Investigating *APOE* DNA methylation in Alzheimer's disease and its relationship to *APOE ε4* genotype

Submitted by Hau Tak Tam to the University of Exeter
as a thesis for the degree of
Masters by Research in Medical Studies
In July 2019

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: …………………………………………………………..

ABSTRACT

Alzheimer's disease is a neurodegenerative condition that is the leading cause of dementia in the elderly, which affected close to 50 million people worldwide in 2017. The disease is characterised by the presence of amyloid beta plaques and neurofibrillary tangles of hyperphosphorylated tau protein in the brain. There are no disease-modifying treatments, largely because the mechanisms underlying disease and pathology are not understood. It is estimated that common genetic variants explain ~33% of disease incidence, although the mechanism behind their action is not clear. Polymorphisms in the *APOE* gene have been widely linked with AD, and the *APOE ε4* variant is the greatest genetic risk factor for sporadic Alzheimer's disease.

We hypothesise that epigenetic differences, namely DNA methylation, in *APOE* may contribute to Alzheimer's disease aetiology and that this may also be related to *APOE* genotype. In this thesis we have investigated DNA methylation in prefrontal cortex and temporal gyrus tissue form individuals with varying degrees of pathology or disease using data from the Illumina 450K methylation array. We identified pathology-associated hypomethylation of *APOE* in four probes that reside in a CpG island in the 3' untranslated region. Using a large cohort of pre-natal and post-natal brain tissue samples from individuals with methylation and matched genotype data we showed that methylation of one site in this region seems to be driven by *APOE ε4* genotype in pre-natal development, although this was not observed in post-natal samples. We did however identify another loci in this island that showed hypomethylation with age in post-natal samples. In the promoter at the 5' end of the gene we observed three adjacent loci near the transcription start site and one loci in the gene body that were significantly hypomethylated with advancing age in Alzheimer's disease samples. Overall *APOE ε4* genotype had little effect on *APOE* DNA methylation in the context of Alzheimer's disease. This project is the first large-scale study of *APOE* DNA methylation with respect to Alzheimer's disease diagnosis and pathology, age and *APOE* genotype and highlights the need for further work in to the role of *APOE* DNA methylation in disease aetiology.

STATEMENT OF WORK

The analyses presented in Chapter 3 of this thesis were conducted under the direct supervision of Dr Rebecca Smith, with assistance from Professor Katie Lunnon. The GEO accession numbers for the Mount Sinai and LBB1 (London Brain Bank) datasets are GSE80970 and GSE59685, respectively. Data for the Arizona cohorts was provided directly by collaborators from Arizona State University, Tempe, AZ, and used with permission from Professor Paul Coleman. All data imputation for genetic data was performed by Gemma Shireby.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

Abbreviation	Term
27K array	Illumina Infinium 27K Methylation BeadChip array
450K array	Illumina Infinium 450K Methylation BeadChip Array
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Aβ	Amyloid-beta
APOE	Apolipoprotein E
APP	Amyloid precursor protein
AD	Alzheimer's Disease
ADRDA	Alzheimer's Disease and Related Disorders Association
ASD	Autism spectrum disorder
BP	Base pairs
CAA	Cerebral amyloid angiopathy
CDC	Centre for Disease Control
CER	Cerebellum
CETS	Cell-epigenotype specific
CGIs	CpG islands
CSF	Cerebrospinal fluid
CTL	Non-demented controls
DEP	Depression
DLB	Dementia with Lewy bodies
DMP(s)	Differentially methylated position(s)
DMR(s)	Differentially methylated region(s)
DNA	Deoxyribose nucleic acid
DZ	Dizygotic
EC	Entorhinal cortex

EOAD	Early-onset Alzheimer's disease
EPIC array	Illumina Infinium EPIC Methylation BeadChip Array
EWAS	Epigenome-wide association study
FACS	Fluorescence-activated cell sorting
fMRI	Functional magnetic resonance imaging
GWAS	Genome-wide association study
LD	Linkage disequilibrium
LOAD	Late-onset Alzheimer's disease
MAF	Minor allele frequency
MCI	Mild cognitive impairment
miRNA(s)	Micro-RNA(s)
mQTL	Methylation quantitative trait loci
Mw	Molecular weight
MZ	Monozygotic
NFT(s)	Neurofibrillary Tau tangle(s)
NINCS	National Institute of Neurological and Communicative Disorders and Stroke
OR	Odds ratio
PCR	Polymerase chain reaction
PET	Positron-emission tomography
PFC	Prefrontal cortex
PSEN	Presenilin
QC	Quality control
QN	Quantile normalization
RNA	Ribose nucleic acid
ROSMAP	Religious Orders Study/Rush Memory and Aging Project
RRHP	Reduced representation hydroxymethylation profiling

SCZ	Schizophrenia
SD	Standard deviation
sMRI	Structural magnetic resonance imaging
SNP	Single-nucleotide polymorphism
STG	Superior temporal gyrus
TET	Ten-eleven translocation dioxygenase
TSS	Transcription start sites
UTR	Untranslated region

CHAPTER 1 : INTRODUCTION

1.1 Alzheimer's disease (AD)

Alzheimer's disease (AD) is a progressive neurodegenerative disease that is the main cause of dementia, accounting for ~60-80% cases worldwide (Ashraf *et al.*, 2016). It is clinically characterized by three primary groups of symptoms encompassing different aspects of the disease; cognitive dysfunction (deterioration of memory, executive function, speech, and other areas of cognition), non-cognitive symptoms (including psychosis, emotional instability and agitation) and decline of instrumental and basic skills for activities in daily life, along with general deterioration of motor control and autonomic functions (Burns and Iliffe, 2009). Eventually, patients become incapable of independent living, and are wholly dependent on caregivers. AD is invariably fatal, often due to comorbid conditions such as dysphagia and aspiration pneumonia, and is the sixth leading cause of death in the United States, afflicting around 5.7 million people and projected to rise to nearly 14 million by 2050 (Alzheimer's Association, 2015) . A steady exponential growth in the incidence of dementia is expected worldwide, with epidemiological studies estimating a doubling of new cases every 20 years up to 2040 and a 15-fold increase in dementia prevalence in the 60-85 year old age group (Mayeux and Stern, 2013). This coincides with the Centre for Disease Control (CDC)'s estimate that the over 65 years population in the US will double over four decades, from 46.5 million in 2014, to 83.7 million (Matthews *et al.*, 2018). This inevitable rise in the geriatric population, and consequently AD prevalence, threatens to overwhelm healthcare services due to the inordinate financial and resource cost of caregiving (Matthews *et al.*, 2018).

1.1.1 AD neuropathology

The neuropathology of AD is defined by the accumulation of dense, insoluble extracellular plaques composed of aggregated amyloid-beta (Aβ) and intracellular neurofibrillary tangles (NFT) of hyperphosphorylated tau protein. Aβ is formed from the amyloidogenic cleavage of the amyloid precursor protein (APP) by β-secretase and then γ-secretase. These Aβ monomers form oligomers and ultimately plaques. The spread of the NFTs through neural structures is measured with the Braak staging system, on a scale of I (mild lesions exclusive to the transentorhinal region) to VI (lesions are visible even in the occipital neocortex) (Braak *et al.*, 2006). The presence of Aβ plaques and NFTs are believed to lead to neuronal cell and synapse loss, especially the cholinergic neurons that connect the hippocampus and cerebral cortex and play a critical function in memory encoding/recall (Niikura, Tajima and Kita, 2006), though recent neuronal cell studies have shown Aβ oligomers to have a pathogenic role in their own right by instigating cell-cycle events (Mukhin, Pavlov and Klimenko, 2017) (Varvel *et al.*, 2008), with further gene-knockout studies suggesting their reliance on tau for signalling (Seward *et al.*, 2013). This process is observed as progressive atrophy and shrinkage within the cerebral cortex and several subcortical regions such as the entorhinal cortex, temporal gyrus, and hippocampus.

1.1.2   Diagnosis of AD

The established criteria for AD diagnosis are confirmed to have high sensitivity and specificity, with an 80% positive predictive value (Cummings, 2012). Central aspects as delineated in NINCDS-ADRDA, the most commonly used set in AD research, include severe cognitive deficits and functional compromise characteristic of dementia shown by neuropsychological testing, confirmation of diagnosis with post mortem examination, and definitive exclusion of other forms of dementia (Cummings, 2012).

However, improved in-vivo biomarker detection has revealed neurological damage insidiously beginning and worsening decades before clinical diagnosis (Huynh and Mohan, 2017), placing earlier and better methods of detection/diagnosis as the foremost priority. Another pertinent problem is accurately determining whether potential dementia patients will develop the condition, and also differentiating those with preclinical dementia into future AD and non-AD sufferers (Knopman and Caselli, 2012) (Nitrini, 2010) (B.J., 2015). One of the earliest potential presentations of AD is mild cognitive impairment (MCI), an intermediate stage where individuals have memory impairment with little impact on general cognition and capability in daily tasks (B.J., 2015) However, it is by itself an unreliable indicator of future development of dementia, as controversy exists over subjective differentiations from normal cognitive dysfunction observed in ageing (Masters, Morris and Roe, 2015) , as well as the lack of guaranteed development of AD in patients (Masters, Morris and Roe, 2015). Furthermore, MCI can be a common precursor shared with other forms of dementia such as vascular dementia.

The majority of pathological facets of AD remain difficult to quantify clinically ante mortem, which creates a divide between symptom-based and pathological diagnosis (Hane, Lee and Leonenko, 2017), especially as multiple aggregated protein species can cause similar disease phenotypes (Villemagne *et al.*, 2018). The gold standard remains to be post mortem histological analysis of neuropathology, looking for the presence of Aβ plaques and quantifying the spread of NFTs using Braak staging (Perl, 2010), to confirm a clinical diagnosis. However, even this direct method is fallible due to possible co-existance of other proteins with tau and Aβ (Villemagne *et al.*, 2018).

Cerebrospinal fluid (CSF) collection and blood sampling has been employed to directly measure biomarker levels in the nervous system and blood, respectively during life (Reitz & Mayeux, 2014). Recent advances in neuroimaging have also allowed alternative non-invasive visualisation of changes within the brain itself, which are also highly specific. Structural and functional magnetic resonance imaging (sMRI/fMRI) have been employed to identify gross changes in brain regions and metabolism, respectively, although these are not particularly useful at identifying pre-clinical AD-related changes in brain structure (Johnson *et al.*, 2012). Positron emission tomography (PET) has been combined with new markers in recent years, to allow the assessment of amyloid (Herholz and Ebmeier, 2011) and Tau (Wang *et al.*, 2016) levels directly within the brain. However, there are specific limitations to using

neuroimaging, such as incomplete visualisation of certain aspects of pathology for each method and expensive costs, which hinder their use clinically for diagnosis (Johnson *et al.*, 2012).

1.1.3 Treatments for AD

As of yet, there are no effective treatments that can halt or reverse dementia development. Current medications simply treat the symptoms of the disease, but do not reverse the underlying pathology. This is in part because the disease is currently diagnosed only when clinical symptoms of the disease emerge, by which point considerable pathology has already occurred during the preclinical stage, and in part because we do not know the exact cause of the majority of AD cases (Hane, Lee and Leonenko, 2017). The main class of drugs currently prescribed for AD are acetylcholinesterase inhibitors (e.g. Donepezil, Galantamine, Rivastigmine), often in conjunction with memantine, a glutamate antagonist (Herrmann *et al.*, 2013). The former reduce symptoms such as memory deficits by decreasing breakdown of the neurotransmitter acetylcholine, compensating for damage to cholinergic neural networks in AD, with the majority of patients experiencing mild to moderate improvement with few side effects (Casey, Antimisiaris and O'Brien, 2010) (Mehta, Adem & Sabbagh, 2012), though drug therapies are ultimately highly individualised with no universal points of initiation and discontinuation (Hogan, 2015). Research efforts have focused on developing new pharmaceutical treatments such as immunotherapy drugs against Aβ or tau, which would aim to galvanise the immune system to clear these neuropathological hallmarks in the CNS, although their root cause is still undetermined (Herrmann *et al.*, 2013). Clinical trials have been unsuccessful so far as lost neural networks cannot recuperate merely by eliminating pathology, though more promise has been shown in early treatment of mild cases, leading to a shift in preventative treatment (Hane, Lee and Leonenko, 2017).

1.1.4 Types of AD

The elderly are most at risk of developing the disease in the form of late-onset AD (LOAD), which affects 13% of people 65 years and older, and its incidence is generally agreed to expand exponentially with age (Naj, A.C. *et al.*, 2014). LOAD tends to be sporadic, with no defined aetiology. A small proportion of AD cases are familial (1-6%) (Bekris *et al.*, 2010), and are caused by autosomal dominant mutations in *APP*, Presenilin (*PSEN*) 1 or *PSEN2*, which encode key components of the γ-secretase enzyme, which cleaves APP. Familial AD tends to occur in patients below 65 years of age, which are referred to as early-onset AD (EOAD) (Bekris *et al.*, 2010).

1.1.5 A role for genetic mechanisms in AD

The molecular mechanisms underlying sporadic AD aetiology are poorly defined. Quantitative genetic analyses, such as twin-studies have demonstrated high heritability estimates (58%-79%) for AD in monozygotic (MZ) twins and higher concordance in MZ than in dizygotic (DZ) twins (Gatz *et al.*, 2006), which prompted initial approaches to AD aetiology to focus on uncovering a genetic contribution to disease susceptibility.

Genome-wide association studies (GWAS) have been critical in identifying genetic variation linked with disease, with many common-sequence variations in risk genes being classified. A recent meta-analysis of AD GWAS has highlighted a number of susceptibility loci that reach genome-wide significance (**Table 1.1**). Many of these play important roles in neural repair, immunological and metabolic pathways, and have also been implicated in the development of other neurodegenerative diseases (Lambert *et al.*, 2013). Currently, the most significant and well-established LOAD-risk GWAS loci is the polymorphism of the apolipoprotein E gene (*APOE*) (Foraker *et al.*, 2015), whose product is a glycoprotein that plays a key role in lipid metabolism as a transporter of lipoprotein and cholesterol. The gene has three main alleles; *APOE* ε2, ε3, and ε4, of which ε4 is the most well-established genetic risk factor for development of LOAD, with a frequency of ~40% in patients (Liu *et al.*, 2013). Individuals with one APOE ε4 allele have a high odds ratio for developing AD (2.6-3.2), which is further boosted to 14.9 by a second ε4 allele (Liu *et al.*, 2013). Interestingly, *APOE* ε2 is known to be protective, with individuals with two *APOE* ε2 alleles having an OR for AD of only 0.6 (Liu *et al.*, 2013). Although the three isoforms of the allele differ by only one or two amino acids (**Table 1.2**), possession of the ε4 allele has been linked to decreased amyloid fragment clearance in neurons, and subsequently a greater likelihood of plaque formation, along with increased probability of cerebral amyloid angiopathy (CAA) and age-related cognitive decline during normal ageing. The mechanisms behind amplified disease risk with the *APOE* ε4 protein isoform are currently poorly understood, and cannot be identified by DNA sequence analysis alone. New hypotheses suggest its role in AD pathology may be related to its modulation of amyloid precursor protein (APP), such as catalysing its cleavage by β- and γ- secretase and promoting its transcription (Huang *et al.*, 2017). Although *APOE* is the most well-established common genetic variant for AD, genetic analyses cannot completely account for the problem of 'missing heritability' in disease. Collectively all common variants only account for about a third of disease incidence (Ridge *et al.*, 2016). Recent studies have used whole genome or exome sequencing to identify rare variants with a larger effect size, and such studies have nominated single nucleotide polymorphisms (SNPs) in gene such as TREM2 (Jonsson *et al.*, 2013) (Guerreiro *et al.*, 2013) and PLD3 (Blanco-Luquin *et al.*, 2018). One particular issue with the SNPs that have been associated with AD to date is that for the majority of genes no clear mechanisms behind their actions has been established and mutations are often found in non-coding regions of the genome (Giri, Zhang and Lü, 2016) (Tansey, Cameron and Hill, 2018).

Table 1.1 - Summary table of SNPs associated with AD, which reached genome wide significance in a recent meta-analysis

(Taken from Lambert et al., 2013). Abbreviations: Chr. = Chromosome

SNP	Chr:Position	Closest Gene	Odds Ratio
rs6656401	1:207692049	CR1	1.18
rs6733839	2:127892810	BIN1	1.22
rs10948363	6:47487762	CD2AP	1.1
rs11771145	7:143110762	EPHA1	0.9
rs9331896	8:27467686	CLU	0.86
rs983392	11:59923508	MS4A6A	0.9
rs10792832	11:85867875	PICALM	0.87
rs4147929	19:1063443	ABCA7	1.15
rs3865444	19:51727962	CD33	0.94
rs9271192	6:32578530	HLA-DRB5– HLA-DRB1	1.11
rs28834970	8:27195121	PTK2B	1.1
rs11218343	11:121435587	SORL1	0.77
rs10498633	14:92926952	SLC24A4 / RIN3	0.91
rs8093731	18:29088958	DSG2	0.73
rs35349669	2:234068476	INPP5D	1.08
rs190982	5:88223420	MEF2C	0.93
rs2718058	7:37841534	NME8	0.93
rs1476679	7:100004446	ZCWPW1	0.91
rs10838725	11:47557871	CELF1	1.08
rs17125944	14:53400629	FERMT2	1.14
rs7274581	20:55018260	CASS4	0.88

Table 1.2 - Single nucleotide polymorphisms (SNPs) that give rise to specific APOE alleles (Taken from Foraker et al., 2015)

	rs429358	rs7412	
ε2	GTGC	GTGC	-1 CpG
ε3	GTGC	GCGC	
ε4	GCGC	GCGC	+ 1 CpG

1.2 Epigenetics

Epigenetics refers to the reversible, mitotically and meiotically heritable regulation of gene expression, independent of DNA sequence variation (Lord, Lu and Cruchaga, 2014). Established epigenetic mechanisms include DNA modifications, modifications of histone proteins of nucleosomes and microRNAs (miRNAs). The most highly characterised and stable epigenetic modification is DNA methylation of the pyrimidine ring of cytosine to form 5-methylcytosine (5mC) in CpG dinucleotides, which are often clustered together in CpG islands (CGI) within the genome. These marks are regulated by DNA methyltransferase enzymes (DNMTs), with DNMT1 responsible for maintenance of patterns and DNMT3A and B responsible for de-novo modification (Deaton and Bird, 2011). They were traditionally thought to be inhibitory of transcription as the addition of the methyl group was thought to prevent the binding of transcription factors and draw methyl-binding proteins to alter chromatin structure, shifting it into a transcriptionally-inactive compact state (heterochromatin). However, recent evidence suggests that although this is the case for CpG rich regions in the promoters of constitutively expressed housekeeping genes (Deaton and Bird, 2011) (Vavouri and Lehner, 2012), DNA methylation can have alternative effects on gene expression. DNA methylation has been shown to have downstream effects on gene splicing; exons have higher levels of DNA methylation than flanking introns, with 22% of the splicing of alternative exons being epigenetically regulated as RNA polymerases bind to methyl groups (Lev Maor, Yearim and Ast, 2015).

The presence of hydroxymethylated cytosines (5hmC), produced via catalysis of 5mC by ten-eleven translocation (TET) proteins was initially thought to be an intermediate marker of demethylation. However, recent research suggests it may be an independent marker with unique transcription regulatory roles of its own (Kroeze, van der Reijden and Jansen, 2015). New methodologies such as oxidative bisulfite sequencing (Booth *et al.*, 2013) and reduced representation hydroxymethylation profiling (RRHP) (Ellison, Bradley-Whitman and Lovell, 2017) have facilitated the measurement of 5hmC in isolation.

1.2.1 Methylation Quantitative Trait Loci (mQTL)

Genetic variation such as SNPs at specific loci may influence patterns of DNA methylation across genomic regions, both in *cis* and in *trans* (though primarily the former). They are referred to as methylation quantitative trait loci (mQTLs) and are enriched in regulatory chromatin domains and transcription-factor binding sites, altering gene expression via transcription

enhancement/suppression, and potentially alternative mRNA splicing, through methylation (Gutierrez-Arcelus *et al.*, 2015). As epigenetic associations with gene expression have been found to be more tissue-specific compared to genome sequence variation (Gutierrez-Arcelus *et al.*, 2015) and mQTLs are heavily associated with genetic variants (Hannon *et al.*, 2018), they may be one mechanism by which genetic variants in non-coding regions may exert functional changes by tissue-specific epigenetic mediation. Additionally, the elaborate network of pleiotropic associations between complex traits, DNA methylation sites and gene variants (Hannon *et al.*, 2018) present an intriguing opportunity to dissect complex disease phenotype development and refine GWAS analyses for risk loci. They have also become a subject of interest in neuromedicine, as brain mQTLs show significant overlap with genetic variants associated with gene-expression in the brain (Hannon *et al.*, 2015), with many psychiatric-disease linked SNPs revealed to cause methylation differences dysregulating genes critical for neurological development/function (Ciuculete *et al.*, 2017). They are thus implicated in magnifying risk of developing schizophrenia and other psychiatric disorders (Hannon *et al.*, 2015) (Ciuculete *et al.*, 2017). An example of an mQTL is shown in **Figure 1.1.**

Figure 1.1- An example schematic of an mQTL. In this example individuals homnozygous for allele 1 have less DNA methylation at this specic CpG than individuals who are heterozygous, or homozygous for allele 2.
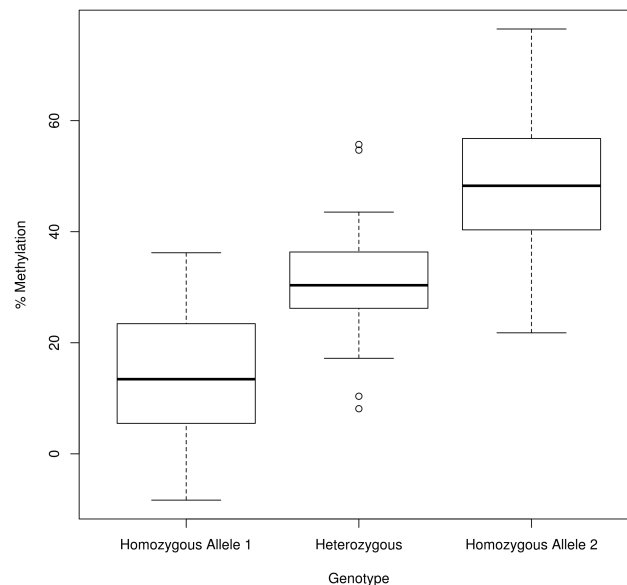
1.2.2   Epigenome-wide association studies (EWAS) of AD

Epigenome-wide association studies (EWAS) are a recent development allowing the profiling of genome-wide DNA methylation patterns in AD (**Table 1.3**). The first EWAS in LOAD was conducted with the Illumina Infinium 27K Methylation BeadChip array (27K array) and identified 948 CpG sites, mapping to 918 unique genes that were associated with disease in the frontal cortex (Bakulski *et al.*, 2012). The most significant loci resided in *TMEM59*. Similar analyses of hippocampal tissue with

varying degrees of Braak-stage pathology demonstrated DNA methylation alterations in four loci distributed between the *DUSP22*, *CLDN15* and *QSCN6* genes (Sanchez-Mut *et al.*, 2014). More recent studies have utilised the Illumina Infinium 450K Methylation BeadChip array (450K array) to determine disease-associated differential DNA methylation at >485,000 CpG sites. Lunnon et al used this technology to assess differential methylation associated with Braak stage across a number of tissues, including cortex (entorhinal cortex (EC), prefrontal cortex (PFC) and superior temporal gyrus (STG)), cerebellum and blood in ~120 individuals (Lunnon *et al.*, 2014). Meanwhile, De Jager and colleagues used this technique to identify differential DNA methylation associated with neuritic plaque burden in PFC in >700 individuals (De Jager *et al.*, 2014). Interestingly, both of these studies highlighted significant differentially methylated positions (DMPs) in the same four CpG sites in the genes *ANK1*, *RPL13*, *RHBDF2* and *CDH23* (Lord, Lu and Cruchaga, 2014). A more recent EWAS has highlighted significant DNA hypermethylation associated with increasing Braak stage across a 48kb region of the *HOXA* gene cluster (Smith *et al.*, 2018). Recently, DNA hydroxymethylation patterns in AD have also been assessed by EWAS, one study of the prefrontal cortex identified 517 plaque-associated and 60 NFT-associated differentially hydroxymethylated regions (Zhao *et al.*, 2017). Another recent novel development has been EWAS of DNA methylation alterations in AD in isolated neuronal and non-neuronal nuclei, allowing the identification of cell type-specific methylation patterns (Gasparoni *et al.*, 2018).

