a spin column and membrane. The membrane was washed to remove all cellular components not bound to the genome. Finally, the genome was obtained by washing the membrane twice with 100 ul of elution buffer. Once obtained, the genomic DNA was quantified using a Nanodrop apparatus and again with a Qubit apparatus.

**Ammonium Acetate and Isopropanol Precipitation of Genome**

Initial genomic DNA yields as determined by Qubit analysis were particularly low. For that reason, genome extraction was repeated ten times for each sample of brain and combined together to form 1 ml of total elute. Once these samples were combined, the total elute was concentrated by adding 100 ul of 3M sodium acetate followed by 1.0 ml of isopropanol to precipitate the DNA. Once a cloudy precipitate was observed in each sample, the solution was centrifuged at 14,000 rpm for 5 minutes to pellet the DNA. The supernatant was aspirated and the DNA pellet was washed twice with 750 ul of 70% ethanol. The pellet was resuspended in 100 ul of ultra pure water and analyzed again with a Qubit apparatus.

**Dot Blot Analysis**

In order to confirm the presence of genomic DNA, a dot blot was performed using an anti-5hmC primary antibody. A dot blot is a quantitative method for detecting levels of a
target antigen using a primary antibody specific for the antigen of choice. In this experiment, antibodies for 5mC, 5hmC, 5fC, and 5caC were used. This experiment was performed first by obtaining an aliquot of each genome sample and sonicating it for 10 seconds to break up the genome into smaller segments. Two samples of genomic DNA from exactly 120 ng of each sonicated genome sample solution were obtained and placed in a 96 well PCR plate. The solution in each well was diluted to 60 ul using TE buffer.

Four two-fold serial dilutions for each sample were made across the PCR plate. To each well, 20 ul of 1M NaOH/25mM EDTA was added. The plate was then placed in a PCR machine for 10 minutes at 95˚ C in order to separate the strands of the double helix. 50 ul of cold ammonium acetate was then added to each well.

At the same time, a 0.2 um nitrocellulose membrane was cut and prepared. It was placed inside a Bio-Rad dot blot apparatus and hydrated with TE buffer. Genomic DNA from each well on the PCR plate was added to their respective wells on the dot blot apparatus and was bound to the membrane using vacuum filtration of the genomic solution. The membrane was washed with SSC solution to enhance genomic DNA binding. Once the DNA had been bound, the membrane was left in a 56˚ C incubator and left to dry for 20 minutes. Immediately after, it was placed under UV light for 20 minutes to promote cross linking of the genomic DNA to the membrane. After, it was placed for one hour in a 5% non-fat milk and Phosphate buffered solution + Tween (PBS-T) in order to prevent non-specific antibody interactions. The membrane was then left overnight in a 5% milk PBS-T solution with a rabbit 5hmC antiserum. The next day, it was washed three times with PBS-T and then reacted with a goat anti-rabbit serum. After washing three times with
PBS-T, the membrane was reacted with 5 ml of Western Lightning ECL luminol solution and 5 ml of Western Lightning ECL oxidizing solution. In a dark room, the membrane was used to expose medical x-ray film for 20 seconds. This procedure was repeated with primary rabbit antibodies specific to 5mC, 5fC, and 5caC. Due to limitations in the amount of DNA that could be responsibly devoted to the dot blot analysis, this procedure was performed once for each primary antibody.

**Illumina Dye Preparation of Genomic DNA**

*Fragmentation of DNA*

To generate a Next Generation Sequencing Library (NGS) obtained DNA was quantified using Qubit analysis. 50 ul of 20 ng/ul gDNA of each sample was obtained. The samples were placed in a 96 0.3 ml PCR plate and sonicated. 80 ul of magnetic sample purification beads were added to each well mixing evenly by pipetting several times. The samples were then left to incubate for five minutes. Then, using a magnetic stand, the samples were left sitting for eight minutes while the beads were pulled out of solution. Completion of this process was marked by the liquid becoming translucent. The supernatant was then removed with the PCR plate still on the magnet.

*Purification of fragmented DNA*

The beads were then washed twice by adding 200 ul of 80% ethanol, incubating the samples for 30 seconds and then removing the supernatant. The ethanol was allowed to evaporate. With the plate still on the magnetic plate, 52.5 ul of resuspension buffer was added to each sample. The plate was then removed from the magnetic plate. The solution
was pipetted repeatedly to resuspend the beads. At this point, the nucleic acid will elute from the beads. The plate was placed again on the magnetic plate. The supernatant from each sample was then removed and then placed in wells of a new 96 0.3 ml PCR plate.

