Shewale, Shantanu J., Epigenetic Regulation of Gene Expression in Alzheimer's disease. Doctor of Philosophy (Biomedical Sciences, Molecular and Medical Genetics), April 2016, 162pp., 13 tables, 21 illustrations, bibliography, 225 titles

Abstract

Alzheimer's disease (AD) is the most common form of age-related neurodegenerative dementia, and it is estimated that over 5 million people currently have AD within United States. AD can either be early onset or Late Onset AD (LOAD). Early onset AD has an age of onset below 60 years, and LOAD has an age of onset of above 65 years. Early onset AD accounts for <5% of the cases and the genetic variants responsible have been well documented and are inherited in a Mendelian dominant manner. Even though Late Onset (LOAD) accounts for >95% of AD cases, and numerous genetic loci have been linked to LOAD; these loci have small effect sizes, and explain only 50% of AD risk. We hypothesize that epigenetic mechanisms are responsible for a significant portion of this missing heritability. The impact of epigenetic mechanisms on AD risk and progression are relatively unexplored, and should be considered when addressing a portion of the ***remaining missing heritability.***

Within this project, post mortem frontal cortex brain tissue from 11 AD patients and 12 age matched controls were used to investigate DNA methylation and differential gene expression in AD. Since post mortem human tissue was used, preliminary analysis showed presence of degraded RNA, most likely due to post mortem intervals. To combat degraded RNA, a novel library preparation process was utilized prior to performing RNA sequencing. DNA methylation was investigated using two methods. For site specific investigation, the Illumina® Infinium HumanMethylation450 BeadChip array was utilized. To investigate differential methylated

regions, a Methyl-Binding Protein capture approach was used to precipitate out methylated regions of the genome. This precipitated DNA was then analyzed for methylated regions by using high throughput sequencing.

The Differentially Expressed Genes (DEGs) found within our RNA-seq dataset all elucidate the importance of some previously suspected pathways involved in the pathogenesis of AD. Gene Ontology (GO) analysis performed indicate that DEGs implicate numerous genes correlated with neurological disease, and collectively effect regulation of synaptic transmission, cell-cell signaling, neurotransmitter transport, genes involved in the inflammatory response, and Amyloid Precursor Protein (APP) processing. The overlap of 32 DEGs and differentially methylated CpGs was observed. GO analysis demonstrated the same GO terms (synaptic transmission & cell-cell signaling) impacted within both, RNA and DNA datasets. This indicates a link between CpG methylation and differential gene expression.

Epigenetic Regulation of Gene Expression in Alzheimer's Disease

DISSERTATION

Presented to the Graduate Council of the

University of North Texas Health Science Center

At Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Shantanu J. Shewale, BS

Fort Worth, TX

April , 2016

Table of Contents

Acknowledgements

I would like to express my first and foremost gratitude to Dr. Kirk C. Wilhelmsen for his mentorship and overwhelming support during this project. I am grateful to him for his patience, contributions of his time, and stimulating discussions for the past years. I am also thankful for him sharing his ideas, knowledge, and experience with me.

I sincerely thank Dr. John V. Planz, my Major Professor, for taking me under his wing and for his wholehearted support. I appreciate his mentorship from the very first day of my graduate education career. In addition, I am indebted to Dr. Robert C. Barber for introducing me to the field of Neurobiology and capturing my interest in Methylation and Alzheimer's disease. We have had many engaging discussions, and I must express that Dr. Barber is an excellent coach. Dr. Barber taught me the value of perspective, has always provided me guidance in my times of need, both in and outside of graduate school.

To my committee member, Dr. Arthur J. Eisenberg, I really appreciate the time you have taken out of your schedule to always make me a priority. I have enjoyed our many conversations, and would not be where I am today if it wasn't for your support and guidance. I owe a huge thanks to my entire committee, Drs. Allen, Barber, Cunningham, Eisenberg, Planz, and Wilhelmsen for holding me to highest standards, and giving me the confidence and freedom to become an independent thinker.

I would like to express my gratitude towards my colleague, Dr. Alejandro Colaneri, as his guidance was crucial towards the completion of this project. I would also like to thank my colleague Dr. Fengshen Kuo for taking time out his schedule to work with me on one of my datasets. His was my main mentor in bioinformatics. I am really grateful towards the entire Wilhelmsen family for allowing me to stay with them during my numerous visits and prolonged

stays at UNC, Chapel Hill. I would like to thank my parents Jaiprakash and Manisha Shewale, for their unconditional love throughout my entire life. Both of them have been my biggest fans, and I am truly lucky to have such a supportive family. Lastly, I would like to thank my fiancé, Shradha Kachi, and my friends for being there for me during my graduate school experience.

List of Figures

List of Tables

List of Abbreviations

5mC	5-Methyl-Cytosine
Aβ	Amyloid-Beta
AD	Alzheimer's disease
APP	Amyloid Precursor Protein
bp	Base Pair
cDNA	Complimentary DNA reverse transcribed from RNA
CpG	Cytosine phosphodiester bonded to Guanine
CTL	Control
DEG	Differentially Expressed Genes
DMP	Differentially Methylated CpG Site
DNA	Deoxyribonucleic Acid
ENCODE	Encyclopedia of DNA Elements
FAD	Familial Alzheimer's disease
GO	Gene Ontology
GWAS	Genome Wide Association Studies
IPA	Ingenuity Pathway Analysis
LOAD	Late Onset Alzheimer's disease
MBD	Methyl Binding Domain
MCI	Mild Cognitive Impairment
NFT	Neurofibrillary Tangles
PCR	Polymerase Chain Reaction
PE	Paired End Reads
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing

Chapter I

INTRODUCTION AND LITERATURE REVIEW

The Alzheimer's disease problem

Alzheimer's disease (AD) is the most common form of age-related neurodegenerative dementia, a serious health problem in the industrialized world, and is currently the 6th leading cause of death in the United States[1]. AD is the most common type of dementia, which is a term that describes a wide range of symptoms such as trouble with memory, language, ability to focus, reasoning skills, and visual perception[2]. AD is an heterogeneous, progressive disease that is fatal given that no other cause of death intervenes[3].

The current figures from the Alzheimer's Association state that one in three seniors will die with AD or another form of dementia. One in eight people 65 years of age and older have AD[1], and when examining the people that are 85 years old or older, the incidence of AD increases to one in two individuals[1]. It is estimated that currently within United States, over 5 million people have AD, and its prevalence is expected to rise to over 13 million by the year 2050[1]. Currently, it is estimated that the care provided by family, and other unpaid caregivers of people with dementia is valued at about $210 billion annually[3].

AD occurs in both a familial and a sporadic form, also known as early onset (familial) or late onset Alzheimer's disease (sporadic), respectively. Familial Alzheimer's Disease (FAD) is an autosomal dominant disorder that is inherited in a Mendelian fashion[4]. FAD accounts for a minority of the total AD cases (<5%)[5,6], and has an earlier age of onset (<60 years)[7]. In

comparison, late onset Alzheimer's disease (LOAD) typically has an age of onset above 65yrs.[7]. LOAD results from various genetic and non-genetic factors, and is hypothesized that pathogenesis begins anywhere between 15-20 years before symptom onset.

Early warning signs of AD include mild cognitive impairment (MCI), which is characterized by individuals having memory problems as well as a decline in cognition, such as impaired judgment, vision and spatial issues. Some, not all individuals, progress towards AD, and others develop some form of dementia. Once a person develops the clinical symptoms of AD, they are classified into three categories: mild, moderate, and severe AD. Mild AD symptoms include increases in memory loss and changes in cognitive abilities, such as getting lost, taking longer to complete daily tasks, and personality changes. Moderate AD includes the loss of ability to learn new things, problems recognizing family, and potentially hallucinations. Severe AD includes an inability to communicate, ultimately resulting in complete reliance on others for care. Clinically, severe AD includes wide distribution of plaques and tangles throughout the brain, and a significant amount of atrophy of brain tissue. Clinically, AD is characterized by the loss of memory, inability to learn new things, loss of language function, depression, delusions, among other manifestations. AD is ultimately fatal within 5 to 10 years, where affected individuals usually die of complications of chronic illness[8].

Alzheimer's disease pathogenesis

AD is characterized by cortical atrophy, synapse loss, and neuronal cell death, and is pathologically defined by the abnormal accumulation of extracellular amyloid βeta (Aβ), and intracellular neurofibrillary tangles (NFTs). The disease was first characterized by Dr. Aloysius "Alois" Alzheimer in 1907, and was based upon his observations and treatment of a 51 year old patient named August 'D'[9], The patient showed symptoms of short term memory loss, unusual behavior and the neuropathological characteristics that have become the hallmarks of Alzheimer's disease[9]. The two hallmarks of AD are extracellular Aβ plaques and NFTs[10,11]. Aβ plaques are formed from cleavage of Amyloid Precursor Protein (APP), which is an integral membrane protein that is expressed throughout the body and is particularly concentrated in neuronal synapses. The primary function of APP is not fully understood, but it has been implicated in neurite extension and synaptic plasticity[3]. Beta (β) and Gamma (γ) secretases cleave APP to produce fragments that aggregate together to form the Aβ plaques. β-secretase is an integral membrane aspartyl protease encoded by the β-site APP-cleaving enzyme 1 (*BACE1*) gene[12]. γ-secretase, is composed of 4 subunits: Presenilin 1(*PSEN1*), Presenilin 2 (*PSEN2*), nicrastin, and *APH1*[2,3] , where the active site consists of presenilin[2] . Normally, Aβ is degraded by peptidases such as neprilysin, insulin-degrading enzyme, and endothelin-converting enzyme[2].

A widely held theory of AD pathogenesis is the amyloid cascade hypothesis[13], which states that accumulation of the amyloid beta peptide in the brain is the initiating event in disease pathology[14]. This hypothesis postulates that the disease is the result of an imbalance between the production & degradation of Aβ[15], and (Aβ) in a variety of forms, triggers a cascade that harms synapses and ultimately neurons, producing the pathological presentations of Aβ plaques, tau

tangles, synapse loss and neurodegeneration, leading to dementia. Aβ accumulation is thought to initiate AD pathology by destroying synapses, causing formation of NFTs, and subsequently inducing neuron loss. Recently there has been a shift toward defining soluble Aβ oligomers as the toxic agent, rather than plaques, but the theory and the way data is interpreted have remained largely the same, i.e. Aβ accumulation as oligomers or plaques triggers AD. A large, growing literature embraces the amyloid hypothesis (Figure 1)[13]. This central theory has strong support, from work beginning with Alois Alzheimer[9] and continuing through deduction of the steps that generate the amyloid beta protein[16] and cloning of mutations in *APP*[17,18], *PSEN1* and *PSEN2* genes[19,20] that cause familial AD. A recent development that significantly strengthened the amyloid hypothesis was the discovery by Jonsson *et al.*[21] of an *APP* mutation that reduces production of Aβ plaques and is protective against AD, as well as age-related cognitive decline. It has been hypothesized that Aβ plaques protein deposition precedes NFTs[22], cell loss, and vascular damage[23]. In transgenic murine models, Aβ plaque deposition developed prior to NFTs[24]. Working in a transgenic mouse model, Xu *et al.*[22] described an accumulation of Aβ plaques precipitated a loss of solubility of intracellular cytosolic proteins such as glycolytic enzymes and members of the chaperone family. Aβ plaques have also been thought to induce neuronal oxidative stress, resulting in phospholipid peroxidation and protein oxidation in AD brain[25].

The second hallmark of AD is the presence of NFTs that arise due to hyper-phosphorylated microtubule associated protein tau (MAPT). The tau protein is primarily expressed in neurons[26] and has been shown to be involved with tubulin polymerization as well as acting to stabilize microtubules against depolymerization [27], stabilize microtubules responsible for axonal transport [26], increase neuritic stability, impact the rate of neurite elongation, and

increase net microtubule stability[2,28]. Different isoforms of the tau protein are expressed due to alternative RNA splicing, and each of these isoforms displays a varied degree in phosphorylation. All isoforms are capable of forming the NFTs that are a hallmark of AD[29]. A balance between kinases (ex. GSK-3Beta, CDK5) and phosphatases (ex. PP-1, PP2) plays a role in regulating tau phosphorylation[2]. Tau hyper phosphorylation leads to disassembly of microtubules causing disruption in axonal transport, and impaired synaptic & neuronal function[2,30]. Hyperphosphorylated tau aggregates into filaments, causing an inability to bind and stabilize microtubules [31], and subsequent formation of NFTs[2,31].

This neuronal atrophy has been documented by magnetic resonance imaging (MRI) and positron emission tomography (PET), as an individual progresses from mild cognitive impairment to Alzheimer's disease [32,33]. Presence of Aβ plaques and NFTs is often used, by neuropathologists, to diagnose an individual as having AD post-mortem, and currently is the only way AD can be confirmed. Using PET, with radio tracers that bind to Aβ plaques, one can now quantify pathological changes within the human brain. Currently, only a few radio tracers are available, and new Aβ tracers are under development[34]. These techniques still make diagnosing AD is difficult due to variability in the pathology, where some individuals have high Aβ load, but do not show symptoms.

Synapse loss is also an early feature of AD, where there is a strong correlation between the extent of synapse loss and the severity of dementia, and significant synapse loss also occurs in MCI individuals[35]. Synapse loss is a pathological correlate of cognitive dysfunction in AD, suggesting that synaptic changes are crucial for AD pathogenesis[36–39]. Synapse loss is prevalent near senile plaques, suggesting that plaques may be a reservoir of synaptotoxic molecules. The cognitive symptomatology of AD has a temporal correlation with the progression of

neurofibrillary tangles in the frontal cortex[35]. Synaptic dysfunction may be a result of toxicity caused by the Aβ-amyloid plaques or NFTs. Mouse AD-models show loss of memory and cognitive impairments before accumulation of Aβ plaques, suggesting that synapse dysfunction may precede and cause neurodegeneration within AD. The nature of many peptidic cerebrospinal biomarkers for early detection of Alzheimer's suggest that presynaptic failures affecting neurotransmission are early events in the progression of this disease. AD may be a disease of synaptic dysfunction and synapse loss that then progresses to include widespread neuronal loss and network failure of neurons. Aβ could be a key player in the induction of synaptic failure, as Aβ activates a variety of molecular cascades that culminate in synapse dysfunction, shrinkage, collapse and loss[35].

Pathogenic mechanisms such as oxidative stress, inflammation, cell-cycle abnormalities, and mitochondrial dysfunction[40,41] have been reported to precipitate neuropathological changes that cause degeneration of neurons and synapses in the cerebral cortex and subcortical regions of the brain[2]. Loss of neurons results in atrophy of the affected regions of the brain, including degeneration in the temporal and parietal lobes, as well as parts of the frontal cortex[42]. In addition to these two hallmarks of AD, pathology and clinical symptoms have been correlated with oxidative stress, inflammation[43,44], obesity[45], cardiovascular disease[46], traumatic brain injury[47–51], and diabetes[52].

AD is characterized by cortical atrophy, synapse loss, and neuronal cell death, and is pathologically defined by the abnormal accumulation of extracellular amyloid βeta (Aβ), and intracellular neurofibrillary tangles (NFTs). The disease was first characterized by Dr. Aloysius "Alois" Alzheimer in 1907, and was based upon his observations and treatment of a 51 year old

patient named August 'D'[9], The patient showed symptoms of short term memory loss, unusual behavior and the neuropathological characteristics that have become the hallmarks of Alzheimer's disease[9]. The two hallmarks of AD are extracellular Aβ plaques and NFTs[10,11]. Aβ plaques are formed from cleavage of Amyloid Precursor Protein (APP), which is an integral membrane protein that is expressed throughout the body and is particularly concentrated in neuronal synapses. The primary function of APP is not fully understood, but it has been implicated in neurite extension and synaptic plasticity[3]. Beta (β) and Gamma (γ) secretases cleave APP to produce fragments that aggregate together to form the Aβ plaques. β-secretase is an integral membrane aspartyl protease encoded by the β-site APP-cleaving enzyme 1 (*BACE1*) gene[12]. γ-secretase, is composed of 4 subunits: Presenilin 1(*PSEN1*), Presenilin 2 (*PSEN2*), nicrastin, and *APH1*[2,3] , where the active site consists of presenilin[2] . Normally, Aβ is degraded by peptidases such as neprilysin, insulin-degrading enzyme, and endothelin-converting enzyme[2].

A widely held theory of AD pathogenesis is the amyloid cascade hypothesis[13], which states that accumulation of the amyloid beta peptide in the brain is the initiating event in disease pathology[14]. This hypothesis postulates that the disease is the result of an imbalance between the production & degradation of Aβ[15], and (Aβ) in a variety of forms, triggers a cascade that harms synapses and ultimately neurons, producing the pathological presentations of Aβ plaques, tau tangles, synapse loss and neurodegeneration, leading to dementia. Aβ accumulation is thought to initiate AD pathology by destroying synapses, causing formation of NFTs, and subsequently inducing neuron loss. Recently there has been a shift toward defining soluble Aβ oligomers as the toxic agent, rather than plaques, but the theory and the way data is interpreted have remained largely the same, i.e. Aβ accumulation as oligomers or plaques triggers AD. A large, growing literature embraces the amyloid hypothesis (Figure 1)[13]. This central theory has strong support,

from work beginning with Alois Alzheimer[9] and continuing through deduction of the steps that generate the amyloid beta protein[16] and cloning of mutations in *APP*[17,18], *PSEN1* and *PSEN2* genes[19,20] that cause familial AD. A recent development that significantly strengthened the amyloid hypothesis was the discovery by Jonsson *et al.*[21] of an *APP* mutation that reduces production of Aβ plaques and is protective against AD, as well as age-related cognitive decline. It has been hypothesized that Aβ plaques protein deposition precedes NFTs[22], cell loss, and vascular damage[23]. In transgenic murine models, Aβ plaque deposition developed prior to NFTs[24]. Working in a transgenic mouse model, Xu *et al.*[22] described an accumulation of Aβ plaques precipitated a loss of solubility of intracellular cytosolic proteins such as glycolytic enzymes and members of the chaperone family. Aβ plaques have also been thought to induce neuronal oxidative stress, resulting in phospholipid peroxidation and protein oxidation in AD brain[25].

The second hallmark of AD is the presence of NFTs that arise due to hyper-phosphorylated microtubule associated protein tau (MAPT). The tau protein is primarily expressed in neurons[26] and has been shown to be involved with tubulin polymerization as well as acting to stabilize microtubules against depolymerization [27], stabilize microtubules responsible for axonal transport [26], increase neuritic stability, impact the rate of neurite elongation, and increase net microtubule stability[2,28]. Different isoforms of the tau protein are expressed due to alternative RNA splicing, and each of these isoforms displays a varied degree in phosphorylation. All isoforms are capable of forming the NFTs that are a hallmark of AD[29]. A balance between kinases (ex. GSK-3Beta, CDK5) and phosphatases (ex. PP-1, PP2) plays a role in regulating tau phosphorylation[2]. Tau hyper phosphorylation leads to disassembly of microtubules causing disruption in axonal transport, and impaired synaptic & neuronal

function[2,30]. Hyperphosphorylated tau aggregates into filaments, causing an inability to bind and stabilize microtubules [31], and subsequent formation of NFTs[2,31].

This neuronal atrophy has been documented by magnetic resonance imaging (MRI) and positron emission tomography (PET), as an individual progresses from mild cognitive impairment to Alzheimer's disease [32,33]. Presence of Aβ plaques and NFTs is often used, by neuropathologists, to diagnose an individual as having AD post-mortem, and currently is the only way AD can be confirmed. Using PET, with radio tracers that bind to Aβ plaques, one can now quantify pathological changes within the human brain. Currently, only a few radio tracers are available, and new Aβ tracers are under development[34]. These techniques still make diagnosing AD is difficult due to variability in the pathology, where some individuals have high Aβ load, but do not show symptoms.

Synapse loss is also an early feature of AD, where there is a strong correlation between the extent of synapse loss and the severity of dementia, and significant synapse loss also occurs in MCI individuals[35]. Synapse loss is a pathological correlate of cognitive dysfunction in AD, suggesting that synaptic changes are crucial for AD pathogenesis[36–39]. Synapse loss is prevalent near senile plaques, suggesting that plaques may be a reservoir of synaptotoxic molecules. The cognitive symptomatology of AD has a temporal correlation with the progression of neurofibrillary tangles in the frontal cortex[35]. Synaptic dysfunction may be a result of toxicity caused by the Aβ-amyloid plaques or NFTs. Mouse AD-models show loss of memory and cognitive impairments before accumulation of Aβ plaques, suggesting that synapse dysfunction may precede and cause neurodegeneration within AD. The nature of many peptidic cerebrospinal biomarkers for early detection of Alzheimer's suggest that presynaptic failures affecting neurotransmission are early events in the progression of this disease. AD may be a disease of

synaptic dysfunction and synapse loss that then progresses to include widespread neuronal loss and network failure of neurons. Aβ could be a key player in the induction of synaptic failure, as Aβ activates a variety of molecular cascades that culminate in synapse dysfunction, shrinkage, collapse and loss[35].

Pathogenic mechanisms such as oxidative stress, inflammation, cell-cycle abnormalities, and mitochondrial dysfunction[40,41] have been reported to precipitate neuropathological changes that cause degeneration of neurons and synapses in the cerebral cortex and subcortical regions of the brain[2]. Loss of neurons results in atrophy of the affected regions of the brain, including degeneration in the temporal and parietal lobes, as well as parts of the frontal cortex[42]. In addition to these two hallmarks of AD, pathology and clinical symptoms have been correlated with oxidative stress, inflammation[43,44], obesity[45], cardiovascular disease[46], traumatic brain injury[47–51], and diabetes[52].

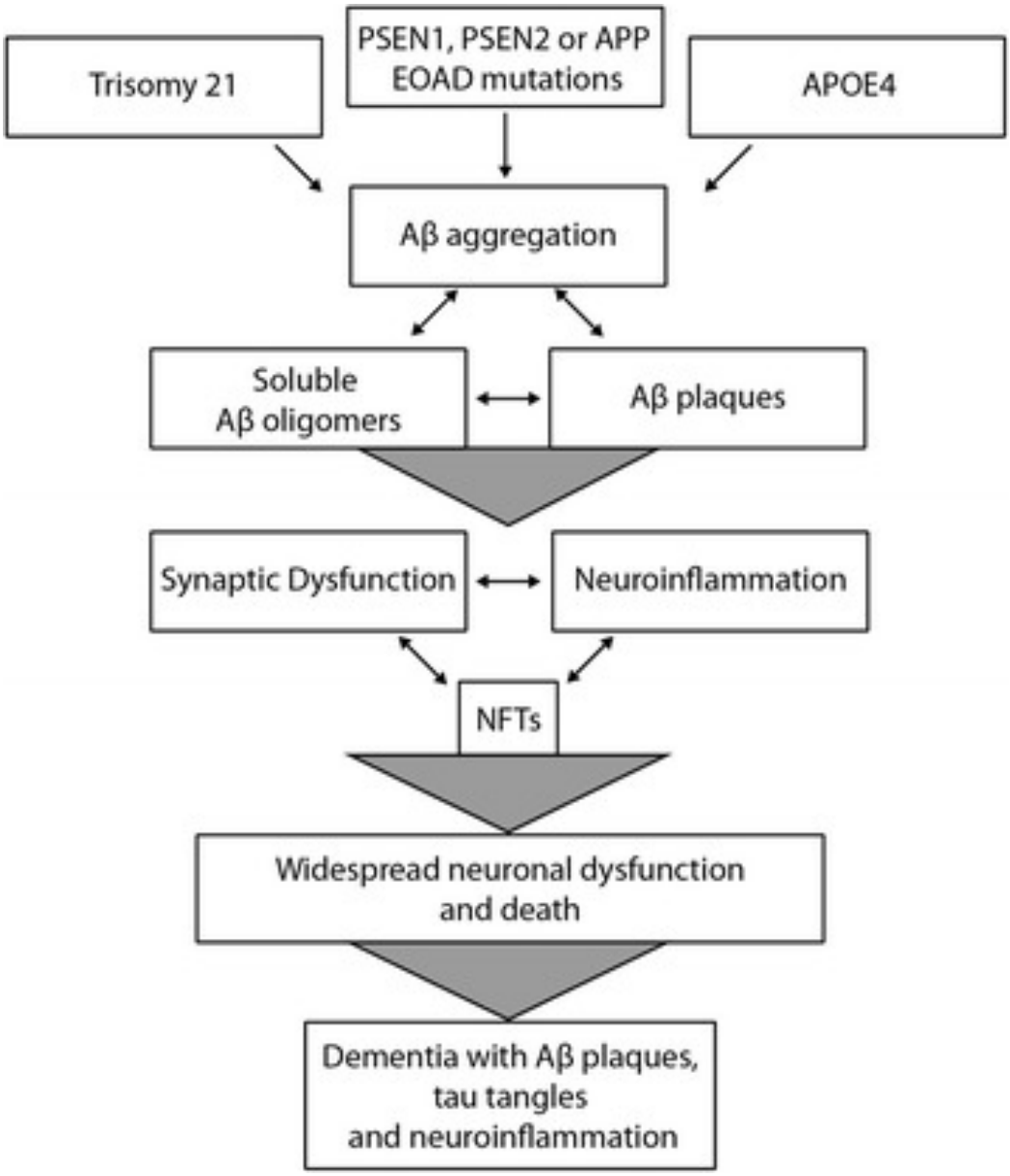


Figure 1. **The Amyloid Hypothesis.**(Adapted from Morris *et al*[13].): Alzheimer's disease is the result of an imbalance between the production and degradation of Aβ[15], and Aβ in a variety of forms, triggers a cascade that harms synapses and ultimately neurons, producing the pathological presentations of Aβ plaques, tau tangles, synapse loss and neurodegeneration, leading to dementia. Aβ accumulation is thought to initiate AD pathology by destroying synapses, causing formation of NFTs, and subsequently inducing neuronal loss.

Genetics of Alzheimer's disease

Familial Alzheimer's Disease (FAD) is an autosomal dominant disorder that is inherited in a Mendelian fashion[4]. FAD accounts for a minority of the total AD cases (<5%)[5,6], and has an earlier age of onset (<60 years)[7]. FAD has been associated with mutations in the *APP*, Presenilin 1 (*PSEN1*), and Presenilin 2 (*PSEN2*) genes[7,14,29,53,54]. Mutations in the *APP* gene that cause FAD are clustered near the **α**, **β**, and **γ**- secretase cleavage sites, where most of these mutations increase cleavage by **γ**-secretase[55]. So far, 24 mutations for *APP*, 185 mutations for *PSEN1*, and 13 mutations for *PSEN2* mutations have been found. All of these mutations, except one, are inherited in a Mendelian autosomal dominant fashion, and are fully penetrant[7]. The early onset forms of AD fit the amyloid cascade hypothesis, where *APP, PSEN1* and *PSEN2* mutations increase production of Aβ[53,54].

LOAD results from various genetic and non-genetic factors. The strongest known genetic risk factor is carriage of the epsilon (**ε**) 4 allele at the Apolipoprotein-E (*APOE*) locus. The protein product of this gene combines with lipids to form lipoprotein molecules that are involved in packaging cholesterol and other fats as well as their transport into blood[3]. In AD pathology *APOE* is believed to play a role in the clearance of Aβ[7]. There are three different alleles of *APOE* known as **ε**2, **ε**3, and **ε**4. These alleles code for three isoforms of the protein that differ among each other for amino acid residues at position 112 and 158. The **ε**3, and **ε**4 alleles code for cysteine/cysteine, cysteine/arginine, and arginine/arginine residues respectively[7]. Individuals that are homozygous for the **ε**4 allele are 10 to 20 times more likely to develop AD in comparison to **ε**4 negative individuals, and the presence of the **ε**2 allele in an individual has a decreased risk for AD[3]. In addition to APOE, genome wide association studies (GWAS) have identified 22 other loci that are associated with LOAD. These genes are *HLA-DRB5-DRB1, SORL1, PTK2B,*

*SLC24A4, ZCWPW1, NME8, CASS4, INPP5D, MEF2C, CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1, SORL1, ATXN*1 and *ABCA7*[7,56–59]. However, combining the effects of all known genetic variants that are associated with the disease does not fully explain the presumably genetic component[60]. Taken together, these genetic loci account for only 50% of the genetic variation among LOAD cases[57], leaving a large portion of the heritability still unidentified. ***This is the missing heritability***. The remaining unexplained heritability within individuals that develop LOAD may be explained through a variety of mechanisms. I hypothesize that a significant portion of this missing heritability is due to epigenetic factors.

Given that AD is such a complex and heterogeneous disease that is impacted by various factors, there arises a need to better understand the complexities of AD that will lead to prevention methods or a cure.

Overview of Epigenetics

The term epigenetics was coined by Conrad Waddington in 1942 to describe "the branch of biology which studies the casual interactions between genes and their products, which bring the phenotype into being"[61]. Today the term broadly applies to changes in gene regulation and cellular phenotype without changes to the DNA sequence itself, as the phenotype of a cell is determined by its expression profile[62]. Epigenetic marks drive much of this expression and provide diversity to this phenotype via chromatin alteration that affects gene transcription. Epigenetics is the molecular phenomenon by which phenotypic changes are transmitted from one generation to another with no apparent alterations to the sequence of DNA itself[63]. Epigenetics incorporates several aspects that include DNA methylation, histone modifications, and microRNAs regulation. Epigenetic modifications regulate gene expression pre-transcriptionally, whereas miRNAs suppress gene expression post-transcriptionally. Epigenetic status is modified by environmental exposures such as nutrition, social status, chemical and emotional environment, pregnancy conditions, infertility, contraception, and different modalities of pharmacological intervention[63]. Epigenetic status is also influenced by genotype, genetic variation in genes encoding numerous enzymes[63,64]. Furthermore, DNA methylation contributes to natural human variation[63,65].

An epigenome is the chromatin state found across the genome at a certain time point and cell type, and therefore thousands of epigenomes can exist for a single given genome[66]. Even though there is no alteration in the DNA sequence itself, epigenetic marks, chromatin activity, and histone modifications[67] are heritable during cell division, keeping these epigenetic marks intact and passed on to dividing cells[68]. Some epigenetic modifications are stabilized and maintained throughout the life of an organism, while others change over time due to intrinsic or

environmental factors[69]. The epigenome can be influenced by many aspects that impact DNA or the histones that are bound to DNA. An example of an epigenetic modification to DNA in mammals is the covalent addition of a methyl group to the 5-carbon of cytosine in CpG dinucleotides to form 5-MethylCytosine (5mC) [67,68] (Figure 2)[70]. Similar to DNA, histones can be methylated, or acetylated. Histone methylation and demethylation are are performed by enzymes termed histone methyltransferases, and demethylatses. Histone acetylation and deacetylation is performed by histone acetyltransferases (HATs) and histone deacetylates (HDACs). Further, regulation of histones can occur via modifications that involve phosphorylation, ADP-ribosylation, ubiquitylation, sumoylation, crotonylation, propionylation, deiminiation and O-GlcNAcylation, which are all controlled by a et of enzyme complexes[71].