1.2.3   Epigenetic differences in APOE in AD

The *APOE* genomic region contains a single well-defined CGI spanning 880 base pairs (bp) in length with 90 CpG sites, overlapping with the last exon of the gene, spanning from the 3' end of intron 3 and and ending in the 3' untranslated region of exon 4 (**Figure 1.2**). This area contains the two SNPs that define *APOE* allele type, rs429358 and rs7412 (**Table 1.2**). Another SNP, rs405509, is located in the promoter region, and has been suggested to increase DNA methylation in light of its inhibitory effect on gene transcription, with these findings supported by higher methylation levels for carriers of the rs405509 A and ε4 alleles (Ma *et al.*, 2015). Differential methylation of the *APOE* CGI has been observed in AD-affected brain tissue (Foraker *et al.*, 2015), where it may play a crucial role in pathology. Interestingly, this does not seem to be restricted to AD, and a recent study has shown altered DNA methylation in *APOE* in Dementia with lewy bodies (DLB) (Tulloch, Leong, Chen, *et al.*, 2018). As previously discussed, Smith et *al*, recently published an EWAS that identified significant Braak-associated DNA hypermethylation of the *HOXA* gene cluster in PFC (Smith *et al.*, 2018). Interestingly, the authors also profiled the STG in the same individuals, and noted two CpG sites (cg05501958 and cg21879725) in the *APOE* gene that showed significant Braak-associated hypermethylation of the *APOE* gene (Lunnon, Personal communication). These two loci reside close to the SNPs that determine *APOE* genotype.

Given that *APOE* currently represents the genetic variant associated with sporadic AD with the greatest disease risk, we hypothesise that altered DNA methylation in AD may be one mechanism by

which *APOE* genotype exerts a biological effect. There are 13 probes on the 450K array that map to the *APOE* gene. To address my hypothesis I will investigate an association of *APOE* DNA methylation at these probes with Braak stage or disease diagnosis in all available AD EWAS. Subsequently, I will perform a mQTL analysis to examine the data with respect to *APOE* genotype and to determine whether any disease-associated differences may be caused by genetic differences between individuals with AD and controls. Finally, I will access matched GWAS and EWAS data collected from ~1000 individuals across the life span (from foetal development to centenarians). This will allow me to investigate the mQTL in individuals with *APOE* ε4 alleles and no AD, and also to investigate whether there is any association with the mQTL and advancing age.

This project is of particular interest as it will be the first large-scale study to investigate the mechanisms by which *APOE* genotype may affect gene regulation. Furthermore, given that epigenetic differences are potentially reversible, it could potentially identify novel targets for future pharmacological intervention.

Figure 1.2 - Gene structure of APOE

(Foraker et al., 2015)
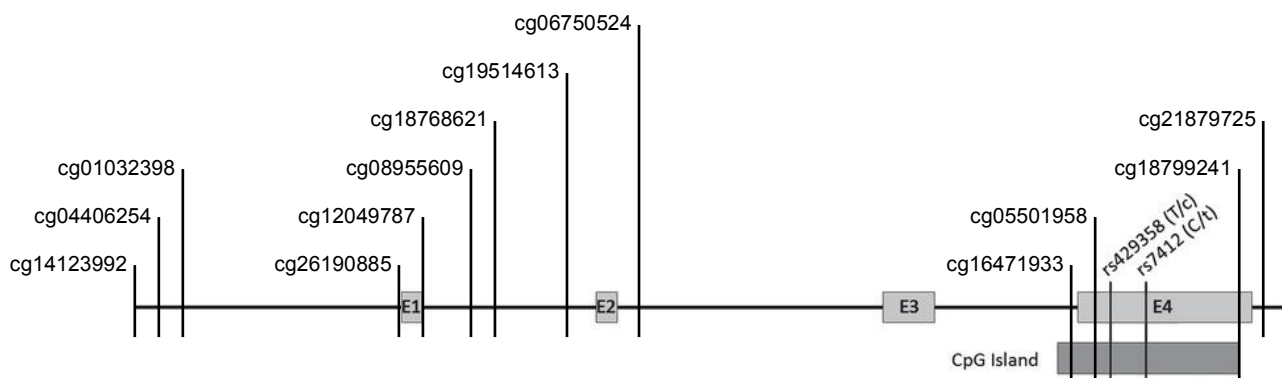
Aims

1. To analyse DNA methylation in the *APOE* gene in all available DNA methylomic studies of AD

2. To determine whether DNA methylation differences in *APOE* in AD are due to a mQTL in the gene

3. To investigate whether *APOE* DNA methylation differences due to the mQTL are stable, or altered over the life course

CHAPTER 2 : MATERIALS AND METHODS

2.1 Available EWAS datasets

2.1.1 AD EWAS datasets

Previously generated DNA methylation data was obtained from four independent EWAS datasets of AD, which all used the 450K array. Although for some cohorts data was available from multiple tissues, we focused our analyses on data from the PFC or temporal gyrus (TG) (STG or middle temporal gyrus (MTG)), as either one or both of these regions was available for all four datasets. Our cohorts were as follows:

1. The "London" cohort consisted of EWAS data from 117 individuals generated by Lunnon et al (Lunnon *et al.*, 2014). This included data generated in both the PFC and STG from individuals archived in the MRC London Neurodegenerative Disease brain bank and is available under gene expression omnibus (GEO) accession number GSE59685.

2. The "Mount Sinai" cohort consisted of EWAS data from 147 individuals generated by Smith et al (Smith *et al.*, 2018). This included data generated in both the PFC and STG from the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank and is available under GEO accession GSE80970.

3. The "ROS/MAP" cohort consisted of EWAS data from 740 individuals by De Jager et al (De Jager *et al.*, 2014). This was generated in the PFC from the Religious Orders Study and the Rush Memory and Aging Project and was provided direct from the authors.

4. The "Arizona" cohort consisted of EWAS data from 404 individuals generated by Professor Paul Coleman (Personal Communication). This included data generated in the MTG from individuals archived in the Banner Sun Health Research Institute Brain and Tissue Bank. This data is currently unpublished data and was provided to us as part of an ongoing collaboration.

In total we utilized data from 1408 unique individuals, which equated to 996 PFC samples and 665 TG samples, and for 259 donors we had matched data from both the PFC and TG (**Table 2.1**). We used Tri-allelic Exact test for Hardy Weinberg Equilibrium (HWE) to investigate whether *APOE* genotypes were different to the general population. We saw a nominally significant difference to the general population in the Arizona cohort (P = 0.047), but not the London, ROSMAP or Mount Sinai cohorts (P >0.05). Demographics for each individual cohort, and information such as age, sex and Braak stage can be found in **Table 2.2.**

Table 2.1 - AD EWAS sample numbers available for each cohort by tissue.

Source of EWAS data	PFC EWAS data	TG EWAS data	Total independent donors
London (Lunnon *et al.*, 2014)	114	117	117
Mount Sinai (Smith *et al.*, 2018)	142	144	147
ROS/MAP (De Jager *et al.*, 2014)	740	-	740
Arizona (Coleman et al, Unpublished)	-	404	404
TOTAL	996	665	1408

Table 2.2 - Demographics for individual cohorts used in the AD study.

Shown for all four cohorts are the number of control samples (Braak 0-II), and AD samples (Middle stage AD (Braak III-IV) and Late stage AD (Braak V-VI)). Shown for each group are mean age at sampling ± standard deviation (SD), number of males (M) and females (F), Braak stage, Number of APOE ε4 alleles (0,1,2 or unknown [N/A]) and brain regions investigated.

Cohort	London			Mount Sinai			ROS/MAP			Arizona		
Disease stage	Control	Middle Stage	Late Stage	Control	Middle Stage	Late Stage	Control	Middle Stage	Late Stage	Control	Middle Stage	Late Stage
Number of samples	29	18	66	60	43	44	146	413	161	81	144	176
Age at death (±SD)	77.6 (12.80)	88.5 (5.20)	85.4 (8.13)	82.0 (7.56)	82.7 (6.55)	88.0 (7.53)	83.0 (6.08)	86.9 (4.09)	87.9 (3.45)	80.8 (7.64)	86.8 (6.39)	82.6 (8.29)
Sex (M/F)	13/16	7/11	26/40	32/28	13/30	12/32	73/73	146/267	45/116	54/27	72/72	76/100
Braak stage	0-II	III-IV	V-VI	0-II	III-IV	V-VI	0-II	III-IV	V-VI	0-II	III-IV	V-VI
No APOE ε4 alleles (0/1/2/NA)	19/6/1/3	8/7/0/3	13/36/11/6	44/10/0/6	25/11/0/7	21/18/5/0	121/25/0/0	323/84/6/0	86/72/3/0	64/14/2/1	98/38/3/5	68/83/22/3
Brain regions investigated	PFC, TG			PFC, TG			PFC			TG		

2.1.2 Life course ("Big Brain") EWAS dataset

The Complex Disease Epigenomics Group at the University of Exeter has generated EWAS data on the 450K array for individuals across the life course from their ongoing studies into DNA methylation in a range of complex disorders, including autism spectrum disorder (ASD), schizophrenia, depression, MCI and age-matched controls. In addition, they have also generated 450K data on foetal brain samples (Spiers *et al.*, 2015). For the purpose of the current study we utilised EWAS data from PFC for individuals in these various studies, and combined it with the foetal brain data and data from the AD study, to allow us to look at differences in methylation with age and in each cohort with genotype. Together this gave me EWAS data from 136 individuals (***Table 2.3***).

Table 2.3 - Demographics for individual cohorts used in the life-course study.

Shown are the number of control samples, number of males (M) and females (F), mean age ± standard deviation (SD), number of APOE ε4 alleles (0,1,2 or unknown [N/A]). *Please note for age, this is shown as days post conception for foetal brain EWAS, and in years for the other EWAS. Abbreviations: SCZ (schizophrenia), DEP (Depression), MCI (Mild cognitive impairment), SD (Standard deviation).

	Foetal Brain	ASD cohort		Schizophrenia cohort		Depression cohort		MCI cohort
		Control	ASD	Control	SCZ	Control	DEP	
No. Samples	136	24	21	45	39	20	20	618
Sex (M/F)	77/59	18/6	18/3	35/10	26/13	16/4	15/5	237/881
Age* (±SD)	92.1 (24.66)	45.1 (22.5)	25.5 (19.5)	56.1 (19.7)	54.6 (18.1)	39.4 (19.5)	48.6 (20.8)	87.0 (4.96)
No. *APOE* ε4 alleles (0/1/2/NA)	85/43/3/0	18/6/0/0	13/7/1/0	31/12/2/0	29/8/0/0	18/2/0/0	18/2/0/0	442/136/9/31

2.2 DNA Methylomic profiling

2.2.1 450K array

The Illumina Infinium Human Methylation BeadChip array is a high-throughput, cost-effective method of quantifying genome-wide levels of DNA methylation at specific genomic loci. The 450K array is capable of assaying 485,764 CpG sites with a combination of the type I probes used in its predecessor (27K array) and a new type II probe. All EWAS datasets described in section 2.1 were generated on the 450K array. The EWAS data was all generated prior to the start of this project. However, the methods for generating this data were the same across all studies and were as follows. First, DNA was extracted from brain tissue, and the quantity and quality assessed using the Nanodrop Spectophotometer and gel electrophoresis, respectively. Only high molecular weight (Mw) DNA was then taken forward to the 450K arrays. Second, 500ng DNA was bisulfite treated using the Zymo EZ-96 DNA Methylation-Gold TM Kit (Cambridge Bioscience, cat no.: D5007) and eluted in 20µl buffer. This process converts unmodified cytosines to uracils whilst modified cytosines remains as cytosines. The sequence can then be amplified via subsequent PCR, resulting in all uracil residues being amplified as thymine and only modified cytosine residues being amplified as cytosine. Quality control (QC) checks were then performed using PCR to check bisulfite conversion. Third, all samples were processed using 450K array (Illumina Inc, CA, USA) according to the manufacturer's instructions, with minor amendments and quantified using an Illumina HiScan System (Illumina, CA, USA). In this process the denatured DNA undergoes whole-genome amplification to uniformly increase the amount of DNA by several orders of magnitude while preventing amplification bias. The samples are then fragmented and deposited on the BeadChip arrays and incubated in an Illumina Hybridisation oven to stimulate hybridising of probes to the chip, which is then washed to remove any nonspecifically and unsuccessfully hybridised DNA. The oligonucleotides are then extended and stained in capillary flow-through chambers to apply fluorophores. These are excited by the laser of the Illumina HiScan system(Illumina, San Diego, CA, USA), which then records high resolution images of the fluorescent intensity.

2.2.2 450K array normalisation and QC

Signal intensities were imported into R using the methylumi package (Davis and Bilke, 2018) as a methylumi object. The DNA modification level at each studied CpG site is determined by calculating the ratio of fluorescent intensity for modified and unmodified probes, producing a beta value for each site. The range of values spans from 0 (i.e. no modified cytosines at the site) to 1, (i.e. all cytosines are modified at that site), and can be interpreted as a percentage. Initial QC checks that were undertaken in the methylumi package for the various EWAS datasets included multi-dimensional scaling of both X and Y chromosomes to assess concordance between sex as recorded on post-mortem records and predicted sex using 65 control SNPs probes on the 450K array. The 65 SNPs were also used to confirm that matched PFC and TG data were sourced from the same individual.

One of the most important steps of analysing 450K array data when studying complex polygenic disorders is quantile normalisation (QN), in order to ensure optimal sensitivity when detecting differential DNA methylation due to technological shortcomings such as differences in accuracy and reproducibility between Infinium II and I probes within the 450K array. Beta values undergo this process to replace intensity scores with the mean of features from a single rank from each array, generating identical array-wide data distributions. All QN and data QC steps were performed using R 3.4.3 (R Core Team 2018). This was carried out using the dasen function in the R package wateRmelon (Pidsley *et al.*, 2016). Within the package the pfilter function is used to remove samples where 5% of sites had a detection p-value > 0.05. Probes with common (minor allele frequency (MAF) > 5%) SNPs in the CpG or single base extension position, or probes that were nonspecific or miss-mapped (Chen *et al.*, 2013) were removed from downstream analyses. For the purpose of the current project normalised beta values for 13 probes that mapped to the *APOE* gene (***Table 2.4*, Figure 2.1**) were taken forward for statistical analysis. For the Life course ("Big Brain") EWAS dataset the probe cg18768621 did not pass QC and therefore we only examined the remaining 12 CpG sites in this dataset.

Table 2.4 - A list of all CG probes mapping to the APOE gene on the 450K array. Chromosomal positions (hg19), Illumina annotation (APOE), gene region, and respective CpG islands are shown.

Probe	Position	Illumina Annotation	Region	CpG Island
cg14123992	19:45,407,868	APOE	TSS1500	chr19:45411720-45412600
cg04406254	19:45,407,945	APOE	TSS1500	chr19:45411720-45412600
cg01032398	19:45,408,121	APOE	TSS1500	chr19:45411720-45412600
cg26190885	19:45,409,005	APOE	TSS200	chr19:45411720-45412600
cg12049787	19:45,409,080	APOE;APOE	1stExon;5'UTR	chr19:45411720-45412600
cg08955609	19:45,409,353	APOE	5'UTR	chr19:45411720-45412600
cg18768621	19:45,409,440	APOE	5'UTR	chr19:45411720-45412600
cg19514613	19:45,409,713	APOE	5'UTR	chr19:45411720-45412600
cg06750524	19:45,409,955	APOE	Body	chr19:45411720-45412600
cg16471933	19:45,411,802	APOE	Body	chr19:45411720-45412600
cg05501958	19:45,411,873	APOE	Body	chr19:45411720-45412600
cg18799241	19:45,412,599	APOE	3'UTR	chr19:45411720-45412600
cg21879725	19:45,412,647	APOE	3'UTR	chr19:45411720-45412600

Figure 2.1 - A gene map of APOE showing the position of its main 3' CpG island, along with location of all CpG probes mapping to the APOE gene on the 450K array.
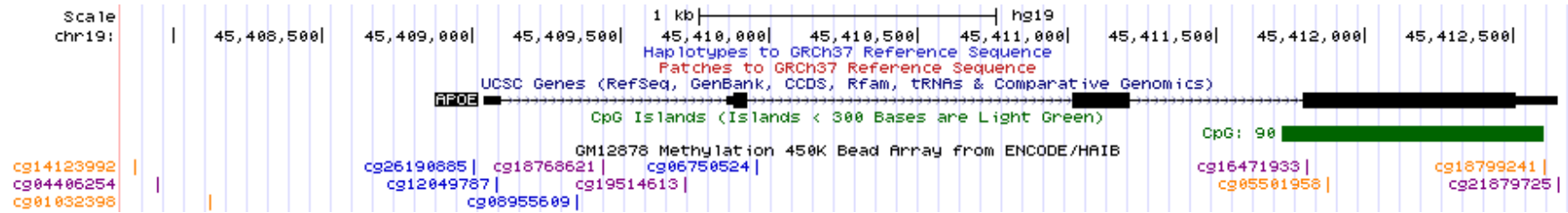
2.3 *APOE* genomic profiling

For the Arizona and ROS/MAP AD cohorts *APOE* genotype data was supplied from the data supplier as they had previously genotyped these samples for rs429358 and rs7412. For the London and Mount Sinai AD cohorts, and all the longitudinal EWAS datasets, *APOE* genotype was determined by imputing this from GWAS data collected in the same samples using Michigan imputation server using Minimac3 (Das *et al.*, 2016). The 1000 Genomes Phase 3 (Version 5) was used as the reference panel for imputation (1000 Genomes Project Consortium, 2015). It has been demonstrated that imputing rs429358 and rs7412 using reference panels is very accurate and generates comparable results to directly genotyped samples (Radmanesh *et al.*, 2014).

2.4 Statistical analyses

All data analysis scripts can be found in Appendix 1.

2.4.1 Analysis of AD EWAS data

All statistical analyses in this project were performed using R 3.4.3. To investigate DNA methylation differences at the 13 CpG sites in AD we performed a linear regression analysis in each AD EWAS cohort and within each tissue separately, looking for an association of DNA methylation with the variables of Braak score or AD (case/control) diagnosis, controlling for the co-variates of age, sex and cell-proportion. Cell proportion was estimated from the genome-wide methylation data for each EWAS cohort from the R package *CETS* (Guintivano, Aryee and Kaminsky, 2013). CpG methylation was expressed as percentages of methylated cells, derived from adjusted beta values. Effect sizes (coefficients) of variables were also calculated from these models, and used to quantify degrees of methylation of CpGs. "Nominal" significance was deemed to be when P <0.05, whilst our Bonferroni significance threshold was deemed to be $P < 3.85E^{-3}$, to control for multiple testing across 13 CpG sites. Bonferroni assumes independence between all tests performed. However, due to the high correlation between DNA methylation at adjacent CpG sites and multiple tissues from the same individuals, the tests are not fully independent. Therefore, by using a Bonferroni significance threshold we are being overly conservative. However, we acknowledge that by performing tests across multiple cohorts and tissues, the significance threshold could have been adjusted for this. A Fisher's combined probability test was conducted across all the AD cohorts within each tissue separately using the "MetaDE.pvalue" function in the package MetaDE (Wang *et al.*, 2012).

2.4.2 Analysis of Lifecourse "Big Brain" EWAS data

To investigate DNA methylation differences at the 13 CpG sites with respect to age in brain samples over the life course we used a linear regression model to look for an association of DNA methylation with age, whilst controlling for the co-variates of sex, cell-proportion, brain region, diagnosis and brain bank. This was performed (a) within the foetal brain cohort and then (b) within all other (postnatal) samples. In this cohort the probe cg18768621 failed QC, and so we only analysed the remaining 12 probes. As such, for this analysis our Bonferroni significance threshold was deemed to be $P < 4.17E^{-3}$, to control for multiple testing across 12 CpG sites.

2.4.3 Analysis of mQTLs

The mQTL analyses used a linear regression analysis to look for an association of DNA methylation (at each probe individually) with the number of *APOE* ε4 alleles (0,1 or 2), whilst controlling for co-variates. In the first mQTL analysis we looked in the PFC and STG separately for an association of DNA methylation with number of *APOE ε4* alleles in the AD meta-analysis cohort, looking for an association (regardless of phenotype) at all 13 CpG sites individually, whilst controlling for the co-variates of sex, cell proportion, brain region, diagnosis and cohort. Positive/negative differences between adjusted beta values of samples were used to represent corresponding effects on methylation by selected variables. We then repeated this analysis but stratifying the data by AD/control status to look to see if this was more apparent in AD or control samples. Next, we looked in the "Big Brain" EWAS data to look for an association of DNA methylation with the number of *APOE ε4* alleles, individually at each of the 12 probes available in this cohort, whilst controlling for the co-variates of sex, cell proportion, brain region, diagnosis and brain bank. We analysed the pre-natal and post-natal cohorts separately.

2.4.4 Power

Concerning Braak stage, we have 95% power to detect a 2.5% methylation difference between high and low Braak stage at a significance level of $P < 3.85E-03$ (Bonferroni), in the London cohort, with 29 samples per group, and 100% power in the Mount Sinai cohort, with 44 samples per group. In the ROSMAP cohort, with 146 samples per group, we have 100% power and in the Arizona cohort, with 81 samples per group, we have 100% power to detect differences of 2.5% between high and low Braak stage at a significance level of $P < 3.85E-03$.

In the London cohort, with 113 samples, we have 80% power to detect r of 0.34 at a significance level of $P < 3.85E-03$ and 90% power to detect r of 0.38 at the same significance threshold. With 146 samples in the Mount Sinai cohort, we have 80% power to detect r of 0.30 at a significance level of $P < 3.85E-03$ and 90% power to detect r of 0.33 to the same significance. For the ROSMAP cohort, with 720 samples, we have 80% power to detect r of 0.14 at a significance level of $P < 3.85E-03$, and 90% power to detect r of 0.15 to the same significance. For the Arizona cohort, with 401 samples, we have 80% power to detect r of 0.18 at a significance level of 3.85E-03 and 90% power to detect r of 0.21 to the same significance.

Regarding APOE genotype, we have 100% power to detect differences of 2.5% between genotype groups at a significance level of $P < 3.85E-03$, in the London cohort, with 12 samples per group (number of samples with two E4 alleles), 99% power in the Mount Sinai cohort, with 5 samples per group (number of samples with two E4 alleles), we have 72% power, and 100% power in the ROSMAP cohort, with 9 samples per group (number of samples with two E4 alleles), and in the ROSMAP cohort, with 27 samples per group (number of samples with two E4 alleles).