**Repair of the 3’ Overhangs**

To each sample well in the PCR plate, 10 ul of Resuspension buffer followed by 40 ul of end repair mix 2 were added. The PCR plate was then placed in a thermocycler where it was preheated to 100˚ C and then run at 30˚ C for 30 minutes. This PCR reaction acts to seal any 3’ overhangs in the fragmented DNA so that all DNA fragments have blunt ends.

**Removal of Significantly Larger or Smaller than 350 bp**

Each sample was run on

**Adenylation of 3’ Ends**

In order to prevent fragments from ligating to each other and forming chimera fragments upon addition of the adaptors, each fragment was given a 3’ poly adenosine tail. To each sample, 2.5 ul of resuspension buffer followed by 12.5 ul of A-tailing mix was added. The PCR plate was then place in a thermocycler and preheated to 100˚ C. It was then run at 37˚ C for 30 minutes, 70˚ C for 5 minutes and finally at 4˚ C for 5 minutes.

**Ligation of Adapters**

To each sample, 2.5 ul of resuspension buffer followed by 2.5 ul of ligation mix 2 were added. Then 2.5 ul of DNA Adapter was added. The PCR plate was placed in a thermocycler and preheated to 100˚ C. The samples were incubated at 30˚ C for 10 minutes and then cooled to 4˚ C. 5 ul of Stop Ligation Buffer was added to each well following incubation. The fragments were then washed using 42.5 ul of magnetic sample
purification beads. The beads were once again washed by placing the samples on a magnetic bar, removing the supernatant and adding 200 ul of 80% ethanol twice. Once the ethanol was removed the second time, the samples were resuspended in 52.5 ul of resuspension buffer. This final Next Generation Sequencing Library (NGS) was then sent to Illumina Dye Sequencing in Shanghai, China.

**Oxidative Bisulfite Preparation of Genomic DNA**

*End Ligation of Genomic DNA*

Obtained genomic DNA was quantified using Qubit analysis. Obtained genomic DNA was treated with an end repair kit in order to seal any 3’ overhangs of the obtained genomic DNA.

*DNA Oxidation*

The sample of remaining NGS library was split in half with one half designated for DNA oxidation and one sample designated for mock DNA oxidation. Using the sample designated for DNA oxidation, a Bio-Rad P6 Micro-Bio spin column was prepared by flicking the column and centrifuging at 1000 x g in order to settle the spin column matrix. The top cap and bottom were removed and the column was placed in a 2.0 ml wash tube. Once placed in the tube, the column was centrifuged for 120 seconds at 1000 x g. The column was washed three times with 500 ul of Ultra Pure water and centrifuged at 1000 x g for 120 seconds a total of four times. 20 ul of gDNA of each sample was added its own respective column and centrifuged at 1000 x g for 120 seconds.
The gDNA from both the DNA oxidation samples and the DNA mock oxidation samples was then denatured by adding 1.25 ul of denaturing solution and enough Ultra Pure water to bring the final solution to 24 ul for each sample. This solution was added to a 0.2 ml tube and placed in thermocycler and incubated at 37˚ C for 30 minutes. After being allowed to cool on ice, each sample from the DNA oxidation designation was reacted with 1 ul of oxidant solution and placed back in the thermocycler for 30 minutes at 40˚ C. The mock oxidation samples were treated with 1 ul of ultra pure water and placed on the same thermocycler program. These samples were then centrifuged at 14000 x g for 10 minutes in order to remove any black precipitate.

*Bisulfite Conversion*

A bisulfite reagent solution was prepared by dissolving Bisulfite Reagent in 700 ul of Bisulfite Diluent for each sample. The solution was then incubated at 60˚ C for 15 minutes. This solution was added to 25 ul of the oxidized DNA solution and 5 ul of bisulfite additive and place in to a 0.2 ml PCR tube and centrifuged. These tubes were placed in a thermocycler. The thermocycler program followed the following program: 5 minutes at 95˚ C followed by 20 minutes at 60˚ C, five minutes at 95˚ C followed by 40 minutes at 60˚ C, 5 minutes at 95˚ C followed by 165 minutes at 60˚ C, 5 minutes at 95˚ C followed by 20 minutes at 60˚ C, 5 minutes at 95˚ C followed by 40 minutes at 60˚ C, and finally 5 minutes at 95˚ C followed by 20 minutes at 60˚ C. Then, the samples were centrifuged at 14000 x g to remove and salt contaminants. The supernatant was mixed with 310 ul of desulfonation buffer and transferred to an Amicon collection tube provided in the kit. The tube was centrifuged at 14,000 x g and the elute was discarded three times.
and then washed three times with wash buffer.