In mammalian genomic DNA, methylation of the cytosine residue within DNA is achieved by the addition of a methyl group from S-adenosyl-L-methionine(SAM). This residue is added by a subset of enzymes known as DNA methyltransferases (DNMT1, DNMT3a, and DNMT3b). DNMT1 is highly expressed in neurons and has high affinity to hemimethylated CpGs, thereby acting as a maintenance methyltransferase[72]. The other two transferases, DNMT3a and DNMT3b, methylate previously unmethylated cytosines of CpG dinucleotides and therefore are *de novo* methyltransferases. Modification of DNA via 5mC is considered crucial in mammals, as genome-wide disruption of DNA methylation results in embryonic lethality[72]. In mammals, the 5mC is mostly restricted to CpG dinucleotides that are confined to short genomic regions called CpG Islands (CGIs) and to promoters linked to 70% of all genes, which contain high numbers of hypomethylated CGIs. CGIs are short stretches of DNA where the presence of the CpG sequence is higher than in other regions[73]. The frequency of CpG islands varies greatly between mammalian genomes[74]. CGIs tend to be in GC rich regions and are methylated in

various cell types under certain pathological conditions[73]. They are rarely larger than 5kbp and overlap with the promoter regions for 50% to 60% of human genes[73]. Most housekeeping and widely expressed genes have a CpG island covering the transcriptional start site[75], except when they are associated with imprinted genes. These islands tend to be unmethylated for housekeeping and tissue specific genes at developmental stages[76]. *De novo* methylation of these CGIs is associated with gene-specific or tissue-specific gene expression regulation and is influenced by composition of individual cis-regulatory regions within chromatins. Compared with histone modification, DNA methylation is a more stable mark, especially in differentiated cells, and is essential in gene expression control and cell type maintenance. Regulation of gene expression doesn't always directly correlate to the global genome levels of 5mC. DNA methylation characteristics of a given locus contribute to whether the locus would be subjected to specific regulatory mechanisms in gene expression. In addition to the 5mC-mediated gene regulation, the density and distribution of 5hmC in the genome show cell- and tissue-type specificity that further contributes complexity to the epigenetic regulation of gene expression.

The machinery involved in the methylation of DNA in mammals consists mainly of two components, a DNA methyltransferases (DNMTs) and methyl-CpG binding proteins (MBD)[68]. DNMTs establish and maintain DNA methylation patterns, whereas MBDs read methylation marks. The maintenance enzyme DMNT has subclasses where DMNT1 is responsible for methylation of hemi-methylated CpG dinucleotides in nascent strands of DNA following DNA replication. One important function of DMNT1 includes maintaining DNA methylation patterns in proliferating cells[67]. DNMT3a and DNMT3b are involved in *de novo* methylation and for establishing new DNA methylation patterns in cells during development[67]. Both enzymes also interact with histone deacetylase to repress transcription[77]. Mouse models have shown that

DNMT1 and DNMT3b are essential for embryonic development and that DNMT3a deficient mice die within a few weeks of birth[68].

The methylation of DNA is usually associated with the silencing of gene expression by directly blocking transcription regulatory factors from binding to their target sequences[67]. Methylation of promoter regions has shown to be associated with lower levels of gene expression[78], and methylation within the gene body has been associated with increased gene expression[79]. The proposed mechanism is that the methylation of DNA causes recruitment of binding proteins that recognize the methylated DNA and associate with histone deacetylase and chromatin remodeling complexes to cause the stabilization of condensed chromatin[80]. These proteins that recognize methylated DNA contain a methyl-binding domain (MBD) that recognize methylated cytosine residues. These binding proteins play a role in chromatin modification and remodeling and do not act in isolation; evidence has shown that they often interact with each other by forming large protein complexes[67]. The main components of chromatin modification and remodeling can be divided into four sub categories. The first category consists of DNA cytosine methyl-transferases such as DMNT1, DMNT3a and DMNT3b. The second are Methyl-CpG-binding proteins that bind methylated DNA. There are six such known proteins that have been identified in mammals. These proteins can act as transcriptional repressors or can recruit further chromatin modifying proteins and transcription-regulatory complexes upon binding methylated DNA[67]. The third category includes Histone-modification enzymes that modify core histones (H2A, H2B, H3 and H4) at their amino-terminus. Modifications to the histones include phosphorylation, methylation, and acetylation which also regulate gene transcriptional activity[67]. Methylation of the Lysine4 residue of the H3 complex has been associated with gene repression, while methylation of the Lysine9 of H3 complex has been associated with transcriptional

silencing[67]. The last subsets of complexes are ATP-dependent remodeling complexes. These enzymes utilize ATP hydrolysis to modify super helical torsion in nucleosomal DNA, which modifies nucleosomes and alters the accessibility of chromatin to various proteins that further control downstream transcription, replication, or recombination events[67]. The term epigenetics incorporates all of the mechanisms described above and involves complex interplay between these mechanisms.

DNA methylation is a major epigenetic mechanism that has been shown in eukaryotes to play an important role for gene control, cell differentiation during development[73], embryonic development, chromatin structure, X chromosome inactivation, chromosome stability and genomic imprinting[68]. DNA methylation is crucial for numerous cellular processes, where altered DNA methylation has been linked to many common human diseases[81]. Numerous lines of evidence indicate that DNA methylation plays a direct role in carcinogenesis. Defects in enzymes involved in epigenetic modification have been linked to various types of tumor formation and leukemia; elevated levels of DNMTs and MBD-containing proteins have been observed in human tumors[81]. Both hyper- and hypo-methylation have been observed in cancer cells and the loss of methylation from repetitive regions of the genome results in genomic instability and is a hallmark of some tumors[68]. Outside of cancer, DNA methylation has been shown to play an important role in neurodevelopmental disorders, neurodegenerative, neurological, and autoimmune diseases.

Further, the finding that global 5mC decreases with age in various tissues of mammals, has made DNA methylation a considerable topic of interest in age related neurological disorders. Johansson *et al.*[82] found that methylation plays an important role in the process of aging. Testing 476,366 sites within peripheral blood of individuals whose age ranged from 14 to 96 years, it

was shown that 29% of the sites were affected due to age, where 60.5% of the sites become hypomethylated and 39.5% of sites become hypermethylated as age increases. Epigenome wide analysis of CpG nucleotide methylation was performed at 27,000 loci within the frontal cortex, temporal cortex, pons and cerebellum of 387 individuals aged from 1 to 102 years[83]. The authors report that some loci showed differential DNA methylation with increasing age. DNA methylation is impacted by multiple environmental factors like alcohol consumption, body mass index, smoking, folate intake, among others, which is why an individual's lifestyle can play an important role in terms of epigenetic factors impacting numerous diseases. Evidence for this is given that methylation status varies among monozygotic twins since epigenetic differences arise throughout the lifetime of an individual. Fraga et. al.[69] examined locus specific and global DNA methylation as well as histone acetylation patterns in a cohort of monozygotic twins. They found that during their early life, twins are epigenetically indistinguishable, but the overall epigenetic profile is different monozygotic twins' age. Investigating this dynamic epigenetic process, that, is within the infancy phase in the field of neurological diseases.

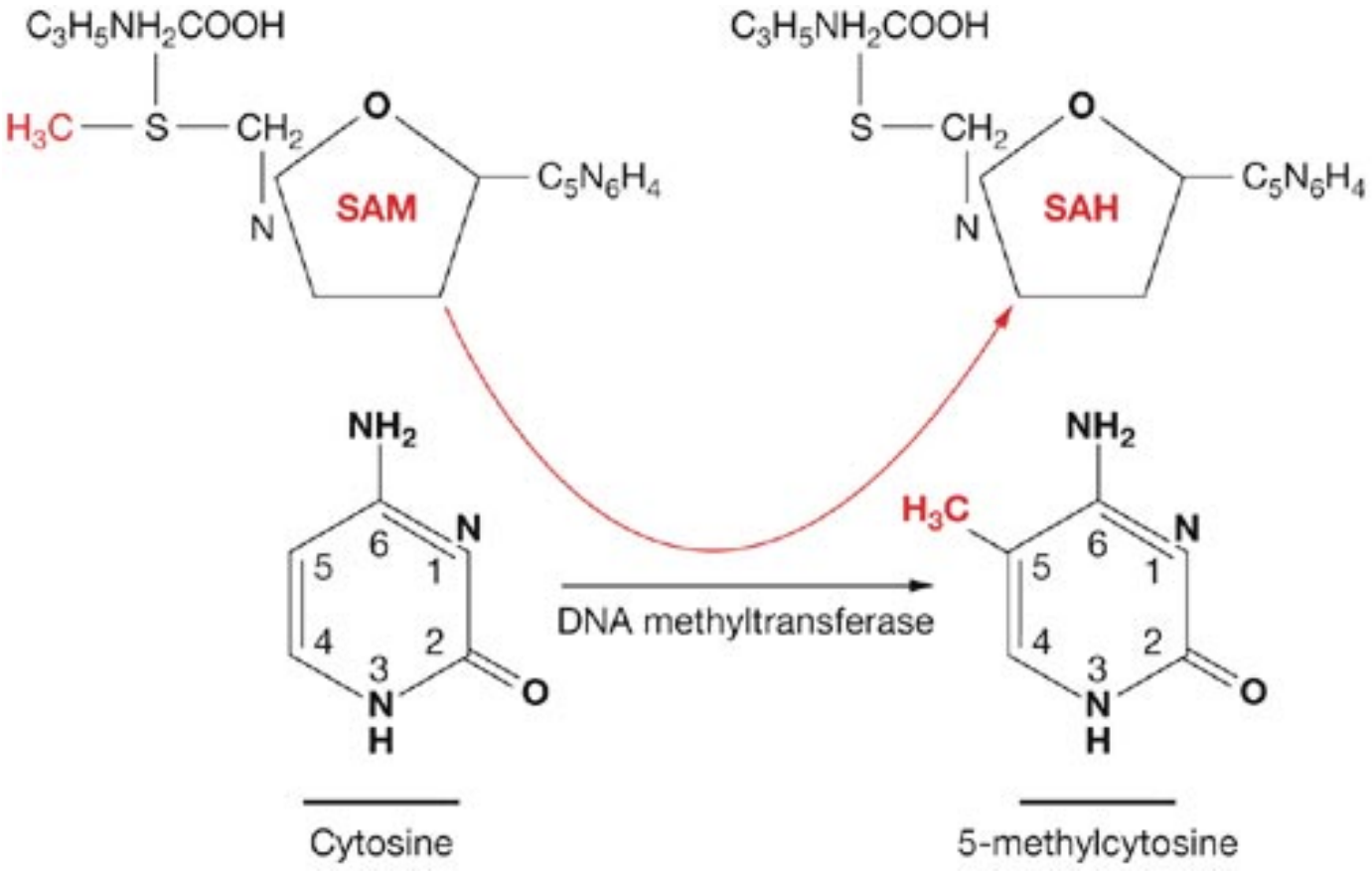


Figure 2: Transition of Cytosine to 5-MethylCytosine in Mammals (Adapted from Richardson *et. al.*[70]) DNA methyltransferases catalyze the transfer of the methyl group from SAM to the 5 position of cytosine in DNA, producing 5-methylcytosine and SAH. Accumulation of SAH suppresses the reaction. The lines under the cytosine bases indicate that these are not free molecules, they are components of DNA. Abbreviations: SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

DNA Methylation and Alzheimer's disease

The field of epigenetics has had an impact on understanding of brain function and neurological disorders. Environmental modulation of epigenetic mechanisms is implicated with many neurological conditions such as depression, Parkinson's disease, Huntington's disease, multiple sclerosis, dementia, and Alzheimer's disease. When examining factors that contribute to LOAD, a large network of interactions has to be taken under consideration. Genomic mutations and epigenetics combine in a complex and dynamic cellular phenotype that impacts important neurobiological functions such as homeostasis, stress response, neuroplasticity, neuroprotection, and neurodegeneration. It is also known that neuronal activity is associated with DNA methylation and histone modifcation[84,86–89].

Various epigenetic modifications have also been reported in disorders of synaptic plasticity and cognition[90] and DNA Methylation has also been suggested as an important molecular mechanism in the maintenance of memory[87]. Studies have shown that in brain, about 80% of CpG sites are methylated, and about 25% of this methylation occurs outside the context of CpG methylation; occurring in Cytosine residues that are not next to Guanine[71,84]. Recent studies show significant associations between alterations of genomic DNA methylation patterns and distinct phenotypes in neuronal systems. Deletion of *DNMT1*, the enzyme responsible for transferring methyl residues to cytosine, in neural precursor cells causes DNA hypomethylation within Central Nervous System (CNS), that disrupts neural control of breathing at birth, and ultimately causes neonatal lethality in mice. In humans, mutations within the *DNMT1* gene is linked to the hereditary sensory and autonomic neuropathy type 1 (*HSAN1*). DNMIT1 mutations cause global hypomethylation, but near CpG Isalnds, hypermethytlation is observed that can potentially contribute to neurodegeneration that manifests clinically with *HSAN1*. In addition, *de*

novo mutations of methyl-CpG binding protein-2 (*MeCP2*) are linked to sporadic cases of Rett syndrome (RTT). RTT is an X-linked, progressive neurodevelopmental disorder caused by mutations in MECP2, a nuclear protein that binds specifically to methylated DNA. Deficiency in MeCP2 protein causes neuronal dysfunction of mature neurons. In mice showing an RTT-like phenotype, MeCP2 has been shown to regulate the expression of a wide range of genes within the hypothalamus by functioning as both an activator and a repressor of transcription. Within humans, a strong correlation between the frequency of *MeCP2* mutations and distinct clinically diagnosed RTT phenotypes has been found within a cohort of patients. Additional studies show that neuronal activity can cause genome-wide changes to 5mC residues. Guo *et al.*[89] demonstrated that activating dentate granular neurons within the hippocampus via electro stimulation resulted in specific modifications in 5mC residues. In these experiments, *de novo* methylation occurred at new CpG sites and caused demethylation of CpG sites that were previously methylated.

In terms of the two hallmarks of AD, the *APP* promoter is estimated to have a GC content of 72% and the rate of the CpG dinucleotides is five times higher than what is normally observed in other eukaryotic promoters [91]. Analysis of methylation status in healthy brain tissue failed to detect the presence of methylcytosines in the 460bp-275bp region of the *APP* promoter. However, the 500bp upstream region showed brain tissue specific profiles of methylation that were associated with *APP* expression[91] . Studies have also suggested that age related demethylation may impact β-Amyloid deposition in the brain[91] . **β** - secretase (*BACE1*), and Presenilin 1 (*PSEN1*) have also been shown to be regulated by methylation[91]. In addition to methylation of cytosines, acetylation/deacetylation of histones has also shown to impact neurological disorders. The balance between Histone Acetyltransferases (HATs) and Histone

Dacetylase(HDAC) expression plays a role in neurological disorders, where the imbalance has been showing to play a role in neuronal apoptosis[92]. Recently it has been found that malfunction of the HAT-CREB-binding protein causes changes in chromatin acetylation status and this loss of function is associated with neurodegenerative disease[92]. The dysregulation of histone Acetyltransferases (HATs) can play a role in AD[93] and has been linked to clinical disorders, and inhibitors of histone acetyltransferases have been studied for use in treating neurodegenerative disorders such as Huntington's disease, depression, and schizophrenia[90]. After initial cleavage of *APP*, γ secretase activity generates Aβ and an intracellular tail fragment. This intracellular fragment has been found to recruit HAT-Tip60, and may play a role in the expression of certain genes[90]. The protein HDAC6 positively correlates with tau burden, and a decrease in HDAC6 promotes tau clearance, making HDAC6 a key factor in regulation of tau protein levels[31]. HDAC inhibition could be an avenue potential therapeutic approach for the treatment of a range of nervous system disorders[93].

Environmental factors such as oxidative stress and its impact on epigenetic modifications have been studied in human neuroblastoma cell lines. Gu *et al.*[94] found that oxidative stress increased intracellular Aβ levels, and *BACE1* expression. An increase in *BACE1* expression and a decrease in DNA methyltransferases was also correlated with demethylation of the *BACE1* promoter region. Oxidative stress has also been observed to induce an increase in HAT expression and a decrease in HDAC expression. Caffeine has also been speculated as an environmental agent that is protective against AD progression, potentially as an epigenetic modulator[95]. Going further than just cell line studies, there have been studies that examined the physiological benefits of environmental factors on aging and AD within animal models. Exercise and environmental enrichment (EE) is found improve learning and memory, and leads to an

increase in neurogenesis and angiogenesis within the hippocampus of aged mice[96]. At the same time, EE and exercise has shown to slow the progress of aging within the brain in rodents[96]. Physical exercise itself has shown to delay cognitive deterioration that is observed in AD patients. The epigenetic changes caused by certain environmental factors on CNS are not well known. In mammals, specifically monkeys, exposure to lead (Pb) at a young age and its impact on *APP and BACE1* expression has been studied. The authors reported that exposure of lead caused an increase in *APP*, Aβ, and *BACE1* gene expression; a decrease in DNMT1, DNMT3a, MECP2 expression, and also observed higher levels of oxidative damage to DNA[97].

Another study has shown that global levels of 5mC inversely correlate with the presence of NFTs[72], a hallmark of AD. The study examined 5mC levels in post mortem cortical tissues in AD patients and age matched controls. A recent study demonstrated that immunoreactivity of 5mC in the neurons of postmortem cortical tissue from AD patients is significantly less than in age-matched controls, proposing that a significant global loss of 5mC takes place within AD. Another finding showed the promoter region of APOE ε4, was hypermethylated in brain tissue samples of AD patients. In addition, studies have shown a loss of DNA methylation in the entorhinal cortex, and hippocampus of AD individuals[98,99]. This illustrates the complex relationship that has to be considered when looking at DNA methylation within the AD brain.

It is clear that epigenetics plays an important role in AD. Evidence thus far illustrates the multifaceted role of DNA methylation in the neurodegenerative process. Investigating this dynamic epigenetic modification is still within the infancy phase in the field of Alzheimer's disease. Understanding how DNA methylation contributes to AD pathogenesis is unclear given the problem of DNA methylation being such a dynamic process that is impacted by various external processes and highly variable within individuals.

RNA and Alzheimer's disease

The one gene and one protein hypothesis was first stated by Beadle & Tatum[116], and the central dogma of molecular biology[117] provides the classic view of how DNA is translated to protein via an RNA intermediate. However, as sequencing techniques have advanced, our knowledge of molecular biology has tremendously increased and it has become lucid that the pathway from gene to protein is a very intricate and complicated process. It is known that there is no correlation between genome size, number of genes, and the complexity of an organism. This seems counter intuitive as it would be assumed that increased organism complexity would correlate to a larger number of protein coding genes. The human genome contains about 30,000 protein coding genes, compared to the salamander, which has 20 times the number of genes as does a human[118]. One trend that does exist among higher eukaryotes and indeed humans is an increase in alternative splicing events. It has been thought that up to 98% of the transcriptional output of the genome is made up of non-coding RNAs(ncRNAs)[119]. This would seem as an immense waste of cellular energy, but it is now known that these ncRNAs play major regulatory roles and are involved in chromatin remodeling, RNA-DNA, RNA-RNA, and RNA-protein interactions, as well as other ways of regulation that are not yet understood[120].

The transcriptome of a cell reflects cellular activity within a tissue at a given point in time. Transcriptome profiling can help provide an unbiased approach for investigating the pathogenesis of complex diseases like AD. Transcriptome analyses has been performed using transgenic animal models of AD and patient-derived cell lines. In contrast to these approaches, post-mortem human brain tissue is difficult to obtain, and some RNA quality concerns exist that might potentially influence transcriptome studies. Nevertheless, post-mortem brain tissue, being identical to the tissue affected by the disease, is the gold standard against which all other model

systems are evaluated. Transcriptome studies of AD utilizing brain tissue have however generated mostly discordant results[109].

In transcriptome studies performed so far for neurodegeneration (ND) and neuropsychiatric disorders, source of RNA used has been the mRNA isolated from transgenic animal models and, recently, patient-derived cell lines. Even though post-mortem brains have been frequently reported as the 'gold standard', there are definite difficulties that arise as a result of the nature of using postmortem human brain tissue and the fragile nature of isolated RNA render transcriptome studies quite difficult. Microarray analysis, widely used for ND and neuropsychiatric disorders, provided much information about the transcriptional profiles in pathological states, and results that are not concordant have been reported[109]. This non-concordance may be due to microarray drawbacks, but also due to the variable quality/integrity of RNAs, affecting the measure of gene expression levels. As ND patients have prolonged agonal state in brain tissue (strongly correlated with pH alterations), differences in RNA integrity may, to some extent, account for aberrant gene expression profiles[121].

Some recent papers have pointed out the great advantages of using RNA-seq to profile the transcriptome of brain tissue affected by ND. Nonetheless, to date a few studies have been published with direct use of RNA-Seq on AD patient brains[109,122,123]. Most others have utilized animal models and human neurons derived from induced pluripotent stem cells proposing an ideal system for further studies on defective neurogenesis in patients. The study of Twine *et al.*[109] has provided, for the first time, an extensive transcriptome analysis of postmortem frontal and temporal lobes of AD patients, highlighting a differential expression of known causative genes and also of previously unannotated expressed regions. It should be considered that given the high-level complexity of the human brain, achieved with the same number of genes as those

of less evolved organisms, some of its complexity may probably be due to alternative splicing and alternative promoter usage. Such events have been described in this study and possibly associated to the progression of neurodegeneration in patients. Twine *et al.*[109] utilized RNA-seq as their tool for analysis of differentially expressed genes (DEG) within postmortem human brain tissue and tissue of normal individuals and individuals affected by AD. Significant differences in gene isoform expression levels, alternated use of promoters and transcription start sites between normal and AD brain tissue were reported, however, it's important to note that their cohort did not contain enough biological replicates nor did it contain age matched controls. Mills *et al.* [122], have also utilized RNA-seq on normal and AD parietal lobes of postmortem human brain tissue, where differentially expressed genes were revealed in pathways related to lipid metabolism.

Transcriptional profiling has been performed in mouse models that overexpress the human tau protein. These studies have shown that when compared with non-NFT-bearing CA1 neurons obtained from normal control brains, NFT-bearing CA1 neurons from AD brains significantly under express genes for cytoskeletal elements, dopamine (DA) receptors, glutamate receptors (GluRs), protein phosphatase subunits 1α and 1γ, kinases, APP and synaptic markers including synaptophysin, synaptotagmin, synapsin I, α-synuclein and β-synucleing[124–126]. These neurons where NFT was over expressed in AD hippocampi also show overexpression of protein phosphatase 3CB[127].

Transcriptomic studies have also implicated the possible role of genes related to synaptic function in AD pathogenesis. These studies emphasize genes coding for proteins associated with presynaptic vesicles or with the postsynaptic binding machinery[39,126,128–131]. Synaptic failure has been thought to be an affiliated feature, along with amyloid or tau pathology, that contributes to the cognitive decline observed in AD[13,132]. It has also been found that the expression of genes

coding for synaptic markers including synaptophysin, synaptotagmin, synapsin I, α-synuclein and β-synuclein[124–126] are down-regulated in NFT-bearing CA1 neurons. Reduction in synapsin I gene expression was also found in the entorhinal cortex[133].

Only a handful of studies[109,122,134,135] have utilized RNA directly from postmortem brain tissue of humans for transcriptome analysis. These studies either **lack enough biological replicates**, a key entity that increases the power to detect differentially expressed genes[136], or **lack enough sequencing depth** as recommended by ENCODE[137]. These studies, limited as they are, show DEGs that cluster within synaptic transmission, transmission of nerve impulse, immune response, lipid metabolism, mitochrondrial dysfunction, and neurotransmitter transport when performing gene ontology analysis[109,122,134,135].

The complexity of humans is not only because of the number of genes present within the entire genome, but largely due to the complex intertwined connection between regulation of these genes and their expression. Such a fine-tuned system is subject to certain disturbances that result in disease. Hence, it is important to examine gene expression differences between cases and controls to understand the changes that could be contributing to disease pathology, rather than focusing on finding on a one gene causative approach.

For the approach used within this project, the source of RNA is from postmortem brain tissue made available by the University of Texas Southwestern Medical Center at Dallas. We utilize postmortem brain tissue as it is considered the gold standard for these studies, since one is able to directly look at tissue impacted by the disease. However, this has considerable challenges, as RNA is fragile and easily degrades due to the high availability of RNAses. These studies are very limited because brain tissue is the only majorly affected tissue in these neurodegenerative diseases, and the availability of samples is scarce. The fragility of RNA, and post mortem

intervals, along with the complexity of RNA degradation, accompany post mortem human samples, and is something that cannot be controlled for. A novel probe-based capture approach that combats RNA degradation and allows for the use of RNA-seq even in poor quality samples was used within this project.

Chapter II

MATERIALS AND METHODS

All methods were performed at the High Throughput Sequencing Facility, located at University of North Carolina Chapel Hill.

Project Overview

Alzheimer's disease (AD) is a rising problem; it is estimated that the affected U.S. population will reach 13 million by 2050. Gaps in knowledge still exist in terms of understanding Late Onset AD pathogenesis, progression, and diagnosis. In terms of pathogenesis, two hypotheses currently dominate within the field: the amyloid beta (Aβ) cascade hypothesis and the presence of neurofibrillary tangles (NFTs).

The purpose of this project is to investigate DNA methylation and RNA expression between age-matched controls and AD patients. DNA methylation was assessed using a probe-based microarray method that investigates over 450,000 individual CpG dinucleotides, and using a CpG Methyl Binding Protein capture approach followed by deep sequencing. RNA expression was assessed by developing cDNA libraries using a novel probe-based capture approach followed by deep sequencing.

***Hypothesis*:** Late Onset Alzheimer's disease (LOAD) is a heterogeneous disease where a large portion of the heritability is unknown. We hypothesize that DNA Methylation alters gene regulation and contributes towards the AD pathogenesis

This hypothesis was tested in collaboration with Dr. Kirk C. Wilhelmsen at University of North Carolina at Chapel Hill. Human post-mortem prefrontal cortex brain tissue of 11 AD individuals and 12 age-matched controls, were acquired from the brain bank located at University of Texas Southwestern Medical Center at Dallas, Texas. Individuals were classified as controls if a score of four and lower was achieved, and classified as an AD patient if they achieved a score of five on the Braak staging scale[138,139].

DNA and RNA extracts from postmortem frontal cortex brain tissue were investigated for three measures:

(1) Change in RNA expression using a novel RNA sequencing approach;

(2) Site specific genomic DNA methylation using the Illumina® 450K methylation microarray; and

(3) Analysis of genomic DNA for differentially methylated regions using a protein capture approach to enrich for methylated regions, followed by deep sequencing.

Sample Description:

The subjects used for these studies were acquired following protocols approved by University of Texas Southwestern Medical Center at Dallas. Freshly frozen human postmortem Frontal Cortex samples of 11 AD individuals and 12 age-matched controls, were acquired from the UT Southwestern brain bank. Individuals were classified as having AD if they achieved a score of V on the Braak staging scale[138,139]. A Braak staging of V is achieved when the NFT distribution is found within the neocortical region of the brain, with a clinical impression of having AD. The samples utilized in this study, and the patients' age, sex, Post Mortem Interval in hours (PMI), the cortical region of the brain from where the tissue was acquired, and the respective Braak Staging Score is outlined in Table 1.

Table 1: Samples utilized within this study. Sample ID, Disease Category (Control (CTL) or Alzheimer's disease (AD)), Patient's Age, Sex, Post Mortem Interval in hours (PMI), Cortical Region of the brain from where the tissue was acquired, and the respective Braak Staging Score of each patient.

Sample ID	Category	Age	Sex	PMI(hrs)	Cortical Region	Braak
24381	CTL	68	Male	6.3	Frontal	II
24385	CTL	73	Male	4.8	Frontal	0
24430	CTL	72	Male	8	Frontal	I
24509	CTL	67	Male	9.5	Frontal	0
25666	CTL	96	Female	5	Frontal	II
36082	CTL	69	Male	14.8	Frontal	0
36359	CTL	84	Male	23.5	Frontal	IV
42133	CTL	100	Female	12.2	Frontal	IV
45116	CTL	79	Male	10.3	Frontal	III
46202	CTL	77	Male	19.5	Frontal	II
25667	CTL	75	Female	3.7	Frontal	II
42990	CTL	84	Female	14	Frontal	I
Average		**78.67**		**10.97**		
40482	AD	81	Male	18	Frontal	V
41176	AD	79	Male	9	Frontal	V
35289	AD	74	Female	12.7	Frontal	V
41639	AD	69	Female	11.5	Frontal	V
41969	AD	86	Female	25	Frontal	V
43045	AD	69	Male	10.2	Frontal	V
43172	AD	91	Male	8.5	Frontal	V
43192	AD	82	Male	24.8	Frontal	V
43509	AD	79	Male	5	Frontal	V
44624	AD	95	Female	10.5	Frontal	V
45392	AD	65	Male	16.8	Frontal	V
Average	**AD**	**79.09**		**13.82**		

RNA Methods

Overview of RNA Sequencing

Deep sequencing is rapidly becoming the method of choice for transcriptional profiling experiments. Unlike RNA microarrays, sequencing RNA (RNA-seq) allows identification of novel transcripts, avoids background noise associated with fluorescence quantification as used in microarrays. RNA-seq goes beyond just providing information on gene expression, by incorporating data on alternative splicing events and the presence of non-coding RNAs. RNA-seq allows genome-wide analysis of transcription at single base pair resolution and allows identification of alternative splicing events, and post-transcriptional RNA editing. RNA-Seq also detects new splicing variants, and allows for precise quantitative determination of exon and splicing isoform expression.

Next Generation Sequencing (NGS) techniques have several advantages over current microarray technologies. RNA-Seq has a low frequency of false-positive findings and is highly reproducible compared with microarray technology. RNA-Seq produces a digital signal from the cDNA template sequence, whereas microarray analyses must overcome the issue of nonspecific probe hybridization. Validation techniques such as quantitative PCR (qPCR) and spike-in RNA have demonstrated that RNA-seq is extremely accurate[109–111]. A false positive rate <2% has been demonstrated for this technique. As recently reported by Marioni *et al.*[112], qPCR results agreed more closely with Illumina® sequencing results than with microarrays[112]. Furthermore, RNA-Seq has a much higher resolution of transcript structure than microarrays, which allows for the identification of the transcript boundaries at the level of a single base. These qualities make

RNA-Seq particularly useful for studying complex transcriptomes, such as those found in the human brain[113].

RNA-seq is not only highly suited to investigations of the complex human brain tissue but it can potentially overcome technical issues inherent to case–control comparisons of postmortem brain tissue in neurodegenerative diseases. RNA-seq experiments follow an overall similar protocol. Total RNA is isolated from a sample of interest, which may be purified to enrich for messenger RNAs (mRNAs), microRNAs (miRNAs) or long non-coding RNAs (lncRNAs) before preparing an RNA library that is created by reverse transcription to cDNA. Sequencing can produce one read in a single-end sequencing reaction, or two ends separated by an unsequenced fragment in paired-end reactions. Together, RNA-seq has allowed an unparalleled view of the transcriptome in normal and pathological processes and has revealed that the transcriptome is more complex than previously described.

Transcriptome analysis using microarray-based methods is responsible for most of the current understanding of gene expression in development and disease. One disadvantage of microarray analysis is the limitation of detection of only known transcripts. Microarray hybridization-based detection suffers from many disadvantages such as poor sensitivity, and low specificity. Reports that compare microarrays and RNA-seq, do state that RNA-seq is superior to microarrays, when considering low frequency of false positive signals and high reproducibility of the method. Van Bakel *et al.*[114] report that the transcript analysis of intragenic regions unambiguously showed the hybridization signals from microarrays can lead to false positive signals of transcripts that are in low abundance. RNA-seq on the other hand, reduces these limitations. Non-specific hybridization or cross-hybridization is also of concern when interpreting microarrays, especially with closely related gene family members with highly

similar sequences. RNA-seq addresses this by eliminating the ambiguity of sequence detection by mapping reads to a reference genome; however, a single sequenced read may map to multiple locations. To avoid this dilemma, paired-end sequencing solves this problem, and this additional information of a paired read allows for more accurate mapping.

One disadvantage of RNA-seq could be the introduction of bias that PCR introduces due to GC content and length of the amplicon. A number of analysis tools accordingly correct for this. Since the total number of reads per transcript is proportional to the level of a transcript multiplied by transcript length, a long transcript will be sequenced more often than a short transcript when expressed at equivalent levels. Since statistical power is closely linked to sample size, a long transcript is more likely to be found differentially expressed than a short transcript. To get around this problem, expression levels are frequently expressed by calculating the number of reads as fragments per kilobase per million reads (FPKM). FPKM transformation enables direct comparison of RNA transcript expression levels between two libraries with different sequencing depth and determination of relative expression levels between two or more transcripts in a single library.