CHAPTER 3 : RESULTS

3.1 Loci in *APOE* are differentially methylated in the cortex in donors with Alzheimer's disease

We were interested to investigate whether any of the 450K probes that mapped to the *APOE* gene (N=13 probes) displayed differential DNA methylation in the cortex (PFC or STG) in donors with AD, compared to elderly non-demented controls. To address this we performed individual linear regressions for the 13 *APOE* probes in all four AD EWAS dataset cohorts (see **Table 2.1**), in each individual tissue. Our models looked for an association of disease diagnosis with DNA methylation whilst controlling for the co-variates of age, sex and neuronal cell proportion. We used a nominal significance level of $P<0.05$ and a Bonferroni significance level of $P<3.85E-03$ (*i.e.* 0.05/13 tests). To identify probes showing a Bonferroni-significant difference in methylation in AD across all cohorts we conducted a Fisher's meta-analysis to identify consistent effects

When we examined the individual cohorts we identified a number of loci that were nominally significantly hypomethylated methylated in AD in the PFC (**Table 3.1**), but with only one loci exceeding Bonferroni significance in only one cohort (cg21879725 [London cohort: $P = 6.88E-04$) (**Figure 3.1**). However, we did identify a number of hypomethylated loci in AD in the STG that exceeded Bonferroni significance (cg16471933 [Arizona cohort: $P = 3.79E-06$], cg05501958 [Mount Sinai cohort: $P = 6.23E-05$; Arizona cohort: $P = 1.14E-13$], cg18799241 [Arizona cohort: $P = 3.31E-05$], cg21879725 [Mount Sinai cohort: $P = 4.02E-05$; Arizona cohort: $P = 6.71E-07$]) (**Figure 3.2**).

We were also interested to investigate whether similar patterns of methylation in disease were seen across the cohorts within tissues (regardless of significance). We observed that methylation patterns were not similar in the PFC (**Figure 3.3**), but showed similarity in the STG (**Figure 3.4**)

A meta-analysis of all three PFC cohorts revealed seven nominally significant differentially methylated loci (**Figure 3.5**; **Figure 3.6**; **Figure 3.8**), with two loci that exceeded Bonferroni significance (cg05501958: $P = 3.64E-03$; cg21879725: $P = 1.31E-03$), whilst in the STG we identified five nominally significant differentially methylated loci (**Figure 3.5**; **Figure 3.7**; **Figure 3.8**), with four neighbouring loci reaching Bonferroni significance (cg16471933: $P = 1.05E-05$; cg05501958: $P = 1.67E-15$; cg18799241: $P = 3.26E-05$; cg21879725: $P = 7.85E-09$)

Table 3.1 - Significant diagnosis-associated DMPs in the APOE gene in the prefrontal cortex (PFC) and superior temporal gyrus (STG) of AD subjects and non-demented controls (CTL).

Shown for all of the 13 CG probes in the APOE gene are P values from a Fisher's meta-analysis, percentage of DNA methylation, and associated P value derived from a linear regression analysis adjusting for the covariates of age, gender and cell proportion, in each individual tissue in each individual cohort for CTL and AD subjects. All p-values reaching Bonferroni significance (<3.85E-03) are highlighted in bold.

| | Fisher's test | | London cohort | | | | | | Mount Sinai cohort | | | | | | Arizona cohort | | | ROSMAP cohort | | |
| | | | STG | | | PFC | | | STG | | | PFC | | | STG | | | PFC | | |
Probe	STG p value	PFC p value	p value	CTL % met	AD % met	p value	CTL % met	AD % met	p value	CTL % met	AD % met	p value	CTL % met	AD % met	p value	CTL % met	AD % met	p value	CTL % met	AD % met
cg14123992	5.71E-02	1.63E-02	0.27	54.30	55.87	4.51E-02	66.24	68.21	0.36	63.68	64.30	0.11	54.44	55.65	2.26E-02	59.45	60.24	0.39	56.62	56.32
cg04406254	0.26	5.23E-02	0.14	57.36	59.07	7.47E-03	72.75	75.11	0.66	66.34	66.62	0.64	63.06	63.35	0.24	65.83	66.23	0.38	58.36	58.08
cg01032398	0.64	0.31	0.32	83.06	83.84	0.35	86.36	86.78	0.41	80.34	80.62	0.35	81.49	81.18	0.91	84.49	84.51	0.27	82.98	83.23
cg26190885	0.48	0.11	0.26	18.10	17.59	8.60E-02	17.02	16.36	0.57	16.38	16.53	0.10	17.59	17.13	0.44	22.18	22.36	0.54	80.47	81.36
cg12049787	0.98	0.76	1.00	11.32	11.33	0.90	10.86	10.91	0.60	10.69	10.87	0.91	15.72	15.75	1.00	14.70	14.70	0.22	16.17	15.74
cg08955609	0.80	0.86	0.43	6.13	6.24	0.36	6.26	6.04	0.64	7.28	7.18	0.77	7.43	7.48	0.77	6.38	6.41	0.49	9.16	9.07
cg18768621	0.22	0.33	0.14	11.71	12.25	0.20	13.85	13.17	0.13	12.72	13.24	0.63	14.94	15.10	0.91	14.51	14.54	0.38	12.68	12.45
cg19514613	0.18	0.16	0.87	13.65	13.68	6.91E-02	15.72	14.77	0.27	11.48	11.87	0.46	15.12	15.39	4.87E-02	16.08	16.65	0.58	21.01	21.18
cg06750524	0.39	2.77E-02	0.91	42.38	42.65	2.45E-02	42.50	40.67	0.97	40.00	40.02	0.16	39.88	40.62	4.82E-02	45.86	46.43	0.58	54.47	54.74
cg16471933	**1.05E-05**	9.14E-02	0.63	76.60	76.49	5.88E-02	75.29	76.95	2.85E-02	74.00	72.73	**4.32E-03**	70.44	68.67	**3.79E-06**	80.49	79.16	0.10	71.63	72.70
cg05501958	**1.67E-15**	**3.64E-03**	0.26	82.98	82.42	0.32	86.28	87.23	**6.23E-05**	80.68	78.42	0.20	80.89	80.14	**1.14E-13**	85.44	83.16	3.00E-02	71.56	72.68
cg18799241	**3.26E-05**	**3.39E-03**	0.39	84.05	83.45	7.73E-02	82.88	84.14	1.92E-02	82.25	81.22	0.30	79.72	79.27	**3.31E-05**	86.76	85.86	1.12E-02	83.11	84.19
cg21879725	**7.85E-09**	**1.31E-03**	0.90	83.95	84.59	**6.88E-04**	83.23	86.64	**4.02E-05**	76.99	74.78	0.26	77.76	77.04	**6.71E-07**	86.98	85.56	2.28E-02	28.82	29.00

Figure 3.1 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and disease diagnosis in the PFC in three cohorts.

In our linear regression analyses, cg16471933 in the Mount Sinai cohort, and cg21879725 in the Mount Sinai and ROSMAP cohorts exceeded the Bonferroni significance threshold (black line: P < 3.85E-03). The three PFC cohorts are denoted as different shapes. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
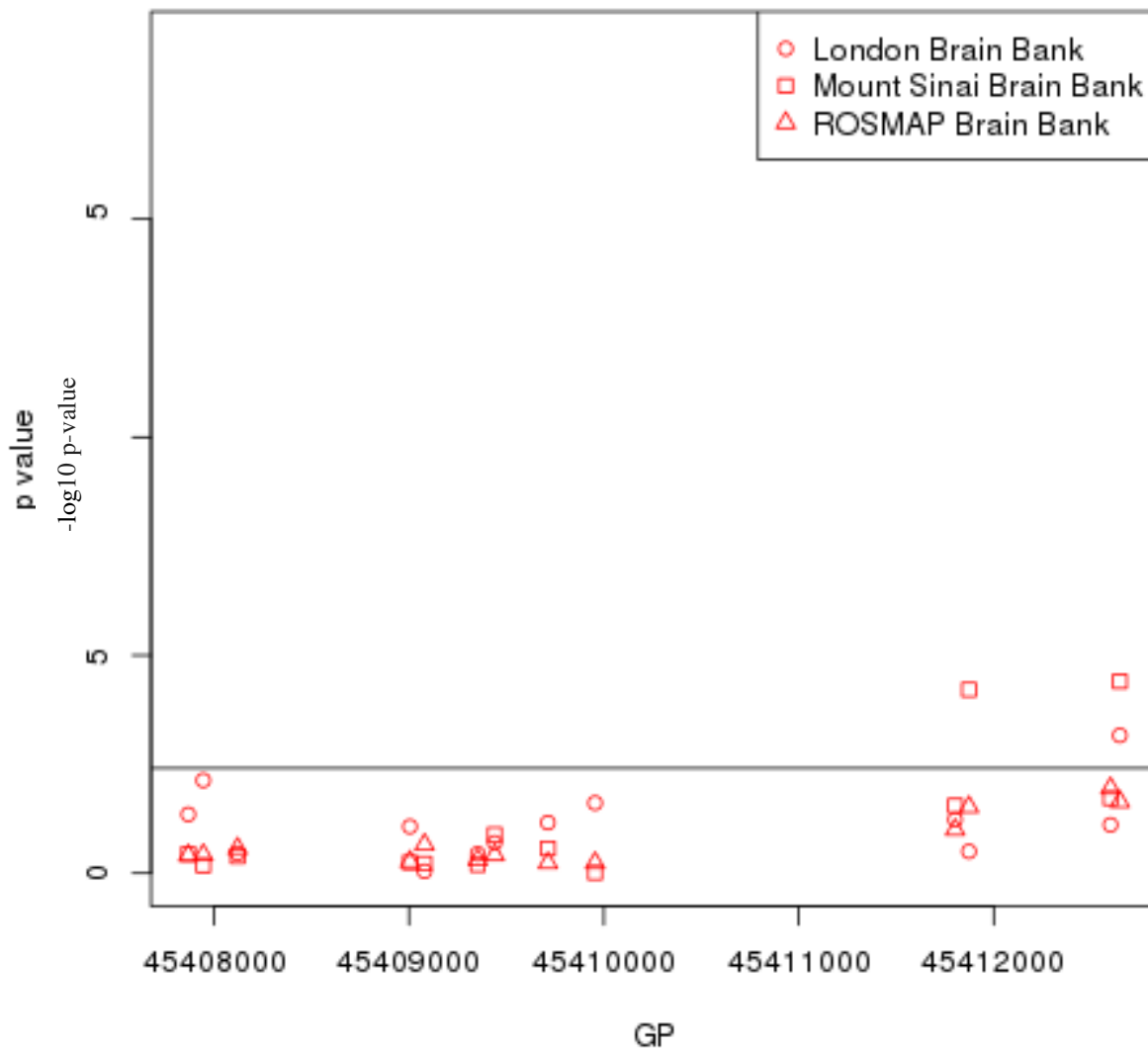
Figure 3.2 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and diagnosis in the STG in three cohorts.

In our linear regression analyses, four CpG sites (cg16471933, cg05501958, cg18799241, cg21879725) passed our Bonferroni significance threshold (black line: P < 3.85E-03). The three STG cohorts are denoted as different shapes. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
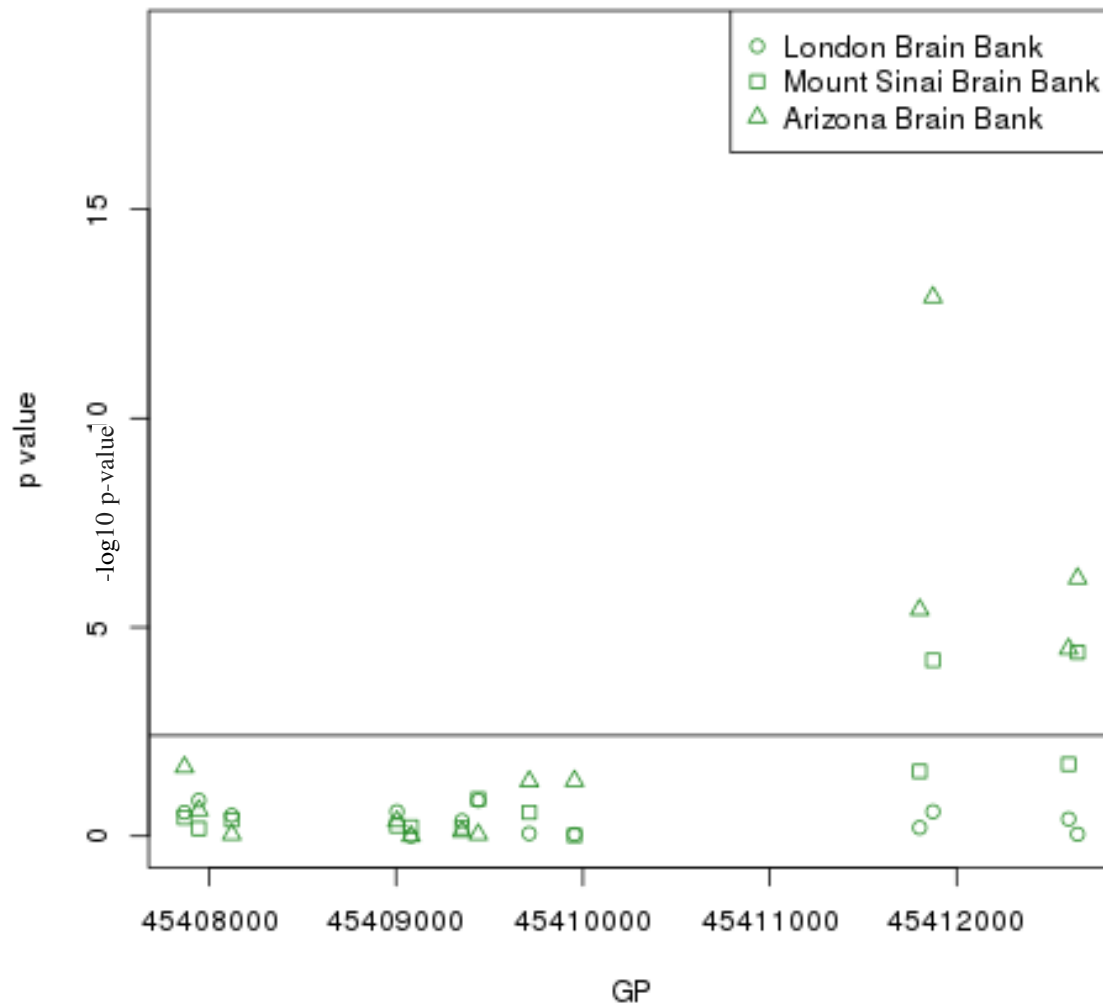
Figure 3.3 - A Graph showing effect-size coefficient (methylation difference between control and AD samples) across the APOE gene in all three cohorts in the PFC samples.

Adjusted beta value difference between AD and control are represented as methylation effect size. The different cohorts are denoted by different line colours.
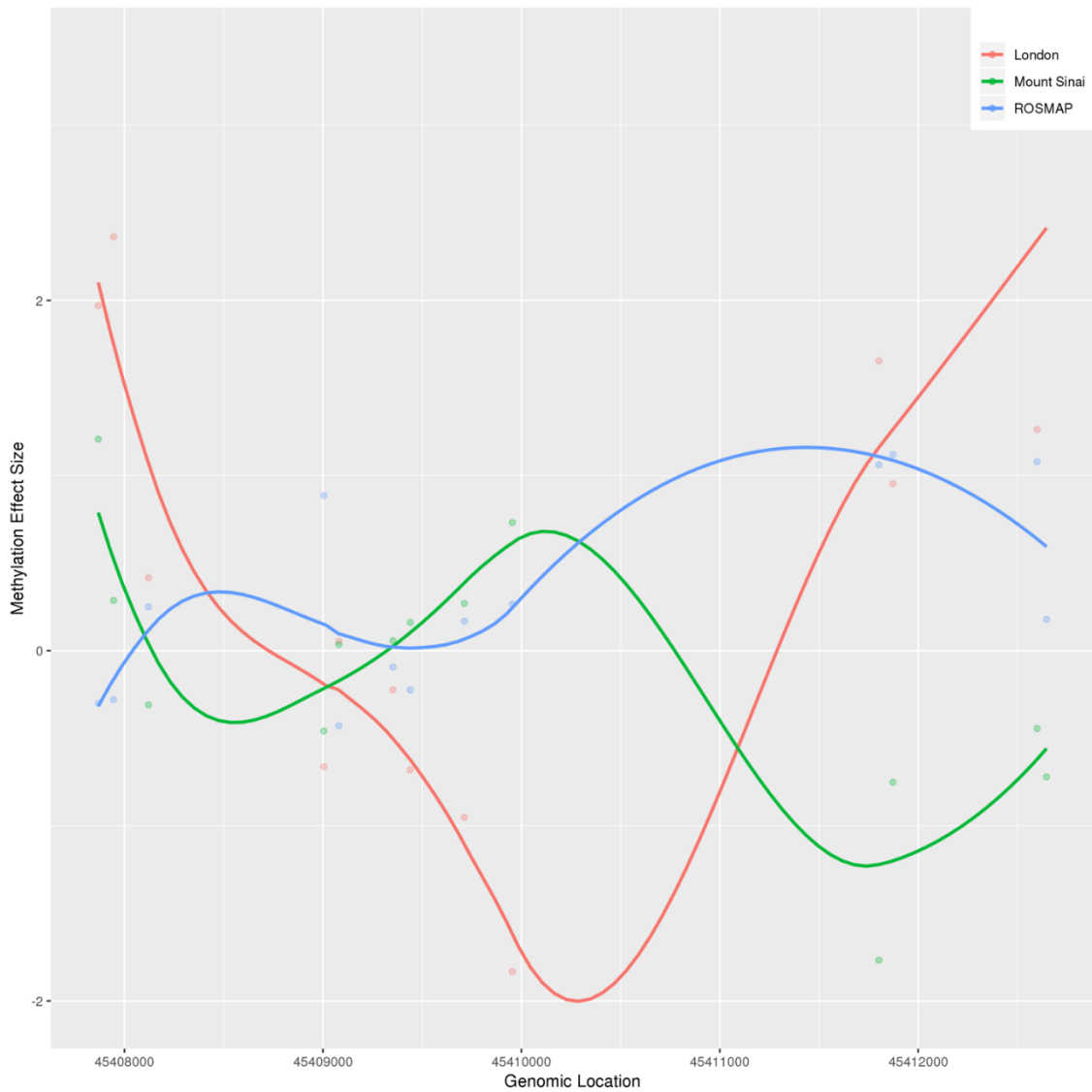
Figure 3.4 - A Graph showing effect-size coefficient (methylation difference between control and AD samples) across the APOE gene in all three cohorts in the STG samples.

Adjusted beta value difference between AD and control are represented as methylation effect size. The different cohorts are denoted by different line colours.
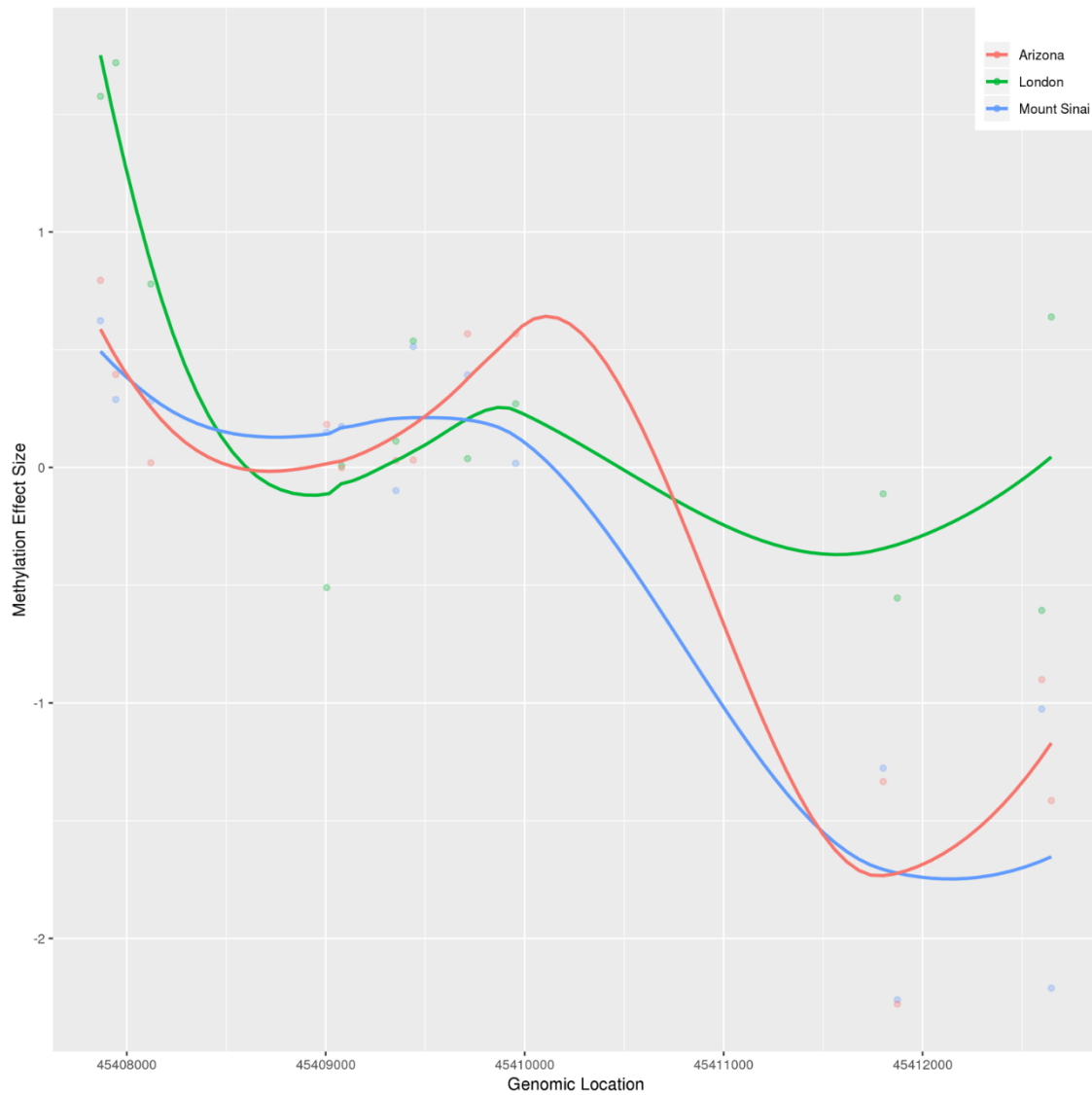
Figure 3.5 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and diagnosis in the STG and PFC in our meta-analysis.