**PCR Amplification**

The converted DNA was then amplified using TrueMethyl polymerase buffer, dNTPs, primers specific to the NGS adaptors and Truemethyl polymerase. The tube was then placed in a thermocycler and run at 95°C for 30 seconds, followed by 60°C for 20 seconds followed by 72°C for 45 seconds fifteen times.
RESULTS

Barak Scores of Tissue Donors

Table 1: Identification of Brain Sample Tissue. Eight human post-mortem brain samples were obtained from the Boston University School of Medicine Center for Alzheimer’s Disease with known diagnosis and Braak stage of the tissue.

<table>
<thead>
<tr>
<th>N</th>
<th>PIN</th>
<th>Diagnosis</th>
<th>Braak Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10640</td>
<td>Control</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10499</td>
<td>Control</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10427</td>
<td>AD</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>21345</td>
<td>AD</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>21400</td>
<td>AD</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>21423</td>
<td>AD</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>10544</td>
<td>AD</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>21298</td>
<td>AD</td>
<td>6</td>
</tr>
</tbody>
</table>

N=8 samples of frozen prefrontal ctx (1 each of Braak I-IV, 2 Braak V, and 2 Braak VI) PMI unspecified

Table 2: Concentrations, volumes, and Total Obtained Genome from Samples. This data is obtained from samples before any efforts to concentrate them were executed. Oxidative bisulfite sequencing requires concentrations of at least 20 ng/ul and no above values match that requirement

<table>
<thead>
<tr>
<th>Identification #</th>
<th>Concentration (ng/ul)</th>
<th>Volume (ul)</th>
<th>Total DNA (ug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10640</td>
<td>8.07</td>
<td>1000</td>
<td>8.07</td>
</tr>
<tr>
<td>10499</td>
<td>13.5</td>
<td>1000</td>
<td>13.5</td>
</tr>
<tr>
<td>10427</td>
<td>15.5</td>
<td>1000</td>
<td>15.5</td>
</tr>
<tr>
<td>21345</td>
<td>15.9</td>
<td>1000</td>
<td>15.9</td>
</tr>
<tr>
<td>21400</td>
<td>18.3</td>
<td>1000</td>
<td>18.3</td>
</tr>
<tr>
<td>21423</td>
<td>7.18</td>
<td>1000</td>
<td>7.18</td>
</tr>
<tr>
<td>10544</td>
<td>8.05</td>
<td>1000</td>
<td>8.05</td>
</tr>
<tr>
<td>21298</td>
<td>8.82</td>
<td>1000</td>
<td>8.82</td>
</tr>
</tbody>
</table>
Brain samples arrived to the lab stored on dry ice. No information was known about the donors besides their AD status based on the presence or absence of AD-like symptoms and diagnosed Braak stage at the time of death. Other factors that we would have wished to know and control including gender, age at death, racial background, and classical genetic predisposition to AD were not disclosed to us.

Qubit Analysis

<table>
<thead>
<tr>
<th>Identification #</th>
<th>Concentration (ng/ul)</th>
<th>Volume (ul)</th>
<th>Total DNA (ug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10640</td>
<td>31.3</td>
<td>100</td>
<td>3.13</td>
</tr>
<tr>
<td>10499</td>
<td>32.3</td>
<td>100</td>
<td>3.23</td>
</tr>
<tr>
<td>10427</td>
<td>29.4</td>
<td>100</td>
<td>2.94</td>
</tr>
<tr>
<td>21345</td>
<td>35.2</td>
<td>100</td>
<td>3.52</td>
</tr>
<tr>
<td>21400</td>
<td>44.8</td>
<td>100</td>
<td>4.48</td>
</tr>
<tr>
<td>21423</td>
<td>23.8</td>
<td>100</td>
<td>2.38</td>
</tr>
<tr>
<td>10544</td>
<td>26.8</td>
<td>100</td>
<td>2.68</td>
</tr>
<tr>
<td>21298</td>
<td>56.6</td>
<td>100</td>
<td>5.66</td>
</tr>
</tbody>
</table>

Table 3: Concentrations, volumes, and Total Obtained Genome after Concentration. This data was obtained from samples after efforts to concentrate them using ammonium acetate and isopropanol precipitation of DNA. Though the concentrations were greatly improved by the process, a significant amount of genome for each sample was lost in the process.