There have been some concerns regarding sense and anti-sense transcription and its biological relevance. Even though the field tends to agree that if the goal is to determine differential gene expression, the strand information will not affect the data generated, a stranded library approach was pursued anyway which is standard within the Illumina® TruSeq® RNA Access Library Preparation protocol. When making stranded libraries, during the 2nd strand cDNA synthesis, uracil is incorporated instead of thymine. Illumina® library prep continues as normal, but after adapter ligation and before PCR amplification, Uracil-DNA glycosylase is used to degrade the 2ndstrand. This results in all reads starting in the same orientation so you can

determine which strand was being transcribed within the sample. Figure 3 (Adapted from Zhao *et al.*[115]) illustrates non-stranded versus stranded RNA-seq protocol. The disadvantage of using a stranded library approach is that you lose the information about antisense transcription. It been realized non-coding RNAs are more than just artifacts of erroneous transcription and play vital regulatory roles at the genomic, transcriptional and translational level. Transcription of DNA sense strand produces these antisense RNA transcripts. This antisense transcription results in the production of non-coding RNAs that are complementary to their associated sense transcripts, and some studies that show production of protein from some antisense transcripts. It seems that antisense transcriptional 'hot spots' are located around nucleosome-free regions such as those associated with promoters, indicating that it is likely that antisense transcripts carry out important regulatory functions, and does occur within mammals.

Lastly, there are some benefits for utilizing paired-end reads (PER) instead of single-end reads when performing deep sequencing. Having pairs of reads improves read alignment and allows for detection of chromosomal rearrangements, like deletions and insertions. Specifically, within RNA, PER allow for detection of different isoforms. When utilizing paired-end reads, one knows roughly the range of the distance of two reads. This helps when mapping reads back to a reference genome. In the case of splicing junction events, two exons maybe joined together. If paired-end read 1 and read 2 are mapped to a genome with 1kb between them and the expected distance between them in the transcript is only about 500bp at maximum, then one may conclude that an intron of around 500bp may be removed. This information is lost when using only single-end reads. Paired-reads also improve directional sequencing accuracy.

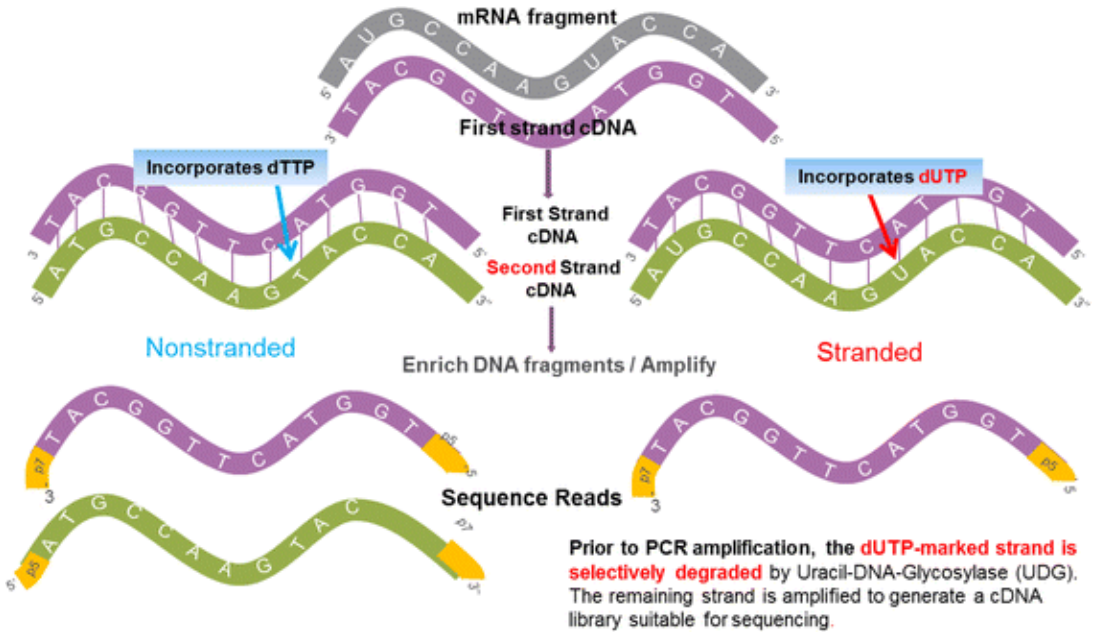


Figure 3: Non-stranded versus stranded RNA-seq protocol (Adapted from (Zhao *et al.*[115]). The stranded protocol differs from the non-stranded protocol in two ways. First, during cDNA synthesis, the second-strand synthesis continues as normal except the nucleotide mix includes dUTPs instead of dTTPs. Second, after library preparation, a second-strand digestion step is added. This step ensures that only the first strand survives the subsequent PCR amplification step and hence the strand information of the libraries.

RNA extraction, library preparation, and sequencing:

RNA extraction was performed by utilizing the RNeasy Lipid Tissue Mini Kit (Qiagen, Catalog No. 74804). This kit was chosen due to its inclusion of the QIAzol Lysis Reagent (Qiagen, Catalog No. 79306), that allows for high yields of RNA from fatty tissues, since the brain is comprised of a large amount of lipids. It also includes a clean-up column step to minimize the carryover of any compounds used within the organic extraction method that may interfere with the downstream library preparation process. Approximately, 95 μg of freshly frozen tissue was placed in 1ml of QIAzol Lysis Reagent. The tissue was homogenized within the QIAzol solution for 60 seconds at 20,000 rpm using the TissueRuptor (Qiagen, Catalog No. 9001271). After homogenization, the manufacturer recommended protocol was followed for RNA extraction. RNA quality was assessed by using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Catalog No. 5067-1511). The Agilent 2100 Bioanalyzer (Agilent Technologies, Catalog No. G2940CA) system was used to perform the analysis. This system allows for the sizing, quantitation, and quality assessment of RNA. The system comes with the proprietary 2100 Expert Software that is used for RNA integrity number (RIN) analysis. This tool was designed to help scientists estimate the integrity of total RNA in samples. The expert software automatically assigns an integrity number to any eukaryote total RNA sample. The RIN algorithm uses the entire electrophoretic trace, and not just by the ratio of the 18S and 28S

ribosomal RNA, which makes up ~80% of most RNA samples. This tool allows for a robust and reliable prediction of RNA integrity[140]. A wide range of RINs was observed within our samples, from 3.7 to 6.8, most likely due to the variation in post mortem intervals before sample collection. Sample purity of RNA and RNA quantification was done utilizing the NanoDrop 2000 UV-Vis Spectrophotometer (ThermoFisher). Table 2 provides the RIN, quantity of RNA in ng/μL and sample purity using the A260/280 and A260/230 ratios as given by the NanoDrop for the test samples. RNA extractions were repeated three times for all 23 samples, and the samples with the highest RIN among the three extraction trials were chosen for library preparation. The average RIN for the samples advanced to the library preparation process was 5.6.

Table 2: List of RNA samples characteristics that were advanced to library preparation.

Sample IDs, quantity of RNA sample obtained (ng/µl), A260/230, A260/280, and RNA Integrity Number is given for each sample.

Controls	Nanodrop quantification (ng/µl)	A260/230	A260/280	RIN from Agilent
24381	322	2.07	2	6.7
24385	228	0.99	1.98	4.5
24430	254	2.1	1.99	6.4
24509	183	2.21	1.97	6.2
25666	227	1.27	1.99	5.7
36082	245	1.76	1.96	6.1
36359	260	2.17	1.99	5.5
42133	349	2.2	1.98	6.1
45116	226	2.07	1.96	6.2
46202	197	1.49	1.93	5.1
25667	268	1.2	2.02	6
42990	203	1.47	1.93	5.8
AD Cases				
40482	231	2.18	2	5.6
41176	273	2.13	1.98	5
35289	217	0.88	1.99	4.3
41639	242	1.46	2.01	3.7
41969	248	2.08	2	5.5
43045	265	2.24	1.9	5.6
43172	209	1.51	2.02	4.5
43192	247	2.14	1.95	5.4
43509	240	1.56	2.02	6.2
44624	260	1.8	2.01	5.9
45392	314	2.22	1.92	6.8
Average	248.17	1.79	1.98	5.6

A novel probe based capture method was used for the library preparation process in contrast to the traditional approach. Traditional approaches for mRNA sequencing involves the use of ribosomal RNA (rRNA) reduction methods or poly-A mRNA enrichment followed by the TruSeq® Stranded Total RNA (Illumina®) library preparation method. However, rRNA-reduced whole transcriptome analysis, requires deep sequencing to sufficiently cover both coding and non-coding transcripts, which tend to be very costly. For our mRNA library preparation, the TruSeq® RNA Access Library Prep Kit by Illumina® (Catalog Number RS-301-2001) was used. This kit provides an exon-capture approach that is effective for difficult samples, such as RNA isolated from formalin- fixed paraffin-embedded (FFPE) tissues. FFPE sample archives tend to yield low RINs, as observed within our samples, and the use of this kit helps combat this low quality RNA issue. The TruSeq® RNA Access Library Prep Kit overcomes these challenges by capturing the coding regions directly rather than a poly(A) tail pull-down. Poly(A) capture is problematic in cases of highly fragmented RNA. A combination of over 425,000 probes that target over 98% of the known RefSeq exome are used. The probe set was designed to capture > 214,000 targets, spanning 21,415 genes of interest as seen in Table 3.

Table 3: TruSeq® RNA Access Coverage Details. Table lists the characteristics of the RNA Access library preparation kit.

RNA Access Coverage Details	
Number of Target Genes	21,415
Number of Targeted Exonic Regions	214,126
RefSeq exome covered	98.30%
Number of Probes	425,437

A tremendous advantage of utilizing RNA-seq is the ability to get high quality data at a fraction of the cost that's associated with traditional methods. Illumina® demonstrates much deeper coverage of exons using the TruSeq® RNA Access method, at 1/10 th the number of reads as using the TruSeq® Stranded total RNA Kit[141]. Sequencing of paired FFPE lung tumor and normal samples prepared using TruSeq® RNA Access revealed that > 85% of the bases covered were within the coding and UTR regions of RNA. Figure 4 displays a direct comparison between the TruSeq® Stranded total RNA kit and TruSeq® RNA Access Kit. Samples using the TruSeq® Stranded RNA kit were used to sequence at 250 Million reads, and samples using the TruSeq® RNA Access kit was sequenced at 25 Million reads. One can note the large difference between the amount of coding regions obtained at a fraction of the coverage depth, which means more information per sample at a fraction of the cost.

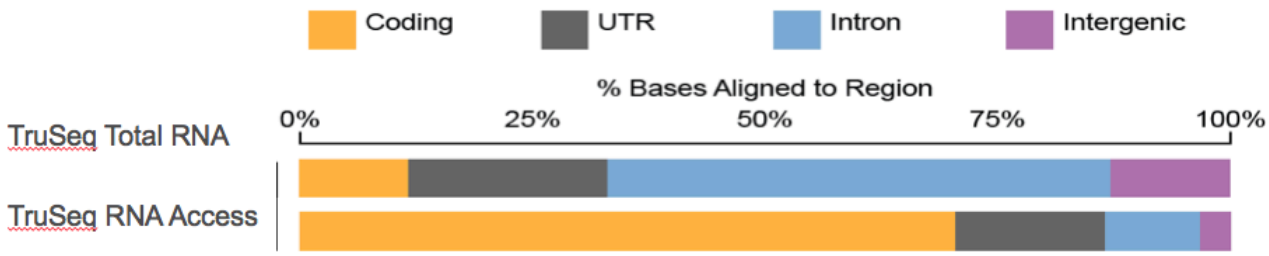


Figure 4: TruSeq® RNA Access vs TruSeq® Total RNA: The same Lung Tumor FFPE sample is sequenced with both the TruSeq® Stranded Total RNA and TruSeq® RNA Access. These samples were sequenced on a Illumina® HiSeq at 250 million reads for Total RNA and 25 million reads for RNA Access. This demonstrates that less than 20% of reads align to the coding region when using TruSeq® Total RNA, in comparison to 70% when using TruSeq® RNA Access. Figure adapted from the Illumina® RNA Access Application Note[141].

The RNA sample libraries were prepared following the manufacturer recommended protocol. Stranded RNA-Seq libraries were prepared using the Illumina® TruSeq® RNA Access chemistry. Figure 5 depicts an overview of the TruSeq® RNA Access Capture Chemistry. During the first step, unique oligonucleotides were added to each library, tagging them for downstream pooling (Figure 5-A). This allows multiple samples to be loaded on a single sequencing run. After libraries were pooled, they underwent a series of capture steps that produce a targeted library, depleted of ribosomal RNA and intronic and intergenic regions. Pooled libraries were hybridized to biotin-labeled probes specific for coding RNA regions (Figure 5-B). Specific targets within the library were then captured by adding streptavidin beads that bind to the biotinylated probes (Figure 5-C) and then magnets pulled the bound RNA fragments from the solution (Figure 5-D). Figure 6 depicts the library preparation workflow that was followed. The library preparation process was completed after a series of capture reactions and PCR reactions as depicted in Figure 6. Validation of library was performed to observe a PCR product, which demonstrated successful completion of library preparation. The average cDNA library fragment size observed was 277bp. The library was then submitted for cluster generation and subsequent sequencing.

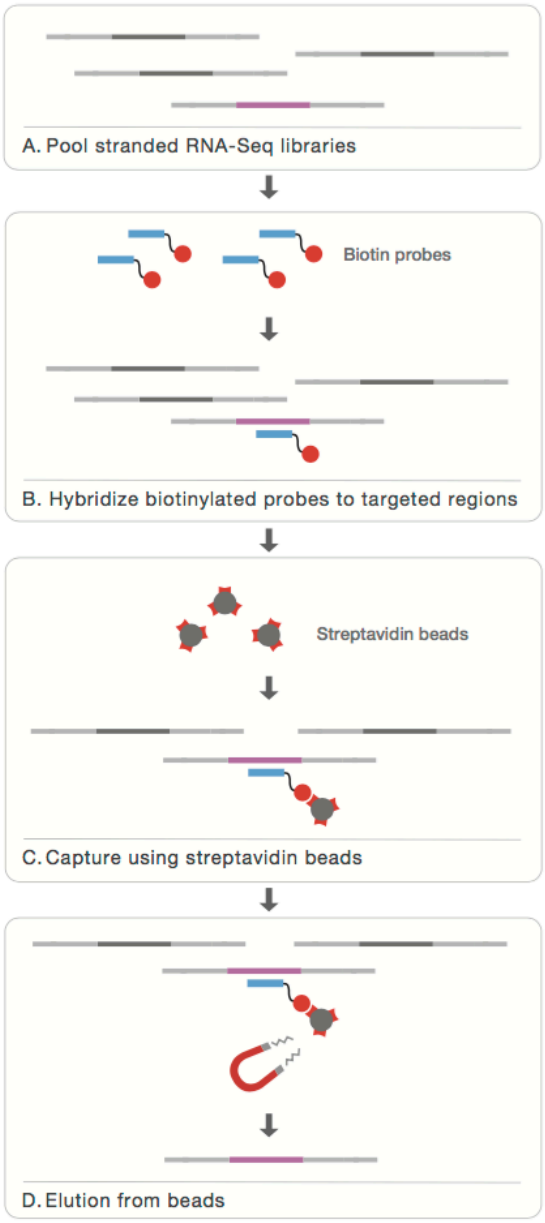


Figure 5: Overview of the Truseq® RNA Access Capture Chemistry. (Adapted from the Illumina® Truseq® RNA Access Application Note[141]). Overview of the TruSeq® RNA Access Capture Chemistry. **5-A:** Unique oligonucleotides added to each library, allows for multiplexing of samples. **5-B:** Pooled libraries were hybridized to biotin-labeled probes specific for coding RNA regions. **5-C:** Streptavidin beads bind to the biotinylated probes to capture specific fragments. **5-D**: RNA fragments separated and eluted from the magnetic beads.

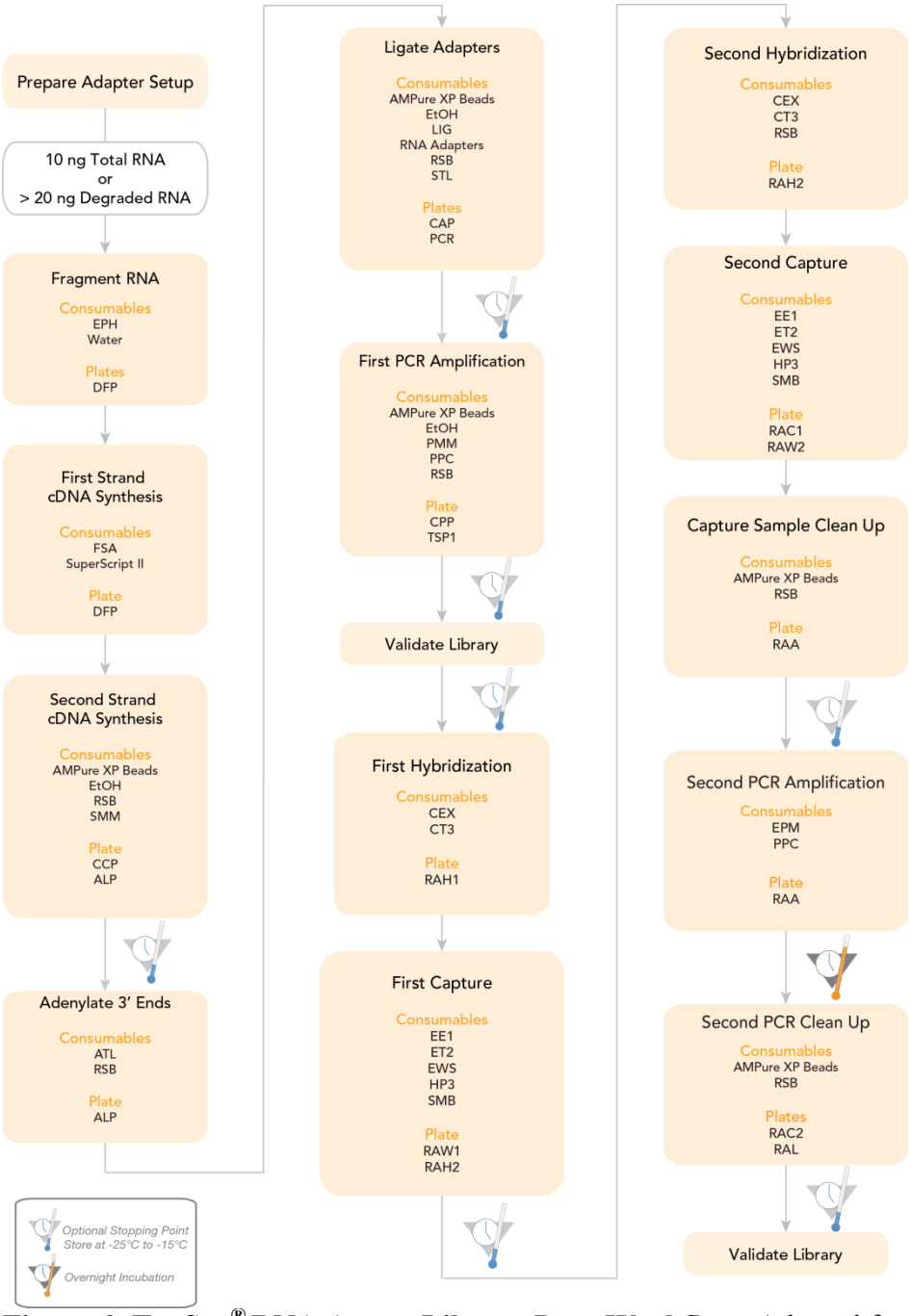


Figure 6: TruSeq® RNA Access Library Prep Workflow: Adapted from Truseq® RNA Access Library Preparation Guide(Revision B)[142]. Work flow of the Truseq® RNA Access Library preparation process that involves a series of capture, amplification and clean up steps.

Sequencing was performed on the Illumina® Hiseq 2500 instrument, in a paired end 2 x 100bp manner. A total of three lanes were utilized for RNA-seq, where two lanes consisted of 8 samples, and one lane received 7 samples (n=23).

RNAseq Data Analysis:

RNA-seq experiments must be analyzed with robust, efficient and statistically principled algorithms. The bioinformatics community has successfully been able to develop ideas into tools that utilize mathematics, statistics and computer science for RNA-seq analysis. FASTQ files generated by the HiSeq were first processed using TopHat[143,144]. TopHat is a fast splice junction mapper for RNA-seq reads. It aligns reads to mammalian-sized genomes using the short read aligner Bowtie[145], and then analyzes the mapping results to identify splice junctions within exons. TopHat finds splice junctions without a reference annotation by first mapping RNA-Seq reads to the genome, and identifies potential exons, since many RNA-Seq reads will contiguously align to the genome. Using this initial mapping information, TopHat builds a database of possible splice junctions and then maps the reads against these junctions to confirm them. All the FASTQ files generated were aligned to the UCSC RefSeq (hg19) annotation. After alignment, a list of read alignments called 'accepted hits' are stored in a .bam file (binary format for storing sequence data) file. "BioC2014: RNA-Seq workflow for differential gene expression", a guide published by Love *et al.*[146] was followed to analyze DEGs. This workflow uses DeSEQ2[146] as the package to run the pipeline. DeSEQ2 is an available R/Bioconductor package that is widely used for differential gene expression analysis.

RNA-seq counts also show a gene length bias: the expected number of reads mapped on a gene is proportional to both the abundance and length of the isoforms transcribed from that gene.

Indeed, longer genes produce more reads than shorter ones, causing higher power for DEG detection. This is corrected by Mortazavi *et al.'s*[147] approach that proposes to summarize mapped reads as 'Reads Per Kilobase of exon model per Million mapped reads' (RPKM), that is calculated by dividing the number of reads aligned to gene exons by the total number of mapped reads and by the sum of exonic bases. RPKM helps reduce both differences in library size and length bias. Some DEG packages use RPKM method for analysis. It's a simple method is to count reads overlapping each region, dividing by the length of the region of interest to accommodate differences in gene length. A problem with this approach is that reads are not sampled uniformly across genes, so gene length (the 'PK' part of RPKM) is not a good proxy for expression level. Each read represents an observation, and contributes to the certainty with which a gene is measured as 'expressed'. The summary RPKM measure fails to incorporate uncertainty, which is where a particular value of RPKM may result from alignment of one or 100 reads. This contrasts with a simple count of the number of reads in the region of interest. Count data has known statistical properties that can be exploited in statistical analysis. Therefore, it is stated the most useful for assessing differential expression is read count.

DeSEQ2 uses a generalized linear model to evaluate differential expression while accounting for biological variance and uses a Wald test statistic to evaluate significance. A DESeqDataSet object is created using HTSeq counts and the DeSEQ wrapper function is called to perform differential analyses. The fold change is determined by dividing the average normalized read counts of AD samples over control samples for each transcript. P-values are corrected using the Benjamini and Hochberg False Discovery Rate[148].

DeSEQ2, uses raw count data obtained from .bam files. These count values are raw counts of sequencing reads. This allows DeSEQ2 to use a data-driven approach to provide more

robust dispersion estimates, and rely on strategies to moderate per-gene estimates with more robust local estimates derived from genes with similar expression values. The estimation of the parameters for the respective statistical model is followed by the test for differential expression, the calculation of the significance of change in expression of gene between two conditions. DeSEQ2 uses a variation of the Fisher exact test; hence, they return exact P values computed from the derived probabilities. DeSEQ2 uses raw counts as input and has its own normalization methods.

For the RNA-seq dataset in this project, the use of multiple samples within the same disease category (i.e. AD vs Control) served as biological replicates within each comparison category. Genes were identified as differentially expressed if they were observed at ± 0.6 logarithmic fold change (lfc) and significance expressed as q-value (FDR-adjusted p-value < 0.001) as determined by DeSEQ2. These stringent thresholds were chosen, based on what is accepted within the RNA-seq community, and due to the large number of replicates we have within our dataset for each treatment.

DEGs from DeSEQ2 were imported and a core analysis was performed utilizing QIAGEN's Ingenuity® Pathway Analysis software (IPA®), (Qiagen, Redwood City, www.qiagen.com/ingenuity). Core analysis performed by IPA®, helps understand complex transcriptonomics data at multiple levels by providing insight into the molecular and chemical interactions and disease processes within the RNA-seq data. IPA® identifies regulators, relationships, mechanisms, functions, and pathways relevant to changes observed in our dataset. Analytics go beyond pathway analysis to understand experimental results within the context of biological systems. The same parameters, P_{adj} <0.001, lfc ± 0.6 were used. Conservative settings for the Core Analysis were used: direct relationships only, excluded casual & indirect interaction

networks, direct interaction networks only with 35 molecules per network & 25 Networks per analysis, using all data sources, confidence setting of experimentally observed, human species only, and excluded all tissues or cell lines.

DNA Methods

Overview of Methylated DNA Enrichment:

All current methods have certain drawbacks that prevent a truly high-throughput, unbiased, and detailed profiling of genomic cytosine methylation. Whole genome bisulfite sequencing is the gold standard for cytosine methylation analysis, and provides single base pair resolution of methylation patterns throughout the entire genome. Bisulfite treatment of genomic DNA chemically converts unmethylated cytosine to uracil, but leaves methylated cytosines intact. Uracil corresponds to thymine with respect to its basepairing behavior, and after sequencing, DNA-methylation status differences are implied by sequence differences mapped to a genome. To be able to obtain a complete DNA methylome, a minimum of 30X coverage depth (90Gigabases) is the standard set for each biological replicate by the scientific community. In addition, a considerable amount of unconverted DNA is usually spiked in during sequencing to compensate for the lack of diversity of bases in bisulfite-treated DNA; this further increases the cost of performing whole genome bisulfite sequencing. Alternative approaches are based on specific enrichment of methylated portions of the genome. Methylation-sensitive restriction enzyme digestion allows the enrichment of highly methylated regions of the genome [100,101]. This method introduces recognition site biases, gives poor resolution, and is prone to false positives due to incomplete enzymatic digestion. Anti-5-methyl-cytosine antibody immunoprecipitation, known as MEDIP-seq, captures any DNA fragment containing one or more methylated cytosines[100,101]. This causes sporadically methylated DNA fragments to make up a significant portion of the data.

Alternatively, relatively inexpensive bead array-based methods have been developed for bisulfite treated DNA Such as the Illumina® 450K Infinium BeadChip. Though this covers a major part of the human DNA-methylome, it lacks genome-wide coverage[102]. The 450K microarray includes CpG and CNG sites, CpG islands/shores/shelves/open sea, non-coding RNA (microRNAs and long non-coding RNAs) and sites surrounding the transcription start sites (-200 bp to -1,500 bp, 5'-UTRs and exons 1) for coding genes, also for the corresponding gene bodies and 3'-UTRs, in addition to intergenic regions derived from GWAS studies. It covers over 96% of CpG islands, and is a widely used and well-documented tool for epigenome analyses[102].

The alternative for bisulfite-treated DNA characterization is the capture of methylated DNA fragments followed by sequencing, which allows for a cost-efficient genome-wide approach. To overcome these limitations, a method that uses the human Methyl CpG Binding Domain 2 protein to capture methylated genomic DNA, can be used. Following capture using a Methyl Binding Domain(MBD), the DNA undergoes deep sequencing, known as MBD-seq, and allows for high-throughput analysis of multiple samples. Genomic DNA is sheared, followed by the use of the MBD protein construct to precipitate densely methylated sequences. In vivo, MBD2 has shown to bind specifically to methylated CpGs via its Methyl Binding Domain (MBD) and facilitates gene silencing through its transcriptional repression domain and also thru the recruitment of additional transcription inhibitors[100,103] . It is important to note that MBD binds with increasing affinity to multiple methylated cytosines in a close proximity, hence predominantly precipitating biologically relevant, multiply methylated fragments as opposed to sporadically methylated CpGs of uncertain biological relevance[100,104]. Random shearing of the genome by sonication minimizes sequence-specific fragmentation, as compared to restriction enzyme digestion. Aberg *et al.*[105–107] also demonstrated how using MBD sequencing is a cost

effective tool when considering large scale studies.

It was recently demonstrated that, when coverage is sufficient, MBD-seq is generally more sensitive than MeDIP-seq and methylation specific microarrays. A greater sensitivity for MBD-seq compared to MEDIP-seq was also confirmed in a microarray based study[108]. Therefore, until further optimization of sequencing technologies allows for a cost-efficient whole-genome sequencing of bisulfite-treated DNA or direct detection of methylated cytosines at base-resolution, MBD-seq is the better alternative.

DNA extraction, sonication and size selection

Genomic DNA was extracted by using an organic extraction method following the UNTHSC organic extraction protocol[149]. Stain Extraction Buffer (SEB) was freshly prepared using an in house method (Pg. 117). Approximately 90 µg of freshly frozen post mortem brain tissue was placed in 600 µL of working SEB solution. 5 µL of proteinase K (proK) (20mg/mL, ThermoFisher Catalog no. AM2548) was added to this solution. After 12 hour incubation at 65°C, another 5 µL of proK was added. Total incubation time was 24 hours at 65°C. After organic extraction, DNA was ethanol precipitated as stated within the protocol, and pasteurized by incubation at 65°C for two hours. Post extraction, samples were stored at -20°C until further use.

Illumina® 450k Methylation Bead Array

Bisulfite conversion was conducted on 2 μg of DNA for all samples. Conversion was performed following the manufacturer recommendations by using the Zymo Research EZ-96 DNA Methylation™ Kit (Catalog Number D5004). The Illumina® Infinium HumanMethylation450 BeadChip (Illumina®) was processed on the Illumina® HiScan® system.

Illumina® 450k Methylation Bead Array Data Analysis

The following data analysis was performed on the Kure Cluster located at UNC Chapel Hill.

RnBeads[150], is a widely used tool that was developed for analysis and interpretation of DNA methylation data. It builds on prior bioinformatics methods, and allows the implementation of a pipeline with options to customize workflow using custom R scripts. R[151] is a free software environment for statistical computing and graphics, where multiple packages are published and available for data analysis use. The script utilized for this analysis is located within the Appendix (Pg 128). Data analysis starts from raw intensity data (IDAT) files produced from Illumina® Infinium microarrays. RnBeads performs necessary steps such as data normalization, addresses quality control, covariate adjustment, cell type heterogeneity adjustment, and performs differential methylation analysis. Methylation levels (β-values) were estimated as the ratio of signal intensity of the methylated alleles to the sum of methylated and unmethylated intensity signals of the alleles (β-value=C/(T+C)). The β-values vary from 0 (no methylation) to 1 (100% methylation).

Covariate adjustment for the dataset included, age and sex for each sample. The R package, CETS[152], was used to estimate neuronal cell populations within our dataset. CETS is

capable of quantifying neuronal proportions and generating *in silico* neuronal profiles from DNA methylation data. CETs analysis determined that there was less than 1.7% difference in neuronal cell populations between the Control and AD group. Cell type heterogeneity of profiled samples is a source of confounding in DNA methylation profiling studies. A reference free method to adjust for cell type heterogeneity issues using a reference free approach was used. Given that there has been some interest in effects of cell mixture on the measurement of DNA methylation, where small perturbations in cell mixture proportions may register as changes in DNA methylation; where the changes in DNA methylation may arise due to difference in cell mixtures among samples rather than actual change in DNA methylation. To address this problem, Houseman *et al.*[153] published a reference-free approach for cell mixture adjustment within DNA methylation data, which is an option incorporated within RnBeads. Adjustments usually are made after the incorporation of existing reference datasets, that may not always be available for a given tissue.