We identified four CpG sites (cg16471933, cg05501958, cg18799241, cg21879725) that passed the Bonferroni significance threshold (black line: P < 3.85E-03) in the STG, and one (cg21879725) in the PFC. The STG is shown in green and the PFC is shown in red. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
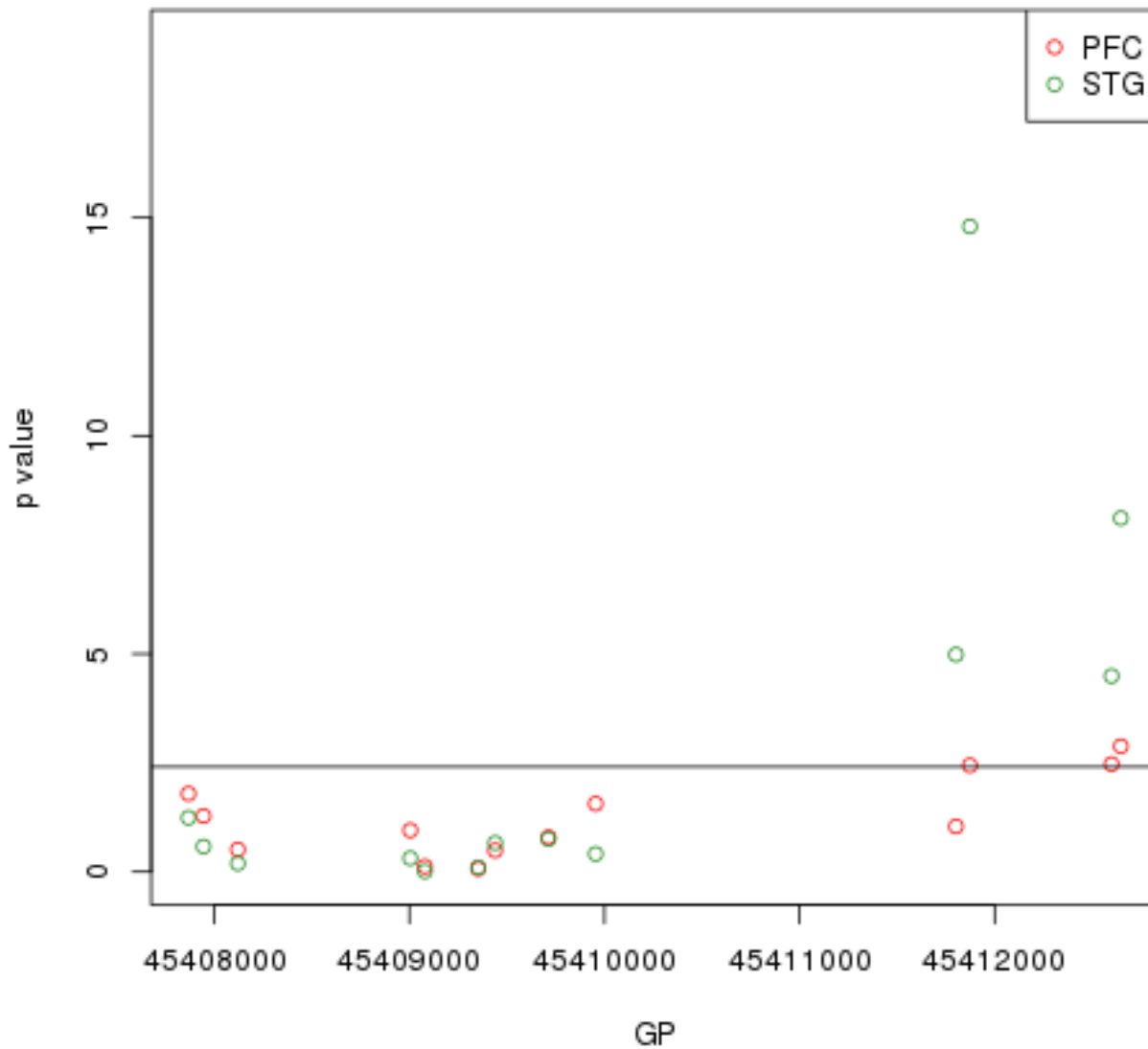
Figure 3.6 - A box and whisker plot demonstrating average corrected DNA methylation levels for the 13 CpGs in APOE in AD (red) and CTL (pink) samples in the PFC.

All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P<0.05), *** = Bonferroni significance (P<3.85E-03)
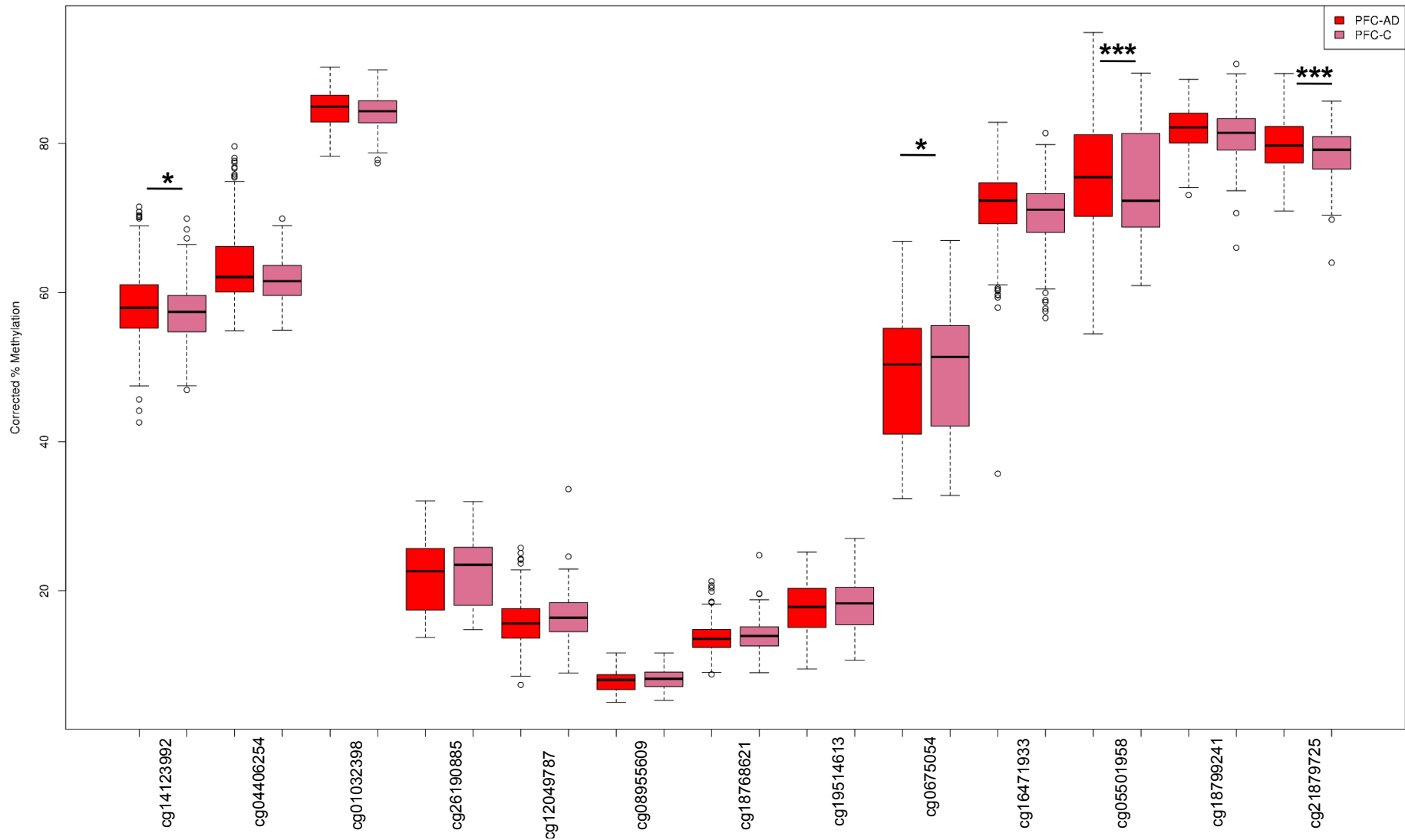
Figure 3.7 - A box and whisker plot demonstrating average corrected DNA methylation levels for the 13 CpGs in APOE in AD (green) and CTL (light green) samples in the STG.

All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P<0.05), *** = Bonferroni significance (P<3.85E-03).
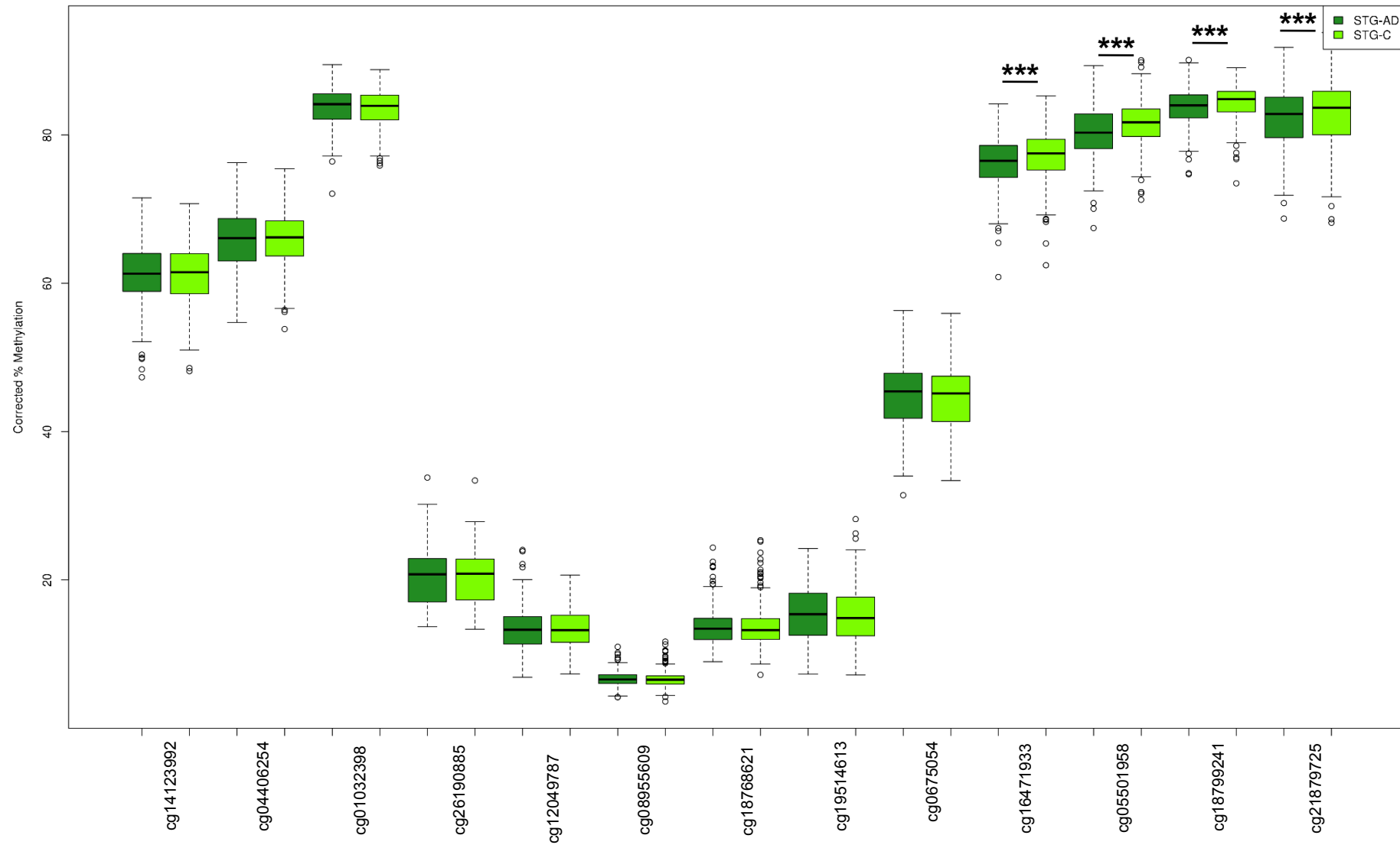
*All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P<0.05_, *** = Bonferroni significance (P<3.85E-03)*
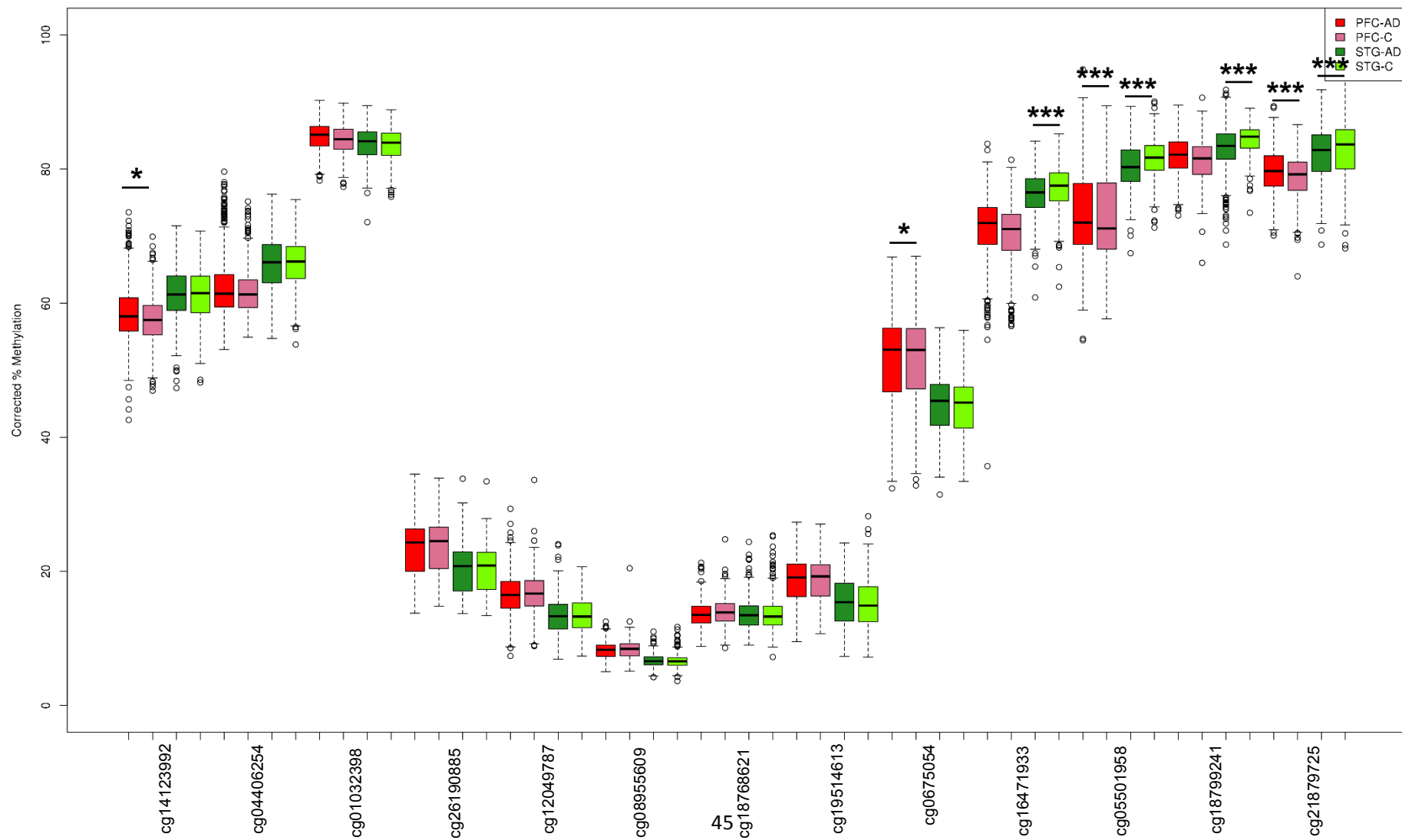
3.2 Loci in APOE are differentially methylated with respect to Alzheimer's disease neuropathology

As we had identified disease-associated differentially methylated loci in *APOE* we were also interested to investigate whether we could identify any loci that were differentially methylated with respect to disease neuropathology measures. To address this we again performed individual linear regressions for the 13 *APOE* probes in all four AD EWAS dataset cohorts in each individual tissue, but looking for an association of Braak stage (as a measure of tau neuropathology) with DNA methylation whilst controlling for age, sex and neuronal cell proportion. We then performed a Fisher's meta-analysis to identify consistent effects across cohorts.

When we examined the individual cohorts we identified a number of hypomethylated loci with respect to Braak stage that reached nominal significance in the PFC and STG (**Table 3.2**). However, only one loci reached Bonferroni significance in the PFC (cg18799241 [ROS/MAP cohort: P = 3.08E-03]) (**Figure 3.9**), whilst in the STG four probes reached Bonferroni significance (cg16471933 [Arizona cohort: P = 1.76E-06], cg05501958 [Mount Sinai cohort: P = 7.33E-05; Arizona cohort: P = 1.36E-12], cg18799241 [Arizona cohort: P = 1.17E-03], cg21879725 [Mount Sinai cohort: P = 3.58E-05; Arizona cohort: P = 2.97E-06]) (**Figure 3.10**). Interestingly, these four probes were the four closest to the 5' end of the gene. When we looked at the pattern of methylation difference between Braak 0 and Braak VI samples (i.e. effect size), we saw little similarity between cohorts in the PFC (**Figure 3.11**), but a very similar pattern between cohorts in the STG (**Figure 3.12**).

The meta-analysis across all three PFC cohorts highlighted six nominally significant loci (**Figure 3.13**; **Figure 3.14**; **Figure 3.16**), with two Bonferroni significant loci (cg18799241: P = 2.56E-03; cg21879725: P = 3.15E-04), both at the 5' end of the gene. In the meta-analysis of the STG we identified five nominally significant loci (**Figure 3.13**; **Figure 3.15**; **Figure 3.16**), with four of these loci being Bonferroni significant (cg16471933: P = 7.01E-07; cg05501958: P = 4.20E-14; cg18799241: P = 1.57E-03; cg21879725: P = 2.46E-08). A Forest plot highlighting the DNA methylation difference in disease in each cohort and tissue for these four loci can be found in **Figure 3.17**.

Together, these results may indicate multi-regional differential methylation of the *APOE* gene in AD, particularly in the STG, which correlated with greater degrees of neuropathology and diagnosis.

Table 3.2 - Significant Braak-associated DMPs in the APOE gene in the prefrontal cortex (PFC) and superior temporal gyrus (STG).

Shown for all of the 13 CG probes in the APOE gene are P values from a Fisher's meta-analysis, percentage of DNA methylation in Braak 0 and Braak VI individuals, and associated P value derived from a linear regression analysis adjusting for the covariates of age, gender and cell proportion, in each individual tissue in each individual cohort. All p-values reaching Bonferroni significance (< 3.85E-03) are highlighted in bold.

| | Fisher's Test | | London cohort | | | | | | Mount Sinai cohort | | | | | | Arizona cohort | | | ROSMAP cohort | | |
| | STG | PFC | STG | | | PFC | | | STG | | | PFC | | | STG | | | PFC | | |
Probe	p value	p value	p value	Braak 0 % met	Braak VI % met	p value	Braak 0 % met	Braak VI % met	p value	Braak 0 % met	Braak VI % met	p value	Braak 0 % met	Braak VI % met	p value	Braak 0 % met	Braak VI % met	p value	Braak 0 % met	Braak VI % met
cg14123992	7.22E-03	4.92E-02	0.15	57.35	59.58	0.01	64.76	67.42	0.02	64.97	67.55	0.16	54.63	56.35	0.06	60.16	61.41	0.91	57.76	57.82
cg04406254	0.10	0.10	0.04	61.46	63.43	5.63E-03	70.46	73.10	0.38	66.79	67.71	0.94	62.98	63.05	0.30	66.30	66.96	0.95	60.62	60.58
cg01032398	0.34	4.11E-02	0.06	85.14	85.97	0.22	86.29	86.85	0.65	80.42	80.67	0.39	81.43	80.97	0.84	84.58	84.51	0.02	85.66	86.47
cg26190885	0.74	0.08	0.18	17.16	16.39	0.04	17.43	16.60	0.98	16.37	16.38	0.28	17.57	17.07	0.96	22.46	22.49	0.37	25.86	26.27
cg12049787	0.89	0.29	0.90	12.19	12.11	0.39	11.18	10.79	0.39	10.91	11.37	0.11	15.38	14.55	0.90	14.65	14.70	0.58	17.24	16.96
cg08955609	0.87	0.87	0.53	6.72	6.83	0.63	6.48	6.36	0.69	7.23	7.09	0.80	7.38	7.30	0.78	6.48	6.43	0.57	8.60	8.47
cg18768621	0.35	0.48	0.47	11.66	11.93	0.17	13.38	12.63	0.15	13.06	13.84	0.61	14.98	15.26	0.51	14.14	14.48	0.61	13.40	13.21
cg19514613	0.31	0.09	0.44	12.66	12.42	0.02	15.92	14.59	0.72	11.52	11.73	0.40	15.21	15.70	0.09	16.46	17.38	0.61	20.18	20.42
cg06750524	0.31	0.12	0.29	42.05	43.06	0.02	43.26	41.08	0.94	39.96	39.91	0.82	39.53	39.34	0.11	46.25	47.13	0.45	55.46	55.98
cg16471933	**7.01E-07**	**4.08E-03**	0.27	76.85	75.24	0.32	74.07	75.08	**7.08E-03**	72.87	70.36	**4.35E-03**	70.01	67.14	**1.76E-06**	79.58	76.97	0.13	71.26	72.69
cg05501958	**4.20E-14**	8.45E-03	0.57	83.32	81.99	0.29	83.06	84.09	**7.33E-05**	79.13	75.53	8.28E-03	80.21	77.70	**1.36E-12**	83.64	79.50	0.03	71.12	72.83
cg18799241	**1.57E-03**	**2.56E-03**	0.67	84.37	83.59	0.03	81.86	83.55	0.03	81.59	80.06	0.44	79.68	79.14	**1.17E-03**	85.72	84.38	**3.08E-03**	83.62	85.37
cg21879725	**2.46E-08**	**3.15E-04**	0.79	85.39	84.97	7.89E-03	80.79	83.68	**3.58E-05**	75.44	71.87	0.04	77.19	75.00	**2.97E-06**	85.76	83.24	0.01	80.58	82.05

Figure 3.9 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and Braak stage in the PFC in three cohorts.

In our linear regression analyses, cg18799241 in the ROSMAP cohort exceeded the Bonferroni significance threshold (black line: P < 3.85E-03). The three PFC cohorts are denoted as different shapes. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
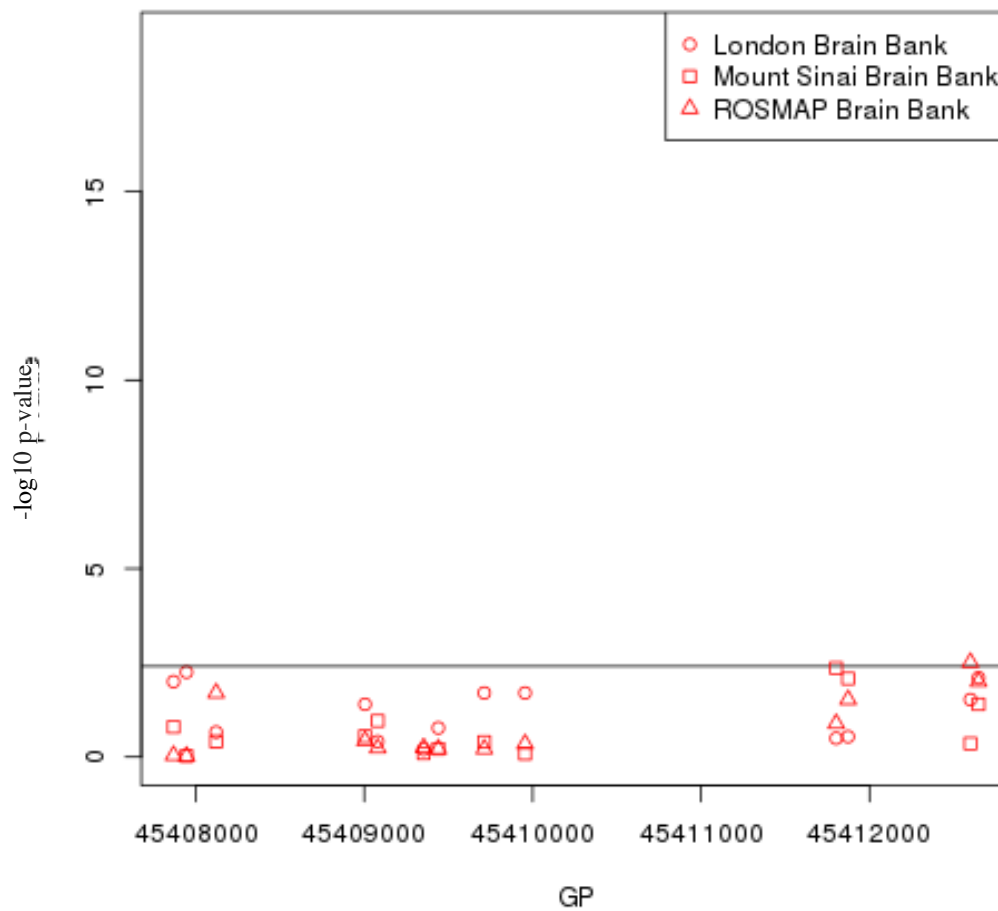
Figure 3.10 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and Braak stage in the STG in three cohorts.