Qubit analysis revealed initial difficulty of obtaining high yields of genomic DNA from the brain tissue. DNA concentrations were originally determined using NanoDrop technology. However, the recorded concentrations fell outside of the accurate reading range of NanoDrop technology. Original genomic DNA levels were very low (see table 2) The reason for the low yields of genomic DNA was originally thought to be due to defects in the Qiagen DNA tissue extraction kit. However, once the extraction was
repeated using a phenol/chloroform extraction, no visible pellet of precipitated DNA was observed, leading to the possibility that the tissue contained low levels of genomic DNA. (See table 2) Several consecutive extractions yielded high volumes of low-concentration DNA. The samples were then concentrated to yield low volumes of highly concentrated DNA. (See table 3) The total amount of DNA available for the experiment was a large limitation.

**Dot Blot Analysis**

We report that dot blot analysis has shown that all eight samples of brain genome contain high levels of 5hmC and 5mC but undetectable levels in 5fC and 5caC. However, we report a present but mostly statistically insignificant drop in 5hmC as the progression of the disease ensues. This is consistent with previously reported data suggesting that 5hmC decreases during AD but otherwise increases with age. We report no major change, significant or otherwise, in 5mC levels during the progression of the disease. Due to the scarcity of the genomic DNA for each Alzheimer’s sample, only one dot blot could be run for each sample for each primary antibody.

**5hmC**

The 5hmC dot blot revealed a visible trend of decreasing 5hmC levels with the progression of the disease. See figure 6 below.
The 5hmC blot was exposed for 25 seconds and scanned for analysis using the ImageJ program. Since the program measures the intensity of white contrasted against black, the image was converted to a negative and the average 5hmC expression per cell on the blot was measured using the white pixilation as a reference. High levels of expression corresponded with the lightest areas on the blot. Because the top row of the chart was overexposed, the next three rows were compiled together to generate the data. The expression levels were normalized by dividing each expression value by the largest value for each dilution. This way all dilutions were able to provide normalized data. The mean and the standard error were determined using the data sorted into four categories: healthy,
early stage, intermediate stage, and late stage. All normalized data across each row was used. See figure 7 for results. Although the graph appears to demonstrate a small inverse correlation between 5hmC levels and the progression of the disease, a two-tailed t-test was run in order to determine any significant statistical difference between healthy levels and the values for the different categories. This t-test compared 5hmC levels between the healthy control and different points in the disease progression. The tests revealed no significant difference between healthy and early stage 5hmC (p=.211). There was a difference between 5hmC levels of healthy controls and those of late stage Alzheimer’s disease. There appears to be a decrease in 5hmC in the late stage. However, these differences were not statistically significant (p=.093) There was, however, a significant difference

![Normalized 5hmC Presence](attachment:image.png)

**Figure 7: Normalized presence of 5hmC.** Values of 5hmC were obtained from the ImageJ analysis of the corresponding dot blot. The program recognized the average brightness of each well as a function of total pixilation over the area of the well. Pixilation values were normalized by dividing all values for a particular sample by the largest value corresponding to that dilution with 1 being the highest possible expression.
between healthy controls and the intermediate stage (p=.038). There is a statistical drop in 5hmC levels in the intermediate samples.

**Figure 8: Dot blot of 5mC expression.** All columns correspond to different samples. Samples are arranged from left to right in order of increasing progression of the disease. Theses samples were sorted into four categories with each category containing two samples. The first two correspond to healthy control. The second two (Braak III+BraakIV) are early stage Alzheimer’s. The next two (Braak V and Braak V) correspond to intermediate stage, and the final two (Braak VI and Braak VI) correspond to late stage Alzheimer’s. Rows on the blot correspond to levels of dilution of the DNA sample. Each descending row represents a 1:2 dilution.