DNA sonication, size selection, and enrichment of Methylated DNA

After extraction, DNA was quantified using the Qubit™ dsDNA Broad Range kit (ThermoFisher Catalog no. Q32853). After quantification, proper dilutions were made to obtain 5 µg of DNA. DNA was then sonicated using the Covaris® E220 (Covaris, Wobrun, MA). 5 µg of DNA was suspended within 100 µL of elution buffer (Qiagen, Catolog number 19086). This solution was placed within the Covaris microTUBE AFA Fiber Crimp-Cap 6x16mm tube (Catalog number 520052), which was placed within the E220 sonicator. The following settings were used for sonication of whole genomic DNA: Time 200 seconds, Duty 10.0, Peak Incident Power 175, Cycles per Burst 200, Amplitude 0, Velocity 0, and Dwell 0. Random shearing of the genome by sonication minimizes sequence specific fragmentation, as compared to restriction enzyme digestion. Sonicated DNA samples were run on the Bio-Rad Experion™ Automated Electrophoresis System to confirm successful sonication. The distribution of sheared DNA ranged between ~160 base pairs (bp) to ~450bp. Figure A.1 (Pg 109), provides an example electropherogram of a sample that was 68uantifie using the settings stated above. Following sonication, DNA was vacuum evaporated down to 30µL by using a Eppendorf Vacufuge® Plus (Eppendorf, Hauppuage, NY), providing the maximum volume allowed in the downstream size selection step. DNA was size selected using the Pippen Prep (Sage Science, Beverly, MA) instrument according to manufacturer recommendations. A 5µg of DNA was chosen as the initial quantity, as it was the recommended maximum input when utilizing the automated size selection instrument. 2% Ethidium-free Agarose Gel Cassettes (Catalog number CEF2010) were used for size selection of DNA. The size selection range was set from 160bp to 220bp on the instrument. For all samples, the entire sonication and size selection process was repeated with an additional 5 µg. This was due to the low recovery yield observed from the use of the automated size selection

method. For every 5 μg of sample used for sonication, total recovery after size selection ranged from ~80ng to 110 ng. It should be noted that some loss occurs during the vacufuge evaporation step, but majority of the loss occurs during size selection step, as >90% sample loss was observed. After size selection, both batches of size selected DNA were combined for each sample into one batch. For all steps mentioned below DNA Lo-Bind 1.7mL tubes (BiooExpress, catalog number C-3228-1) were used to minimize loss of fragmented DNA.

Within this project, the MethylMiner™ Methylated DNA Enrichment Kit (Life Technologies, Catalog Number ME 10025) was used for enrichment of methylated DNA according to manufacturer's recommendation. It allows for the enrichment of double-stranded DNA based on CpG methylation density, with increased sensitivity over antibody-based methods[100,107,154]. First, dynabeads containing streptavidin are bound to the MBD protein biotin construct. Then fragmented DNA is introduced into the capture reaction and allowed to incubate for one hour. Following this incubation, a salt solution is used to denature the protein and isolate methylated genomic DNA fragments (Figure 7). In this project, sonicated and size selected DNA was first quantified using the Qubit™ dsDNA High Sensitivity kit (ThermoFisher Catalog no. Q32854), and then introduced to the MBD protein for methylated DNA capture. After 60 minutes of incubation, the supernatant, which contains unmethylated DNA, is removed. Then a single, high concentration salt solution is added to release captured methylated DNA fragments from the MBD proteins. The only changes that were made to the manufacturers protocol was the use of GlycoBlue™ Coprecipitant (ThermoFisher, Catalog Number AM 9515) during the ethanol precipitation step and precipitation at -80°C for 24 hours. The incorporation of GlycoBlue™ was done for the identification of the DNA pellet after ethanol precipitation, and the time frame of 24 hours was chosen to maximize the amount of DNA recovery within the

ethanol precipitation step. Following capture, DNA was quantified using the Qubit™ dsDNA High Sensitivity kit.

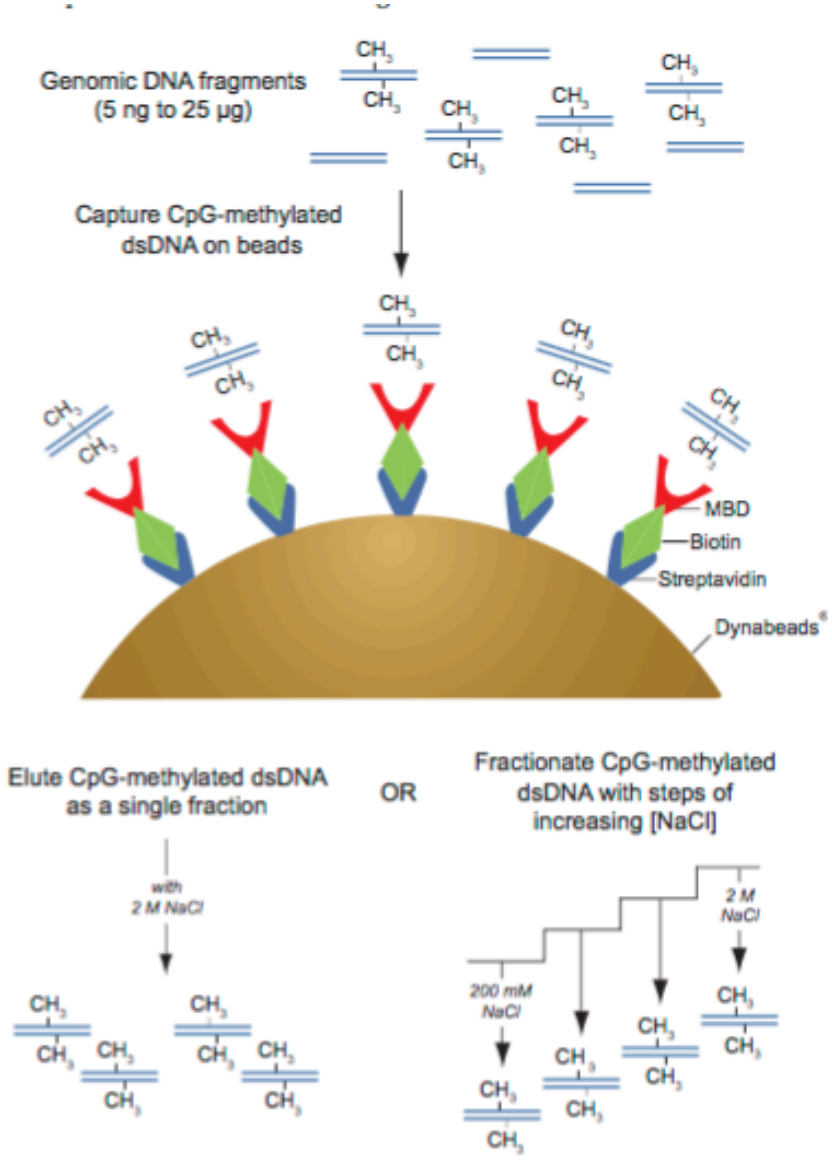


Figure 7: Illustration of the Methylated DNA Capture Process.(Adapted from MethylMiner™ User Guide[155]) Dynabeads containing streptavidin are bound to the MBD protein biotin construct. Fragmented DNA is introduced where methylated fragments are captured and retained by the MBD protein. Following incubation, a salt solution is used to denature the protein and isolate methylated genomic DNA fragments.

The MethylMiner™ kit includes a control to ensure enrichment of methylated DNA is occurring properly. Control capture reactions were prepared and performed following manufacturer's protocol. The control reaction involves the use of a non-methylated DNA control duplex and a methylated DNA control duplex. Each control is spiked into the capture reaction process, and PCR is used to verify the enrichment of methylated DNA and absence of non-methylated DNA within the captured fraction. Similarly, absence of methylated DNA, and abundance of non-methylated DNA is observed within the supernatant. Figure 8 depicts a gel that demonstrates the proper functionality of the capture process. DNA was amplified using PCR with primers that are specific for methylated and non-methylated controls. The amplified products were electrophoresed on a 4% agarose gel (ThermoFisher Catalog number G501804).

Figure 8, below, depicts the proper functionality of the MethylMiner™ kit. Controls included in the kit were PCR amplified, and gel electrophoresis was used to identify the presence or absence of control methylated and non-methylated DNA. Lane M contains a ladder. Lane 1 contains the primers for the control methylated dsDNA. Presence of a band indicates successful enrichment of the methylated control DNA present within the captured methylated fraction. Lane 2 contains primers for the non-methylated control DNA; presence of a band confirms the presence of non-methylated control DNA within the non-methylated supernatant. Lane 3 contains primers for the non-methylated control DNA, within the methylated capture fraction; absence of a band indicates the absence of non-methylated control DNA within the methylated captured fraction. Lane 4 contains primers for control methylated dsDNA. The presence of a barely visible band indicates that there is very little product; indicating that there were minuscule amounts of methylated control dsDNA product present within the non-methylated supernatant. In comparison to lane 1, lane 4 indicates that a miniscule amount of methylated DNA escapes the

capture process. Table 4 (Appendix, Pg 111) lists the amount of DNA that was input into the MBD capture reaction for each sample, the amount of DNA captured within the methylated fraction for each sample, and the amount of DNA present post MBD-capture. The Qubit™ failed to quantify the amount of DNA that was present, possibly due to the amount present being below the range of quantification of the Qubit™ instrument. A decision was made to proceed with the library preparation process.

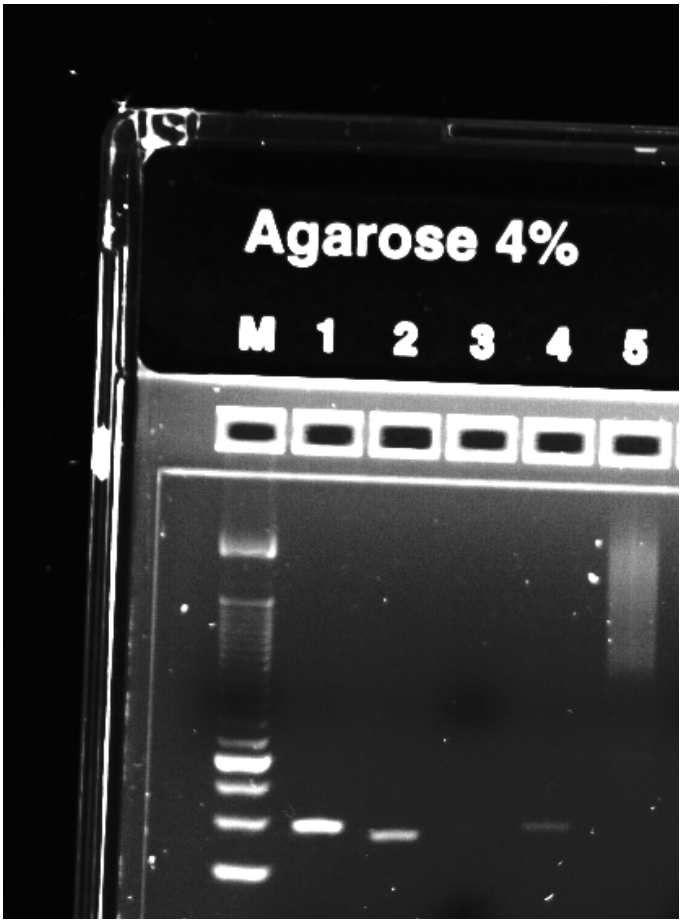


Figure 8: PCR products generated from control DNAs electrophoresed to verify kit integrity. Lane M contains a ladder, where the last three markers are 75bp, 50bp, and 25bp. Lane 1 contains the primers for the control methylated dsDNA. Presence of a band indicates successful enrichment of the methylated control DNA present within the captured methylated fraction. Lane 2 contains primers for the non-methylated control DNA, present within the supernatant ; presence of a band confirms the presence of non-methylated control DNA within the non-methylated supernatant. Lane 3 contains primers for the non-methylated control dsDNA, within the methylated capture solution; absence of a band indicates the absence of non-methylated control DNA within the methylated captured fraction. Lane 4 contains primers for control methylated dsDNA. The presence of a barely visible band indicates that there is very little product; this indicates that there were minuscule amounts of methylated control dsDNA product present within the unmethylated supernatant. In comparison Lane 1, it demonstrates negligible amounts.

DNA library preparation, pooling and sequencing

Post methylated DNA enrichment, library preparation was performed using the KAPA Hyper Prep Kit from KAPABiosystems (Catalog Number: KK8502). The library preparation process consists of End Repair & dA-Tailing, Adapter Ligation, Post-Ligation Cleanup, Library Amplification, and Post-Amplification Cleanup. The protocol was followed according to manufacturer recommendations with one exception. The total amount of time for the ligation of adapters was extended from 15 minutes to 2 hours. After post-amplification cleanup, samples were quantified using the Qubit™ dsDNA High Sensitivity kit (Table 4, Pg 111), and samples were pooled together based on stoichiometric calculations. A calculation was made in order to ensure equimolar concentrations are pooled together of each sample, The calculation involved using the fragment size present within each sample, and its quantity in ng/µL. Each pool was calculated to have a final concentration of about 10 picoMolar. Samples were then sequenced in a paired-end manner, 2 x 100bp on the Illumina® HiSeq 2500 instrument, over 3 lanes.

Methylated DNA Capture Data Analysis

Data Analysis of methylated DNA Capture was performed on a computing cluster located at Renaissance Computing Institute, Chapel Hill, NC.

MBD capture data was analyzed using MEDIPS[156,157], a R package available within Bioconductor. MEDIPS provides functions for quality control and analysis of data derived from MBD-seq experiments. Bowtie2[158] was first used to align all paired-end reads to the HG19_UCSC genome annotation. The following parameters were used: Paired-end alignment, minimum insert size for valid paired-end alignment: 0, maximum insert size for valid paired-end alignment: 250, HG19_UCSC indexed reference genome, end to end alignment, preset: Very Sensitive. The alignment produced BAM files that are used by MEDIPS for further analysis.

Alignment files were imported within the MEDIPS working environment and differential coverage analysis was performed to detect regions with differential coverage between Control and AD individuals. Within MEDIPS, multiple settings were adjusted before performing the analysis. The settings include window size, correction for PCR duplicates, and extension of reads. Window size is an adjustment that breaks the entire hg19 into segments of the specified window size (in bp), and maps reads to windows based on that size. The window size was adjusted from 100bp to 14000bp. Increments of 25bp were used when the window size ranged from 100bp – 1000bp, and increments of 250bp from 1000bp to 14000bp. The extend factor, a factor that extends each read within the dataset, was also adjusted at 0bp, 50bp, 100bp, and 300bp. A minimum coverage depth of 10X and a p- value (FDR adjusted) of <0.05 was used for all analysis. In order to adjust for variability that arises from technical and biological variation the edgeR[159] method was used. edgeR is an R package for examining differential expression of replicated count data. It uses an over dispersed Poisson model is used to account for variability.

An Empirical Bayes method is used to moderate the degree of over dispersion across transcripts, improving the reliability of inference. A sample MEDIPS R script is given within the Appendix (Pg 118).

Table 5: MBD-seq read depth. Read Counts for each sample, and their sequence quality scores. The average number of reads across samples was above 22 million, with an average quality score above 34.

	Number of Reads	%GC	Quality at 70bp	Quality at 100bp
D1	8627699	44	38	34
D2	8238267	44	38	34
D3	59873273	50	36	34
D4	15205638	55	38	36
D5	21376259	49	38	34
D6	8663381	48	38	34
D7	14630252	47	38	36
D8	9348623	48	36	34
D9	7366002	50	38	36
D10	90066332	50	38	36
D11	13712897	52	36	34
D12	64798263	49	38	34
D13	12634176	52	36	34
D14	8249772	49	38	34
D15	15481084	50	38	36
D16	17593376	56	36	34
D17	8243723	51	38	36
D18	7672207	50	38	36
D19	8195794	48	38	36
D20	17087516	47	38	36
D21	7784921	48	36	34
D22	8724928	46	38	34
D23	75851113	49	36	34
Average	22148934.61	49.2173913	37.39130435	34.7826087

Use of BEDOPS for DNA 450k methlation array data

BEDOPS[160] was utilized to retrieve the closest gene transcriptional unit to the differentially methylated CpG sites found within the 450k DNA methylation dataset. BEDOPS is an open-source command-line toolkit that performs multiple operations. Specifically, the 'closest-features' program was utilized that associates nearest features between two sorted inputs, based upon genomic distance measures. Differentially methylated CpG sites to sorted and converted to .bed files as required for BEDOPS use. A reference genome was created, as a reference genome file is required to align the differentially methylated CpG sites. The reference dataset was created by downloading a file from the UCSC table browser webpage <https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=487608909_PM0A6FrnmTWkotyIau5vM6biGX9x>.

The settings used to download the reference file were: Clade: Human, Genome: Human, Assembly: feb2009grch37/hg19, Group: Genes and Gene Predictions, Track: RefSeq Genes, Table: refGene, Region: genome, Output format: all fields from selected table. BEDOPS was utilized to retrieve gene names for all differentially methylated CpG sites. Outputs from BEDOPS tools were compared to the DEGs found within the RNA-seq Dataset.

Chapter III

RESULTS

RNA

Alignment with TopHat[144] demonstrated that over 98% of all reads within each sample aligned to the reference sequence, which was the refseq UCSC hg19 annotation. Table 6 lists the number of reads each sample, the percentage of all reads aligned to the refseq annotation, and the percent of unaligned reads.

Table 6: TopHat Alignment statistics for each sample. Demonstrates that over 98% of sequence generated aligned to the known refseq annotation.

Sample	Number of Reads	% Total Aligned	% Unaligned
24381	19,650,528	98.37%	1.63%
24385	6,676,658	98.26%	1.74%
24430	19,137,674	98.48%	1.52%
24509	14,946,369	98.32%	1.68%
25666	16,804,810	98.48%	1.52%
25667	21,154,738	98.54%	1.46%
35289	14,127,867	98.31%	1.69%
36082	16,906,897	98.61%	1.39%
36359	13,604,921	98.32%	1.68%
40482	20,184,336	98.42%	1.58%
41176	13,769,215	98.23%	1.77%
41639	7,885,964	98.69%	1.31%
41969	20,963,791	98.50%	1.50%
42133	17,875,940	98.47%	1.53%
42990	16,206,976	98.41%	1.59%
43045	12,747,830	98.52%	1.48%
43172	12,196,791	98.44%	1.56%
43192	14,556,330	98.45%	1.55%
43509	13,830,603	98.52%	1.48%
44624	14,599,133	98.61%	1.39%
45116	18,400,851	98.51%	1.49%
45392	22,096,478	98.55%	1.45%
46202	10,581,086	98.41%	1.59%

A total of 23,710 genes were identified within our dataset. Completion of the DESeq2 pipeline resulted in an Excel output result listing all genes identified within the dataset, and the following measures for each gene (Table 7). The first column, baseMean, is a just the average of the normalized count values, dividing by size factors, taken over all samples. The column log2FoldChange is the effect size estimate. It tells one how much the gene's expression changed within treatment in comparison to control. This value is reported on a logarithmic scale to base 2: Ex: a log2 fold change of 1.5 means that the gene's expression is increased by a multiplicative factor of $2^{1.5} = 2.82$. Of course, this estimate has an uncertainty associated with it, which is available in the column lfcSE, the standard error estimate for the log2 fold change estimate. It also expresses the uncertainty of a particular effect size estimate as the result of a statistical test. The purpose of a test for differential expression is to test whether the data provides sufficient evidence to conclude that this value is really different from zero. DSEQ2 performs, for each gene, a hypothesis test to see whether evidence is sufficient to decide against the null hypothesis that there is no effect of the treatment on the gene and that the observed difference between treatment and control was merely caused by experimental variability (i. e., the type of variability that you can just as well expect between different samples in the same treatment group). The result of this test is reported as a p-value, and it is found in the column p-value. The p-value indicates the probability that a fold change as strong as the observed one, or even stronger, would be seen under the situation described by the null hypothesis.

Table 7: Example of a result table output from DESeq2. The result table includes a list of genes, and measures as defined above. This is an excerpt from our RNA-seq dataset.

Gene ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
KRT5	24.6058865	-1.969163299	0.31006352	-6.3508383	2.1414E-10	3.9905E-07
NPY2R	24.3731004	-1.605254856	0.33570623	-4.781725	1.738E-06	0.00011849
SST	62.4830052	-1.52961745	0.25793948	-5.9301408	3.0268E-09	1.8604E-06

Genes were identified as differentially expressed if they were observed at ± 0.6 logarithmic fold change (lfc) and significance expressed as q-value (FDR-adjusted p value < 0.001) as determined by DeSEQ2. 386 total genes were found to be differentially regulated (151 up-regulated and 235 down-regulated). To ensure that the changes observed weren't due to large variation in cell populations among our case and control group, changes in genes that serve as markers, as done by Magistri *et al.* [161] within their RNA-seq dataset were investigated. Since cases of AD could be characterized by extensive neuronal cell death and presence of gliosis in the brain; where this could affect DEGs obtained based solely on changes in of cell-type proportions. The expression of known neuronal, astroglial and microglial markers in Control(CTL) and AD from our RNA-seq data. No changes in the expression of neuronal, astroglial, and microglial markers: *DCX, MAP2, NFH, NEFM, RBFOX3, APQ4, ALDH1L1, SLC1A3, PTPRC,* and *AIF1* were observed. The results suggests that, although AD samples analyzed have high Braak score, they were not depleted of neurons or enriched of astroglia and microglia cells and that the differentially expressed genes from our analysis are not the result of an imbalance between the cellular populations between AD and CTL.

Principal component analysis (PCA), performed using DeSEQ2, identifies that the main source of variation, across our whole dataset, is due to mainly differences between AD and CTL samples (Figure 9).

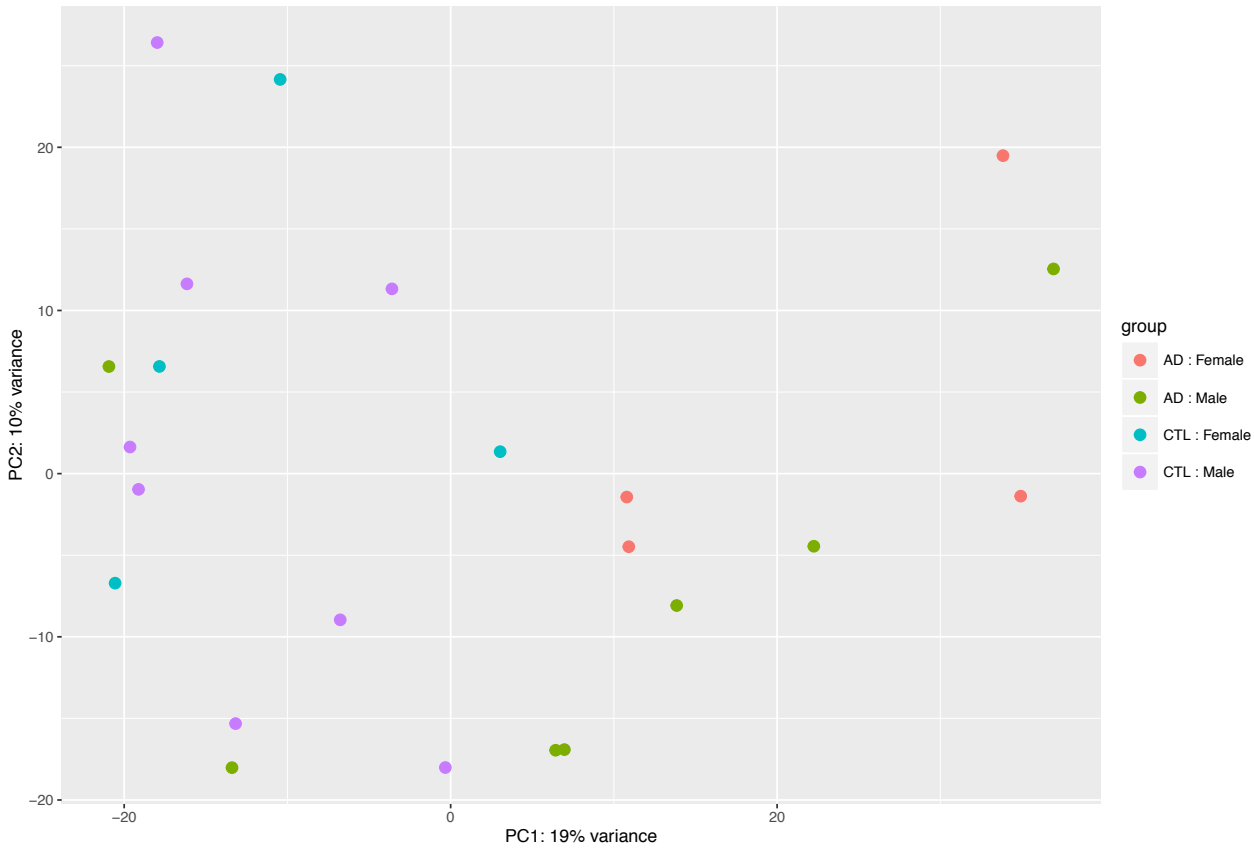


Figure 9: Principal Component Analysis of RNA-seq Dataset. Principal component analysis identifies that the main source of variation, across our whole dataset, is due to differences between AD and CTL samples.

A Mean expression plot was produced using DeSEQ2 (Figure 10). The expression plot displays the log2 fold changes attributable to a given variable over the mean of normalized counts. Points are by default colored red if the adjusted p_{adj} value < 0.001. The blue lines at 0.6 and -0.6 indicate the fold change threshold used to identify DEGs. Therefore, genes above and below the blue line and in red indicate genes identified as differentially expressed within our RNA-seq dataset.

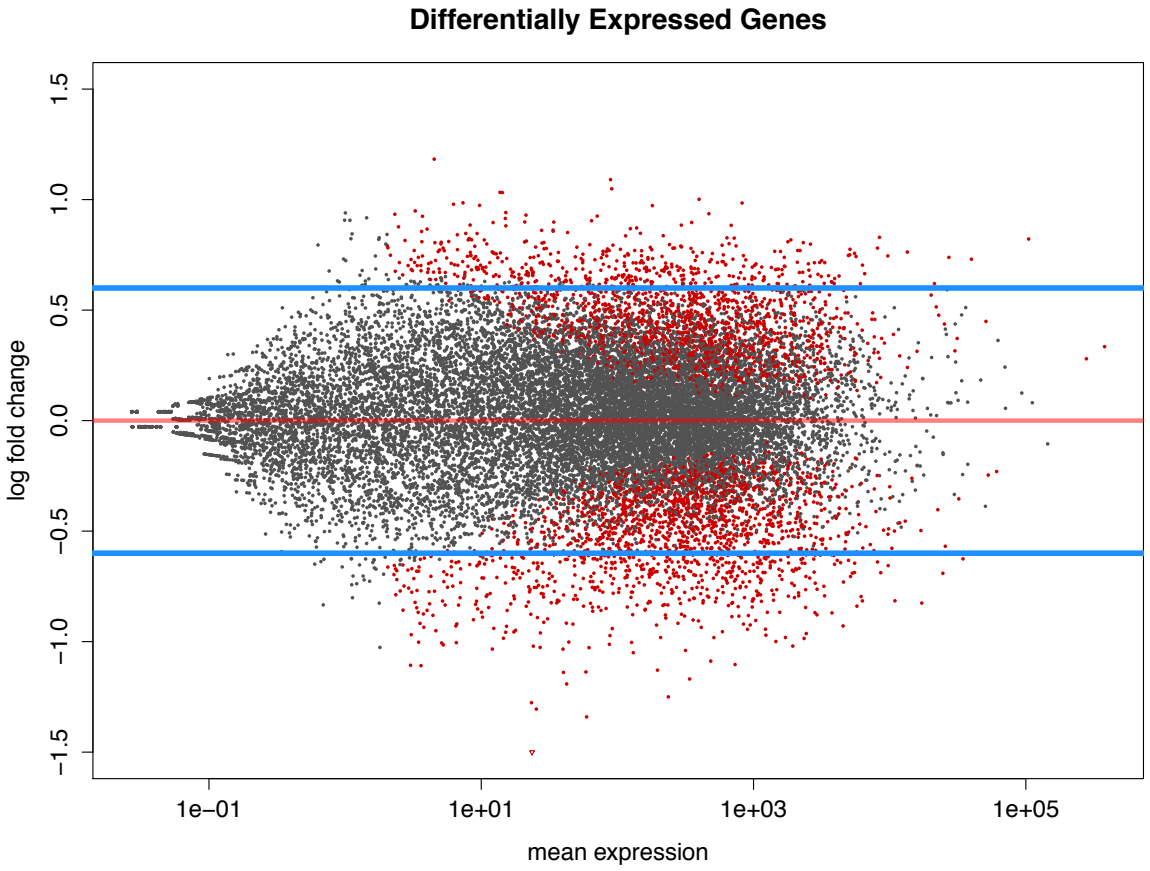


Figure 10: Mean expression plot. Points above and below the blue lines and in red indicate genes identified as differentially expressed within the RNA-seq dataset.

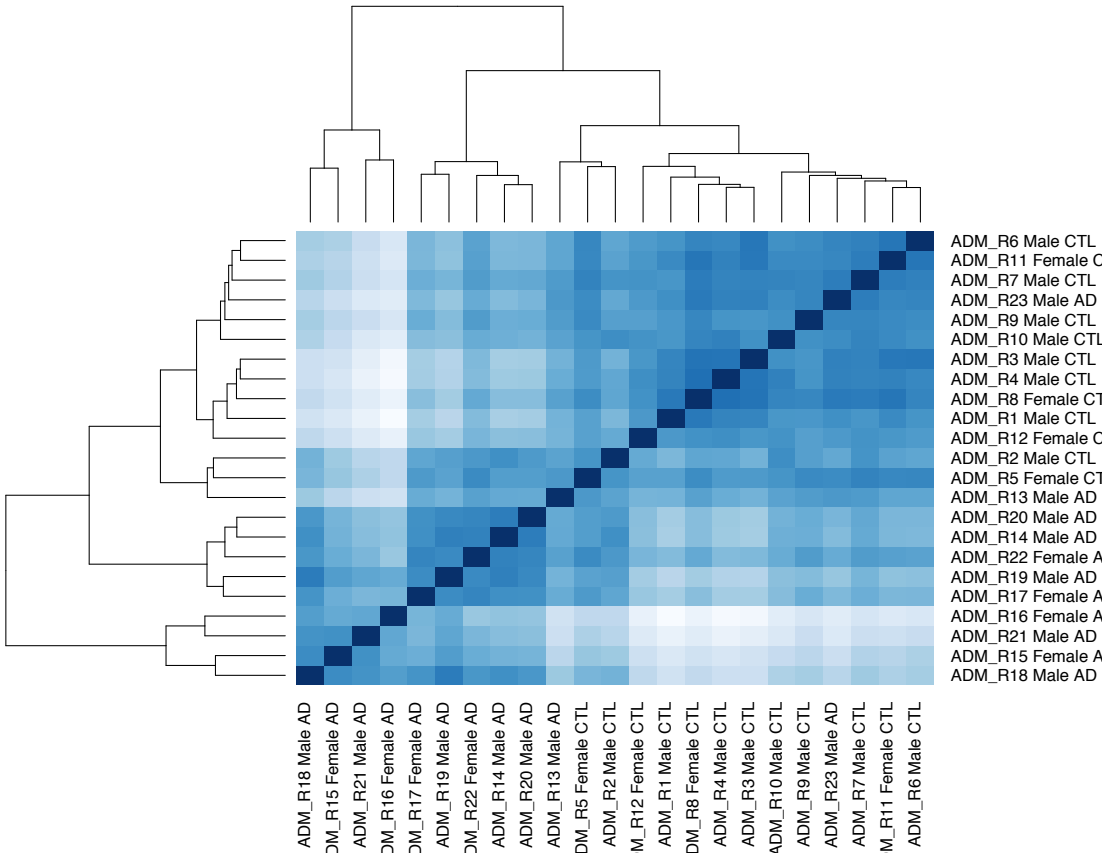


Figure 11: Symmetrical Clustering of Samples. A heatmap displays the Euclidean distances between samples as it is calculated from the regularized log transformation. This clustering of samples gives an overview over similarities and dissimilarities between samples. It also identifies that our samples clustered together closely based on treatment.