In our linear regression analyses, four CpG sites (cg16471933, cg05501958, cg18799241, cg21879725) passed our Bonferroni significance threshold (black line: P < 3.85E-03) in the Arizona cohort, with two of these (cg05501958 and cg21879725) also reaching Bonferroni significance in the Mount Sinai cohort. The three STG cohorts are denoted as different shapes. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
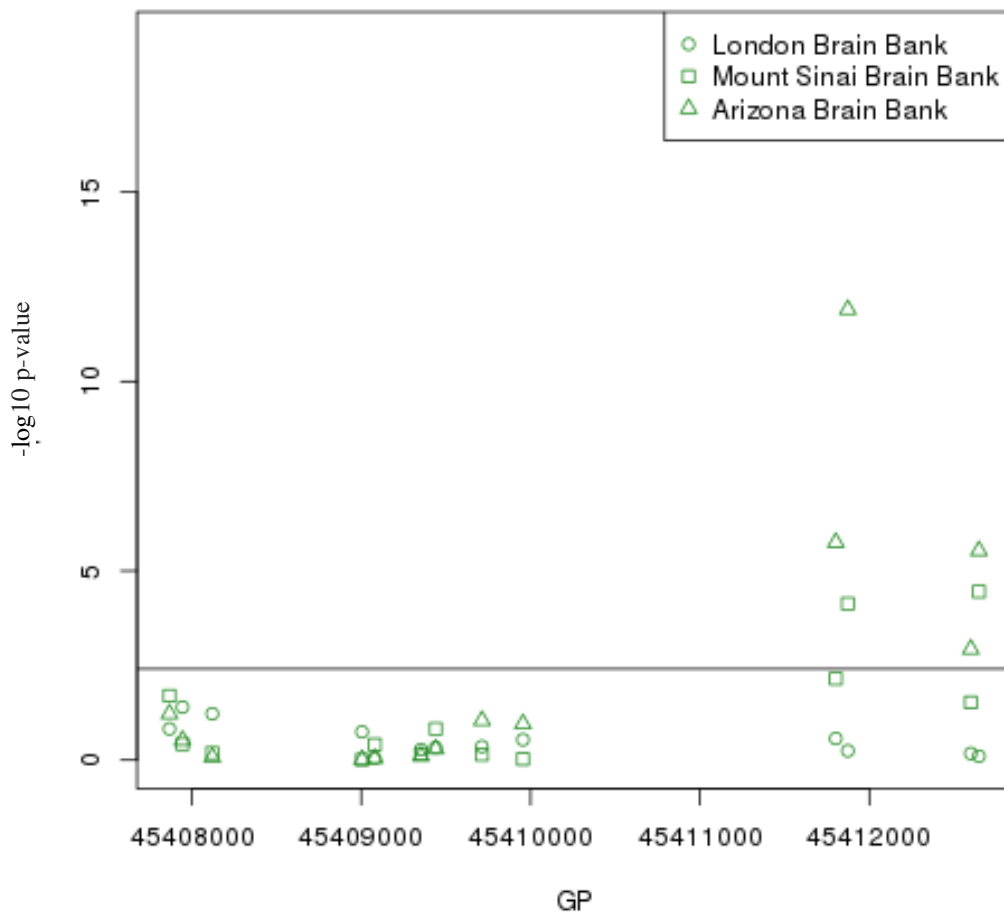
Figure 3.11 - A graph showing effect-size coefficient (methylation difference between Braak 0 and Braak VI samples) across the APOE gene in all three cohorts in the PFC samples.

Adjusted beta value difference between Braak 0 and Braak VI are represented as methylation effect size. The different cohorts are denoted by different line colours.
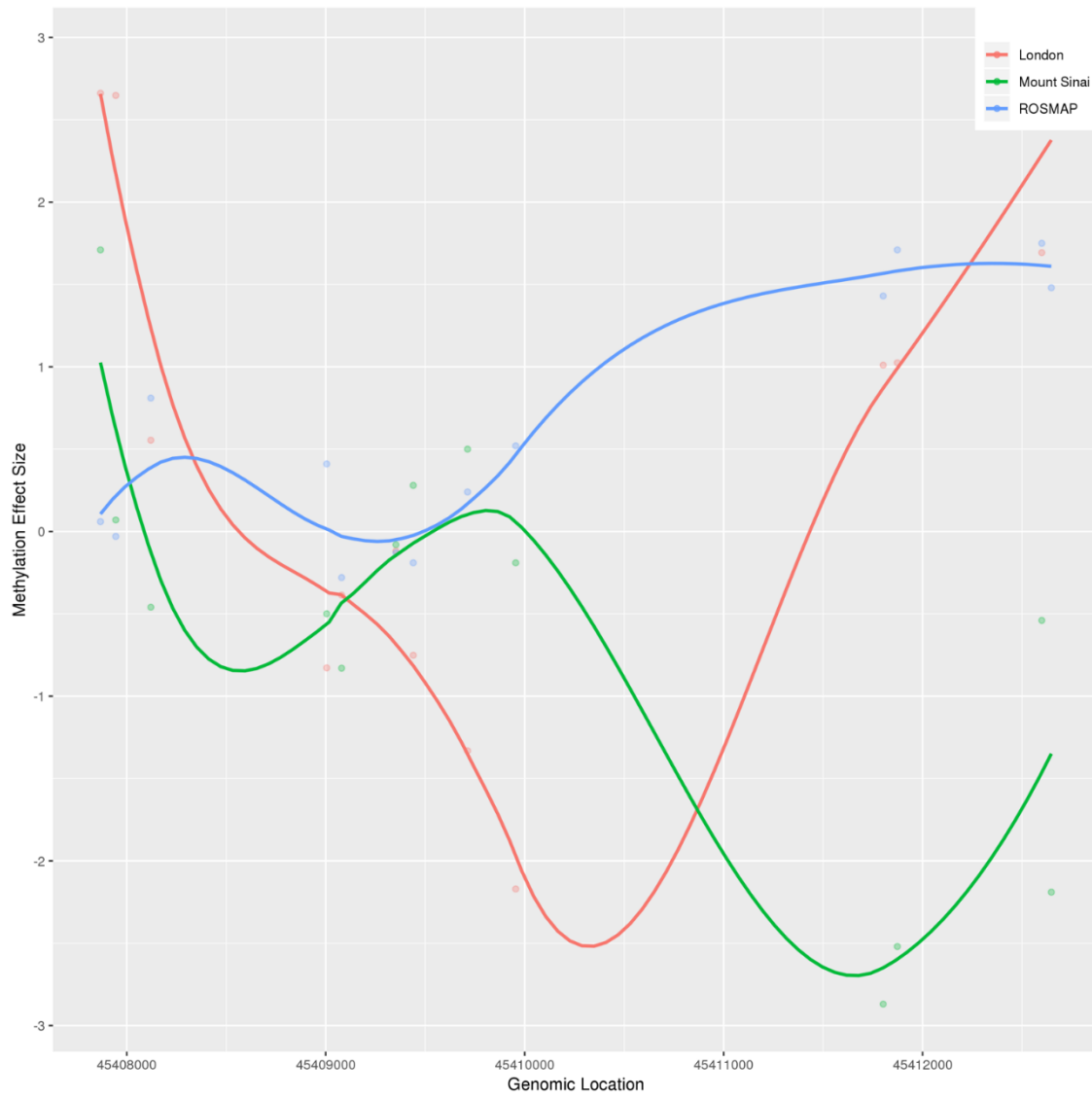
Figure 3.12 - A graph showing effect-size coefficient (methylation difference between Braak 0 and Braak VI samples) across the APOE gene in all three cohorts in the STG samples.

Adjusted beta value difference between Braak 0 and Braak VI are represented as methylation effect size. The different cohorts are denoted by different line colours.
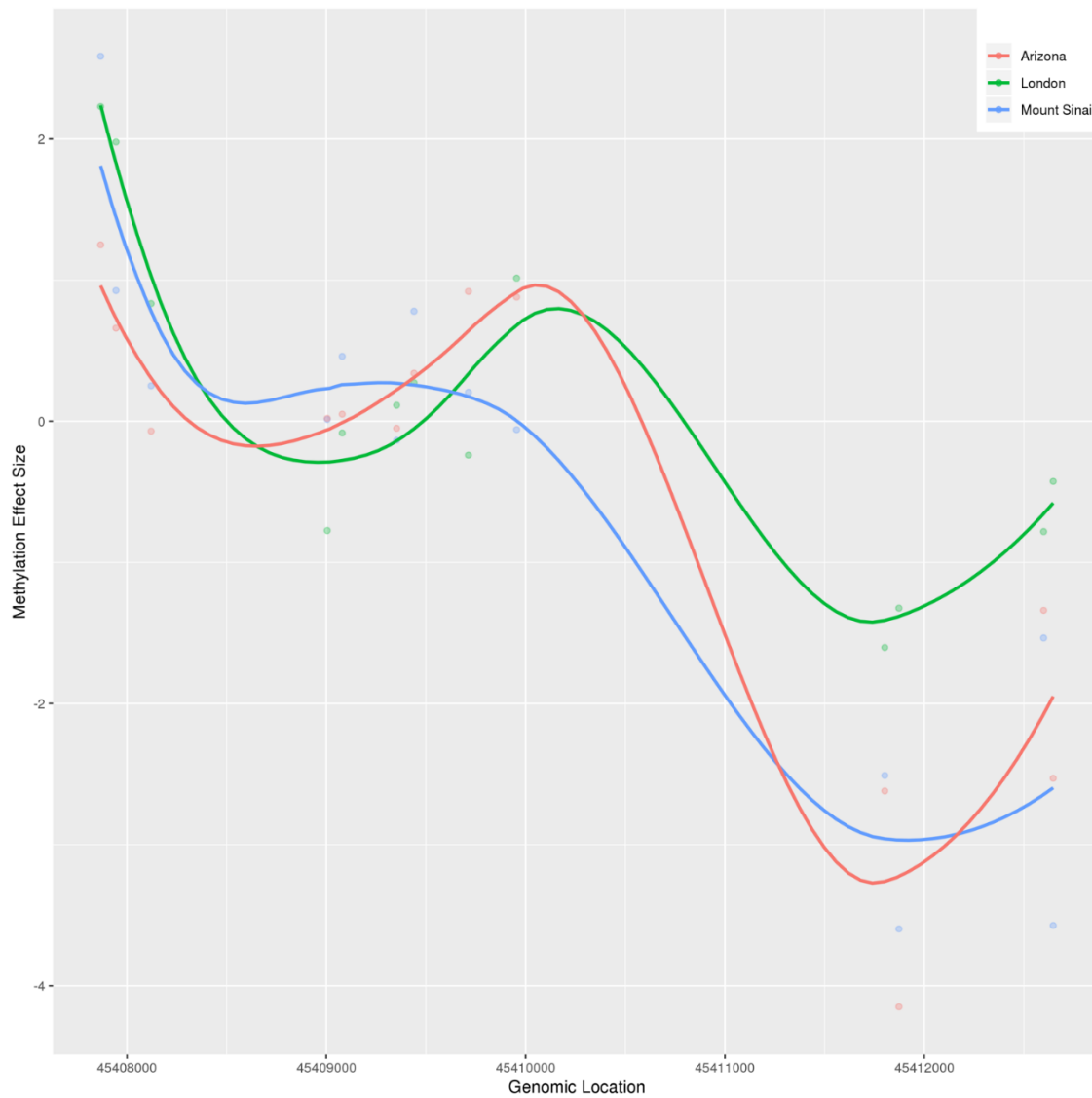
Figure 3.13 - A Manhattan plot of the APOE region, highlighting associations between DNA methylation and Braak stage in the STG and PFC in our meta-analysis.

We identified four CpG sites (cg16471933, cg05501958, cg18799241, cg21879725) that passed the Bonferroni significance threshold (black line: P <3.85E-03) in the STG, and two (cg18799241 and cg21879725) in the PFC. The STG is shown in green and the PFC is shown in red. P value is shown on the X axis and genomic position on chromosome 19 is shown on the Y axis.
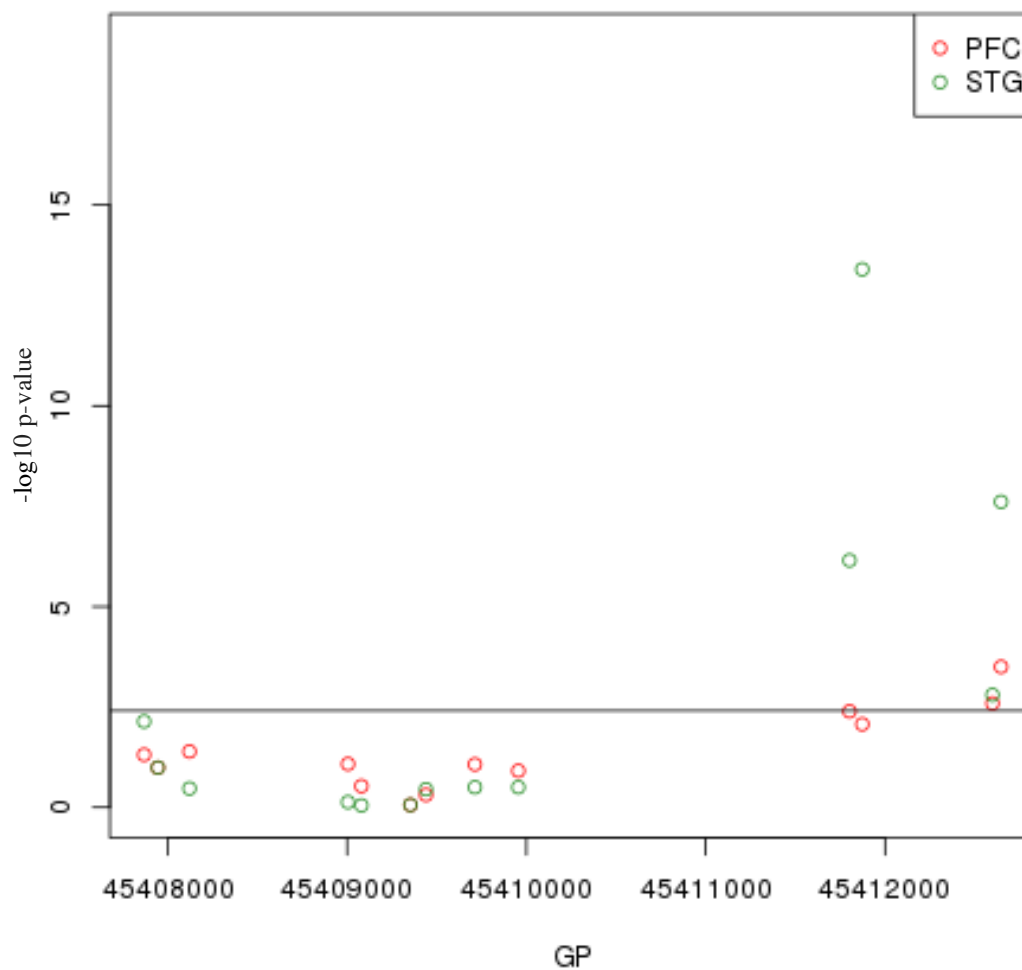
Figure 3.14 - A box and whisker plot demonstrating average corrected DNA methylation levels for the 13 CpGs in APOE in Braak 0 (red) and Braak VI (pink) samples in the PFC.

All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P < 0.05), *** = Bonferroni significance (P < 3.85E-03)
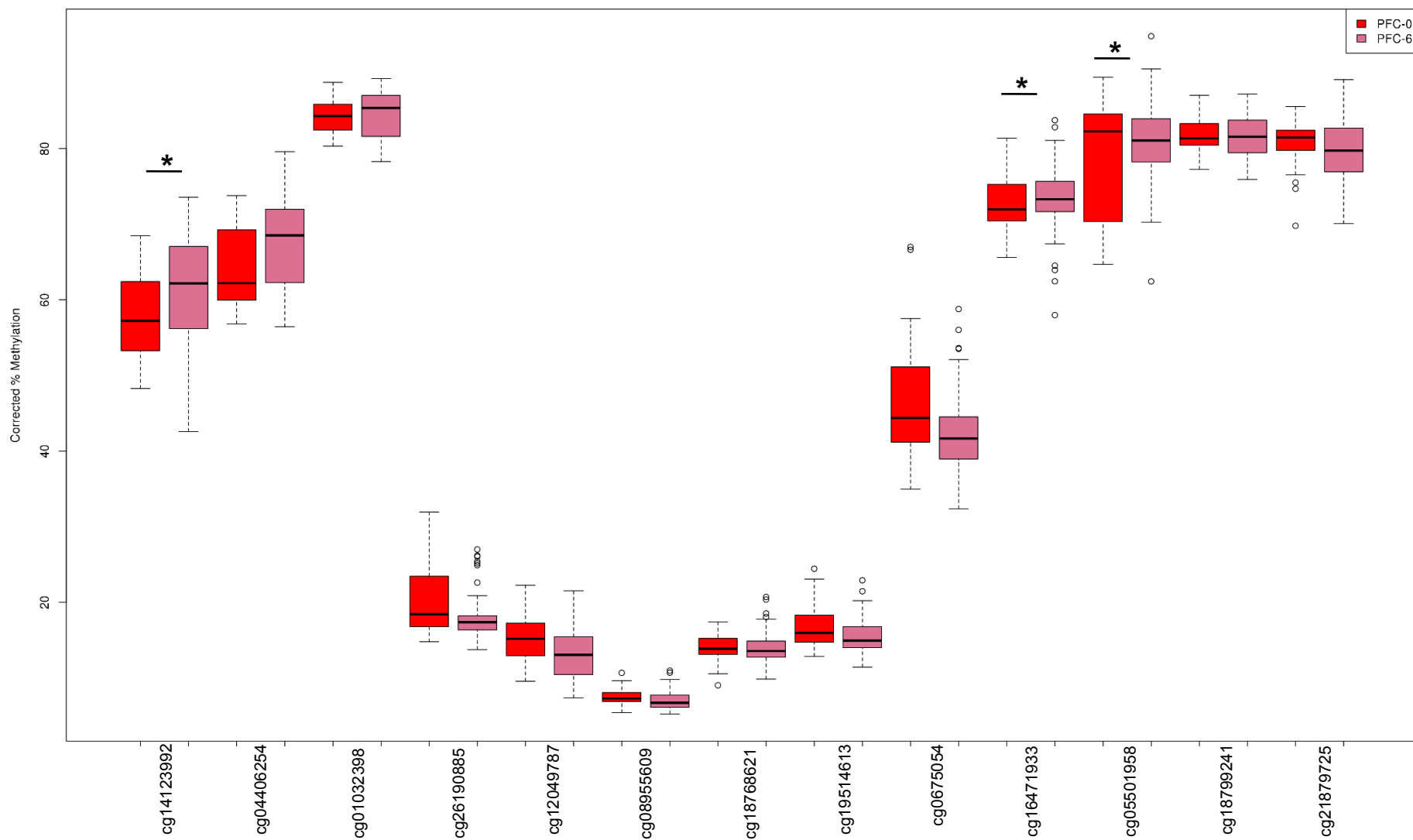
Figure 3.15 - A box and whisker plot demonstrating average corrected DNA methylation levels for the 13 CpGs in APOE in Braak VI (light green) and Braak 0 (dark green) samples in the STG.

*All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P < 0.05), *** = Bonferroni significance (P < 3.85E-03).*
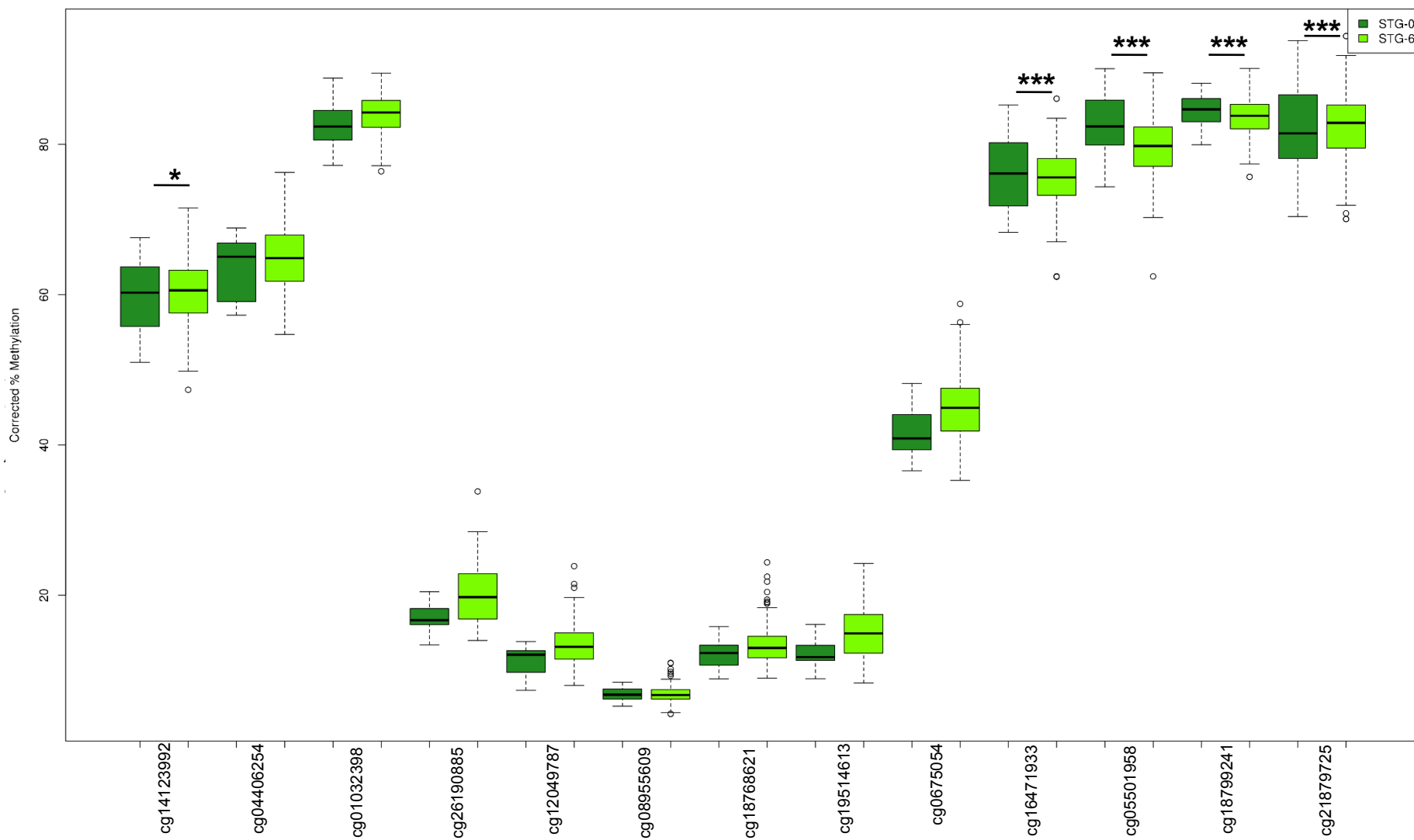
Figure 3.16 - A box and whisker plot demonstrating average corrected DNA methylation levels for the 13 CpGs in APOE in Braak 0 (dark colours) and Braak VI (light colours) samples in the PFC (red shades) and STG (green shades).