**5mC**

The same could not be reliably said about the 5mC data. Even at first glance at the 5mC dot blot, there seems to be little obvious trend. This was confirmed once the expression was measured by ImageJ. One problem associated with the blot is the relatively high background exposure which caused areas not marked by DNA to become slightly darker as well. (see figure 8). Once the data was entered, no statistical significance could be established.
The p-values testing a significant difference between the healthy control group and the early stage, the healthy control group and the intermediate stage, and the healthy control group and the late stage were .450, .296, and .255 respectively. Since none of these numbers indicate significance, there is little that can be interpreted from this data.

5fC and 5caC

The dot blots for both of these DNA modifications appeared blank, even after exposing the film to the membrane for 15 minutes. We can infer that they were present at negligible levels in the genomes of these patients. Though we had predicted a negligible...
presence across the board, we had predicted before that there would be a decrease in 5fC and 5caC during the progression of the disease since 5hmC was predicted to decrease as well and 5hmC is a precursor to 5fC and 5caC. However, this prediction was not supported by the data.

*Sequencing Results*

At the time of the deadline for this thesis, the sequencing results are still pending. If requested by the Boston University School of Medicine, an addendum to this thesis containing the sequencing data will be included and analyzed.
DISCUSSION

At the point of writing and analyzing this thesis, the final results of the sequencing are currently pending. However, we were able to obtain a great deal of relevant evidence relating to the dynamics of DNA methylation and hydroxymethylation that will serve as the basis for future experiments, particularly in the Shi lab. The primary observation that can be drawn from this research is that there seems to be a correlation with the progression of the disease and a decrease in global genomic hydroxymethylation and a global increase in global genomic methylation due to comparisons of dot blots. Of course, it is hard to draw conclusions from the decrease in 5hmC. The only statistically significant drop in 5hmC presence occurs at the intermediate stage in disease progression. However, there actually seems to be a small increase in 5hmC levels as the disease progresses to the late stage. There are two possible explanations for this. One possibility is that the 5hmC levels as reported by intermediate data are artificially low. Indeed, there is an unmistakable artifact in one blot cell in the intermediate category that may be the result of non-uniform vacuuming of the dot blot apparatus. Another explanation is that there is a global increase in 5hmC in the final stage of Alzheimer’s disease resulting from some sort of compensatory action by the cell.

There is little to be said about the 5mC data. No statistically significant difference could be established between any two points in the progression of the disease. We had predicted that 5mC levels would drop slightly. Perhaps this trend will be explained once the sequencing data has arrived. It is almost certainly the case that the role that DNA methylation plays in the progression of the disease is best explained as a shift in location.
of methylation rather than as a global increase or decrease.

The complete lack of visible expression of either 5fC or 5caC is not entirely surprising. Whereas 5hmC in the brain is a stable epigenetic mark, these two DNA modifications serve as intermediates that are quickly removed. They are present in very low numbers and for very small periods of time.

There were several limitations that prevented us from establishing statistical significance in this study. The primary limitation was the relative difficulty of obtaining the number of brain samples required to establish a high power in the study. The successful negotiation of eight samples of human tissue for the sole purpose of promoting this master’s thesis remains one of the successes of the time devoted to this project. However, the division of these eight brain samples into four progression categories (healthy, early stage, intermediate stage, and late stage) meant that each category had only two brain samples from which to draw data. It is difficult to ascertain what if any conclusions about the population or about the disease in general based on such low sample numbers. With only two samples per category, it is no surprise that few trends in the data could reliably be called statistically significant. Should this study be repeated, it is important that more brain samples be procured.

A second limitation of the study was the lack of controls on the brain samples. This study attempted to reveal the effect of disease progression on changes to the methylome and hydroxymethylome assuming all other factors were held constant. However, there was no way to adjust the results for age, race, gender, genetics, lifestyle, or environment of the sample patients. All such factors can alter the epigenome independently of disease
progression. All also may have a role to play in the progression of the disease. The effect that the probable differences in genotype alone had on the differences epigenetic modifications is worth a large and complicated study in of itself. Not only could we not account for such factors due to the low sample size and the lack of data that definitively establishes the epigenetic effects of each of factor, we were not even provided that information at the onset and we had no way of accounting for them.

A third limitation was the relative low mass of each brain sample provided to us as well as the low amounts of obtainable genomic DNA contained in each sample. It remains to be shown if the low yields of genomic DNA is due to the techniques used to extract it or the tissue itself. Since most of the genomic DNA found in neurons in the brain is found in the cell bodies which are largely contained in the grey matter of the brain, it is possible that the location in the prefrontal cortex from which the sample was taken had an effect on our DNA yields.