The Gene Ontology enRIchment anaLysis and visualization tool (GORILLA)[162], was used to identify common functions associated to the deregulated genes and provide an insight to better help understand the etiology of AD. This method identifies, independently for each GO term, the threshold at which the most significant enrichment is obtained. The significance score

is corrected for multiple testing. Consequentially, GORILLA performs the enrichment analysis on thousands of genes and thousands of GO terms. GORILLA was used to generate the best hits (datasets with largest number of matched genes) to get an overview of biological processes, molecular functions, and cellular components that were down-regulated (Figures 12, 13, and 14) within the RNA-seq dataset. These figures display GO terms arranged in a direct acyclic graph (DAG) with a highly intertwined structure, where the color of entities corresponds to a P-value color scale: white = > 10^{-3}, light yellow: 10^{-3} to 10^{-5}, light orange: 10^{-5} to 10^{-7}, orange: 10^{-7} to 10^{-9}, Red: < 10^{-9}. A table with the top ten Gene Ontology (GO) terms, description, P-value, FDR q-value, and an enrichment score is also included (Tables 8, 9, and 10, Pgs 112-113) that represent the data from figures accordingly. Enrichment score (N, B, n, b) is defined as follows: N is the total number of genes, B is the total number of genes associated with a specific GO term, n is the number of genes in the top of the user's input list or in the target set when appropriate, b is the number of genes in the intersection; where the calculation is Enrichment = (b/n) / (B/N).

The most statistically significant ($p<10^{-13}$) entities to be found within Gene Ontology analysis, in terms of effected biological processes altered within our data set, cluster around regulation of synaptic transmission, synaptic transmission, cell – cell signaling, cell communication, neurotransmitter transport, and regulation of neurotransmitter transport & secretion. It's important to note that these analysis were derived from the down-regulated DEGs within the dataset. Submission of up-regulated and down-regulated DEGs to GORILLA had to pass the p-value $<10^{-6}$ threshold. Submission of up-regulated DEGs failed to report any findings.

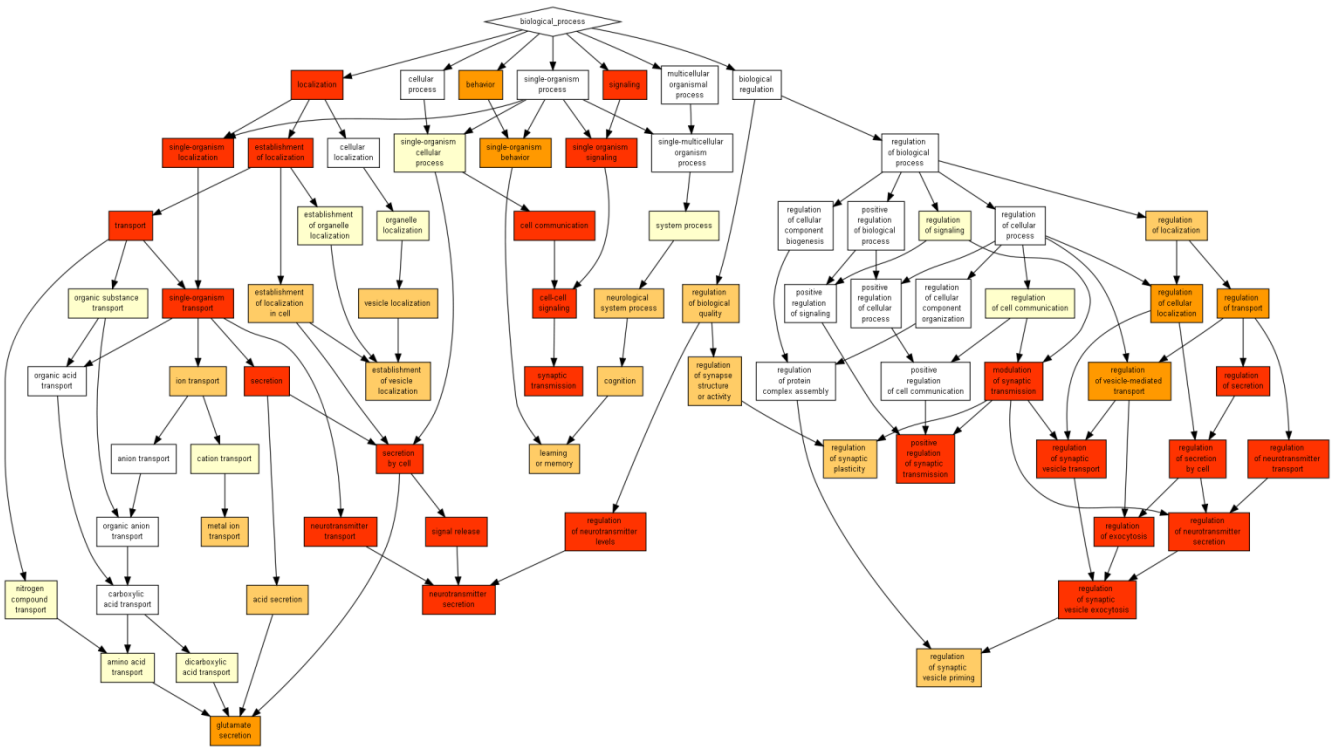


Figure 12: Overview of biological processes to be impacted by down-regulated genes as determined by GO analysis. The most statistically significant ($P < 10^{-13}$) entities cluster around synaptic transmission, cell – cell signaling, cell communication, neurotransmitter transport, and regulation of neurotransmitter transport & secretion.

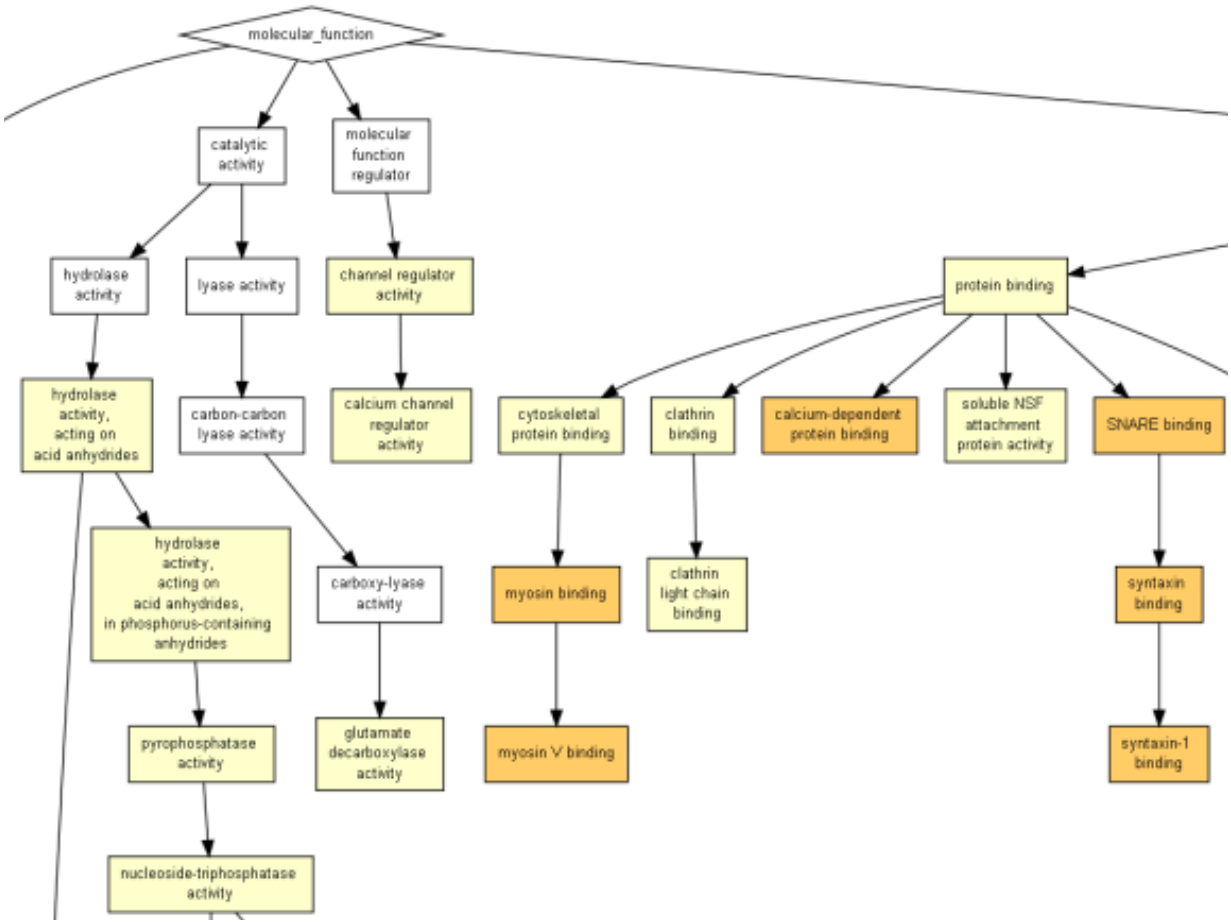


Figure 13: Overview of molecular functions to be impacted by down-regulated genes as determined by GO analysis. The most statistically significant ($P < 10^{-6}$) entities cluster around syntaxin-1 binding, myosin V binding, SNARE binding, myosin binding, and transmembrane transporter activity.

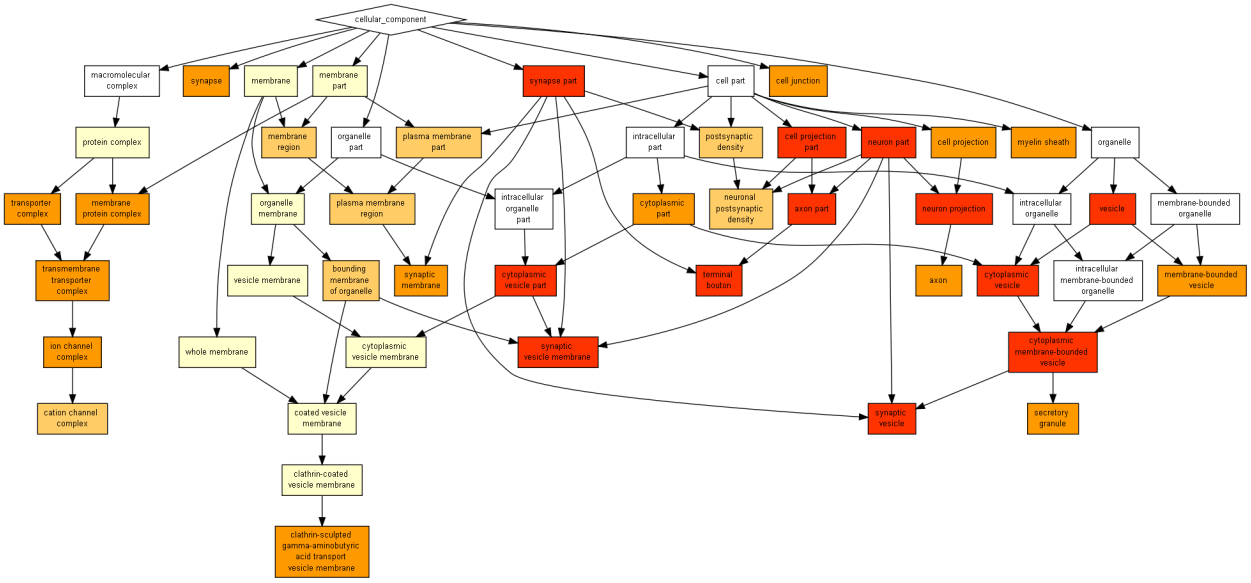


Figure 14: Overview of cellular components to be impacted by down-regulated genes as determined by GO analysis. The most statistically significant ($P < 10^{-13}$) entities cluster around neuron part, synapse part, synaptic vesicles, cell projection part, axon part, and synaptic vesicle membrane.

Core Analysis was performed with Ingenuity Pathway Analysis (IPA®, Qiagen), using the following settings: 35 molecules per network, 25 networks per analysis, casual networks were not included, only direct interaction networks, genes that were experimentally observed only within the human species only, and all tissues and cell lines was not selected. The same parameters were used to define DEGs in IPA®, as was used to define DEGs within the RNA-seq dataset. The analysis showed that the number one disease category that incorporated our DEGs was Neurological Disease, where 176 molecules of the 384 tested fell within the Neurological Disease category. The top four molecular and cellular functions that are affected are Cellular Function and Maintenance, Molecular Transport, Cellular Movement, and Cell-to-Cell Signaling and interaction. 28 specific DEGs within our dataset were specifically linked to AD as determined by IPA®. Table 11 lists these 28 genes as provided by IPA®. Figure 15 displays the DEGs within our dataset that map to a network as determined by IPA®. Down-regulated genes are displayed in green and upregulated genes are displayed in red. This network had the highest score, with the inclusion of 27 focus molecules. Table 12.A (Pg 114) lists genes shown within Figure 15, with a description of each gene.

Table 11: List of specific DEGs linked to AD. DEGs within the RNA-seq dataset that linked to AD as determined by IPA®.

Gene ID	Summary
GFAP	Upregulation of human GFAP protein in temporal lobe is associated with Alzheimer's disease in human
PHYHD1	Upregulation of human PHYHD1 protein in cortical tissue from brain is associated with Alzheimer's disease in human.
LRP4	Upregulation of human LRP4 mRNA in hippocampus is associated with Alzheimer's disease in human.
TLR4	Human TLR4 protein is involved in Alzheimer's disease in human.
PPARA	Docosahexaenoic acid, an agonist of human PPARA protein, in clinical trial for the treatment of Alzheimer's disease in human.
MEGF10	Upregulation of human MEGF10 protein in cerebrospinal fluid is associated with Alzheimer's disease in human.
SOX2	Downregulation of human SOX2 protein in hippocampal dentate gyrus is associated with severe disease stage Alzheimer's disease in human.
ABCA1	Upregulation of human ABCA1 mRNA in hippocampus is associated with Alzheimer's disease in human
PRKCE	Downregulation of human PKC EPSILON [PRKCE] protein in a membrane fraction from brain is associated with Alzheimer's disease in human.
SNCA	Upregulation of soluble human SNCA protein in brain is associated with Alzheimer's disease in human.
SYNJ1	Upregulation of human SYNJ1 protein in cortical tissue from brain is associated with Alzheimer's disease in human.
GABBR2	Upregulation of human GPR51 [GABBR2] mRNA in synaptoneurosomes from prefrontal cortex is associated with incipient Alzheimer's disease in human.
CDK5	Downregulation of human CDK5 mRNA in hippocampal CA1 region is associated with Alzheimer's disease in human.
WASF1	Upregulation of human WAVE [WASF1] protein in neurofibrillary tangles from brain frontal cortex is associated with Alzheimer's disease in human.
UCHL1	Downregulation of human UCHL1 mRNA in hippocampus is associated with Alzheimer's disease in human.
STAR	Upregulation of human STAR protein in cytoplasm from brain hippocampus pyramidal neurons is associated with Alzheimer's disease in human.
ATP6V1G2	Downregulation of human ATP6V1G2 mRNA in hippocampal CA1 region is associated with Alzheimer's disease in human.
GAP43	Downregulation of human GAP43 protein in frontal cortex is associated with Alzheimer's disease in human.
SV2A	Upregulation of human SV2A mRNA in synaptoneurosomes from prefrontal cortex is associated with incipient Alzheimer's disease in human.

SYP	Downregulation of human SYNAPTOPHYSIN [SYP] protein in frontal cortex is associated with Alzheimer's disease in human.
GABRA1	Olanzapine, an antagonist of human GABRA6 protein, is in Phase IV clinical trial for the treatment of Alzheimer's disease in human.
SLC30A3	Downregulation of human ZNT3 [SLC30A3] protein in Brodmann's area 8/9 is associated with Alzheimer's disease in human.
PAK1	Downregulation of human PAK1 protein in cytosolic fraction from human temporal cortex is associated with Alzheimer's disease in human.
GAD2	Valproic acid, an inhibitor of human GAD2 protein, is in Phase III clinical trial for the treatment of Alzheimer disease in human.
GABRA6	olanzapine, an antagonist of human GABRA6 protein, is in Phase IV clinical trial for the treatment of Alzheimer's disease in human.
IL1B	Upregulation of human IL1B protein in brain is associated with Alzheimer's disease in human.

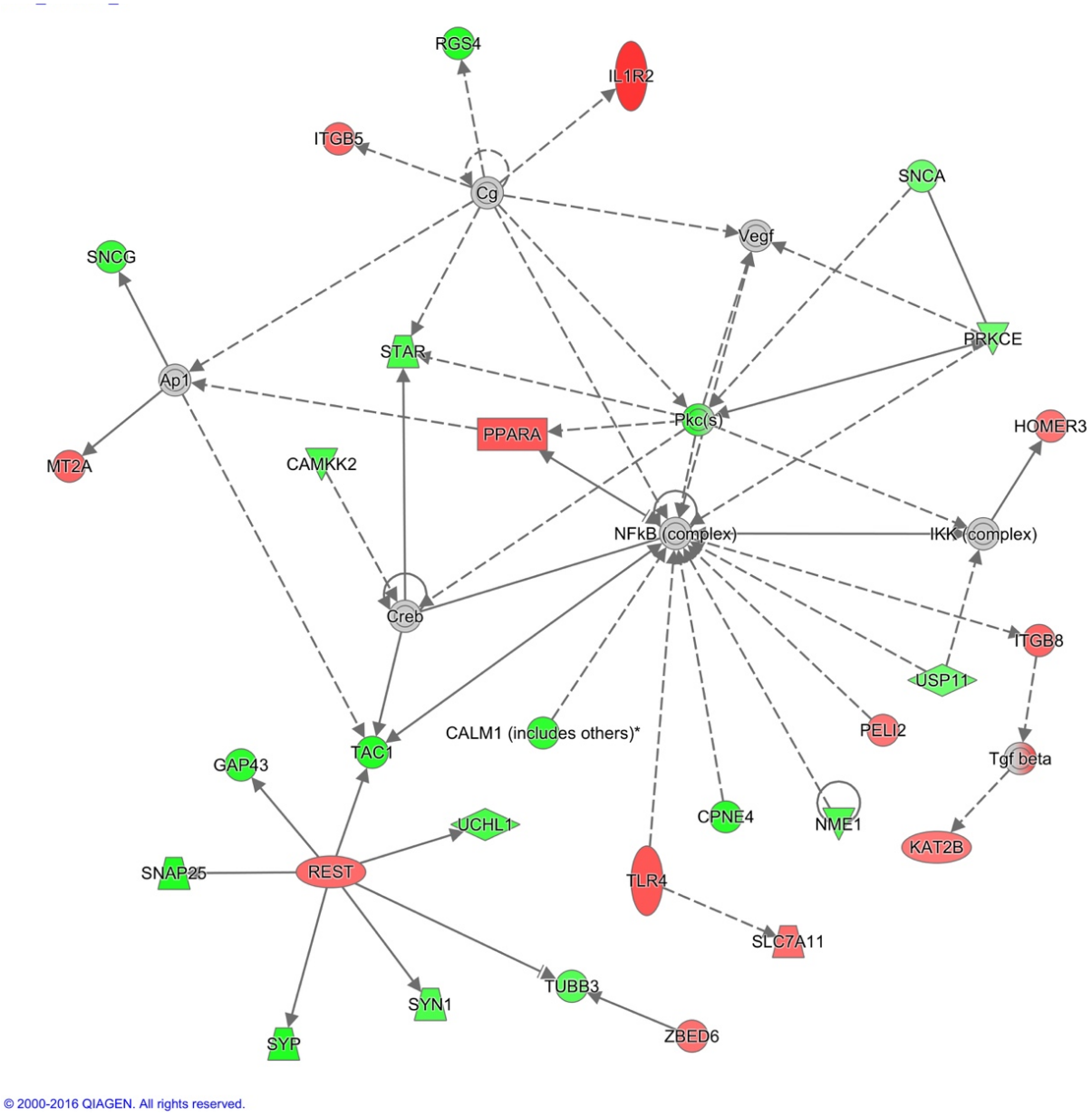


Figure 15: IPA® Network Analysis. DEGs within our dataset that map to a network as determined by IPA®. Down-regulated genes are displayed in green and upregulated genes are displayed in red. This network had the highest score, with the inclusion of 27 focus molecules.

Illumina® 450k Methylation Bead Array Results

Completion of the RnBeads pipeline allowed for the identification of differentially methylated Genes, CpG Islands, Promoters, and individual CpG Sites between the two sample groups, AD and CTL. Differential methylation on the site level was computed based on a variety of metrics. The following quantities were considered for each site: a) the difference in mean methylation levels of the two groups being compared, b) the quotient in mean methylation and c) a statistical test (t-test) assessing whether the methylation values in the two groups originate from distinct distributions. Additionally, each site was assigned a rank based on each of these three criteria. A combined rank is computed as the maximum (i.e. worst) rank among the three ranks. The smaller the combined rank for a site, the more evidence for differential methylation it exhibits.

Control probes for hybridization efficiency, positive & negative controls for bisulfite conversion were analyzed before normalization of data, and SNP & Sex related probes were filtered. In total, 14,376 out of 480,388 probes were removed, and all samples passed quality control standards and therefore included for analysis (Figure 16-A). Background noise & the signal intensity were adjusted for[163], and normalization on the β-values was performed[164]. Figure 16-B demonstrates the distribution of β-values before and after normalization.

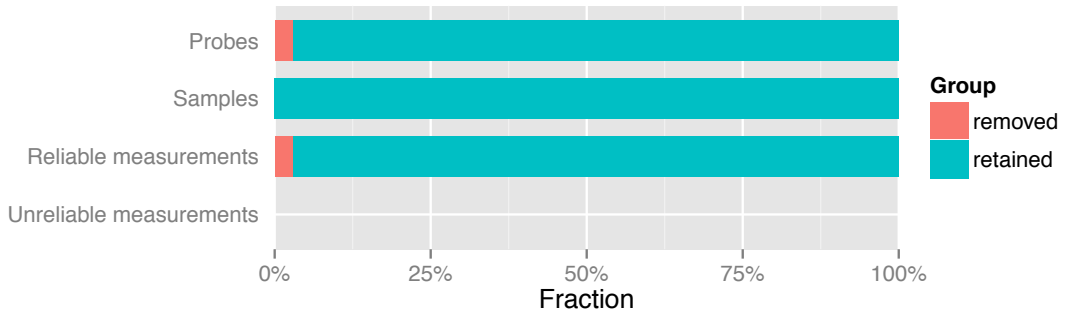


Figure 16-A: Final outcome of filtering procedures. The percentage of probes removed, and the number of samples removed after RnBeads performed quality control of the dataset.

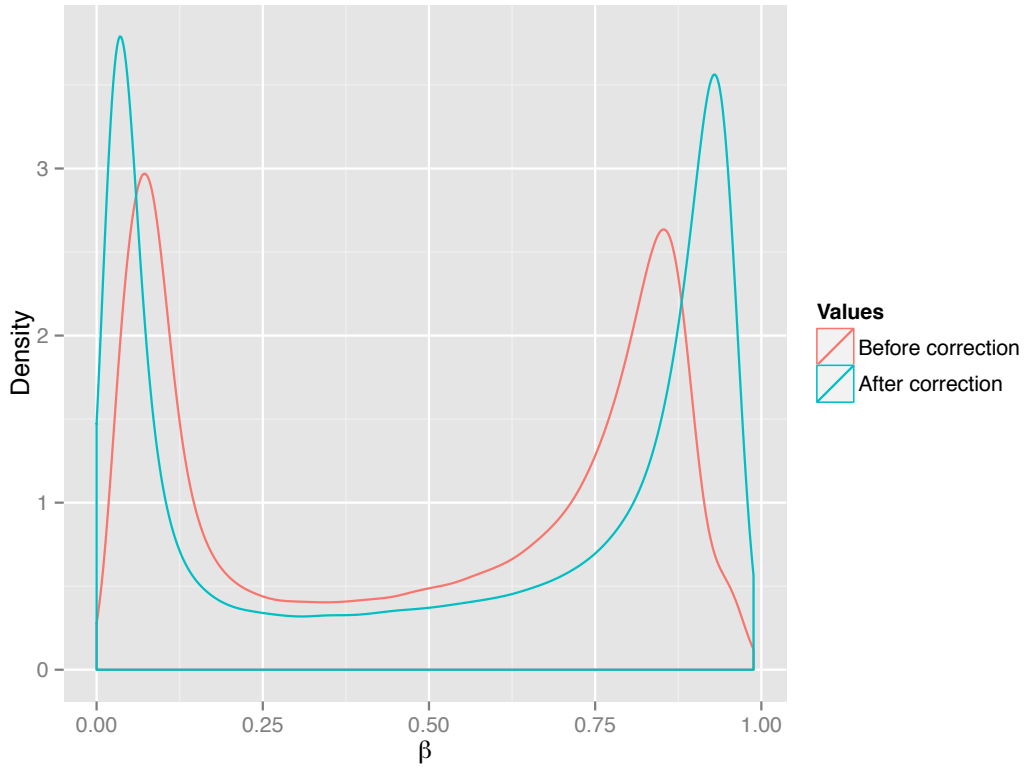


Figure 16-B: Effect of Correction. The influence of the applied normalization procedure on CpG methylation values is displayed. The following figure compares the distributions of the β values before and after performing normalization.

Differential methylation on the region level, for Genes, 5kb Tiles, CpG Islands, and Promoters were also calculated based on a variety of metrics. Using this combined rank as a cutoff within the dataset, 1284 specific CpG Sites, 4 Genes, and 6 Promoters, were considered to be differentially method. The genes identified to be differentially methylated were: *CLUHP4*, *COX6A2*, *TP53TG3D*, and *OSBL10-AS1*. The promoters identified to be differentially methylated corresponded to the *TP53TG3D, HOXA_AS4, GRM2, EEF1A1,* and *LINC00316* genes. Figure 17 displays the Euclidean distance between samples based on methylation values at the 1000 most variable probes. The sample groups, in orange and green, suggests there is some clustering of samples, however it isn't as uniform as observed within the RNA-seq dataset. It should be noted that the Euclidean distance here is mapped based on the variance for 1000 most variable probes across samples. Figures 18 and 20 depict the differentially methylated sites, and genes, respectively, as calculated by RnBeads, in red. The density scatter plots depict differentially methylated probes between Controls and AD in red, as calculated, by RnBeads using the combined rank score.

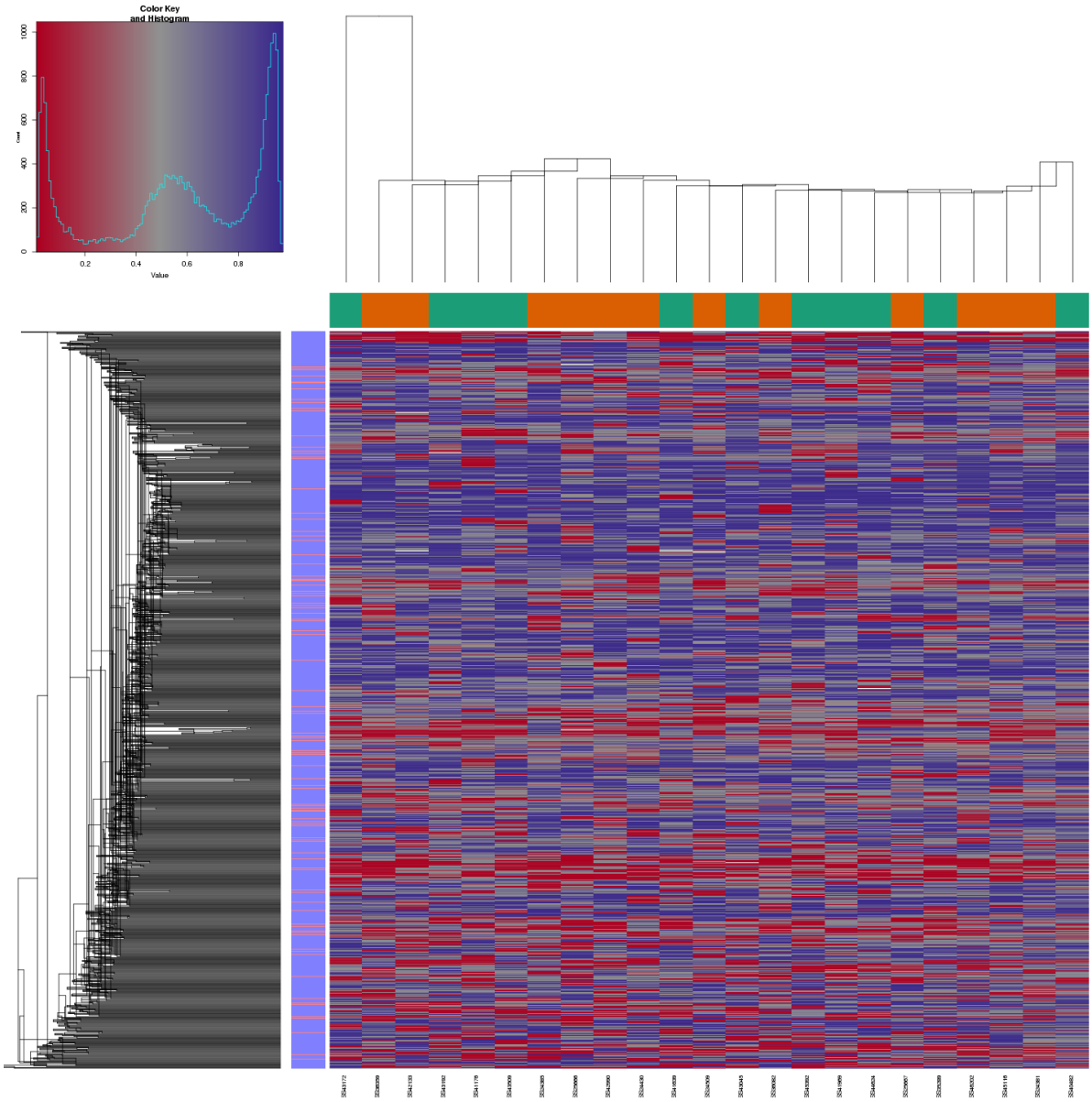


Figure 17: Euclidean distance between samples based on methylation values at the 1000 most variable probes. The sample groups, in orange and green, indicates there is some clustering of samples together when examining the 1000 most variable probes.