All CpG loci are ordered according to genomic position. Box and whisker plots show median, upper and lower quartiles, and maximum and minimum values, with outliers included. Key: * = Nominal significance (P < 0.05), *** = Bonferroni significance (P < 3.85E-03).
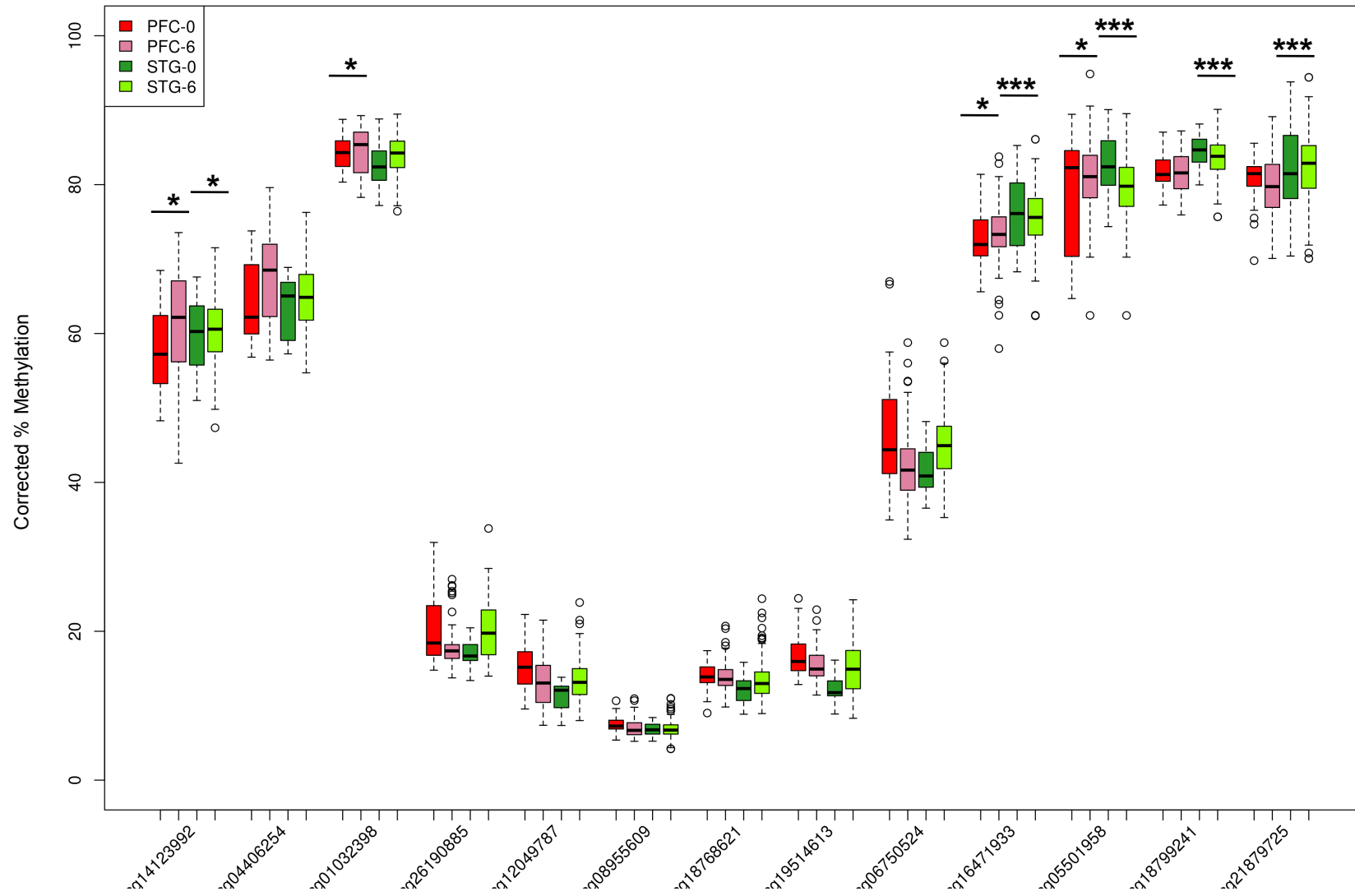
Figure 3.17- Forest plots highlighting the effect size of cg1647933, cg05501958, cg18799241 and cg21879725 across all cohorts and tissues.
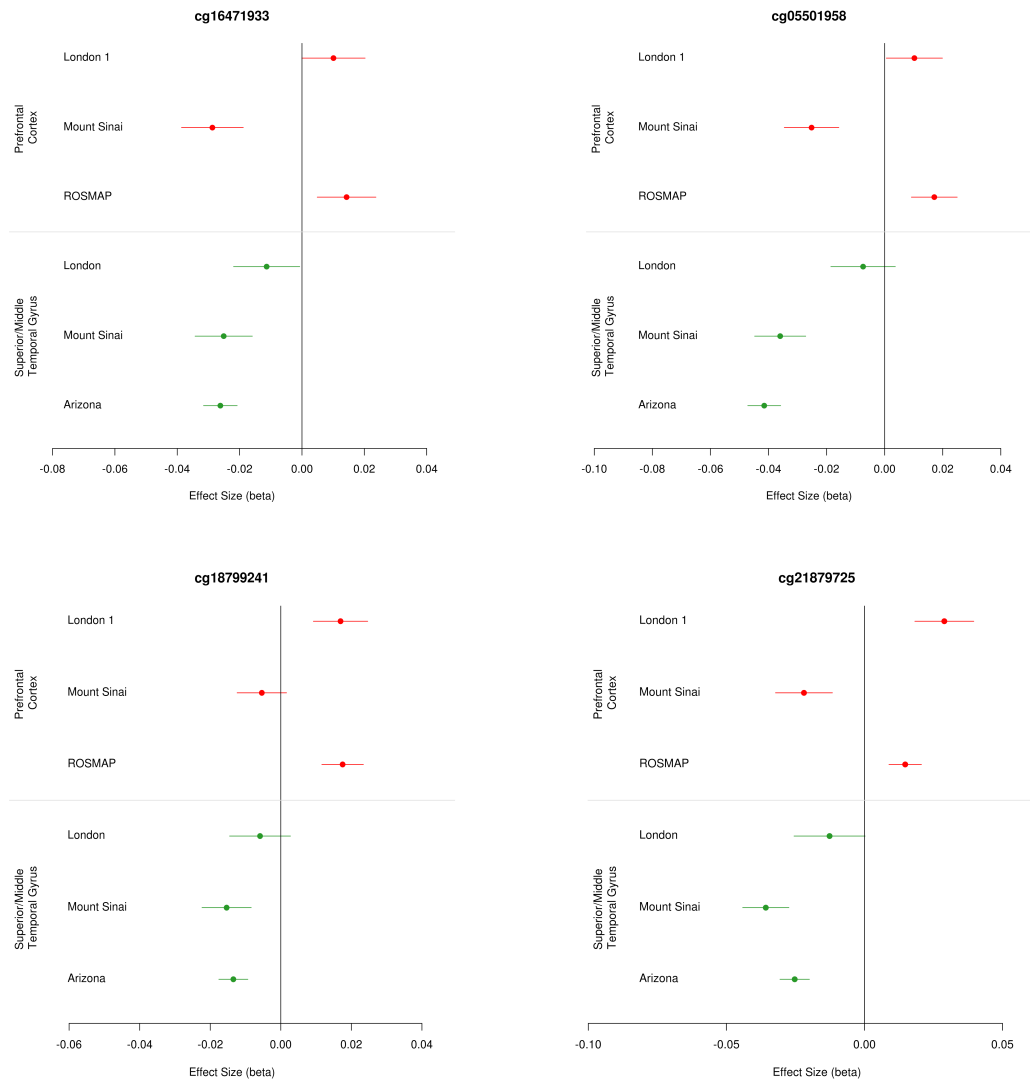
3.3 *APOE* shows hypomethylation with advancing age in AD in the STG

As *APOE* is the strongest common genetic risk factor for late-onset AD, we were interested to investigate whether we observed a mQTL in our data, whereby the number of *APOE ε4* alleles altered DNA methylation within the gene. We were also interested to investigate whether this association differed with age, given that AD is a disease of the elderly. To address this we looked for an association of (a) age, (b) *APOE* genotype (number of ε4 alleles) and (c) an interaction between age and genotype with DNA methylation in our meta-analysis cohort, regardless of disease (**Table 3.3**). We saw that in the PFC cg26190885, cg19514613 and cg06750524 showed a Bonferroni significant decrease in DNA methylation with advancing age (P = 7.02E-05, P = 1.01E-03, P = 2.98E-05, respectively). None of the probes showed any association with *APOE* genotype (**Table 3.3**), in either the PFC (**Figure 3.18**) or STG (**Figure 3.19**).

When we stratified the data by AD diagnosis, we saw no Bonferroni-significant association of (a) age, (b) the number of *APOE* ε4 alleles or (c) an interaction of age and genotype on DNA methylation levels for the 13 probes in the PFC for either AD or control samples (**Table 3.4**). However, we did see a Bonferroni-significant association of DNA methylation of age (but not genotype nor an interaction between the two variables) in the STG for AD, but not control samples (**Table 3.5**) at cg14123992, cg04406254, cg01032398 and cg06750524 (P = 3.84E-04, P = 1.21E-03, P = 4.54E-04, P = 1.40E-04, respectively). This may indicate that the decrease in DNA methylation at these four loci in *APOE* with age is only observed in AD subjects. Interestingly, three of these loci (cg14123992, cg04406254 and cg01032398) are all at the 5' end of the gene, located 1500bp from the transcription start site (TSS).

Table 3.3 - Association of DNA methylation in the APOE gene with (a) age, (b) the number of APOE ε4 alleles and (c) and the interaction between genotype and age in all samples used in the AD meta-analysis (regardless of diagnosis) in the PFC and STG.

Shown for all 13 probes are the effect size (difference in DNA methylation per unit of variable), and associated P value. All P values reaching Bonferroni significance (<3.85E-03) are highlighted in bold. Probes are ordered by genomic position.

| | PFC | | | | | | STG | | | | | |
| | Age | | Genotype | | Age*Genotype | | Age | | Genotype | | Age*Genotype | |
	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
cg14123992	-1.37E-03	3.97E-03	9.98E-03	0.84	2.31E-06	1.00	-9.05E-04	4.30E-03	-3.10E-02	0.32	6.50E-04	8.30E-02
cg04406254	-1.26E-03	9.38E-03	6.29E-02	0.21	-6.89E-04	0.23	-1.02E-03	**2.73E-03**	-5.82E-02	8.23E-02	9.46E-04	1.93E-02
cg01032398	-7.83E-04	2.04E-02	2.12E-02	0.54	-2.81E-04	0.49	-4.09E-04	3.20E-02	-1.33E-02	0.48	2.27E-04	0.31
cg26190885	-1.20E-03	**7.02E-05**	-4.63E-02	0.13	4.78E-04	0.18	-1.64E-04	0.33	-8.92E-03	0.59	1.05E-04	0.60
cg12049787	-7.20E-04	3.02E-02	-4.10E-02	0.23	4.34E-04	0.27	-1.22E-04	0.46	1.03E-02	0.52	-1.30E-04	0.50
cg08955609	-1.59E-04	0.23	-2.79E-02	3.94E-02	3.01E-04	5.49E-02	1.26E-04	9.61E-02	1.29E-02	8.37E-02	-1.60E-04	7.59E-02
cg18768621	4.60E-05	0.853379	1.04E-03	0.97	-1.92E-05	0.95	1.65E-04	0.34	8.15E-03	0.63	-1.26E-04	0.54
cg19514613	-1.14E-03	**1.01E-03**	-2.32E-02	0.51	2.20E-04	0.59	-2.63E-04	0.18	7.63E-03	0.69	-6.47E-05	0.78
cg06750524	-2.40E-03	**2.98E-05**	-7.16E-02	0.22	7.85E-04	0.25	-5.92E-04	2.95E-02	1.45E-02	0.59	-7.15E-05	0.82
cg16471933	2.13E-04	0.75	8.62E-02	0.21	-1.05E-03	0.19	-6.03E-04	0.10	-4.97E-02	0.16	6.24E-04	0.15
cg05501958	1.27E-03	1.80E-02	1.15E-01	3.74E-02	-1.44E-03	2.48E-02	-4.57E-05	0.85	-2.29E-02	0.32	1.36E-04	0.62
cg18799241	-5.28E-04	0.24	-1.67E-02	0.72	1.70E-04	0.75	-4.55E-04	8.20E-02	-2.22E-02	0.39	3.02E-04	0.33
cg21879725	-6.55E-04	0.27	4.06E-02	0.51	-6.41E-04	0.37	-4.60E-04	0.20	-3.24E-02	0.36	4.14E-04	0.34

Figure 3.18 - Probes within APOE show no association with APOE ε4 genotype in the PFC in the combined AD and control cohorts.

Probes are in order of genomic position. Probes are in order of genomic position. Box and whisker plot shows median, upper and lower quartiles, and maximum and minimum values, with outliers included. Blue represents zero ε4 alleles (N = 660), red represents one ε4 allele (N = 266) and green represents two ε4 alleles (N = 26).Zero APOE ε4 alleles n = 660, one APOE ε4 alleles n = 266, two ε4 alleles n = 26.
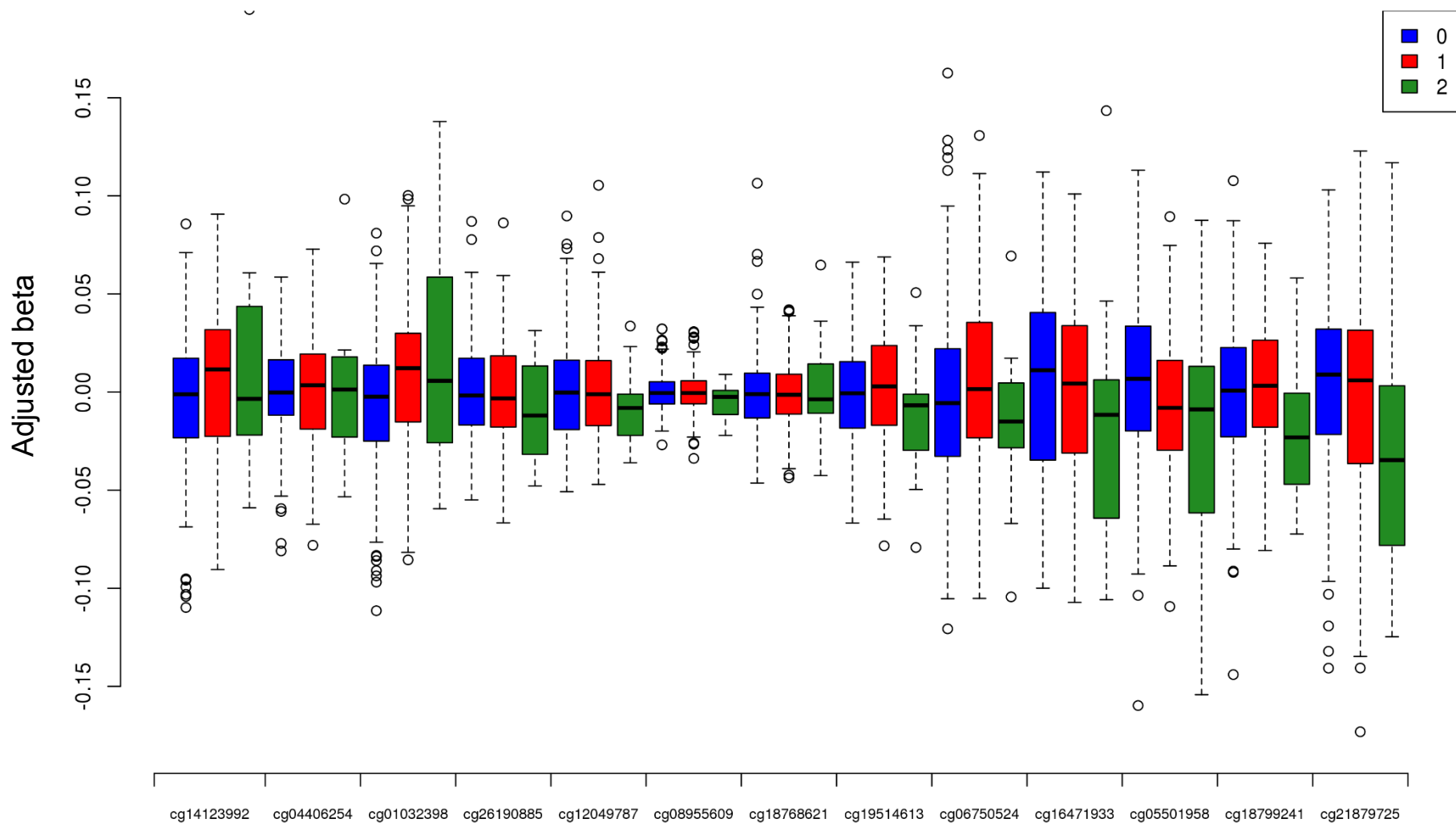
Figure 3.19 - Probes within APOE show no association with APOE ε4 genotype in the STG in the combined AD and control cohorts.

Probes are in order of genomic position. Box and whisker plot shows median, upper and lower quartiles, and maximum and minimum values, with outliers included. Blue represents zero ε4 alleles (N = 360), red represents one ε4 allele (N = 220) and green represents two ε4 alleles (N = 44).

Table 3.4 - Association of DNA methylation in the APOE gene with (a) age, (b) the number of APOE ε4 alleles and (c) and the interaction between genotype and age in AD and control samples independently in the PFC.

Shown for all 13 probes are the effect size (difference in DNA methylation per unit of variable), and associated P value. No probes reached Bonferroni significance

	CONTROL						AD					
	Age		Genotype		Age*Genotype		Age		Genotype		Age*Genotype	
	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
cg14123992	2.95E-04	0.66	-3.26E-03	0.98	3.20E-04	0.87	-1.47E-03	2.37E-02	9.74E-04	0.99	1.23E-04	0.85
cg04406254	9.52E-04	0.13	8.54E-02	0.57	-9.82E-04	0.58	-1.26E-03	0.06	6.60E-02	0.27	-6.96E-04	0.31
cg01032398	3.44E-04	0.44	-2.47E-02	0.82	3.47E-04	0.79	-9.02E-04	0.05	1.71E-02	0.68	-2.25E-04	0.64
cg26190885	-7.88E-04	0.15	-1.62E-01	0.22	1.87E-03	0.23	-1.11E-03	6.85E-03	-3.47E-02	0.34	3.50E-04	0.40
cg12049787	-8.11E-05	0.91	-4.80E-02	0.77	5.87E-04	0.77	-7.02E-04	0.11	-3.61E-02	0.36	3.84E-04	0.40
cg08955609	-2.42E-05	0.92	-5.21E-02	0.36	5.62E-04	0.40	-1.44E-04	0.42	-2.68E-02	0.10	2.92E-04	0.11
cg18768621	1.12E-03	0.07	1.92E-01	0.20	-2.40E-03	0.18	-1.36E-04	0.67	-1.62E-02	0.57	1.91E-04	0.56
cg19514613	-1.25E-04	0.84	1.94E-01	0.20	-2.34E-03	0.19	-9.89E-04	3.44E-02	-1.85E-02	0.66	1.85E-04	0.70
cg06750524	-8.76E-04	0.35	-2.23E-02	0.92	-2.96E-05	0.99	-2.02E-03	9.08E-03	-4.65E-02	0.50	5.46E-04	0.49
cg16471933	-1.26E-05	0.99	3.16E-01	0.39	-3.91E-03	0.37	-4.56E-04	0.59	4.79E-02	0.53	-6.33E-04	0.47
cg05501958	-3.45E-04	0.76	1.26E-01	0.64	-1.74E-03	0.58	4.08E-04	0.55	6.37E-02	0.31	-9.01E-04	0.21
cg18799241	6.86E-04	0.57	1.43E-02	0.96	-3.68E-04	0.92	-1.42E-03	1.09E-02	-6.43E-02	0.20	6.97E-04	0.22
cg21879725	-2.57E-04	0.75	1.30E-01	0.50	-1.78E-03	0.44	-7.73E-04	0.34	3.90E-02	0.60	-6.17E-04	0.46

Table 3.5 - Association of DNA methylation in the APOE gene with (a) age, (b) the number of APOE ε4 alleles and (c) and the interaction between genotype and age in AD and control samples independently in the STG.

Shown for all 13 probes are the effect size (difference in DNA methylation per unit of variable), and associated P value. All P values reaching Bonferroni significance (<3.85E-03) are highlighted in bold. Probes are ordered by genomic position.