Despite some of the setbacks experienced during this project, we are happy to have generated the results that we have and eagerly anticipate the rest of the data when the sequencing results are finished.
FUTURE ENDEAVORS

Since the main focus of this study as described in this thesis began as an attempt to fully sequence the hydroxymethylome and methylome of these specimens but shifted into establishing global trends, there is ample opportunity to finish the sequencing aspect of the project. This will ideally be finished in the next few months.

Even when the sequencing data has returned, there is a severe limitation about what kinds of conclusions can be drawn from sequences of eight samples with two from each general stage in the progression. It lacks statistical power and there are several confounding factors including the age, gender, genetics, and even environment of each individual.

What would have been more useful (if possible, practical, or ethical) would be to take brain samples from the same patients at various points in the progression of their disease in order to remove a variety of confounding epigenetic markers. Although it is impossible to perform such an experiment in vivo, the Shi lab currently is preparing to culture induced pluripotent neuronal stem cells derived from dermal tissue. These cells were obtained from individuals with known AD genotypes such as APOE4. The primary advantages of using these cells are that they are monoclonal and self-renewing. This solves two of the primary disadvantages of this study. The fact that they are monoclonal can help remove confounding factors associated with taking samples from different patients at different stages of the disease. These cells all are derived from one individual, allowing us to ignore the epigenetic differences resulting from the gender, age, race and genotype that likely confounded the results of this study. These cells can be sequenced at various points during the culturing process and this sequencing data can provide a clear
picture of how the methylome and hydroxymethylome change over time. Another excellent advantage of this strategy is that the cells can be chemically manipulated in ways that simulate oxidative stress, lipid, or amyloid build-up. There are several benefits to knowing the direct links between particular external stressors and changes in the epigenome.

Of course, this study design presents an entire new array of problems—namely that the cells need to be cultured in a way that would simulate the environment stimuli to which AD neurons are exposed. Since the study of epigenetics looks for environmental effects on the genome and genome expression, the way that the cells were cultured would have some major confounding effects. If we could fully simulate the cerebral environment of a neuron, there would still be challenges to this model. We have established that cells besides the neurons themselves—namely microglia and astrocytes—are involved with AD pathogenesis and we would need to culture those as well. Despite the limitations of this future study, data generated from such as study could at least shed light on how particular genotypes associated with AD could impact the methylome and hydroxymethylome in affected neurons.

Another possibility is that we could acquire transgenic mice with AD-associated alleles such as APOE4. These mice would be pre-programmed to undergo Alzheimer’s-like dementia after a few months. The primary advantage of this is that it maintains the monoculture that we would get with the induced neuronal stem cells. This could mean that if the mice are controlled for age and gender, all will have almost identical genotypes, lifestyles, and environmental exposures so that we can eliminate other
confounding effects on the epigenome. Also, different individuals from the monoculture could be sacrificed at different stages in the Alzheimer’s progression. The primary disadvantage of this is, like any animal study, the added cost, ethical considerations, and complexity of using live animals.
CONCLUSION

We are happy to report that preliminary analysis of genomes obtained from Alzheimer’s patients revealed useful data. Data showed a statistically significant decrease in 5hmC levels in patients with intermediate stage Alzheimer’s disease compared to healthy individuals. This is consistent with previous literature describing decreasing 5hmC in the AD brain. It can be contrasted with other literature showing that 5hmC increases in a healthy brain. While more data is needed before this information is used as a diagnostic tool for aging patients experiencing possible Alzheimer’s-like symptoms, it does lay the groundwork for future discoveries that may provide medical science with a clearer model of the epigenetic and cellular pathways involved in the disease. This will be particularly evident once the exact sequences are completed.

In addition to obtaining the data reported in this review, the skills learned during this experience have a great personal value. In addition to working with and collaborating with the many renowned researchers at the Endocrinology division at Brigham and Women’s Hospital, I was given the chance to help write multiple grants, develop and finalize an Harvard Medical School IACUC-approved animal protocol, order necessary reagents, and to participate in several ongoing experiments in the Shi lab. Beyond the techniques described in this thesis, I was also able to study and perform laboratory techniques including antibody pull-down of specific HA-flag transfected proteins relating to the epigenetic factors of breast cancer development, the culturing of human cells including human breast cancer cells and induced pluripotent stem cells, western blots, and care for laboratory animals including knockout mice.
REFERENCES