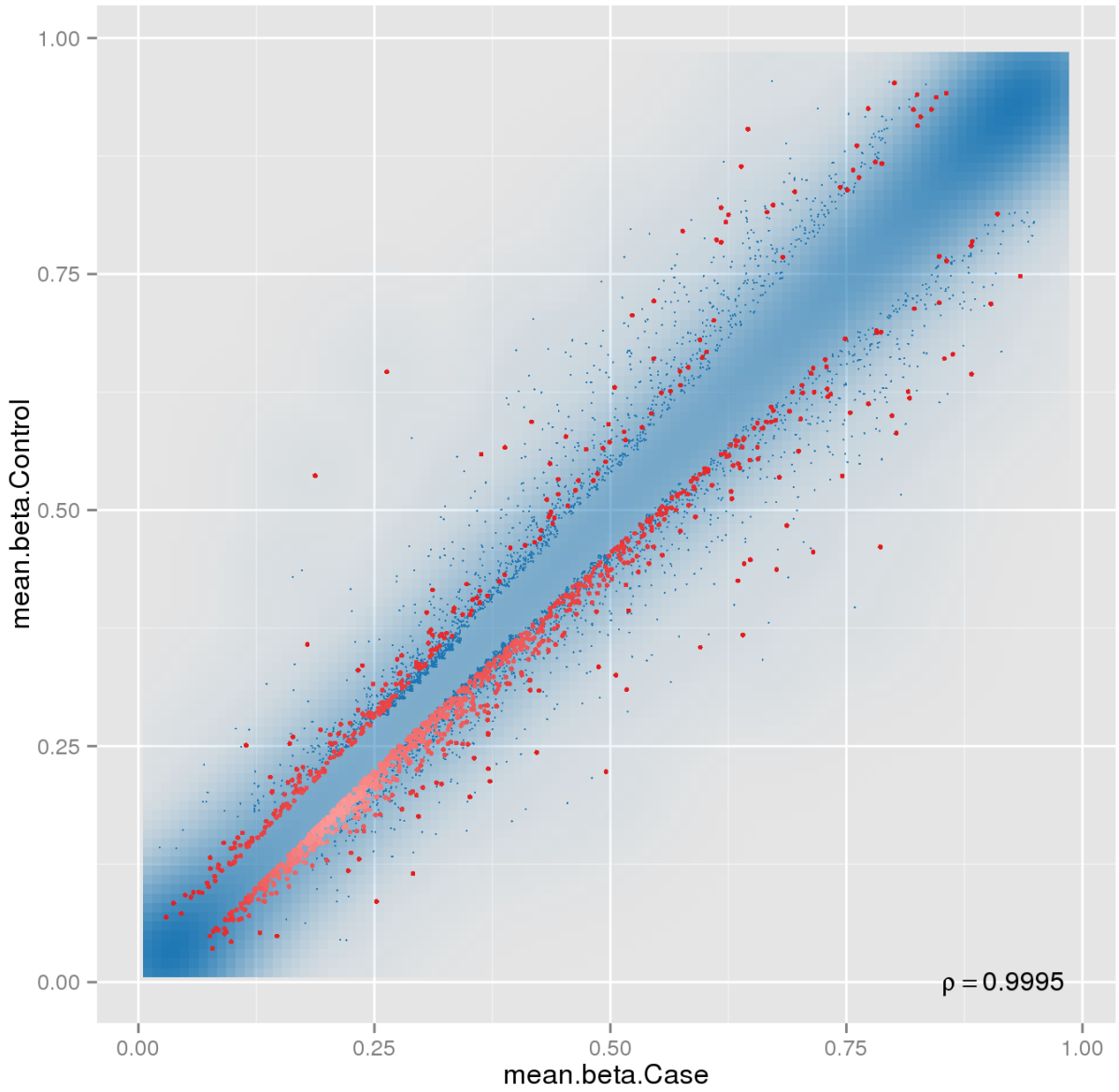


Figure 18: Differentially methylated Probes. Density scatterplot indicating point density and probes differentially methylated between Control and AD are represented in red.

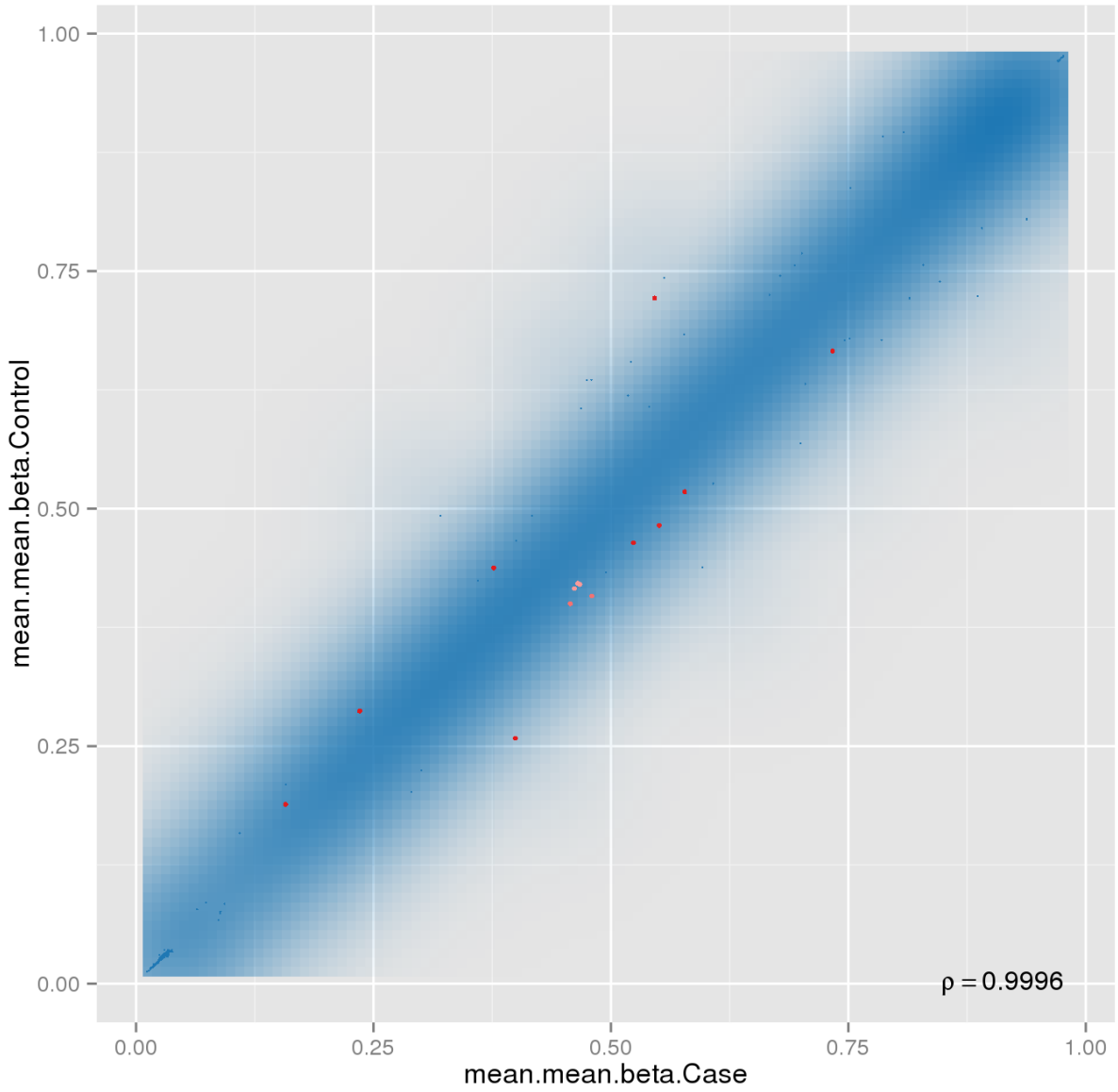


Figure 19: Differentially methylated Genes. Density scatterplot indicating point density and genes determined to be differentially methylated between Control and AD are represented in red.

Methylation levels of differentially expressed genes

Within both our datasets, the only gene identified to be differentially expressed and methylated is *HMP19*. This gene previously has been shown in GWAS analysis to be associated with obsessive-compulsive disorder[165]. Table 12 lists DEGs, their associated differentially methylated CpG Sites, percent change in methylation, and the p-value. Using BEDOPS on genomic Tile regions within 450k methylation array dataset, it was found that the genomic region chr4:139135001-139140000 (Gene *SLC7A11*) was found to be 5.7% hypomethylated within the AD group (pval 0.006).

Table 12: List of CpG sites and the associated DEG. The direction of DEG deregulation within the RNA-seq dataset, percent change in methylation and the p-value is given.

DEG and DMPs				
CpG Site ID	Gene_ID	RNA_seq Expression in AD	Percent Change Methylation in AD	P-value
cg07142667	SALL1	↑	-4%	0.002
cg02455820	SOX13	↑	-8%	0.003
cg14170545	ABCA1	↑	-4%	0.001
cg08598483	HLA-J	↑	-3%	0.003
cg24956452	MYO10	↑	7%	0.004
cg13852822	PHF19	↑	-3%	0.004
cg20225999	TNS1	↑	-5%	0.002
cg09324113	TEX14	↑	-3%	0.003
cg05406923	POU3F2	↑	4%	0.002
cg27152890	PPP1R13L	↑	-5%	5.17E-05
cg03856949	METTL7A	↑	4%	0.003
cg22595230	CENPB	↑	-7%	0.0004
cg16851221	PARD3B	↑	14%	9.80E-05
cg13799302	CYP2J2	↑	-3%	0.0004
cg09668344	PAK1	↓	-4%	0.001
cg10607455	PCSK1	↓	-3%	0.005
cg08072101	RBFOX1	↓	-15%	0.002
cg20414935	ACOT7	↓	-3%	8.70E-05
cg02071463	SNAP25	↓	-3%	0.001
cg17192247	MAPRE3	↓	-3%	0.001
cg13624988	HMP19	↓	-6%	0.001
cg00825252	SMYD2	↓	-5%	0.002
cg01657574	NRN1	↓	5%	0.0006
cg20314918	SCN2A	↓	-5%	0.002
cg13571460	LHX6	↓	-3%	0.003
cg21521518	RASL11B	↓	-5%	0.00008
cg16304656	PRKAR1B	↓	-6%	0.0002
cg00680551	NCALD	↓	-3%	0.003
cg01343363	NDRG4	↓	-8%	0.003
cg18078491	NRXN3	↓	-10%	0.005
cg05841929	FAM81A	↓	-5%	2.25E-05

Methylated DNA capture Results

After analyzing the captured DNA dataset, only one genomic window showed difference in methylation between Controls and AD. The genomic window of 59,5001 to 59,9250 ,was determined to be deferentially methylated (p_{adj} $1.85E^{-02}$), at -7.64 log fold change as determined by edgeR. This genomic window was found to be hypermethylated in controls, and is directly within the region corresponding to the mRNA transcription of the *ERICH1* gene (RefSeq: NM_001303100). This is a glutamate rich protein, and currently, there is no knowledge of pathways or curated interactions for *ERICH1*. GWAS analysis suggested *ERICH1* mutations to be associated with adenoma risk[166] within colorectal cancers.

The dataset generated by using the MBD capture approach was excluded from further analysis. PICARD[167], an open source tool available at GitHub, was utilized to examine coverage depth across hg19. On average, less than 0.3% of the genome was covered at a sequencing depth of 30X or above. This was most likely due to the large amount of sample loss observed when performing the size selection step. The use of 15 cycles of PCR to amplify the adapter ligated product, most likely caused the over representation of a small amount of the genome as demonstrated by PICARD analysis. Future strategy must involve an alternative size selection approach, such as a column based or dual size selection using magnetic beads to avoid the 90% sample loss observed prior to the MBD capture process.

Chapter IV

DISCUSSION

Insights from RNA-seq

Overall, 386 total genes were found to be differentially regulated (151 up-regulated and 235 down-regulated) within the RNA-seq dataset. It's imperative to keep in mind the method of library preparation used prior to RNA-seq. The capture based approach allowed over 98% of the sequences generated to be successfully aligned to the refseq Hg19 annotation, even when the sample RIN was as low as 3.6. In this project, more than 10 biological replicates per group were sequenced, and frozen postmortem brain tissue was used, which is considered the gold standard for analysis. The number of replicates used is important, as within the RNA community it has been suggested that the number of replicates per group improves detection power more than simply increasing sequence depth[136,168]. Although only an average of 15.6 million reads were achieved per sample, using the RNA-Access library preparation approach allowed for over 98% of the reads to be aligned to known RefSeq annotations, where over 75% of the coverage was within coding regions. This is a drastic increase, when compared to traditional RNA-seq approaches where less than 20% of reads typically align to coding regions[169].

Gene Ontology analysis provided an overview of biological processes, molecular functions, and cellular components that were down regulated within the frontal cortex of subjects with AD. The most statistically significant ($p<10^{-9}$) entities under biological processes effected cluster around regulation of synaptic transmission, regulation of synaptic vesicle transport and regulation of neurotransmitter transport & secretion (Table 8, Pgs 112-113). The most statistically significant ($p<10^{-9}$) cellular components effected are synaptic vesicles, neuron

projection, membrane-bound vesicles, synaptic vesicle membrane, axon, and neuron & synapse components (Table 9 Pg 112). The most statistically significant ($p<10^{-6}$) molecular functions impacted by the down-regulated DEGs were SNARE binding, syntaxin binding, myosin biding and transporter activity (Table 10 Pg 113). This again demonstrates that many of the DEGs collectively overlap into a connected biochemical and cellular network.

The genes that were most significant in deregulation of biological processes were clustered around myosin V binding, SNARE (Soluble NSF (N-ethylmaleimide-sensitive fusion protein)-Attachment protein Receptor) binding, transporter activity, and syntaxin-1 binding. The cellular processes impacted by the DEGs also were cell-cell signaling, synaptic dysfunction, and vesicular transport.

Myosins are a family of actin-based cytoskeletal motors that are required for neuronal shape, polarized cargo transport, support synaptic plasticity, and proper synapse function[175]. Nervous system dysfunction is characterized by abnormal myosin V function, and irregular function of myosin motors at neuronal synapses binding has been linked to neurological impairment and severe nervous system dysfunction[175]. Biological functions impacted included syntaxin-1 binding, syntaxin binding, and SNARE binding. This was reported due to the down regulation of the *SXTBP1, SNAP25, SYT1, STX1A, SYN1, SYN2, SYP, STBXP1, STX1B,* and *VAMP2* genes. Syntaxin-1 is a protein that is implicated in docking of synaptic vesicles within the presynaptic plasma membrane. Syntaxin also contains a SNARE domain and binds synaptotagin to interact with voltage gated ion channels[176]. Lastly, SNAREs regulate intracellular trafficking, protein sorting and docking of synaptic vesicles to the plasma membrane during neurotransmitter release[177]. Recently, it was found that in knockout mice of the cysteine-string protein-α (CSPα), a co-chaperone for the SNARE protein SNAP-25, causes

neurodegeneration[178]. Both *CSPa* and *SNAP-25* genes were found to be down-regulated in the AD subjects.

In addition to regulating trafficking, and docking of synaptic vesicles, SNAREs also regulate neuronal Aβ release at pre synaptic terminals[179]. A correctly functioning secretion system for Aβ is critical for maintaining synaptic homeostatic plasticity and any dysregulation of this system could potentially trigger AD pathophysiology. An improperly functioning system could cause an increase in neuronal secretion of Aβ, which itself may have effects on synaptic transmission. Cirrito *et al.*[180] also describe a feedback loop, where an increase in synaptic activity increases Aβ generation, again referring to the point where a proper system is critical and malfunction could be a mechanism that triggers AD pathology. It is estimated that around 70% of extracellular Aβ is derived from the endocytic-exocytic pathway, where APP is endocytosed, processed by **β** and **γ** secretases to form Aβ, and then secreted from the cell. SNAP, and SNARE proteins are thought to be essential for this process, where SNARE proteins allow for the fusion of Aβ -containing exocytosed vesicles, and cause Aβ release. Del Prete D *et al.*[181] have shown APP is cleaved by BACE1 in pre-synaptic vesicles. Additionally, Mukaetova-Ladinska *et al.*[182] showed a decrease in SNAP-25 levels within post mortem brain tissue of individuals that had a lewy body variant of AD. Cell-cell signaling, neurotransmitter transport, modulation of synaptic transmission, and cellular communication were all molecular processes that were shown to be impacted as determined by GO analysis. Terms observed within this category, deregulation of proper synaptic vesicular transport, vesicle docking involved in exocytosis, regulation of cellular localization, synaptic vesicle transport, glutamate secretion, in conjunction with SNARE complex deregulation can all collectively alter secretion of neurotransmitters, permeability of

intercellular junctions, gap junction intercellular communications, formation of focal complexes, and altered morphology of synapses.

Core analysis using IPA® showed the top four molecular and cellular functions that are effected, echoed the same theme as GO analysis, which were Cellular Function and Maintenance, Molecular Transport, Cellular Movement, and Cell-to-Cell Signaling. Figure 15, generated by IPA®, displays correlation between DEGs that had the highest score with the highest number of DEGs mapped within this network. This network correlates our dataset with the inclusion of a network involving NFκB, Creb, Ap1, Cg, and Vegf. NFκB has been known as a ubiquitous transcription factor that controls a wide range of biological functions, including inflammatory and immune functions in both the central and peripheral nervous systems, where inflammation in the central nervous system has been though to play a role in AD[183]. NFκB regulation has also been shown impact long-term changes to adult neuronal function caused by synaptic stimulation[184]. IkB kinase(IKK) is phosphorylated and activates NFkB, where it goes on to translocate into the nucleus and acting as a transcription factor[185]. Although inflammation is a physiological defense mechanism that protects tissues from infection, sustained activation of brain macrophages and glial cells can lead to excess production of various factors that contribute to neuronal injury, including upregulation of proinflammatory chemokines & cytokines and reactive oxidative species; this prolonged inflammatory signaling contributes to cell damage and is observed in progressive degenerative conditions, including AD[186]. Activated microglia accumulate around Aβ plaques in brains of individuals with AD, and have been implicated in neurodegeneration, with GWAS studies identifying risk variants for AD to be implicated in inflammatory response. The inclusion of the transcription factor cAMP-response element binding protein (CREB), hints at synaptic dysregulation as CREB signaling plays a crucial role

for long-lasting changes in synaptic plasticity that mediates the conversion of short-term memory to long-term memory[187]. Also, synaptic efficacy mediating memory storage requires the activation of specific gene expression programs regulated, among others, by the transcription CREB. CREB signaling has been recently involved in several brain pathological conditions including cognitive and neurodegenerative disorders. The Aβ peptide, alters hippocampal-dependent synaptic plasticity and memory and mediates synapse loss through the CREB signaling pathway[187]. AP1 has recently been shown to regulate miRNA-144 that decreases ADAM10, the α - secretase that protects the brain from Aβ build up[188].

IPA® also identified the transcriptional regulator RE1-silencing transcription factor (REST) as predicted to be activated (P 1.67 $^{-7}$, Z-score 2.62). REST was up-regulated in our RNA-seq dataset, and was tied to the down-regulation of 7 genes (*UCHL1, TUBB3, TAC1, SYP, SYN1, SNAP25, GAP43*). Figure 15 displays REST activation in relation with these genes. REST is a gene silencing transcription factor that is widely expressed during embryogenesis, and is a master regulator of neuronal gene expression[189]. REST actively represses a large array of coding and noncoding neuron-specific genes important to synaptic plasticity and structural remodeling, including synaptic vesicle proteins, neuroreceptors and channels, and microRNAs that regulate networks of non-neuronal genes[190]. Proper expression of REST is important as disruption of expression during embryogenesis can result in lethality. REST degradation during terminal neuronal differentiation is essential and is what leads to a neural phenotype of a cell. In mature neurons, REST can be activated in vulnerable hippocampal neurons by ischemic insults.

Expression of REST has been shown to be elevated in aged brains[191]. Lu *et al.*[192], examined REST levels in the pre-frontal cortex of humans and found that REST levels correlate positively with cognitive function. In Alzheimer's disease, REST has been found to be lost from

nuclei of cells and appear in autophagosomes with pathological misfolded proteins. These authors claim that activation state of REST can help distinguish between neuroprotection from neurodegeneration within the brain. Lu *et al.* [192], also found that nuclear REST suppresses several pro-apoptotic genes that encode enzymes involved in the pathology of AD. After treatment of human neurons with hydrogen peroxide, a clear increase in REST levels was observed. They also found that extracts of cortex from AD individuals reduced REST inducing activity compared with age-matched controls. Lu and colleagues went on to show that loss of nuclear REST is associated with a substantial increase in the expression of genes implicated in apoptosis and Alzheimer's pathology in both the least severe and most severe forms of the disease. This somewhat contradicts our findings. Since REST was shown to be up-regulated within our dataset, it could still indicate the loss of nuclear REST, as demonstrated by Lu *et al.*, and the simple explanation is the accumulation of REST in autophagosomes is what is observed within our dataset. It is also important to note that Lu. *et al.*[193] also observed the same trend; they observed that another set of genes targeted by REST — those involved in synaptic transmission and other functions of neural junctions — initially show increased expression in AD, and then again reduced levels in a more severe form of AD. It is plausible that the increased expression of these genes in the beginning may be some form of a compensatory mechanism to maintain neuronal homeostasis[191].

DNA methylation

The differentially methylated CpGs from our study were compared with the top 100DMPs reported by Lunnon *et al.* [194], who examined methylomic profiles using the Illumina® 450k methylation array in Controls and AD patients within the MRC London Brainbank for Neurodegenerative Disease Cohort(n=122), and the Mount Sinai Alzhiemer's Disease Schizophrenia Brain Bank (n=144). When comparing our 450k dataset, only the CpG cg00510320 was found to overlap within our cohort and Lunnon *et al.*[194]. This CpG corresponded with the gene *RLTPR*, as determined using BEDOPS. When comparing to the cohort examined by De Jager *et al.*[195], and their top 72 differentially methylated CpGs, only the CpG cg21806242 overlapped between both datasets. This specific CpG, was associated with gene *ATG16L2*, as determined by BEDOPS. The gene *RLTPR* has a functional domain similar to a protein that binds myosin[196], and *ATG16L2* is an autophagy related gene associated to serve as a potential biomarker in multiple sclerosis[197].

In comparison to the differentially methylated CpGs within the PreFrontal Cortex tissue examined by Lunnon *et al.*[194], the genes *RNF175*(cg07859799, 1.2), *PPARA* (cg06635946, 2.66), *MYO10*(cg07719172, 2.93), and *SOX2*(cg17917241, 2.76) were found to overlap between our RNA-seq dataset and their 450k methylation dataset. All four of these CpGs were more methylated within the AD group, and all three except *RNF175* were up-regulated within our RNA-seq dataset. Next, the DEGs within our dataset were compared to their top 100 cross cortex DMPs, and the genes associated with their DMPs. The down-regulated gene *PCDHAC2*, and up-regulated gene ITGB5 were found to be more methylated within the London cohort and Mount Sinai cohort.

De Jager *et al.*[195] investigated methylation status in prefrontal cortex of 708 subjects and found Differential methylation at 72 CpG sites using the same Illumina® 450k methylation array. Their subjects were derived from the Religious Order Study cohort, and the Memory and Aging Project cohort. Among their 72 CpG sites that were differentially methylated, only one differentially methylated CpG, *MYO10* (cg06742628) overlapped, between our dataset, and *MYO10* was also an DEG within our RNA-seq dataset.

DNA methylation and differentially expressed genes

Comparison of DEGs and DMPs showed there were 32 genes that overlapped between our two datasets (Table 12). To be more stringent, we looked for genes whose percent change in methylation was 5% or greater. Its important to note that SNAP25 was found to have a CpG that was differentially methylated, but the percent change of methylation was only 3%.

MYO10, found to be up-regulated within our RNA-seq dataset, 7% hyper methylated in AD within our 450k methylation dataset, and also overlapped between both the Lunnon *et al.*[194] and De Jager *et al.*[195] dataset. This gene codes for the molecular motor myosin-X, and its expression increases the number of Tunneling nanotubes (TNTs)[198]. TNTs are a type of long-distance intercellular connections that allow for the selective transport of membrane vesicles between cells that contain various signals, and cytosolic materials[198]. Overexpression of MYO10 results in formation of functional TNTs, and increases the number of vesicles transferred between connected cells. Neuronal cells are able to induce TNT formation in response to external signals such as oxidative stress, and MYO10 is up-regulated during nerve regeneration or following peripheral nerve injury[199]. Interestingly it has been hypothesized that the up-regulation of MYO10 can lead to the use of TNTs for intercellular spread of misfolded infectious prion proteins, as shown by Gousset *et al.*[200] and Langevin *et al.*[201]. This allows for the hypothesis that it would be easier for intracellular Aβ -fusion proteins to take advantage of TNTs for intercellular spread. This provides for speculation that up-regulation of MYO10 can contribute towards AD pathogenesis.

The gene NDRG4 was down-regulated in our RNA-seq dataset, and hypomethylated by 8% within the AD group of our methylation dataset. Down-regulation of this gene has been associated in impaired spatial learning and memory function[202]. NDRG4 deficient mice contain

decrease levels of brain-derived neurotrophic factor (BDNF), where BDNF has been shown to play a role in neuroprotection and spatial learning and memory[203].

NRXN3 was down-regulated within our RNA-seq dataset and 10 hypomethylated within AD individuals of our methylation dataset. Neurexins(NRXNS) are a type 1 transmembrane neuronal adhesion receptor that exit and function predominantly at presynaptic terminals. In mammals NRXNs are coded by NRXN1, NRXN2 and NRXN3[204]. The interaction between neurexins and neuroligins promote the formation of functional synaptic structures. GWAS analysis has shown NRXN3 to be associated with AD individuals[205].

Alternative pre-mRNA splicing is an important event that generates several transcripts from one single gene, therefore promoting diversity. RNA-binding protein, fox-1 homolog (RBFOX1), is a neuro-specific splicing factor predicted to regulate neuronal splicing networks clinically implicated in neurodevelopmental disease, and a large network of genes involved in neuronal differentiation and maintenance[206]. RBFOX1 can regulate alternative splicing of APP, and has been shown to exclude exon 7 within APP processing[207]. The APP714 isoform (termed Exon 6-8-9 isoform) was induced in HEK293 and HeLa cells by RBFOX1 overexpression, and this overexpression led to a 10 fold increase in APP714. Alam *et al.*[207] reported that RBFOX proteins promote APP exon 7 skipping via two GCAUG recognition sequences. This suggests RBFOX1 to be involved in exon 7 exclusion. APP isoforms containing exon 7 are increased in brains of AD patients[207,208]. Previous studies suggest that Aβ is preferentially produced from APP770, an isoform containing exon 7, and this isoform was preferentially processed by α and β secretase cleavage in human brain endothelial cells. This leads to the conclusion that deficiency of RBFOX proteins, as observed within our RNA-seq dataset, may contribute to increased

expression of APP770 and therefore an increased production of Aβ. RBFOX1 was found to be down-regulated, and 15% hypomethylated within the AD dataset.

Currently, the LOAD process remains incompletely understood due to the heterogeneity and complexity of the disease. Although there is ample genetic and cellular biology evidence to support the amyloid hypothesis, the recent body of literature points to Aβ not being the sole initiating factor in pathogenesis. It is clear that AD etiology is complex and that Aβ alone is unable to account for all aspects of AD[13,170,171]. The results of this study support the hypothesis that Aβ, alone, does not account for all aspects of AD , and provide novel insights into the pathogenesis of AD. More recent literature expresses an interest in the mechanisms that govern the intracellular traffic in the pathology of AD. There is some experimental data supporting the idea that alterations in vesicle maturation, intracellular trafficking, transport, and secretion are altered in early stages of AD pathology[109,122,123,172–174].

Chapter V

CONCLUSIONS

Conclusions

The DEGs found within our RNA-seq dataset all elucidate the importance of some previously suspected pathways involved in the pathogenesis of AD. Deregulated genes collectively impact myosin binding, intracellular trafficking, proper docking of synaptic vesicles to the plasma membrane, regulation of synaptic transmission, and neurotransmitter transport. Genes deregulated that regulate neuronal Aβ release, along with genes that have been previously implicated with inflammation, and APP processing were differentially expressed. Synapse loss is a pathological correlate of cognitive dysfunction in AD, suggesting that synaptic changes are crucial for AD pathogenesis[35]. Performing GO analysis and submission of differentially methylated genes as a ranked list, based on the combined rank sum score, in GORILLA also indicated deregulation of synaptic transmission (Figure A.2 Pg 110) This suggests a direct link between methylation and differential gene expression. The overlap of multiple DEGs within our RNA-seq dataset and DMPs within the methylation dataset suggest a link between differential methylation and differential gene expression. Genes overlapping both datasets have been shown to be important and play a role in APP expression, formation of tunneling nanotubes, and multiple genes that are important in synaptic transmission & neuroprotection.

Gene Set Enrichment Analysis (GSEA)[209,210]was performed on our DEGs dataset. Overlaps between our up-regulated and down-regulated genes were compared with hallmark, curated and GO gene sets. Out of the 151 up-regulated genes identified within the RNA-seq data set, 52 genes were identified within a gene set that is shown to be up-regulated in brain of patients with AD(FDR q-value $3.59e^{-33}$). Eighty-nine genes that were down-regulated within the

RNA-seq dataset also overlapped with down-regulated genes within the Blalock *et al.*[128] dataset (FDR q-value 3.57e $^{-74}$) that was generated by microarray analysis of hippocampal tissue from 9 controls and 22 AD patients. Given the overlap observed within a curated gene set, we demonstrate a cost effective approach to be able to confidently detect transcriptional changes at a much lower level of sequencing depth than current ENCODE recommendations.

Of the numerous genes examined with a GWAS approach and associated with AD[57–59,211], only the gene *MEF2C* was listed to be down-regulated within the RNA-seq dataset. The MEF2C protein limits excessive synapse formation during activity-dependent refinement of synaptic connectivity and therefore may facilitate hippocampal dependent learning and memory; mutations in this gene has also been linked to synaptic plasticity[59,212,213].

In conclusion, we show overlap between multiple DEGs and their respective differentially methylated CpG sites, suggesting a link between differential methylation and differential gene expression, since these genes impact processes shown to contribute to AD pathogenesis. Our experiment is a final point experiment, and cannot implicitly state whether transcriptional changes are a direct result of differential methylation, and if these changes occur within an early stage of Alzheimer's disease. Understanding the mechanisms of DEGs bring new insight in the development of new therapeutic strategies for preventing the onset of most deleterious symptoms in AD.

Addressing Limitations:

The first RNA-seq analysis of postmortem human brain tissue of age matched controls and AD patients with over 10 biological replicates per group is reported. To combat degraded RNA, a novel probe capture approach to target exonic regions of RNA was used. It is important to note that acidosis in human postmortem brain tissue can be caused by agonal factors such as coma, hypoxia, pyrexia, seizures, dehydration, hypoglycemia, multiple organ failure, head injury, and ingestion of neurotoxic substances, which can affect RNA integrity[214]. Different parameters have been used to assess tissue quality, notably brain pH, gross neuropathological examination, postmortem interval (PMI), and freezer time. Brain pH, has been shown to be related to agonal state and RNA integrity. The acceptable maximum PMI for human studies was reported as 36–48 hours, and it should be noted that our samples are well below this time frame(average PMI = 12.33 hours) [214]. RNA-Seq sample preparation includes multiple procedures (RNA extraction, fragmentation, reverse transcription and amplification), that may be susceptible to experimental bias[121]. We believe the amount of sequence coverage obtained, along with the number of replicates included for each group overall minimize such batch effects that could be observed when only using a low number of replicates.

Although an overlap between the cohort used within our dataset and previously reported studies was observed, its important to note that the method of data analysis varies from one cohort to the other. A relatively new pipeline (RnBeads) was used, and also adjusted for sex, age, and cell type proportions within our data set. This is important to adjust for, as age itself has been shown to impact DNA methylation[82,215,216]. One would expect the amount of differentially methylated CpGs that overlap between multiple cohorts to increase when the same pipeline, and method of analysis is applied uniformly across all datasets. Deviations and non-concordance

within multiple cohorts may arise solely due to variation in data analysis pipelines. Another limitation using the use of Illumina® Infinium HumanMethylation450 BeadChip limits CpG methylation investigation to specific sites. Although these limitations exist, investigation of whole-genome methylation status is not a viable approach for large cohorts, given the current costs.