	CONTROL						AD					
	Age		Genotype		Age*Genotype		Age		Genotype		Age*Genotype	
	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
cg14123992	-4.45E-04	0.34	-8.12E-02	0.13	1.25E-03	7.05E-02	-1.60E-03	**3.84E-04**	-3.62E-02	0.37	6.76E-04	0.16
cg04406254	-7.80E-04	0.12	-1.18E-01	4.36E-02	1.65E-03	2.79E-02	-1.57E-03	**1.21E-03**	-5.63E-02	0.20	8.82E-04	9.11E-02
cg01032398	-1.26E-04	0.67	-1.35E-03	0.97	1.17E-05	0.98	-9.23E-04	**4.54E-04**	-3.66E-02	0.13	4.70E-04	9.60E-02
cg26190885	1.87E-04	0.48	5.18E-03	0.86	-1.09E-04	0.78	-5.81E-04	1.37E-02	-2.72E-02	0.20	3.08E-04	0.22
cg12049787	1.74E-04	0.47	-5.37E-03	0.85	4.04E-05	0.91	-4.97E-04	3.19E-02	4.35E-03	0.84	-7.49E-05	0.76
cg08955609	1.33E-04	0.26	-8.56E-04	0.95	2.05E-05	0.91	2.01E-04	5.49E-02	2.30E-02	1.64E-02	-2.69E-04	1.78E-02
cg18768621	-1.31E-04	0.65	-2.99E-02	0.37	4.31E-04	0.32	1.73E-04	0.45	1.07E-02	0.61	-1.94E-04	0.44
cg19514613	-3.36E-04	0.29	-1.94E-02	0.60	2.79E-04	0.56	-5.26E-04	4.09E-02	1.56E-03	0.95	-3.63E-05	0.90
cg06750524	-8.76E-05	0.82	1.78E-02	0.69	-1.04E-04	0.86	-1.48E-03	**1.40E-04**	-2.77E-02	0.43	3.63E-04	0.38
cg16471933	5.51E-05	0.92	-5.71E-03	0.93	-3.36E-05	0.97	-1.24E-03	1.51E-02	-8.13E-02	8.15E-02	9.92E-04	7.25E-02
cg05501958	5.99E-05	0.87	-2.47E-02	0.55	1.25E-04	0.81	2.34E-04	0.47	-8.52E-05	1.00	-8.16E-05	0.82
cg18799241	7.23E-05	0.85	-8.38E-03	0.85	8.97E-05	0.87	-1.00E-03	7.85E-03	-4.35E-02	0.20	5.44E-04	0.18
cg21879725	8.88E-05	0.88	-3.46E-02	0.60	4.08E-04	0.63	-9.85E-04	5.16E-02	-4.62E-02	0.32	5.72E-04	0.29

3.4 The 5' end of the *APOE gene* shows a cis mQTL in post-natal PFC tissue

Next, we were interested to investigate whether there was an association of *APOE* DNA methylation with age or *APOE* genotype in a larger cohort of post-natal PFC samples collected by our group. To this end we extracted DNA methylation from the "big brain dataset" (see section 2.1.2), which included EWAS data from MCI, ASD, schizophrenia, depression and age-matched controls for 12 *APOE* probes that passed QC. We added this data to the AD and control data we had been analysing in section 3.3 and looked for an association of DNA methylation individually at each probe with (a) age, (b) the number of *APOE ε4* alleles and (c) an interaction of age and genotype, whilst controlling for sex, cell proportion, brain region and brain bank (**Table 3.6**). In this larger cohort we observed Bonferroni significant hypermethylation of cg16471933 with advancing age (P = 3.89E-03) (**Figure 3.20**). We also observed a Bonferroni significant mQTL at cg14123992 (**Figure 3.21**: P = 2.53E-06) and cg04406254 (**Figure 3.22**: P = 8.23E-05), with increased DNA methylation with increased numbers of *APOE ε4* alleles. Both of these two probes are located at the 5' end of the gene, within 1500bp of the TSS.
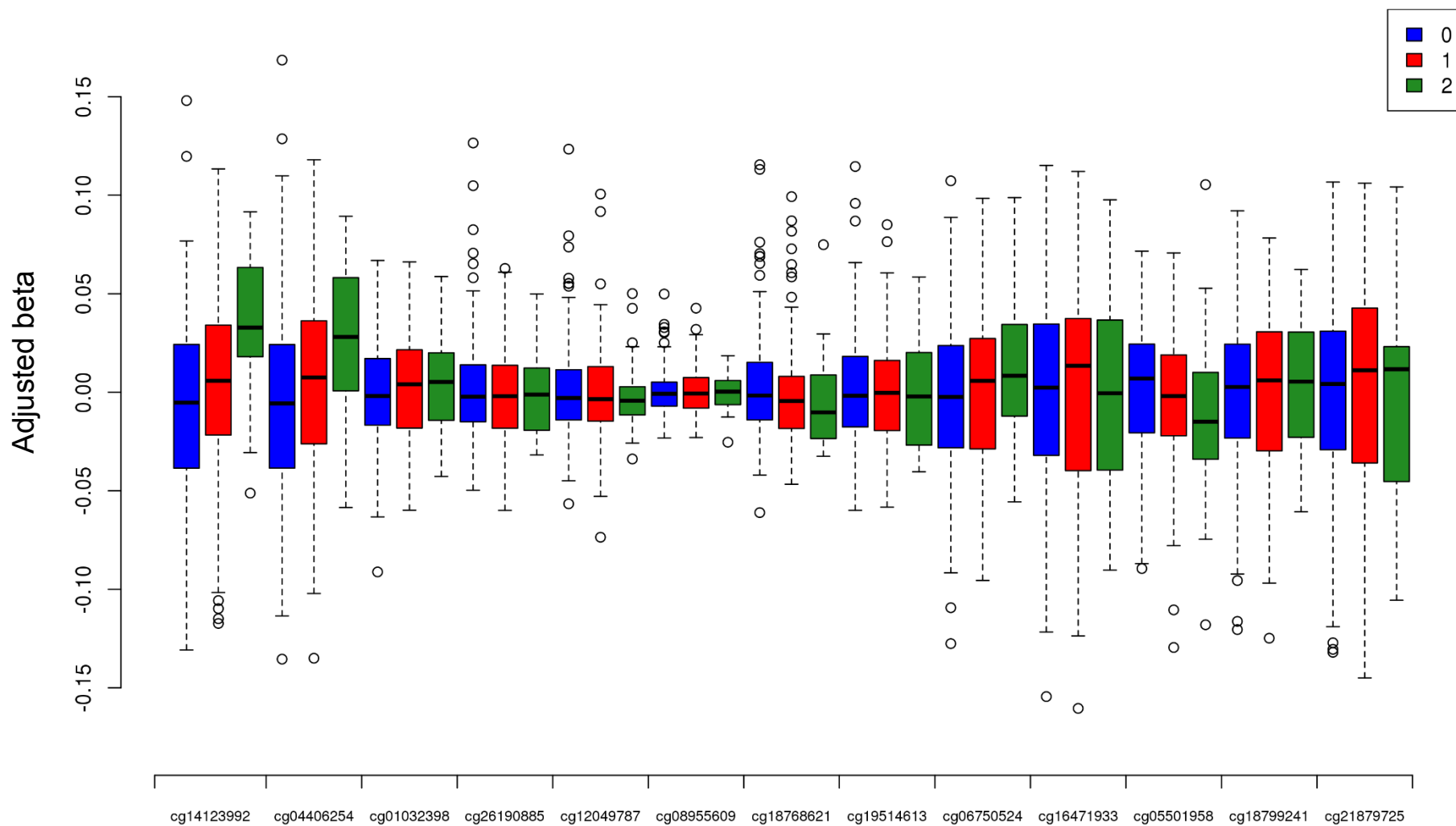
Table 3.6 - Association of DNA methylation in the APOE gene with (a) age, (b) the number of APOE ε4 alleles and (c) and the interaction between genotype and age in post-natal tissue.

Shown for all of 12 probes with data available are the effect size (difference in DNA methylation per unit of variable), and associated P value. All p-values reaching Bonferroni significance (<4.17E-03) are highlighted in bold. Probes are ordered by genomic position.

Probe	Age		Number of *APOE ε4* alleles		Interaction (Age x No. alleles)	
	Effect Size	P value	Effect Size	P value	Effect Size	P value
cg14123992	-1.90E-04	2.65E-02	3.23E-02	**2.53E-06**	-1.58E-04	6.49E-02
cg04406254	-1.68E-04	2.88E-02	2.42E-02	**8.23E-05**	-1.07E-04	0.16
cg01032398	-7.75E-05	0.13	7.67E-03	6.18E-02	-6.39E-05	0.21
cg26190885	-1.38E-05	0.79	2.16E-03	0.60	-4.57E-05	0.37
cg12049787	-1.02E-04	6.36E-02	3.94E-03	0.37	-5.84E-05	0.29
cg08955609	-2.44E-05	0.36	4.62E-03	2.87E-02	-5.23E-05	4.78E-02
cg18768621	-	-	-	-	-	-
cg19514613	-4.97E-05	0.39	6.62E-03	0.15	-7.51E-05	0.20
cg06750524	-9.37E-05	0.27	1.55E-02	2.10E-02	-1.25E-04	0.14
cg16471933	2.88E-04	**3.89E-03**	1.42E-02	7.38E-02	-1.94E-04	5.12E-02
cg05501958	-9.42E-05	0.33	1.88E-03	0.81	-1.23E-04	0.20
cg18799241	1.07E-04	0.13	1.43E-02	1.17E-02	-1.56E-04	2.81E-02
cg21879725	1.41E-04	0.10	1.67E-02	1.64E-02	-2.11E-04	1.54E-02

Figure 3.20 - Cg16471933 shows a Bonferroni-significant association of DNA methylation with increasing age in post-natal PFC brain samples.

Shown are adjusted beta values (corrected for sex, cell proportion, brain region and brain bank) in individual samples plotted against age (in years). Samples with zero APOE ε4 alleles are shown in black, samples with one APOE ε4 allele are shown in red, and samples with two APOE ε4 alleles are shown in green.
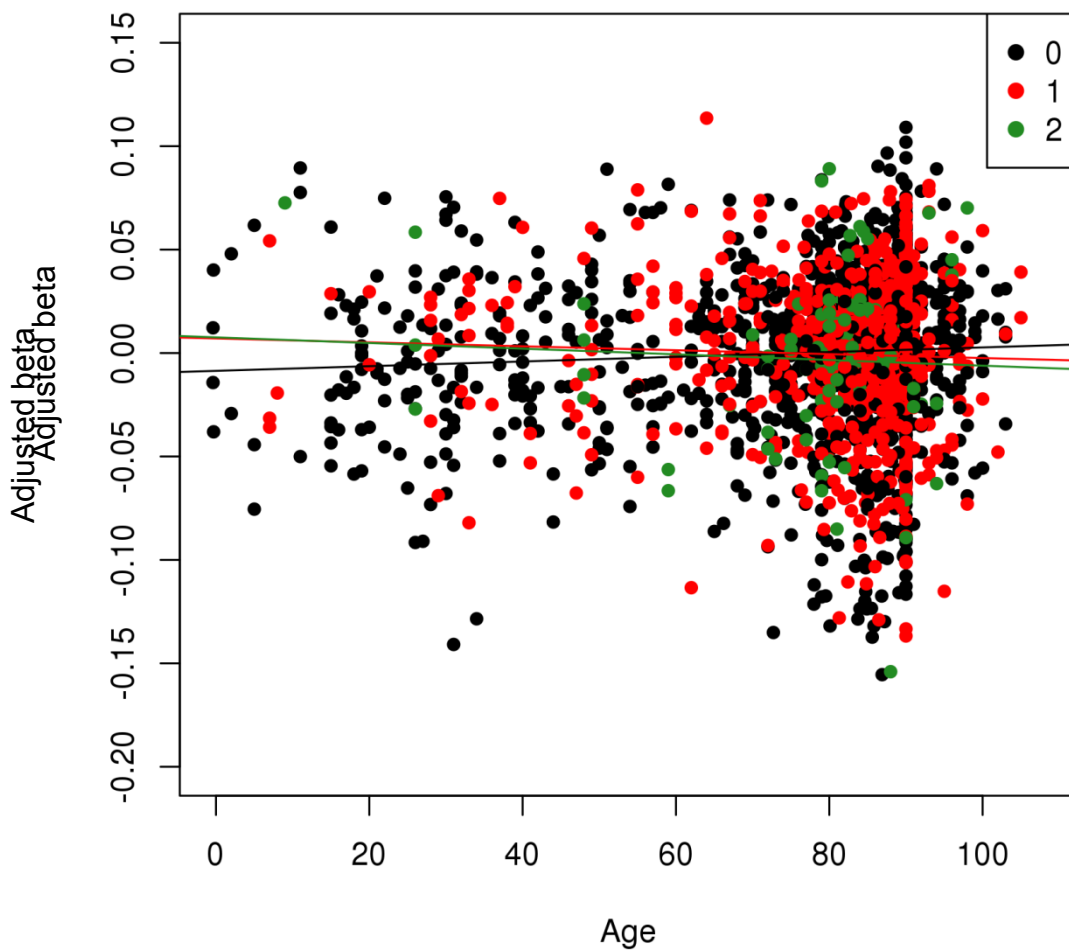
Figure 3.21 - DNA methylation at cg14123992 shows an mQTL with respect to the number of APOE ε4 alleles in post-natal PFC samples, which is evident across all phenotypes.

Box and whisker plot shows median, upper and lower quartiles, and maximum and minimum values, with outliers included. Blue represents zero ε4 alleles, red represents one ε4 alleles and green represents two ε4 alleles.
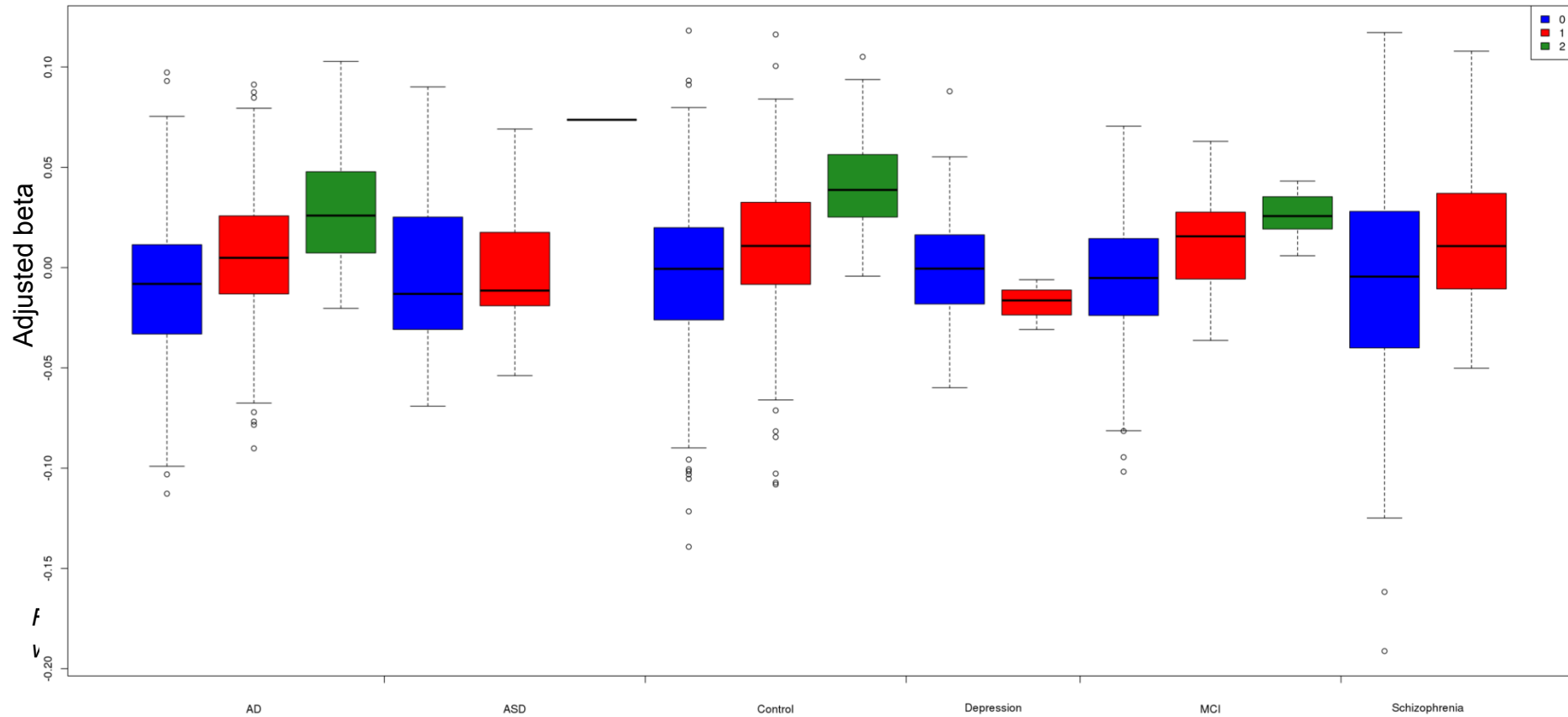
Box and whisker plot shows median, upper and lower quartiles, and maximum and minimum values, with outliers included. Blue represents zero ε4 alleles, red represents one ε4 alleles and green represents two ε4 alleles.
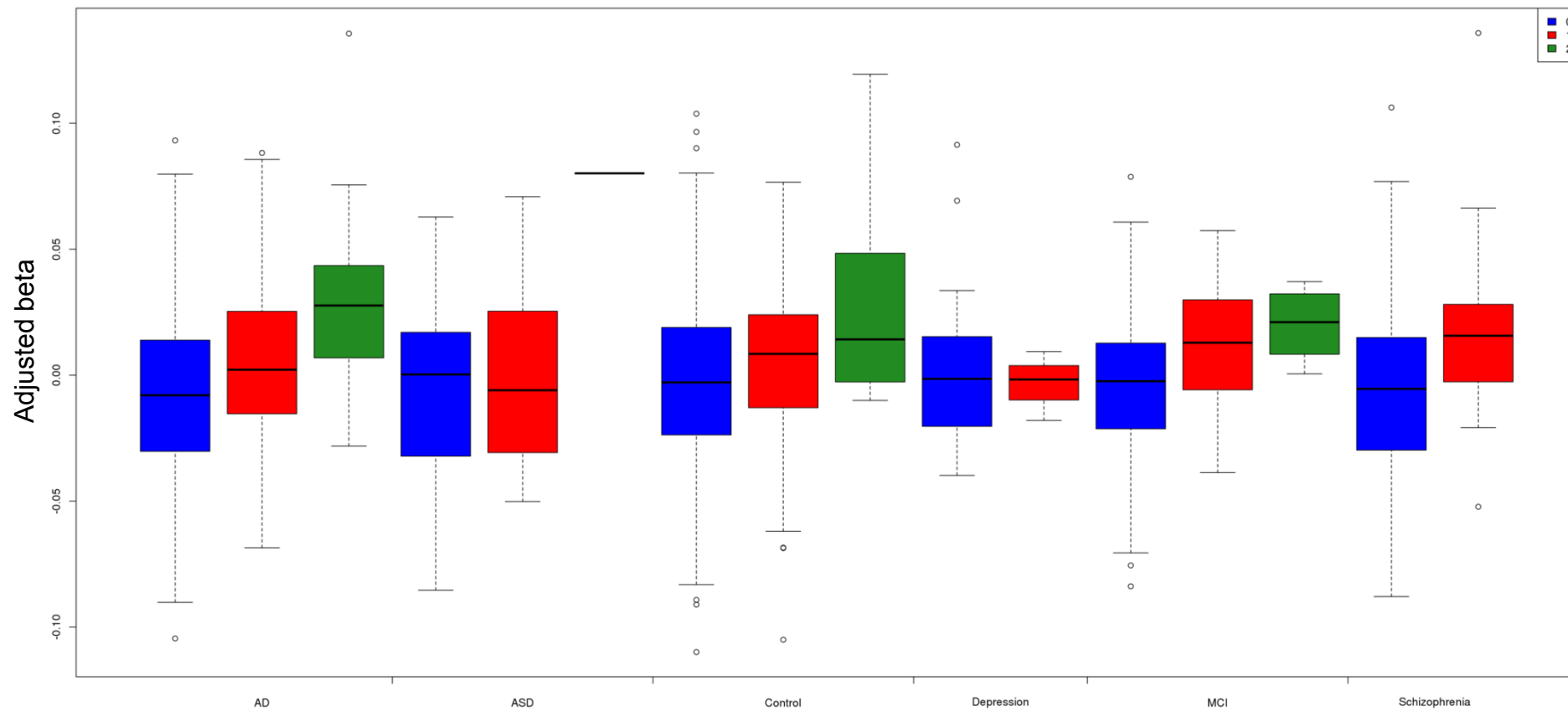
3.5 The 3' end of the *APOE gene* shows hypomethylation during development

Finally, we investigated whether loci in the *APOE* gene show altered DNA methylation with respect to age, or the number of *APOE ε4* alleles in a unique cohort of pre-natal brain samples for which we have access to EWAS and *APOE* genotype data (**Table 3.7**). We observed a Bonferroni-significant association of decreased DNA methylation with increasing foetal age for cg21879725 (**Figure 3.23**: *P* = 8.13E-05). Interestingly, this was one of the loci where we had observed AD-associated hypomethylation in the STG and PFC in our meta-analysis. However, it is worth noting that only three donors were homozygous for *APOE ε4,* and therefore further analysis is required to validate this in a cohort with larger numbers of these individuals.

Table 3.7 - Association of DNA methylation in the APOE gene with (a) age, (b) the number of APOE ε4 alleles and (c) and the interaction between genotype and age in pre-natal tissue.

Shown for all of 12 probes with data available are the effect size (difference in DNA methylation per unit of variable), and associated P value. All p-values reaching Bonferroni significance (<4.17E-03) are highlighted in bold. Probes are ordered by genomic position.

	Age		Number of *APOE ε4* alleles		Interaction (Age x No. alleles)	
Probe	Effect Size	P value	Effect Size	P value	Effect Size	P value
cg14123992	0.037	0.53	-0.007	0.79	-0.048	0.66
cg04406254	0.016	0.81	-0.018	0.56	-0.068	0.56
cg01032398	-0.095	0.11	0.035	0.20	0.135	0.21
cg26190885	0.080	0.10	0.000	0.99	-0.022	0.80
cg12049787	0.021	0.55	-0.030	7.42E-02	-0.119	6.77E-02
cg08955609	0.020	0.62	0.012	0.50	0.041	0.57
cg18768621	-	-	-	-	-	-
cg19514613	0.058	0.20	-0.039	6.62E-02	-0.143	8.31E-02
cg06750524	0.053	0.46	0.037	0.27	0.120	0.36
cg16471933	-0.019	0.85	-0.027	0.57	-0.119	0.52
cg05501958	-0.047	0.31	0.022	0.31	0.077	0.36
cg18799241	-0.024	0.69	-0.024	0.39	-0.105	0.34
cg21879725	-0.257	**8.13E-05**	0.082	6.82E-03	0.332	4.65E-03

Figure 3.23 - Cg21879725 shows a Bonferroni-significant association of DNA methylation with increasing age in pre-natal PFC brain samples.

Shown are adjusted beta values (corrected for sex, cell proportion, brain region and brain bank) in individual samples plotted against the number of days post conception. Samples with zero APOE ε4 alleles are shown in blue, samples with one APOE ε4 allele are shown in red, and samples with two APOE ε4 alleles are shown in green.
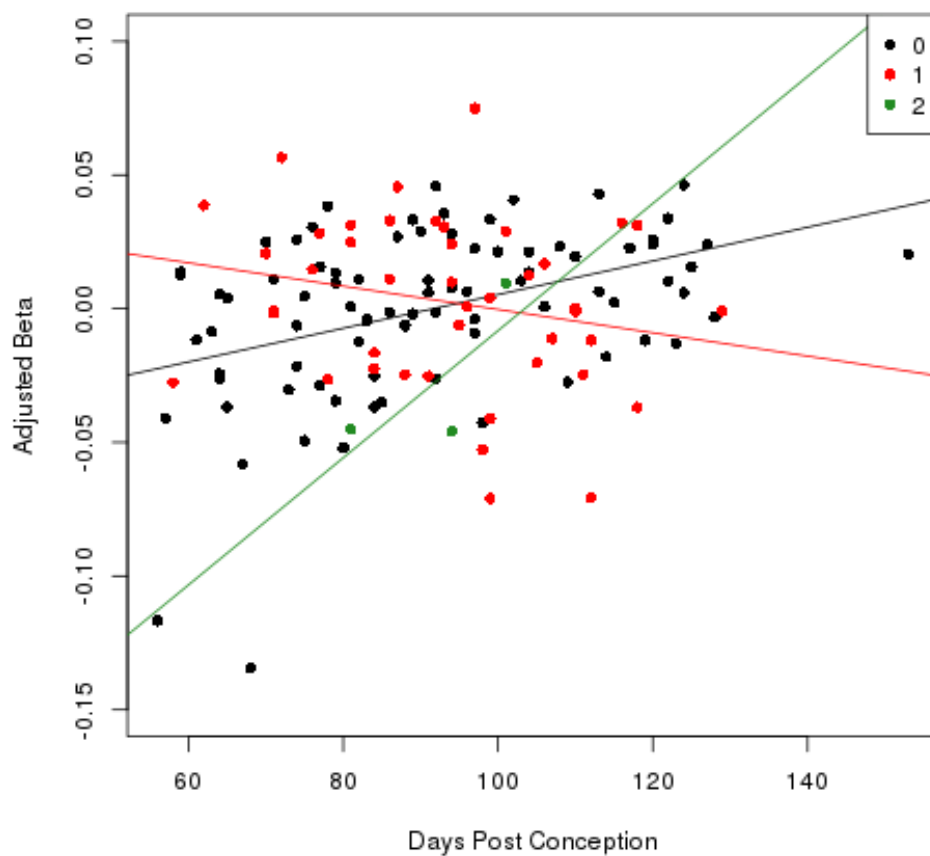
CHAPTER 4 : DISCUSSION

4.1 CpGs in the 3' exon of APOE gene are differentially methylated with respect to AD diagnosis and pathology

This study represents the first large-scale longitudinal study of *APOE* DNA methylation with respect to AD diagnosis, AD pathology, age and *APOE* genotype. The areas of the brain we studied, the STG and PFC, are involved in emotional perception and executive functions/ recall of short-term memory after encoding by the hippocampus respectively.