Loebel, David A. F, Catherine M Watson, R. Andrea De Young, and Patrick P. L Tam. 2003. “Lineage Choice and Differentiation in Mouse Embryos and Embryonic


CURRICULUM VITAE
Michael Allen Smith (1987)
10 Wise St. #2
Jamiaca Plain, MA 02130
smit3933@bu.edu

Educational

Boston University School of Medicine, Boston, Massachusetts 9/12-present
- Enrolled in Masters of Medical Sciences Program class of 2014

University of Minnesota, Minneapolis, Minnesota, 9/07-6/10
- Enrolled in College of Biological Sciences
- Major in Biochemistry, minor in Philosophy

University of California at San Diego, La Jolla, California, 9/06–6/07

Edina High School, Edina, Minnesota, graduated 6/06

Employment

Brigham and Women’s Hospital, Boston, Massachusetts
7/13-present
- Lab manager of the Yujiang Geno Shi Lab.
- Responsible for ordering chemicals reagents, and lab ware, organizing the lab, assisting in the grant writing process, and keeping computerized records of all chemicals in reagents in the lab.
- Responsible for completing a thesis project in epigenetic endocrinology.

The Interamerican School, Quetzaltenago, Guatemala
1/11- 6/12
- High School Science Teacher at a small bilingual school
- Responsible for teaching, planning and grading students in high school environmental science, biology, chemistry, cosmology, and physics.
- Gained strong interpersonal skills and comfort with speaking and explaining scientific concepts in front of groups.
- Succeeded in conveying a love of science to several students.
- Responsible for also teaching one semester of creative writing and physical education
- Responsible for giving two motivational talks in front of the high school students twice every semester.

**Academia Europea, Quetzaltenango, Guatemala**

9/10- 1/12

- English Teacher
- Responsible for teaching English to Basic, Intermediate, and Advanced students.
- Teacher of the year 2011
- Gained valuable experience in speaking enthusiastically in front of an unfamiliar audience.

**Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota,**

9/09- 09/10

- Undergraduate research student
- Responsible for maintaining glassware, media, plastic ware, etc.
- Responsible for performing numerous experiments on behalf of graduate students.
- Other tasks assigned by CGE personnel and staff
- Gained knowledge of practical research in molecular biology
- Worked on a summer-long project involving the synthesis of two new zinc finger endonucleases.

**Hennepin County Medical Center, Minneapolis, Minnesota, 5/09–9/09**

- Research Intern, Emergency Room
- Responsible for assessing and recruiting ER patients for participation in various research studies

*Laboratory Experience*

**University of Minnesota**

- Organic chemistry, Spring, 2008
- Ecology and Field Biology, Itasca Biological Station and Labs, Summer, 2008
- Microbiology, Fall, 2008
- Protein biochemistry, Spring, 2009
- Computational biology, Spring, 2009
- Molecular biochemistry, Summer, 2009
- Virology/Immunology, Spring 2010
- Practical molecular biology, Fall 2009 to Fall 2010

**University of California at San Diego**

- General chemistry, Spring Quarter, 2007
Extracurricular

Primeros Pasos, Quetzaltenago Guatemala 06/12-08/12

- Medical Volunteer
- Responsible for taking vital signs, conducting patient interviews, distributing medicine, and giving classes on proper hygiene and sexual health to children and adolescents of rural Guatemala
- Exclusively performed in Spanish

North Dakota Medical Mission, Quetzaltenago, Guatemala 07/11 and 07/12

- Medical volunteer for group of doctors and dentists from North Dakota and Canada conducting annual free clinic in underserved areas of Guatemala.
- Responsible for taking vital signs and translating between English and Spanish.

San Marcos Episcopal Church, Quetzaltenango, Guatemala 9/10-1/11

Volunteer during lunch food kitchen for the community’s senior residents.

Emergency Medical Services, 12/10-present

Trained Emergency Medical Technician

University of Minnesota Bands, 9/07-1/10

- Play trumpet in University Marching Band, Men’s Basketball Pep Band, Men’s Hockey Pep Band

University Episcopal Community, University of Minnesota

- President, 2009-2010 academic year
- Planned and participated in service trip to Guatemala, Spring, 2008

United Episcopal/Methodist Association, University of California at San Diego

- Planned and participated in service trips to Tijuana, Mexico and El Salvador, Winter and Spring, 2007

Boy Scouts of America

- Troop leader; attained rank of Eagle Scout