Future directions:

Unlike the genome of an organism, the epigenome is largely malleable and reactive to environmental factors[72]. The dynamic profile of 5mC and 5hemi-methlyC(5hmC) demonstrate this point clearly. Epigenetic regulation of gene regulation is a complex network that involves histone modifications, lncRNAs, miRNAs, along with 5hmC and 5mC[217–220]. High levels of 5mC and 5hmC in the CNS further highlight the functional importance of these epigenetic marks[72]. Understanding the epigenetics of Alzheimer's disease and age-related dementia to give a complete understanding of the disease etiology requires the proper investigation of all factors involved in gene regulation.

Future directions would involve transitioning away from microarray-based approaches to investigate cytosine methylation. A microarray-based approach is biased towards sites limited to the hybridization array, and fails to interrogate whole genome methylation. Methylation across the genome can have effects that are not limited too the vicinity of promoter and gene bodies. As RNA-seq has been demonstrated to be more effective than microarray based methods, whole genome bisulfite sequencing will allow for an unbiased approach to survey the entire methylome. The validation of DEGs across different cohorts is important to understand the contribution to AD pathogenesis, but current sequencing costs limit the amount of samples that can be analyzed in a cost effective manner. However, our library preparation method demonstrates that DEGs can be identified, at a fraction of the sequencing depth recommended by ENCODE[137]. This opens up the possibility towards large-scale RNA-sequencing studies. In tandem, it would also be beneficial to investigate miRNA and lncRNAs by sequencing. However, this may be difficult to do in postmortem human tissue, as these species of oligonucleotides are very small, which becomes an issue when looking at PMI and RNA degradation.

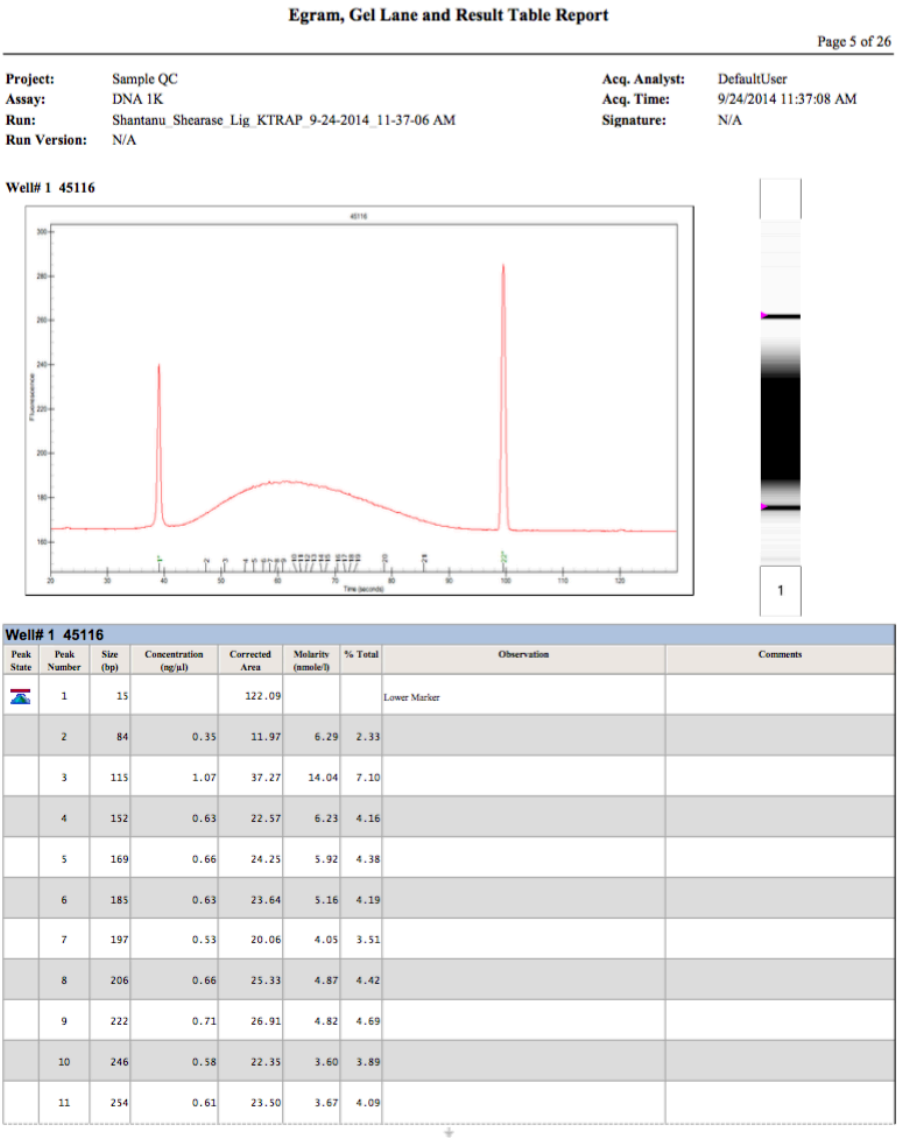
Translational impact

Overall understanding of this dynamic relationship between epigenomics, genetics, and the environment will certainly enhance our understanding of AD pathogenesis and possibly lead towards the development of novel therapeutic targets. Epigenetic modifications are reversible, can effect life span of an individual, and can be targeted by pharmacological interventions and change in diet[63]. Information on pharmacoepigenetic drugs is currently limited, however there are compounds currently approved by the Food and Drug Administration for the treatment of neoplastic processes[221]. Interventions targeting epigenetic regulation may be effective in treating neurodegenerative disorders, effect the hippocampal transcriptome, and reverse age-related cognitive dysfunction[63,221–224] . Drugs that impact epigenetics are often clustered in 6 categories: DNA Methyltransferase inhibitors, Histone deacetylase inhibitors, histone acetyltransferase modulators, histone methyltransferase inhibitors, histone demethylase inhibitors, and drugs that impact non-coding RNAs. AD-related genes do show DNA methylation changes, histone acetylation reduction, and deregulation of several lncRNAs & miRNAs. Pharmacoepigenetics doesn't predict all phenotypic variations in drug response, as individual differences in drug response are associated with genetic and epigenetic variability, along with disease determinants[63]. Currently, the information available on most drugs used for epigenetic alterations is limited, and proper evaluation of efficacy and safety of drugs has to still be evaluated in drug development and clinical trials[63].

Another avenue for a potential therapeutic approach would be the use of RNA interference (RNAi), to improve AD treatment. RNAi can be highly specific for mRNAs, is easy to synthesize and manufacture, but is very difficult to deliver. Over 21 RNAi therapeutics have

been developed for diseases such as cancers, viruses, and multiple genetic disorders. In mouse models, RNAi has been used to knock down *BACE1* expression within mouse brain[225], which has a direct role in Aβ production. Greater understanding of AD pathogenesis could lead to specific RNAi therapeutics as an efficient method for clinical applications.

Appendix A



Egram, Gel Lane and Result Table Report

Page 5 of 26

Project:	Sample QC	**Acq. Analyst:**	DefaultUser
Assay:	DNA 1K	**Acq. Time:**	9/24/2014 11:37:08 AM
Run:	Shantanu_Shearase_Lig_KTRAP_9-24-2014_11-37-06 AM	**Signature:**	N/A
Run Version:	N/A		

Well# 1 45116

Well# 1 45116

Peak State	Peak Number	Size (bp)	Concentration (ng/µl)	Corrected Area	Molarity (nmole/l)	% Total	Observation	Comments
	1	15		122.09			Lower Marker	
	2	84	0.35	11.97	6.29	2.33		
	3	115	1.07	37.27	14.04	7.10		
	4	152	0.63	22.57	6.23	4.16		
	5	169	0.66	24.25	5.92	4.38		
	6	185	0.63	23.64	5.16	4.19		
	7	197	0.53	20.06	4.05	3.51		
	8	206	0.66	25.33	4.87	4.42		
	9	222	0.71	26.91	4.82	4.69		
	10	246	0.58	22.35	3.60	3.89		
	11	254	0.61	23.50	3.67	4.09		

Figure A.1: Electropherogram of DNA after sonication on the Covaris E220. Sonication Target of 200bp was applied, and one can observe the distribution of DNA was between 168bp to 450bp as it is shown using the BioRad Experion DNA 1000 assay. The following settings were used: Time 200 seconds, Duty 10.0, Peak Incident Power 175, Cycles per Burst 200, Amplitude 0, Velocity 0, and Dwell 0.

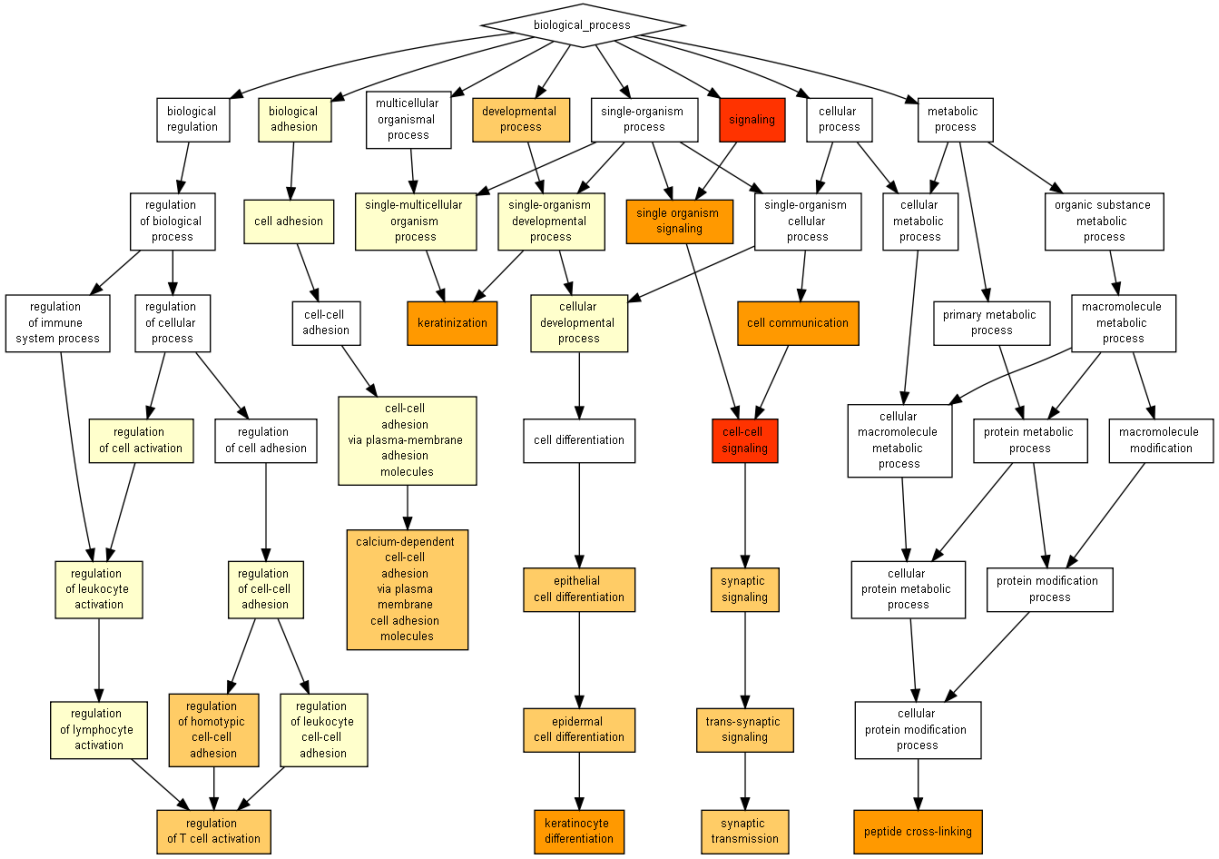


Figure A.2: Biological Processes effected as determined by GO analysis. A ranked list of differentially methylated genes, sorted by p-value was submitted for analysis. Inclusion of synaptic transmission in both, GO analysis of RNA dataset and DNA methylation dataset indicate a link between methylation and differential gene expression.

Table 4: MBD-Seq DNA characteristics. Listing the Total nanograms (ng) of DNA going into MBD-Capture, and final available for pooling after 15 cycles of PCR.

Sample ID	Total ng of human DNA going into MBD-Cap	Total ng availalble for pool after Amplification
24381	160	27.4
24385	175	26
36082	110.8	23.616
42990	121.56	2.44
41176	118.2	9.38
45116	156.6	15.9
25666	162.1	6.96
42113	132.8	32
24430	158.4	10.1
24509	156	30.4
36359	108	8.1
46202	205	31.1
25667	159.2	12.1
40482	207.2	21.6
35289	200.8	9.34
41639	202	3.9
41969	210	22.2
43045	190	11.2
43172	195.2	18
43192	145.2	6.86
43509	159.2	11.4
44624	170.4	27.6
45392	202	35
Average:	210.2782609	167.9191304

Table 8: List of top ten biological functions entities impacted by down-regulated genes as determined by GO analysis. Description, P-value, FDR q-value, and an enrichment score is given.

Description	P-value	FDR q-value	Enrichment
syntaxin-1 binding	1.53E-06	6.53E-03	23.74
myosin V binding	1.53E-06	3.26E-03	23.74
SNARE binding	2.12E-06	3.01E-03	6.84
transporter activity	3.35E-06	3.57E-03	2.3
myosin binding	8.06E-06	6.87E-03	9.58
calcium-dependent protein binding	8.06E-06	5.73E-03	9.58
syntaxin binding	8.70E-06	5.31E-03	7.78
transmembrane transporter activity	2.12E-05	1.13E-02	2.36
small molecule binding	2.51E-05	1.19E-02	1.73
channel regulator activity	3.32E-05	1.42E-02	5.63

Table 9: List of top ten molecular processes impacted by down-regulated genes as determined by GO analysis. Gene Ontology (GO) terms, description, P-value, FDR q-value, and an enrichment score is given.

Description	P-value	FDR q-value	Enrichment
synaptic transmission	1.84E-21	2.56E-17	6.77
modulation of synaptic transmission	9.96E-18	6.92E-14	8.61
single organism signaling	7.20E-17	3.33E-13	4.5
signaling	8.29E-17	2.88E-13	4.48
cell-cell signaling	4.83E-16	1.34E-12	4.48
cell communication	2.24E-15	5.18E-12	3.91
neurotransmitter transport	2.43E-15	4.82E-12	12.97
transport	6.94E-14	1.21E-10	2.11
regulation of synaptic vesicle transport	1.94E-13	2.99E-10	32.29
localization	2.52E-13	3.51E-10	1.99

Table 10: List of top ten cellular components impacted by down-regulated genes as determined by GO analysis. Gene Ontology (GO) terms, description, P-value, FDR q-value, and an enrichment score is given.

Description	P-value	FDR q-value	Enrichment
neuron part	4.76E-30	7.95E-27	4.88
synapse part	9.75E-29	8.14E-26	7.41
synaptic vesicle	1.33E-16	7.39E-14	16.94
cell projection part	3.58E-16	1.49E-13	4.11
axon part	3.55E-14	1.19E-11	9.44
synaptic vesicle membrane	8.11E-14	2.26E-11	19.43
neuron projection	1.02E-13	2.43E-11	4.33
cytoplasmic membrane-bounded vesicle	1.85E-12	3.87E-10	4.03
vesicle	5.32E-12	9.86E-10	2.05
cytoplasmic vesicle	6.00E-12	1.00E-09	3.68

Table 12.A: List of genes mapped to an interaction network. The following list of genes mapped to the number one interaction network as determined by IPA®. This is the same network displayed in Figure 15.

Symbol	Description
Ap1	Transcription Factor AP-1, involved in Cellular proiferation, transformation, and death
CALM1 (includes others)	EF-Hand calcium binding protein
CAMKK2	Calcium/calmodulin-dependent protein kinase kinase2
Cg	Computed gene
CPNE4	Calcium-dependent, phospholipid-binding protein, which may be involved in membrane trafficking, mitogenesis and development.
Creb	Cyclic AMP responsive element binding protein
GAP43	Growth Associated Protein 43 - presynaptic protein playing a key role in axonal growth, and modulating synapse information.
HOMER3	Family of postsynaptic density scaffolding proteins
IKK (complex)	IkB kinase-phosphorylated and activates NFkB
IL1R2	Interleukin 1 Receptor type 2 - This protein binds interleukin alpha (IL1A), interleukin beta (IL1B), and interleukin 1 receptor, type I(IL1R1/IL1RA), and acts as a decoy receptor that inhibits the activity of its ligands.
ITGB5	Integrin Beta 5: integrin complexes mediate cell-cell and cell-extracellular matrix interactions and this complex plays a role in human airway epithelial proliferation.
ITGB8	Integrin Beta 8: integrin complexes mediate cell-cell and cell-extracellular matrix interactions and this complex plays a role in human airway epithelial proliferation.
KAT2B	Lysine Acetyltransferase 2B - associates with p300/CBP. Contains Histone acetyl transferase activity, plays a direct role in transcriptional regulation
MT2A	Metallothionein 2A
NFkB (complex)	NFKB1 (Nuclear Factor Of Kappa Light Polypeptide Gene Enhancer In B-Cells 1)
NME1	Diseases associated with NME1 include anal canal carcinoma and laryngeal carcinoma. Among its related pathways are Integrated Pancreatic Cancer Pathway and Metabolism. This gene (NM
PELI2	Pellino E3 Ubiquitin Protein Ligase Family Member 2, is a Protein Coding gene. Among its related pathways are Immune System and Interleukin receptor SHC signaling.
Pkc(s)	Protein Kinase
PPARA	Peroxisome proliferator-activated receptor alpha is a member of the nuclear

	receptor family of ligand-activated transcription factors that heterodimerize with the retinoic X receptor (RXR) to regulate gene expression.
PRKCE	PRKCE (Protein Kinase C, Epsilon) is a Protein Coding gene. Among its related pathways are Immune System and DAG and IP3 signaling. This kinase has been shown to be involved in many different cellular functions, such as neuron channel activation, apoptosi
REST	REST RE1-Silencing Transcription Factor
RGS4	Regulator Of G-Protein Signaling 4. Among its related pathways are Signaling by GPCR and Activation of cAMP-Dependent PKA. Regulator of G protein s
SLC7A11	SLC7A11 Solute Carrier Family 7. Among its related pathways are Hemostasis and Transport
SNAP25	Synaptosomal-Associated Protein.
SNCA	Synuclein, Alpha. Diseases associated with SNCA include snca-related parkinson disease and parkinson disease.
SNCG	Synuclein, Gamma. Diseases associated with SNCG include synucleinopathy and multiple system atrophy. GO annotations related to this gene include beta-tubulin binding and alpha-tubulin binding.
STAR	Steroidogenic Acute Regulatory Protein. Diseases associated with STAR include lipoid adrenal hyperplasia and classic congenital lipoid adrenal hyperplasia due to star deficiency.
SYN1	SYN1 (Synapsin I) is a Protein Coding gene. Diseases associated with SYN1 include epilepsy, x-linked, with variable learning disabilities and behavior disorders and x-linked epilepsy - learning disabilities - behavior disorders. Among its related pathways
SYP	Synaptophysin.
TAC1	Tachykinin, Precursor 1. Diseases associated with TAC1 include neurotrophic keratopathy and neuroschistosomiasis.
Tgf beta	This gene encodes a member of the transforming growth factor beta family of cytokines, which are multifunctional peptides that regulate proliferation, differentiation, adhesion, migration, and other functions in many cell types.
TLR4	Toll-Like Receptor 4.
TUBB3	This gene encodes a class III member of the beta tubulin protein family. Beta tubulins are one of two core protein families (alpha and beta tubulins) that heterodimerize and assemble to form microtubules.
UCHL1	Ubiquitin Carboxyl-Terminal Esterase L1. This gene is specifically expressed in the neurons and in cells of the diffuse neuroendocrine system.
USP11	Ubiquitin Specific Peptidase 11. Among its related pathways are Transport to the Golgi and subsequent modification and Protein folding.
Vegf	This gene is a member of the PDGF/VEGF growth factor family.

ZBED6	Zinc Finger, BED-Type Containing 6.

Table 13: List of top ten biological processes impacted by GO analysis. A ranked gene list was submitted, sorted based on combined ranked score, generated from the 450k methylation array dataset. Gene Ontology (GO) terms, description, P-value, FDR q-value, and an enrichment score is given.

Description	P-value	FDR q-value	Enrichment
cell-cell signaling	4.82E-10	6.83E-06	2.09
signaling	8.03E-10	5.69E-06	1.87
single organism signaling	1.69E-09	7.98E-06	1.99
keratinization	6.56E-09	2.32E-05	9.58
cell communication	1.13E-08	3.20E-05	1.87
peptide cross-linking	4.62E-08	1.09E-04	13.22
keratinocyte differentiation	6.38E-08	1.29E-04	8.36
epidermal cell differentiation	7.70E-07	1.36E-03	6.86
epithelial cell differentiation	7.96E-07	1.25E-03	2.75
calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	8.71E-07	1.23E-03	8.6
regulation of homotypic cell-cell adhesion	1.85E-06	2.38E-03	2.91
regulation of T cell activation	4.78E-06	5.64E-03	2.9
synaptic signaling	6.37E-06	6.94E-03	2.03
trans-synaptic signaling	6.37E-06	6.45E-03	2.03
synaptic transmission	6.37E-06	6.02E-03	2.03
developmental process	9.97E-06	8.83E-03	1.18

Organic Extraction Protocol

Preparation of Solution Extraction Buffer(SEB)

1 Liter of stock SEB solution was made with the following compounds:

10ml Tris (pH8.0)

20 mL of 5M NaC1

20 mL of 0.5M EDTA

200 m1 of 10% SDS

600 mL of diH2O

This was then titrated to pH8.0 with concentrated HCl. A 50mL of working solution was made by adding 300mg of 1,4-Dithiothreitol to 50 mL of SEB stock solution. DNA extraction was performed for all samples following the UNTHSC organic extraction protocol[149].

MEDIPS R Script

This script was utilized to perform MEDIPS analysis within the MBD-seq dataset.

```
#! /projects/sequence_analysis/vol3/tools/R-3.1.3/bin/Rscript --save
library(MEDIPS)
library(MEDIPSData)
library(BSgenome.Hsapiens.UCSC.hg19)
# Get commandline arguments
args <- commandArgs(trailingOnly = TRUE)
# arg 1 = project name
# arg 2 = BAM file list file full path (must in this format *.merged.rg.deduped.bam)
# arg 4 = output dir
pjname <- args[1]
bamFile1 <- args[2]
bamFile2 <- args[3]
outdir <- args[4]
winSize <- as.numeric(args[5])
test_type <- args[6]
adj_type <- args[7]

if(winSize <= 0)
{
 winSize <- 500
}
```

```
if(test_type == '')
{
 test_type <- 'edgeR'
}

if(adj_type == '')
{
 adj_type <- 'bonferroni'
}
# Use chr22 for testing for now
chr.select <- "chr22"
# Assuming the hg19 genome used for alignment has no difference with the one used by
 BSgenome.Hsapiens.UCSC.hg19
BSgenome <- "BSgenome.Hsapiens.UCSC.hg19"
#uniq <- 1e-3
#uniq <- 0
extend <- 50
shift <- 0
sig_p <- 0.1

projectname <- file.path(outdir,pjname)
# Get the bam file names
```

```
bam_files1 <- read.delim(bamFile1, header = FALSE, sep = "\n")
bam_files2 <- read.delim(bamFile2, header = FALSE, sep = "\n")
MeDIP_Set1 <- c()
MeDIP_Set2 <- c()
# Setup bam file name list and datasets
bam1_lst <- NULL
for (n in 1:nrow(bam_files1))
{
 bam1_lst[n] <- toString(bam_files1[n,1])
 MeDIP_Set1 <- c(MeDIP_Set1, MEDIPS.createSet(file = bam1_lst[n], BSgenome =
 BSgenome, extend = extend, shift = shift, window_size = winSize, paired=TRUE))
 #MeDIP_Set1 <- c(MeDIP_Set1, MEDIPS.createSet(file = bam1_lst[n], BSgenome =
 BSgenome, extend = extend, shift = shift, window_size = winSize, chr.select = chr.select,
 paired=TRUE))
}
bam2_lst <- NULL
for (n in 1:nrow(bam_files2))
{
 bam2_lst[n] <- toString(bam_files2[n,1])
 MeDIP_Set2 <- c(MeDIP_Set2, MEDIPS.createSet(file = bam2_lst[n], BSgenome =
 BSgenome, extend = extend, shift = shift, window_size = winSize, paired=TRUE))
 #MeDIP_Set2 <- c(MeDIP_Set2, MEDIPS.createSet(file = bam2_lst[n], BSgenome =
 BSgenome, extend = extend, shift = shift, window_size = winSize, chr.select = chr.select,
```

```
 paired=TRUE))
}
# Creating coupling vector
CS <- MEDIPS.couplingVector(pattern = "CG", refObj = MeDIP_Set1[[1]])
# Run analysis
#mr.edgeR2 = MEDIPS.meth(MSet1 = MeDIP_Set1, MSet2 = MeDIP_Set2, CSet = CS, p.adj =
 "bonferroni", diff.method = "edgeR", MeDIP = T, CNV = F, minRowSum = 10)
mr.edgeR <- MEDIPS.meth(MSet1 = MeDIP_Set1, MSet2 = MeDIP_Set2, CSet = CS, p.adj =
 adj_type, diff.method = test_type, MeDIP = T, CNV = F, minRowSum = 10)
# Save total result
output.file <- paste0(projectname, '_MEDIPS_All.txt')
write.table(mr.edgeR,file=output.file, sep='\t', quote=FALSE, row.names=FALSE)
# Filter by adjusted P value
mr.edgeR.s <- MEDIPS.selectSig(results = mr.edgeR, p.value = sig_p, adj = T, ratio = NULL,
 bg.counts = NULL, CNV = F)
if(nrow(mr.edgeR.s) > 0)
{
 output.file <- paste0(projectname, '_MEDIPS_Sig.txt')
 write.table(mr.edgeR.s,file=output.file, sep='\t', quote=FALSE, row.names=FALSE)

 # Find the rows with edgeR.logFC value > 0 which means Set1 gain
 mr.edgeR.s.gain <- mr.edgeR.s[which(mr.edgeR.s[, grep("logFC", colnames(mr.edgeR.s))] >
 0), ]
```

```
if(nrow(mr.edgeR.s.gain) > 0)
{
    # Save total result
    output.file <- paste0(projectname, '_MEDIPS_Sig_Gain.txt')
    write.table(mr.edgeR.s.gain,file=output.file, sep='\t', quote=FALSE, row.names=FALSE)
    # Merge adjacent windows
    mr.edgeR.s.gain.m <- MEDIPS.mergeFrames(frames = mr.edgeR.s.gain, distance = 1)
    if(nrow(mr.edgeR.s.gain.m) > 0)
    {
        # Save total result
        output.file <- paste0(projectname, '_MEDIPS_Sig_Gain_Mrg.txt')
        write.table(mr.edgeR.s.gain.m,file=output.file, sep='\t', quote=FALSE,
row.names=FALSE)
    }
}
# Find the rows with edgeR.logFC value < 0 which means Set1 lost
mr.edgeR.s.lost <- mr.edgeR.s[which(mr.edgeR.s[, grep("logFC", colnames(mr.edgeR.s))] < 0),
]
if(nrow(mr.edgeR.s.lost) > 0)
{
    output.file <- paste0(projectname, '_MEDIPS_Sig_Lost.txt')
    write.table(mr.edgeR.s.lost,file=output.file, sep='\t', quote=FALSE, row.names=FALSE)
    # Merge adjacent windows
```

```
        mr.edgeR.s.lost.m <- MEDIPS.mergeFrames(frames = mr.edgeR.s.lost, distance = 1)
        if(nrow(mr.edgeR.s.gain.m) > 0)
        {
                output.file <- paste0(projectname, '_MEDIPS_Sig_Lost_Mrg.txt')
                write.table(mr.edgeR.s.lost.m,file=output.file,          sep='\t',          quote=FALSE,
 row.names=FALSE)
        }
 }
}
writeLines("Done")
```

DeSEQ2 Analysis R Script

This script was utilized to perform DeSEQ2 analysis within the RNA-seq dataset.

```
#First load the proper tools
library("Rsamtools")
library(DESeq2)
library("GenomicFeatures")
library(GenomicAlignments)
library("org.Hs.eg.db")
setwd("~/Desktop/RNA_BAM")
##Read the sampletable. Using 23 samples.
sampleTable <- read.csv("sampleTableAD_23.csv")
##View the loaded table
sampleTable
## Build the path to the tophat produced BAM files.
bamFiles <- file.path("~/Desktop/RNA_BAM/", sampleTable$dirName,
 sampleTable$bamName)
##View to make sure bamFiles is loaded
bamFiles
seqinfo(BamFile(bamFiles[1]))
##Need to have the hg19_genes.gtf file in our working directory
hse <-makeTxDbFromGFF("hg19_genes.gtf", format="gtf")
exonsByGene <- exonsBy(hse, by="gene")
## Use the function summarizeOverlaps to count reads within the gene
```

```
## This uses library("GenomicAlignments");
se <- summarizeOverlaps(exonsByGene, BamFileList(bamFiles), mode="Union",
 singleEnd=FALSE, ignore.strand=FALSE, fragments=TRUE);
## The "se" object (summarize experiment) have different component parts. The assay slot
 containts all the counts.
## You may see  the se object with the command: head(assay(se)).
head(assay(se))
## Now the comand: colData(se) returns: DataFrame with 23 rows and 0 columns
## Assign sampleTable (containing metadata info) to this object called se
colData(se) <- DataFrame(sampleTable);
## The colData slot contains pertinent sample/phenotypic information for the experiment
## Now Print colData to view.
colData(se)
# DataFrame with 23 rows and 7 columns
# colData(se)
colnames(se) <- sampleTable$dirName;
colData(se)
head(assay(se))
#Generating THE DESeqDataSet (dds), column metadata and design formula for differentil gene
 expression
ddsFull <- DESeqDataSet(se, design= ~treatment);
## Note: In order to benet from the default settings of the package, you should put the variable of
 interest at the end of the formula and make sure the control level is the first level.
```

```
ddsFull$treatment
# [1] CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL AD  AD  AD  AD  AD  AD
 AD  AD  AD  AD
# Levels: AD CTL
#As shown above the control level (CTL) it is not the first level, WE NEED TO RELEVEL
ddsFull$treatment <- relevel(ddsFull$treatment, "CTL");
ddsFull$treatment
# [1] CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL CTL AD  AD  AD  AD  AD  AD
 AD  AD  AD  AD
# Levels: CTL AD
## Now we can see that control is the first level
colData(ddsFull);
#Running the analysis
ddsFull <- DESeq(ddsFull);
## The last command returned a DESeqDataSet object with all the fitted information within it.
 Now we start to
## generate/extract results and tables
res
##Write a results table. "res"
write.table(res, file= "DGE_ADvsCTL_23_treatment", sep="\t", col.names=TRUE,
 row.names=TRUE);
#write.table(res, file= "DGE_ADvsCTL_22_rownamesTrue", sep="\t", col.names=TRUE,
 row.names=TRUE)
```

This rewrote our file, but the column names are one cell off.

Rather than adjusting the script, just shift the first row over one column.

##End of analysis

R Script utilized to perform RnBeads Analysis

This script was utilized to perform RnBeads analysis within the Illumina® 450K Infinium BeadChip dataset.

idat files,the annoation table "sample_annotation_AD23.cvs" and R script file "Finalpaperscript.R" are stored on the Kure Cluster. Login was after a succesful VPN connection, using my UNC Credentials. The directory where the data is stored is /proj/kirklab/users/SJS/RnB/SJS_idat

```
#Time to start analysis
library(RnBeads);
logger.start(fname=NA);
num.cores <- 8;
parallel.isEnabled();
parallel.setup(num.cores);
parallel.isEnabled();
## First you have to set up the analysis environment.
### Define the idat.dir; is the directory where my data is located
idat.dir <- "/proj/kirklab/users/SJS/RnB/SJS_idat"
sample.annotation <-"/proj/kirklab/users/SJS/RnB/SJS_idat/1217_EWAS_23.csv"
###Directory to write output
analysis.dir <- "/proj/kirklab/users/SJS/RnB/SJS_idat/12171000prbs_EWAS_23";
###Directory to write reports
report.dir <- file.path(analysis.dir, "12171000prbs_EWAS_23");
```

```
### running RnBeads
rnb.options(filtering.sex.chromosomes.removal=TRUE, identifiers.column="Sample_ID",
 differential.comparison.columns = "Sample_Group", exploratory.clustering.top.sites=1000,
 differential.site.test.method="refFreeEWAS", export.to.ewasher=TRUE,
 covariate.adjustment.columns = c("Age", "Sex"));
rnb.run.analysis(dir.reports=report.dir, sample.sheet=sample.annotation, data.dir=idat.dir,
 data.type="infinium.idat.dir")

#### In the server, and using AD23_idat as the working directory I ran:
#### bsub -q week -n 4 -R"span[hosts=1]" -o 1217_1000prbs_ewas_23.lsfout Rscript
 1217_1000prbs_EWAS_23.R
##End of analysis script
```

Bibliography:

1. 2012 Alzheimer's disease facts and figures. *Alzheimers. Dement.* **8,** 131–68 (2012).

2. Blennow, K., de Leon, M. J. & Zetterberg, H. Alzheimer's disease. *Lancet* **368,** 387–403 (2006).

3. RC, B. Genetics of Alzheimer Disease. *Scientifica (Cairo).* **2012,** (2012).

4. Bateman, R. J. *et al.* Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimers. Res. Ther.* **3,** 1 (2011).

5. Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* **80,** 1778–83 (2013).

6. Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A. & Evans, D. A. Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch. Neurol.* **60,** 1119–22 (2003).

7. Tanzi, R. E. The genetics of Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2,** a006296– (2012).

8. Selkoe, D. J. Resolving controversies on the path to Alzheimer's therapeutics. *Nat. Med.* **17,** 1060–1065 (2011).

9. Strassnig, M. & Ganguli, M. About a peculiar disease of the cerebral cortex: Alzheimer's original case revisited. *Psychiatry (Edgmont).* **2,** 30–3 (2005).

10. Selkoe, D. J. Alzheimer's disease: a central role for amyloid. *J. Neuropathol. Exp. Neurol.* **53,** 438–47 (1994).

11. Selkoe, D. J. Biochemistry and molecular biology of amyloid beta-protein and the mechanism of Alzheimer's disease. *Handb. Clin. Neurol.* **89,** 245–60 (2008).

12. Vassar, R. *et al.* Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286,** 735–41 (1999).

13. Morris, G. P., Clark, I. a & Vissel, B. Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta Neuropathol. Commun.* **2,** 135 (2014).

14. Karran, E., Mercken, M. & De Strooper, B. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat. Rev. Drug Discov.* **10,** 698–712 (2011).

15. Tanzi, R. E. & Bertram, L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* **120,** 545–55 (2005).

16. Glenner, G. G. & Wong, C. W. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem. Biophys. Res. Commun.* **122,** 1131–5 (1984).

17. Robakis, N. K., Ramakrishna, N., Wolfe, G. & Wisniewski, H. M. Molecular cloning and characterization of a cDNA encoding the cerebrovascular and the neuritic plaque amyloid peptides. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 4190–4 (1987).

18. Tanzi, R. E. *et al.* Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. *Science* **235,** 880–4 (1987).

19. Li, Y. M. *et al.* Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **405,** 689–94 (2000).

20. Kopan, R. A common enzyme connects Notch signaling and Alzheimer's disease. *Genes Dev.* **14,** 2799–2806 (2000).

21. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488,** 96–9 (2012).

22. Xu, G., Stevens, S. M., Moore, B. D., McClung, S. & Borchelt, D. R. Cytosolic proteins lose solubility as amyloid deposits in a transgenic mouse model of Alzheimer-type amyloidosis. *Hum. Mol. Genet.* **22,** 2765–2774 (2013).

23. Hardy, J. A. & Higgins, G. A. Alzheimer's disease: The amyloid cascade hypothesis.

24. Oddo, S., Caccamo, A., Kitazawa, M., Tseng, B. P. & LaFerla, F. M. Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer's disease. *Neurobiol.*

Aging **24,** 1063–70 (2003).

25. Butterfield, D. A. & Lauderback, C. M. Lipid peroxidation and protein oxidation in Alzheimer's disease brain: potential causes and consequences involving amyloid β-peptide-associated free radical oxidative stress1,2 1Guest Editors: Mark A. Smith and George Perry 2This article is part of a seri. *Free Radic. Biol. Med.* **32,** 1050–1060 (2002).

26. Weingarten, M. D., Lockwood, A. H., Hwo, S. Y. & Kirschner, M. W. A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci. U. S. A.* **72,** 1858–62 (1975).

27. Drubin, D. G. Tau protein function in living cells. *J. Cell Biol.* **103,** 2739–2746 (1986).

28. Esmaeli-Azad, B., McCarty, J. & Feinstein, S. Sense and antisense transfection analysis of tau function: tau influences net microtubule assembly, neurite outgrowth and neuritic stability. *J. Cell Sci.* **107,** 869–879 (1994).

29. Avila, J. Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett.* **580,** 2922–7 (2006).

30. Selkoe, D. J. Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiol Rev* **81,** 741–766 (2001).

31. Cook, C. *et al.* Loss of HDAC6, a novel CHIP substrate, alleviates abnormal tau accumulation. *Hum. Mol. Genet.* **21,** 2936–45 (2012).

32. Desikan, R. S. *et al.* Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* **132,** 2048–57 (2009).

33. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* **6,** 67–77 (2010).

34. Vlassenko, A. G., Benzinger, T. L. S. & Morris, J. C. PET amyloid-beta imaging in preclinical Alzheimer's disease. *Biochim. Biophys. Acta* **1822,** 370–9 (2012).

35. Shankar, G. M. & Walsh, D. M. Alzheimer's disease: synaptic dysfunction and Abeta. *Mol. Neurodegener.* **4,** 48 (2009).

36. Koffie, R. M., Hyman, B. T. & Spires-Jones, T. L. Alzheimer's disease: synapses gone

cold. *Mol. Neurodegener.* **6,** 63 (2011).

37. Terry, R. D. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* **30,** 572–80 (1991).

38. DeKosky, S. T., Scheff, S. W. & Styren, S. D. Structural correlates of cognition in dementia: quantification and assessment of synapse change. *Neurodegeneration* **5,** 417–21 (1996).

39. Coleman, P. D. & Yao, P. J. Synaptic slaughter in Alzheimer's disease. *Neurobiol. Aging* **24,** 1023–7 (2003).

40. Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443,** 787–95 (2006).

41. Phillips, N. R., Simpkins, J. W. & Roby, R. K. Mitochondrial DNA deletions in Alzheimer's brains: A review. *Alzheimers. Dement.* (2013). doi:10.1016/j.jalz.2013.04.508

42. Wenk, G. L. Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiatry* **64 Suppl 9,** 7–10 (2003).

43. Akiyama, H. Inflammation and Alzheimer's disease. *Neurobiol. Aging* **21,** 383–421 (2000).

44. Vom Berg, J. *et al.* Inhibition of IL-12/IL-23 signaling reduces Alzheimer's disease-like pathology and cognitive decline. *Nat. Med.* **18,** 1812–9 (2012).

45. Luchsinger, J. A., Cheng, D., Tang, M. X., Schupf, N. & Mayeux, R. Central obesity in the elderly is related to late-onset Alzheimer disease. *Alzheimer Dis. Assoc. Disord.* **26,** 101–5

46. Newman, A. B. *et al.* Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort. *J. Am. Geriatr. Soc.* **53,** 1101–7 (2005).

47. Mortimer, J. A., French, L. R., Hutton, J. T. & Schuman, L. M. Head injury as a risk factor for Alzheimer's disease. *Neurology* **35,** 264–7 (1985).

48. Fleminger, S., Oliver, D. L., Lovestone, S., Rabe-Hesketh, S. & Giora, A. Head injury as a risk factor for Alzheimer's disease: the evidence 10 years on; a partial replication. *J. Neurol. Neurosurg. Psychiatry* **74,** 857–62 (2003).

49. Salib, E. & Hillier, V. Head injury and the risk of Alzheimer's disease: a case control study. *Int. J. Geriatr. Psychiatry* **12,** 363–8 (1997).

50. Horton, A. M. & Reynolds, C. R. Early detection of risk of onset for dementia of the Alzheimer type and subtle executive dysfunction after TBI using the test of verbal conceptualization and fluency during clinical neuropsychological assessment: two case studies. *Appl. Neuropsychol.* **14,** 224–9 (2007).

51. Breed, S. *et al.* Cognitive functioning among individuals with traumatic brain injury, Alzheimer's disease, and no cognitive impairments. *J. Head Trauma Rehabil.* **23,** 149–57

52. Sims-Robinson, C., Kim, B., Rosko, A. & Feldman, E. L. How does diabetes accelerate Alzheimer disease pathology? *Nat. Rev. Neurol.* **6,** 551–9 (2010).

53. Borchelt, D. R. *et al.* Familial Alzheimer's Disease–Linked Presenilin 1 Variants Elevate Aβ1–42/1–40 Ratio In Vitro and In Vivo. *Neuron* **17,** 1005–1013 (1996).

54. Scheuner, D. *et al.* Secreted amyloid β–protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat. Med.* **2,** 864–870 (1996).

55. De Jonghe, C. Pathogenic APP mutations near the gamma-secretase cleavage site differentially affect Abeta secretion and APP C-terminal fragment stability. *Hum. Mol. Genet.* **10,** 1665–1671 (2001).

56. Rogaeva, E. *et al.* The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat. Genet.* **39,** 168–77 (2007).

57. Kamboh, M. I. *et al.* Genome-wide association study of Alzheimer's disease. *Transl. Psychiatry* **2,** e117 (2012).

58. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **41,** 1088–93 (2009).

59. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–8 (2013).

60. Wu, J., Basha, M. R. & Zawia, N. H. The environment, epigenetics and amyloidogenesis. *J. Mol. Neurosci.* **34,** 1–7 (2008).

61. Capell, B. C. & Berger, S. L. Genome-wide epigenetics. *J. Invest. Dermatol.* **133,** e9 (2013).

62. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11,** 191–203 (2010).

63. Cacabelos, R. & Torrellas, C. Epigenetics of aging and alzheimer's disease: Implications for pharmacogenomics and drug response. *Int. J. Mol. Sci.* **16,** 30483–30543 (2015).

64. Haggarty, P. Genetic and metabolic determinants of human epigenetic variation. *Curr. Opin. Clin. Nutr. Metab. Care* **18,** 334–8 (2015).

65. Castegna, A., Iacobazzi, V. & Infantino, V. The mitochondrial side of epigenetics. *Physiol. Genomics* **47,** 299–307 (2015).

66. Johannes, F., Colot, V. & Jansen, R. C. a quantitative genetics perspective. **9,** 883–890 (2008).

67. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3,** 662–73 (2002).

68. Robertson, K. D. DNA methylation and human disease. *Nat. Rev. Genet.* **6,** 597–610 (2005).

69. Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* **13,** 97–109 (2011).

70. Richardson, B. DNA methylation and autoimmune disease. *Clin. Immunol.* **109,** 72–79 (2003).

71. Sanchez-Mut, J. V *et al.* DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. *Brain* **136,** 3018–27 (2013).

72. Irier, H. A. & Jin, P. Dynamics of DNA methylation in aging and Alzheimer's disease. *DNA Cell Biol.* **31 Suppl 1,** S42–8 (2012).

73. Bock, C. *et al.* CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. *PLoS Genet.* **2,** e26 (2006).

74. Han, L., Su, B., Li, W.-H. & Zhao, Z. CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biol.* **9,** R79 (2008).

75. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13,** 1095–1107 (1992).

76. Shen, L. *et al.* Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* **3,** 2023–36 (2007).

77. Burgers, W. A., Fuks, F. & Kouzarides, T. DNA methyltransferases get connected to chromatin. *Trends Genet.* **18,** 275–7 (2002).

78. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22,** 2497–506 (2012).

79. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77–88 (2007).

80. Newell-Price, J., Clark, A. J. & King, P. DNA methylation and silencing of gene expression. *Trends Endocrinol. Metab.* **11,** 142–8

81. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **8,** 286–98 (2007).

82. Johansson, A., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One* **8,** e67378 (2013).

83. Hernandez, D. G. *et al.* Distinct DNA methylation changes highly correlated with

chronological age in the human brain. *Hum. Mol. Genet.* **20,** 1164–72 (2011).

84. Sanchez-Mut, J. V & Gräff, J. Epigenetic Alterations in Alzheimer's Disease. *Front. Behav. Neurosci.* **9,** 347 (2015).

85. Berdasco, M. & Esteller, M. Genetic syndromes caused by mutations in epigenetic genes. *Hum. Genet.* **132,** 359–83 (2013).

86. Gräff, J. & Tsai, L.-H. Histone acetylation: molecular mnemonics on the chromatin. *Nat. Rev. Neurosci.* **14,** 97–111 (2013).

87. Zovkic, I. B., Guzman-Karlsson, M. C. & Sweatt, J. D. Epigenetic regulation of memory formation and maintenance. *Learn. Mem.* **20,** 61–74 (2013).

88. Gupta, S. *et al.* Histone methylation regulates memory formation. *J. Neurosci.* **30,** 3589–99 (2010).

89. Guo, J. U. *et al.* Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14,** 1345–51 (2011).

90. Abel, T. & Zukin, R. S. Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. *Curr. Opin. Pharmacol.* **8,** 57–64 (2008).

91. Zawia, N. H., Lahiri, D. K. & Cardozo-Pelaez, F. Epigenetics, oxidative stress, and Alzheimer disease. *Free Radic. Biol. Med.* **46,** 1241–9 (2009).

92. Anne-Laurence, B., Caroline, R., Irina, P. & Jean-Philippe, L. Chromatin acetylation status in the manifestation of neurodegenerative diseases: HDAC inhibitors as therapeutic tools. *Subcell. Biochem.* **41,** 263–93 (2007).

93. Langley, B., Gensert, J. M., Beal, M. F. & Ratan, R. R. Remodeling chromatin and stress resistance in the central nervous system: histone deacetylase inhibitors as novel and broadly effective neuroprotective agents. *Curr. Drug Targets. CNS Neurol. Disord.* **4,** 41–50 (2005).

94. Gu, X., Sun, J., Li, S., Wu, X. & Li, L. Oxidative stress induces DNA demethylation and histone acetylation in SH-SY5Y cells: potential epigenetic mechanisms in gene transcription in Aβ production. *Neurobiol. Aging* **34,** 1069–79 (2013).

95. Marques, S., Batalha, V. L., Lopes, L. V. & Outeiro, T. F. Modulating Alzheimer's disease through caffeine: a putative link to epigenetics. *J. Alzheimers. Dis.* **24 Suppl 2,** 161–71 (2011).

96. van Praag, H., Shubert, T., Zhao, C. & Gage, F. H. Exercise enhances learning and hippocampal neurogenesis in aged mice. *J. Neurosci.* **25,** 8680–5 (2005).

97. Wu, J. *et al.* Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (Pb): evidence for a developmental origin and environmental link for AD. *J. Neurosci.* **28,** 3–9 (2008).

98. Mastroeni, D. *et al.* Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol. Aging* **31,** 2025–37 (2010).

99. Chouliaras, L. *et al.* Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* **34,** 2091–2099 (2013).

100. Serre, D., Lee, B. H. & Ting, A. H. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* **38,** 391–9 (2010).

101. Hatada, I. *et al.* Genome-wide profiling of promoter methylation in human. *Oncogene* **25,** 3059–64 (2006).

102. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6,** 692–702 (2014).

103. Barr, H. *et al.* Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol. Cell. Biol.* **27,** 3750–7 (2007).

104. Yegnasubramanian, S. Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. *Nucleic Acids Res.* **34,** e19–e19 (2006).

105. Aberg, K. A. *et al.* Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. *Epigenomics* **5,** 367–77 (2013).

106. Aberg, K. A. *et al.* MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case--control samples. *Epigenomics* **4,** 605–21 (2012).

107. Aberg, K. A. *et al.* Evaluation of Methyl-Binding Domain Based Enrichment Approaches Revisited. *PLoS One* **10,** e0132205 (2015).

108. De Meyer, T. *et al.* Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. *PLoS One* **8,** e59068 (2013).

109. Twine, N. A., Janitz, K., Wilkins, M. R. & Janitz, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One* **6,** e16266 (2011).

110. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320,** 1344–9 (2008).

111. Asmann, Y. W. *et al.* 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics* **10,** 531 (2009).

112. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–17 (2008).

113. Sutherland, G. T., Janitz, M. & Kril, J. J. Understanding the pathogenesis of Alzheimer's disease: will RNA-Seq realize the promise of transcriptomics? *J. Neurochem.* **116,** 937–46 (2011).

114. van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.* **8,** e1000371 (2010).

115. Zhang, Z., Theurkauf, W. E., Weng, Z. & Zamore, P. D. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence* **3,** 9 (2012).

116. Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proc. Natl. Acad. Sci. U. S. A.* **27,** 499–506 (1941).

117. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–3 (1970).

118. Cooper, G. M. The Complexity of Eukaryotic Genomes. (2000). at <http://www.ncbi.nlm.nih.gov/books/NBK9846/>

119. Dominic Mills, J., Kawahara, Y. & Janitz, M. Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Curr. Genomics* **14,** 173–181 (2013).

120. Cao, J. The functional role of long non-coding RNAs and epigenetics. *Biol. Proced. Online* **16,** 11 (2014).

121. Costa, V., Aprile, M., Esposito, R. & Ciccodicola, A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur. J. Hum. Genet.* **21,** 134–42 (2013).

122. Mills, J. D. *et al.* RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neurosci. Lett.* **536,** 90–5 (2013).

123. Satoh, J. I., Yamamoto, Y., Asahina, N., Kitano, S. & Kino, Y. RNA-seq data mining: Downregulation of neurod6 serves as a possible biomarker for alzheimer's disease brains. *Dis. Markers* **2014,** (2014).

124. Ginsberg, S. D. *et al.* Regional selectivity of rab5 and rab7 protein upregulation in mild cognitive impairment and Alzheimer's disease. *J. Alzheimers. Dis.* **22,** 631–9 (2010).

125. Ginsberg, S. D., Hemby, S. E., Lee, V. M., Eberwine, J. H. & Trojanowski, J. Q. Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann. Neurol.* **48,** 77–87 (2000).

126. Ginsberg, S. D., Che, S., Counts, S. E. & Mufson, E. J. Single cell gene expression profiling in Alzheimer's disease. *NeuroRx* **3,** 302–18 (2006).

127. Hata, R. *et al.* Up-regulation of calcineurin Abeta mRNA in the Alzheimer's disease brain: assessment by cDNA microarray. *Biochem. Biophys. Res. Commun.* **284,** 310–6 (2001).

128. Blalock, E. M. *et al.* Incipient Alzheimer's disease: microarray correlation analyses reveal

major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2173–8 (2004).

129. Williams, C. *et al.* Transcriptome analysis of synaptoneurosomes identifies neuroplasticity genes overexpressed in incipient Alzheimer's disease. *PLoS One* **4,** e4936 (2009).

130. Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* **5,** 48 (2013).

131. Yao, P. J. *et al.* Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer's disease. *Neurobiol. Dis.* **12,** 97–109 (2003).

132. Selkoe, D. J. Alzheimer's disease is a synaptic failure. *Science* **298,** 789–91 (2002).

133. Ho, L. Altered expression of a-type but not b-type synapsin isoform in the brain of patients at high risk for Alzheimer's disease assessed by DNA microarray technique. *Neurosci. Lett.* **298,** 191–194 (2001).

134. Satoh, J.-I., Yamamoto, Y., Asahina, N., Kitano, S. & Kino, Y. RNA-Seq data mining: downregulation of NeuroD6 serves as a possible biomarker for alzheimer's disease brains. *Dis. Markers* **2014,** 123165 (2014).

135. Sekar, S. *et al.* Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. *Neurobiol. Aging* **36,** 583–591 (2015).

136. Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics* **30,** 301–304 (2014).

137. Encode. ENCODE RNA standards v1.0 1 2011 May. *Encode* 1–7 (2011).

138. Braak, H. & Braak, E. Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* **16,** 271–278 (1995).

139. Braak, H. & Braak, E. Evolution of neuronal changes in the course of Alzheimer's disease. *J. Neural Transm. Suppl.* **53,** 127–40 (1998).

140. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7,** 3 (2006).

141. RNA Access Data Sheet. at <https://support.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet-truseq-rna-access.pdf>

142. RNA Access Protocol. at <http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqrnaaccess/truseq-rna-access-library-prep-guide-15049525-b.pdf>

143. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–78 (2012).

144. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–11 (2009).

145. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

146. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

147. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–8 (2008).

148. Benjamini, Y. & Hochberg, Y. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

149. Procedure for DNA organic extraction. Lab techniques from Intro to Forensic Laboratory Course: Guidelines for DNA Organic Extraction, UNTHSC.

150. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11,** 1138–40 (2014).

151. R: The R Project for Statistical Computing. at <https://www.r-project.org/>

152. Guintivano, J., Aryee, M. J. & Kaminsky, Z. A. A cell epigenotype specific model for the

correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* **8,** 290–302 (2013).

153. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30,** 1431–9 (2014).

154. Lan, X. *et al.* High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. *PLoS One* **6,** e22226 (2011).

155. Invitrogen. MethylMiner ™ Methylated DNA Enrichment Kit. *ReVision* (2009).

156. Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30,** 284–6 (2014).

157. Chavez, L. MEDIPS : genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments . 1–14 (2015).

158. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).

159. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).

160. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28,** 1919–20 (2012).

161. Magistri, M., Velmeshev, D., Makhmutova, M. & Faghihi, M. A. Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid-β Homeostasis, and Deregulated Expression of Long Noncoding RNAs. *J. Alzheimer's Dis.* **48,** 647–665 (2015).

162. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).

163. Bioconductor - methylumi. at

<https://www.bioconductor.org/packages/release/bioc/html/methylumi.html>

164. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30,** 1363–9 (2014).

165. Stewart, S. E. *et al.* Genome-wide association study of obsessive-compulsive disorder. 788–799 (2013). doi:10.1038/mp.2012.85

166. Edwards, T. L. *et al.* Genome-wide association study identifies possible genetic risk factors for colorectal adenomas. *Cancer Epidemiol. Biomarkers Prev.* **22,** 1219–26 (2013).

167. Picard Tools - By Broad Institute. at <http://broadinstitute.github.io/picard/>

168. Rapaport, F. *et al.* Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* **14,** R95 (2013).

169. RNA_Access. at <http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet-truseq-rna-access.pdf>

170. Pimplikar, S. W., Nixon, R. A., Robakis, N. K., Shen, J. & Tsai, L.-H. Amyloid-independent mechanisms in Alzheimer's disease pathogenesis. *J. Neurosci.* **30,** 14946–54 (2010).

171. Pimplikar, S. W. Reassessing the amyloid cascade hypothesis of Alzheimer's disease. *Int. J. Biochem. Cell Biol.* **41,** 1261–8 (2009).

172. Jiang, S. *et al.* Trafficking regulation of proteins in Alzheimer's disease. *Mol. Neurodegener.* **9,** 6 (2014).

173. Tang, B. L. Neuronal protein trafficking associated with Alzheimer disease: From APP and BACE1 to glutamate receptors. *Cell Adhes. Migr.* **3,** 118–128 (2009).

174. Jahn, H. *et al.* Peptide fingerprinting of Alzheimer's disease in cerebrospinal fluid: identification and prospective evaluation of new synaptic biomarkers. *PLoS One* **6,** e26540 (2011).

175. Kneussel, M. & Wagner, W. Myosin motors at neuronal synapses: drivers of membrane

transport and actin dynamics. *Nat. Rev. Neurosci.* **14,** 233–47 (2013).

176. STX1A syntaxin 1A [Homo sapiens (human)] - Gene - NCBI. at <http://www.ncbi.nlm.nih.gov/gene/6804>

177. Ungar, D. & Hughson, F. M. SNARE protein structure and function. *Annu. Rev. Cell Dev. Biol.* **19,** 493–517 (2003).

178. Sharma, M. *et al.* CSPα knockout causes neurodegeneration by impairing SNAP-25 function. *EMBO J.* **31,** 829–41 (2012).

179. Sevlever, D. *et al.* Genetically-controlled Vesicle-Associated Membrane Protein 1 expression may contribute to Alzheimer's pathophysiology and susceptibility. *Mol. Neurodegener.* **10,** 18 (2015).

180. Cirrito, J. R. *et al.* Endocytosis is required for synaptic activity-dependent release of amyloid-beta in vivo. *Neuron* **58,** 42–51 (2008).

181. Del Prete, D., Lombino, F., Liu, X. & D'Adamio, L. APP is cleaved by Bace1 in pre-synaptic vesicles and establishes a pre-synaptic interactome, via its intracellular domain, with molecular complexes that regulate pre-synaptic vesicles functions. *PLoS One* **9,** e108576 (2014).

182. Mukaetova-Ladinska, E. B. *et al.* Lewy body variant of Alzheimer's disease: selective neocortical loss of t-SNARE proteins and loss of MAP2 and alpha-synuclein in medial temporal lobe. *ScientificWorldJournal.* **9,** 1463–75 (2009).

183. Finsen, B. & Owens, T. Innate immune responses in central nervous system inflammation. *FEBS Lett.* **585,** 3806–3812 (2011).

184. Meffert, M. K., Chang, J. M., Wiltgen, B. J., Fanselow, M. S. & Baltimore, D. NF-kappa B functions in synaptic signaling and behavior. *Nat. Neurosci.* **6,** 1072–8 (2003).

185. Karin, M. How NF-kB is activated: the role of the IkB kinase (IKK) complex. *Oncogene* **18,** 6867–6874 (1999).

186. Lukiw, W. J. NF-κB-regulated, proinflammatory miRNAs in Alzheimer's disease. *Alzheimers. Res. Ther.* **4,** 47 (2012).

187. Saura, C. A. & Valero, J. The role of CREB signaling in Alzheimer's disease and other cognitive disorders. *Rev. Neurosci.* **22,** (2011).

188. Cheng, C. *et al.* MicroRNA-144 Is Regulated by Activator Protein-1 (AP-1) and Decreases Expression of Alzheimer Disease-related A Disintegrin and Metalloprotease 10 (ADAM10). *J. Biol. Chem.* **288,** 13748–13761 (2013).

189. Abrajano, J. J. *et al.* REST and CoREST Modulate Neuronal Subtype Specification, Maturation and Maintenance. *PLoS One* **4,** 949–957 (2009).

190. Noh, K.-M. *et al.* Repressor element-1 silencing transcription factor (REST)-dependent epigenetic remodeling is critical to ischemia-induced neuronal death. doi:10.1073/pnas.1121568109

191. Madabhushi, R. a M. A protective factor for the ageing brain. *Nature* **2352,** 10–11 (2014).

192. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507,** 448–54 (2014).

193. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer/'s disease. *Nature* **507,** 448 (2014).

194. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17,** 1164–70 (2014).

195. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* **17,** 1156–1163 (2014).

196. Matsuzaka, Y. *et al.* Identification, expression analysis and polymorphism of a novel RLTPR gene encoding a RGD motif, tropomodulin domain and proline/leucine-rich regions. *Gene* **343,** 291–304 (2004).

197. Yin, L. *et al.* Autophagy-related gene16L2, a potential serum biomarker of multiple sclerosis evaluated by bead-based proteomic technology. *Neurosci. Lett.* **562,** 34–8 (2014).

198. Gousset, K., Marzo, L., Commere, P.-H. & Zurzolo, C. Myo10 is a key regulator of TNT

formation in neuronal cells. *J. Cell Sci.* **126,** 4424–35 (2013).

199. Plantman, S., Zelano, J., Novikova, L. N., Novikov, L. N. & Cullheim, S. Neuronal myosin-X is upregulated after peripheral nerve injury and mediates laminin-induced growth of neurites. *Mol. Cell. Neurosci.* **56,** 96–101 (2013).

200. Gousset, K. *et al.* Prions hijack tunnelling nanotubes for intercellular spread. *Nat. Cell Biol.* **11,** 328–336 (2009).

201. Langevin, C. *et al.* Characterization of the role of dendritic cells in prion transfer to primary neurons. *Biochem. J.* **431,** 189–198 (2010).

202. Yamamoto, H. *et al.* NDRG4 protein-deficient mice exhibit spatial learning deficits and vulnerabilities to cerebral ischemia. *J. Biol. Chem.* **286,** 26158–65 (2011).

203. Mattson, M. P. Glutamate and neurotrophic factors in neuronal plasticity and disease. *Ann. N. Y. Acad. Sci.* **1144,** 97–112 (2008).

204. Alzheimer Disease, by, Bot, N., Schweizer, C., Ben Halima, S. & Fraering, P. C. Processing of the Synaptic Cell Adhesion Molecule Neurexin-3 NL. (2010). doi:10.1074/jbc.M110.142521

205. Martinez-Mir, A. *et al.* Genetic study of neurexin and neuroligin genes in Alzheimer's disease. *J. Alzheimers. Dis.* **35,** 403–12 (2013).

206. Fogel, B. L. *et al.* RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21,** 4171–4186 (2012).

207. Alam, S., Suzuki, H. & Tsukahara, T. Alternative splicing regulation of APP exon 7 by RBFox proteins. *Neurochem. Int.* **78,** 7–17 (2014).

208. Matsui, T. *et al.* Expression of APP pathway mRNAs and proteins in Alzheimer's disease. *Brain Res.* **1161,** 116–123 (2007).

209. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34,** 267–73 (2003).

210. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).

211. Tan, L. *et al.* Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. *Alzheimers. Dement.* **9,** 546–53 (2013).

212. Akhtar, M. W. *et al.* In vivo analysis of MEF2 transcription factors in synapse regulation and neuronal survival. *PLoS One* **7,** e34863 (2012).

213. Le Meur, N. *et al.* MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J. Med. Genet.* **47,** 22–9 (2010).

214. Atz, M. *et al.* Methodological considerations for gene expression profiling of human brain. *J. Neurosci. Methods* **163,** 295–309 (2007).

215. Wang, S.-C., Oelze, B. & Schumacher, A. Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One* **3,** e2698 (2008).

216. Talens, R. P. *et al.* Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* **11,** 694–703 (2012).

217. Sato, F., Tsuchiya, S., Meltzer, S. J. & Shimizu, K. MicroRNAs and epigenetics. *FEBS J.* **278,** 1598–609 (2011).

218. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10604–9 (2005).

219. Alagiakrishnan, K., Gill, S. S. & Fagarasanu, A. Genetics and epigenetics of Alzheimer's disease. *Postgrad. Med. J.* **88,** 522–9 (2012).

220. Langevin, S. M. *et al.* The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics* **6,** 908–919 (2011).

221. Cacabelos, R. Epigenomic networking in drug development: from pathogenic mechanisms to pharmacogenomics. *Drug Dev. Res.* **75,** 348–65 (2014).

222. Cacabelos, R. & Torrellas, C. Epigenetic drug discovery for Alzheimer's disease. *Expert Opin. Drug Discov.* **9,** 1059–86 (2014).

223. Sewal, A. S. *et al.* Experience Modulates the Effects of Histone Deacetylase Inhibitors on Gene and Protein Expression in the Hippocampus: Impaired Plasticity in Aging. *J. Neurosci.* **35,** 11729–42 (2015).

224. Qureshi, I. A. & Mehler, M. F. Epigenetics and therapeutic targets mediating neuroprotection. *Brain Res.* **1628,** 265–72 (2015).

225. Chen, S. *et al.* Advances with RNA interference in Alzheimer's disease research. *Drug Des. Devel. Ther.* **7,** 117–25 (2013).