Of the 13 CpG sites studied, four (cg16471933, cg05501958, cg18799241, cg21879725) were discovered to have Bonferroni-significant differential methylation (hypomethylation) within the STG with respect to AD diagnosis and Braak stage. cg18799241 also showed Bonferroni-significant differential methylation in the PFC with respect to Braak stage, along with cg21879725. Furthermore, the latter probe was differentially methylated in the PFC with respect to diagnosis, alongside cg05501958. The fact we saw two loci that were differentially methylated in both brain regions potentially indicates common epigenetic dysfunction in the *APOE* gene in AD across different areas of the brain. Interestingly, all of these four differentially methylated sites were located in a well-defined CGI towards the 3' untranslated region (UTR) of the APOE gene in the 4th exon spanning 880 base pairs, and were closely grouped within ~100 bases of each other. A previous study has grouped the 13 CpG probes within *APOE* via their methylation levels and location into three distinct regions (Ma *et al.*, 2015). In that study they classed these four CpG probes at the 3' end of the gene as "Group 3" reporting they are in general hypermethylated across all tissues studied (>50% methylation). The other two regions they identified included "Group 1" located in the promoter and consisting of three probes (cg14123992, cg04406254, cg01032398) that are hypermethylated (>50% methylation) in all tissues studied) and "Group 2" consisting of the remaining six probes (cg26190885, cg12049787, cg08955609, cg18768621, cg19514613, cg06750524) that are found in the 5' end of the gene and are generally hypomethylated (<50% methylation) in all tissues studied. The particular CGI in the 3' UTR containing the "Group 3" probes where we identified AD-associated hypermethylation has been previously recognized as an important DMR in a previous cross-tissue study of the AD (Foraker *et al.*, 2015). The authors of that study showed that this CGI is differentially methylated in post-mortem brain tissue in a tissue-specific manner, with the highest levels of methylation in the cerebellum compared to the hippocampus and the lowest levels in the frontal lobe. However, in AD they only observed disease-associated differences (hypomethylation) in the hippocampus and frontal lobe, replicating the same direction of effect that we have seen in the temporal and frontal lobe in our study (Foraker *et al.*, 2015). The replication of their finding highlights the robustness of the results, as they used another technology (bisulfite pyrosequencing) and were limited to only 15 AD and 10 control subjects. Tulloch and colleagues recently assessed DNA methylation of the *APOE* gene in 15 AD and 14 control brain samples by performing bisulfite pyrosequencing in bulk (frontal lobe) tissue and fluorescence-activated cell sorting (FACS) isolated neuronal and non-neuronal (glia) cells (Tulloch, Leong, Thomson, *et al.*, 2018). They showed that hypomethylation of the CGI was observed in AD in glia, but not in neurons; as glia clear neural waste such as protein aggregations as part of their function, this may indicate the methylation of *APOE* plays a critical role in mediating this process.

Our data suggests differential methylation of the *APOE* 3'UTR CGI in AD brain in a tissue-specific manner, with a greater effect size (and more significant P values) in the STG than the PFC. The STG is a structure of the temporal lobe, which includes and is closely associated with the hippocampus (and its main information input, the entorhinal cortex (EC)) as a key processor of audiovisual sensory information to be encoded as long-term declarative memory; it has been previously found that AD pathology strikes first in the entorhinal cortex, before spreading to other cortices and the hippocampus (Devanand *et al.*, 2012) (Huijbers *et al.*, 2014) (Khan *et al.*, 2014). . Interestingly, the 3' UTR of genes is thought to be important in mediating the post-transcriptional processing of proteins. This is because the 3' UTR can possess a variety of cis-acting regulatory elements that act as binding sites for trans-acting elements such as RNA-binding proteins or RNAs themselves, which can mediate gene expression via alternative splicing (Matoulkova *et al.*, 2012). This may imply that the differential methylation we observed in this region of the *APOE* gene may affect the post-transcriptional processing and regulation of the *APOE* protein. However, it is worth considering that the difference in DNA methylation in the current study is very small (average of 1.11% and 1.02% difference in STG and PFC for AD pathology, respectively).It is therefore worth considering whether these are functionally or biologically relevant. At many of the loci where I observed a Bonferroni-significant methylation difference in disease this was equivalent to ~1% increase in disease compared to control, which represents that ~1% of cells that were unmethylated have now become fully methylated at this CpG. Although this is unlikely to have considerable functional relevance this does require further exploration as we saw the same pattern across multiple adjacent CpGs in many different cohorts and tissues. Indeed, it will be of interest in the future to utilise epigenetic editing techniques to alter DNA methylation in *APOE* at these CpG sites and to examine the downstream consequences on cell function, particularly in the presence of Aβ. .

A recent study of the 13 *APOE* CpG sites using Illumina 450 cell methylation data from various cell types obtained from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) and the Encyclopedia of DNA Elements consortium, indicates that DNA methylation is mostly positively correlated with *APOE* expression in three of the four CpG sites in the 3'UTR CG (Ma *et al.*, 2015). This, taken together with our results suggest that methylation of these CpGs are critical to an overall shift towards increased *APOE* gene expression in AD, and possible cis-mediation of the 5' TSS of the CGI by a 3' DMR as a transcriptional enhancer. *APOE* transcriptional upregulation has also been demonstrated to be positively correlated with worse cognitive performance, increased expression of AD biomarkers, and accelerated *Aβ* aggregation and proteinopathy (K Gottschalk and Mihovilovic, 2016). However, contradictory studies showed accelerated breakdown of Aβ with enhanced expression of *APOE* (Jiang et al., 2008), which may provide a strong argument for the impairment of APOE function via methylation-induced post-transcriptional processing. Conversely, there is evidence both for (Shi *et al.*, 2017) and against (Morris *et al.*, 2010) APOE exacerbating the development of Tau pathology, which may indicate the correlation with differential methylation found in this study may

not signify a direct causal link and further research is needed in to the downstream consequences of altering methylation of the 3' UTR CGI of *APOE*.

The two SNPs that confer APOE genotype (rs429358 and rs7412), which are two of the largest genetic risk factors for LOAD, are located between cg05501958 and cg18799241 within the 3' UTR CGI. Interestingly, although we observed Bonferroni-significant AD diagnosis and Braak pathology-associated hypermethylation at both these loci in the STG, and Braak pathology-associated hypermethylation at cg18799241 in the PFC, we did not observe any mQTL in this region in the AD and control samples. This suggests that the association of *APOE* genotype and epigenotype (hypomethylation) with AD are potentially via independent mechanisms. Concerning the influence of genotype, previous mQTL analyses seemed to show genotype only impacted methylation in the 3' CGI in AD when comparing *ε3/ε4* versus *ε4/ε4* (Foraker *et al.*, 2015). Together with our results, this implies the *ε4* allele interacts with AD when epigenetically mediating *APOE*. A recent study used the ROS/MAP data to investigate whether methylation differences in the *APOE* region are related to *APOE ε4* genotype and AD neuropathological (neuritic plaque) burden (Chibnik *et al.*, 2015). They showed that there are large DNA methylation differences that are associated with genotype, but these do not seem to mediate the effect of genotype on outcome measures (Chibnik *et al.*, 2015). In this study they did highlight that one CpG in the 3' UTR CGI (cg18799241) shows increased levels of methylation with greater neuritic plaque burden, but is independent of *APOE ε4.* The authors suggest that non-genetic factors may influence AD pathology through differences in DNA methylation, which could lead to different gene expression.

4.2 *APOE* DNA methylation differs with age in AD samples

In the AD samples we observed a Bonferroni-significant association in the STG of decreased DNA methylation with advancing age for the three "Group 1" CpG probes, located in the promoter (cg14123992, cg04406254, cg01032398) as well as the "Group 2" probe cg06750524. This association was not observed in the PFC or in the control samples. One possible explanation for this association in the AD samples only could be because these individuals are more likely to have an *APOE ε4* allele than control individuals. However, we did not observe any interaction of age with *APOE* genotype on DNA methylation at these CpG sites, nor did we see any effect of *APOE* genotype on DNA methylation at these CpG sites in the post-natal samples, which includes individuals with the genotype, but have not developed disease. It is therefore possible that hypomethylation of these four probes in the AD samples with advancing age may be driven by other disease mechanisms such as, for example, pathology in the STG specifically, and warrants further investigation.

4.3 DNA methylation at the 5' end of the APOE gene is driven by a mQTL in post-natal PFC samples

Our study determined that two probes (cg14123992 and cg04406254), which are "Group 1" probes, situated at the 5' end of the *APOE* gene near the TSS show differential methylation with respect to

APOE genotype in post-natal brain PFC brain samples. At both of these two loci, increased DNA methylation was observed with increasing numbers of *APOE ε4* alleles. Interestingly, although *APOE* genotype is associated with AD diagnosis, we did not observe any AD-associated differential DNA methylation in this 5' region. It is intriguing that we did not observe any mQTLs in the control/AD cohort, or the pre-natal samples which is most likely due to power, given the smaller sample sizes in those analyses, and may also have been affected by the greater distribution of APOE *ε4* homozygous genotypes in our post-natal cohorts. These may also have generated observed genotype frequencies differing from Hardy-Weinberg expected APOE genotype frequencies for the general population, for example the Arizona and Mount Sinai cohorts had a different distribution of *APOE* genotypes compared to the general population, although this was not the case for the London or ROSMAP cohorts In the case of the pre-natal samples it may also suggest a tissue-specific effect, given that that analysis used bulk cortical tissue, whilst our post-natal analyses used PFC. A previous study has shown that cg14123992 shows a tissue-specific mQTL; Smith and colleagues highlighted a significant mQTL in the frontal cortex, temporal cortex and pons, which was not observed in the cerebellum or peripheral blood samples (Smith *et al.*, 2014). Given that the two sites we identified as having a mQTL in post-natal samples are located at the 5' end of the gene, close to the TSS, this may suggest that they could be cis-factors influencing transcription and further research should be undertaken to link the genetic and epigenetic differences we have identified to differences in *APOE* gene expression. These two CpG sites we identified are also located closely to another SNP (rs405509). This SNP is located at the 5' end of the *APOE* gene within the promoter and possession of the T allele is reported to modulate the effect of *APOE ε4* genotype on cognitive performance and brain grey matter structure (Ma *et al.*, 2016). In addition it has been shown to alter the methylation of the nearby 5' CpG cg04406254 as well as cg06750524 and cg16471933 located further towards the 3' end of the gene (Ma *et al.*, 2015). The same study also found a significant correlation between cg04406254 methylation and age. There may thus be a possible cooperative effect in epigenetic mediation of gene expression, and AD pathology by extension that is worsened with age, which is supported by prior observations of rs405509 interacting with *APOE ε4* to impair cognition/neural functional connectivity in non-demented seniors (Ma *et al.*, 2016). In the future it will be of interest to assess the relationship between rs05509 genotype and *APOE* methylation in the context of AD.

Further evidence of negative effects of *APOE* DNA methylation on cognition, include an inverse association noted between methylation of the Group 1 CpGs at the 5' promoter (cg04406254, cg01032398) or the Group 3 CpG (cg18768621) in the 3' UTR in whole blood samples, with cognitive performance as assessed by delayed recall in individuals >60 years (Liu *et al.*, 2018). However, it is worth noting that the published study was undertaken on blood samples, rather than brain tissue, and that they did not observe any associations with other cognitive measures. The authors did use a publically available resource to look at DNA methylation correlations of the *APOE* probes across matched brain and blood tissue and suggest that CpG sites that are highly methylated in blood are also so in many brain regions. Interestingly, the 3 CpGs identified by Liu and colleagues (alongside cg18799241) have been reported to have a significant correlation of DNA methylation with *APOE*

gene expression in ENCODE data, which was not observed for the rest of the 13 CpG sites (Ma *et al.*, 2015).

4.4 Age is associated with APOE methylation at cg21879725 in foetal brain tissue and at cg16471933 in post-natal brain tissue

Finally, we observed a Bonferroni-significant decrease of DNA methylation at cg21879725, located in the CGI at the 3' end of the UTR, with increasing developmental foetal developmental stage, while no Bonferroni-significant interaction between CpG methylation, age and genotype was apparent in post-natal subjects. This CpG also showed a trend towards a mQTL with *APOE* genotype and towards an interaction of age and genotype, which marginally missed out on our stringent Bonferroni significance level. Our analyses highlighted that particularly in individuals possessing two *APOE ε4* alleles there was a positive correlation of DNA methylation with increasing days post conception, which was not seen in individuals with fewer alleles. In the context of our AD results this is interesting as we saw altered DNA methylation in the 3' UTR in AD brain samples. This may implicate *APOE ε4* as a mQTL affecting foetal gene methylation that is conserved into adulthood, albeit with altered function, similar to foetal mQTLs implicated in other neurological diseases such schizophrenia that may exert foetal-specific effects (Hannon *et al.*, 2015), and introduce a new epigenetic biomarker for early prediction of disease risk. One potential reason that we did not quite reach Bonferroni significance in the pre-natal genotype analyses is because of the low number of foetal samples with an *APOE ε4* genotype. It is therefore important that our findings are validated in a larger group of samples. In addition, we observed an inverse relationship in a heterozygous genotype, perhaps signifying the effect of a protective allele such as APOE ε3 (de-Almada *et al.*, 2012). We also observed a Bonferroni-significant effect of age on DNA methylation at cg16471933 in post-natal brain samples, which is also located in the 3' UTR CGI. However, this particular CpG site was not altered in the pre-natal samples with ageing.

A positive association between APOE methylation at the 3' UTR CGI and age / *APOE ε4* genotype has already been noted by studies of mixed age cohorts (Foraker *et al.*, 2015). Also a greater pathological effect has been previously seen to be exerted by *APOE ε4* with increasing age (L.W. *et al.*, 2016), which could be complimented by increased methylation in this region (Ma *et al.*, 2015) although those studies were limited as they did not study methylation within foetal brain tissue. Our study was limited by fewer pre-natal samples than post-natal samples, making the inference of age methylation trends between the pre-natal and post-natal samples less reliable due to statistical power issues.

4.5 Limitations of our study

Our study represents a robust analysis of *APOE* DNA methylation in AD using multiple independent cohorts. However, there were some limitations with our study. First, our study is limited to the 13 CpG sites covered by the array, and it will be of interest to profile other methylation sites in the gene in the

future, including CpH sites. We acknowledge their limitations as part of the Illumina 450k manifest, and assignation based on functional annotation and linkage disequilibrium (LD) could have been more precise. However, our ultimate goal was to develop and confirm findings of our group's previous APOE methylation studies based on the same Illumina probes (Lunnon *et al.*, 2014), which themselves have been employed frequently in the literature and are thus generally accepted. In future, an improved approach could be an examination across APOE's LD block at all CpG sites in the region. Yet another direction could be delving into the specific genomic contexts within which the aforementioned APOE CpGs are altered in AD.

Second, we have used bisulfite-treated DNA, and thus our results reflect differences in total methylated cytosines, which is a sum of DNA methylation and hydroxymethylation. 5-hydroxymethylcytosine (5hmC) is potentially a functionally discrete epigenetic marker that may have different effects on gene expression to 5mC. 5hmC has been found to be highly prevalent (and thus possibly epigenetically significant) in the neural epigenome, though its status as a unique mark or an intermediate state of methylation is controversial (Sanchez-Mut and Gräff, 2015). Interestingly, risk loci for AD have been found to show a trend towards enrichment near DMRs with reduced 5hmC in excitatory neurons and near enhancer regions in oligodendrocytes (Kozlenkov *et al.*, 2018). Levels of 5hmC are also very dynamic during foetal development (Spiers *et al.*, 2017). Together, it will therefore be of interest to assess 5hmC levels in the *APOE* gene in both AD and pre-natal samples.

Third, we used bulk tissue and so did not examine cell-specific gene methylation and it will be of interest to assess *APOE* DNA methylation in different brain cell types. Fourth, we limited our analyses to the STG and PFC and in the future the relationship of *APOE* DNA methylation and AD should be investigated in other brain regions. Interestingly, although we had more samples for the PFC, the most significant differences (and greater effect sizes) were observed in the STG and the biological basis of this should be further investigated. Fifth, our discovered methylation effects on 3' APOE DMPs by diagnosis and disease pathology were minute, but may still exert a notable impact by aforementioned mechanisms, and could be further validated by cell-specific analyses to determine consistency in all types of neuronal cells, alongside studying the proportion/effect of hydroxymethylation in results. The relationship between cognitive decline and pathology development with differential methylation could also be further refined by quantifying any link between the rate of development of both to degrees of differential methylation in biopsied brain tissue. Sixth, we only assessed the relationship between methylation and genotype using the number of *APOE ε4* alleles and it would be of interest to examine the relationship between the number of *APOE ε2* alleles and DNA methylation for the 13 CpG sites, given that this genotype has been widely reported to be protective for AD risk (Wu and Zhao, 2016). However, although it has been associated with intact cognition, it has been reported to increase AD pathology in the oldest old (Berlau *et al.*, 2009). In addition, given the study by Ma and colleagues shows rs05509 also alters DNA methylation, it will be of interest to study all SNPs in the *APOE* gene.

Seventh, our study of the effect of age and genotype on DNA methylation in the pre-natal cohort had limited sample numbers, particularly for *APOE ε4* homozygotes, making inference of age or genotype-associated methylation trends less reliable due to statistical power issues. Seventh, on the subject of statistical significance, Bonferroni tests assume independence between all tests performed. However, due to the high correlation between DNA methylation at adjacent CpG sites and multiple tissues from the same individuals, the tests are not fully independent. Therefore, by using a Bonferroni significance threshold we are being overly conservative. However, we acknowledge that by performing tests across multiple cohorts and tissues, the significance threshold could have been adjusted for this. Further to this, there were several outliers within our pre-natal methylation data for all genotype groups, which could possibly have skewed the methylation trends we observed and made subsequent extrapolations unreliable. Eighth, it would have been of interest to look at other outcomes (to diagnosis or Braak stage), as it is known that pathology does not always go hand in hand with specific cognitive deficits (Bilgel *et al.*, 2018). In a similar vein we did not address the possibility of cooperative epigenetic mediation between Tau pathology and *APOE ε4* genotype. Finally, although we have shown robust differences in *APOE* DNA methylation in the 3' UTR CGI in AD, this does not appear to be driven by genetic variation. Therefore, we currently do not know whether the epigenetic differences we have identified are causal in disease pathogenesis, or are a consequence of the disease process and further research is required to clarify this.

4.6 Conclusions

Overall, our study has provided evidence of AD-associated hypomethylation of the *APOE* 3' UTR CGI in AD cortex. Methylation of one site (cg21879725) in this region seems to be driven by *APOE ε4* genotype in pre-natal development, although this was not observed in post-natal samples. We did however identify another loci in this CGI that showed hypomethylation with age in post-natal samples (cg16471933). In the promoter at the 5' end of the gene we observed three adjacent loci near the TSS (cg14123992, cg04406254, cg01032398) and one loci in the gene body (cg06750524) that were hypomethylated with advancing age in AD samples. Although the same direction of effect was seen in control individuals, this was not significant. In light of our current findings, future directions for our research would include bisulfite and oxidative- bisulfite sequencing in parallel of the *APOE* gene at single nucleotide resolution, to analyze which of these marks is present on the gene across its entire length (including CpH sites). It is also important to relate our epigenetic findings to gene expression and protein measurements in the same samples. Another avenue would be quantification of DNA modifications in specific neuronal/non-neuronal cell populations, utilizing a cell-sorting protocol akin to that used in a recent AD EWAS of purified neurons and glia (Gasparoni *et al.*, 2018). It will be of interest to expand our study to look at other disease phenotypes, such as amyloid load or cognitive outcomes. In addition, it will be of interest to examine other *APOE* genotypes and SNPs for their relationship with methylation, as well as looking at allele-specific methylation. For the methylation sites not driven by genetic variation, it will be important to examine epigenetic causality using epigenetic editing *in vitro*.

CHAPTER 5 : REFERENCES

(EADI), European Alzheimer's Disease Initiative et al. 2013. "Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease." *Nature genetics* 45(12): 1452–58.

Alzheimer's Association. 2015. "Latest Facts & Figures Report | Alzheimer's Association." *Http://Www.Alz.Org/Facts/Overview.Asp*. http://www.alz.org/facts/overview.asp.

Ashraf, Ghulam et al. 2016. "Recent Updates on the Association Between Alzheimer's Disease and Vascular Dementia." *Medicinal Chemistry* 12(3): 226–37.

Auton, Adam et al. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526(7571): 68–74.

B.J., Kelley. 2015. "Treatment of Mild Cognitive Impairment." *Current Treatment Options in Neurology* 17(9). http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L6053 58489%5Cnhttp://dx.doi.org/10.1007/s11940-015-0372-3%5Cnhttp://resolver.lib.washington.edu/resserv?sid=EMBASE&issn=15343138&id=doi:10.1007%2Fs11940-015-0372-3&atitle=Treatment+o.

Bakulski, Kelly M. et al. 2012. "Genome-Wide DNA Methylation Differences between Late-Onset Alzheimer's Disease and Cognitively Normal Controls in Human Frontal Cortex." *Journal of Alzheimer's Disease* 29(3): 571–88.

Bekris, Lynn M., Chang En Yu, Thomas D. Bird, and Debby W. Tsuang. 2010. "Review Article: Genetics of Alzheimer Disease." *Journal of Geriatric Psychiatry and Neurology* 23(4): 213–27.

Berlau, Daniel J., María M. Corrada, Elizabeth Head, and Claudia H. Kawas. 2009. "ApoE E2 Is Associated with Intact Cognition but Increased Alzheimer Pathology in the Oldest Old." *Neurology* 72(9): 829–34.

Bilgel, Murat et al. 2018. "Effects of Amyloid Pathology and Neurodegeneration on Cognitive Change in Cognitively Normal Adults." *Brain* 141(8): 2475–85.

Blanco-Luquin, Idoia et al. 2018. "PLD3 Epigenetic Changes in the Hippocampus of Alzheimer's Disease." *Clinical Epigenetics* 10(1).

Booth, Michael J. et al. 2013. "Oxidative Bisulfite Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine." *Nature Protocols* 8(10): 1841–51.

Braak, Heiko et al. 2006. "Staging of Alzheimer Disease-Associated Neurofibrillary Pathology Using Paraffin Sections and Immunocytochemistry." *Acta Neuropathologica* 112(4): 389–404.

Burns, Alistair, and Steve Iliffe. 2009. "Alzheimer ' s Disease." *bmj* 338(February): 467–71.

Casey, David A, Demetra Antimisiaris, and James O'Brien. 2010. "Drugs for Alzheimer's Disease: Are They Effective?" *P & T: a peer-reviewed journal for formulary management* 35(4): 208–11. http://www.ncbi.nlm.nih.gov/pubmed/20498822

Chen, Yi An et al. 2013. "Discovery of Cross-Reactive Probes and Polymorphic CpGs in the Illumina Infinium HumanMethylation450 Microarray." *Epigenetics* 8(2): 203–9.

Chibnik, Lori B. et al. 2015. "Alzheimer's Loci: Epigenetic Associations and Interaction with Genetic Factors." *Annals of Clinical and Translational Neurology* 2(6): 636–47. http://doi.wiley.com/10.1002/acn3.201 (July 11, 2019).

Ciuculete, Diana M. et al. (2017) 'A methylome-wide mQTL analysis reveals associations of methylation sites with GAD1 and HDAC3 SNPs and a general psychiatric risk score', *Translational Psychiatry*. doi: 10.1038/tp.2016.275.

Cummings, Jeffrey. 2012. "Alzheimer's Disease Diagnostic Criteria: Practical Applications." *Alzheimer's Research and Therapy* 4(4).

Das, Sayantan et al. 2016. "Next-Generation Genotype Imputation Service and Methods." *Nature genetics* 48(10): 1284–87. http://www.ncbi.nlm.nih.gov/pubmed/27571263%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5157836.

Davis, Sean, and Sven Bilke. 2018. *An Introduction to the Methylumi Package.* https://bioconductor.org/packages/release/bioc/vignettes/methylumi/inst/doc/methylumi.pdf (April 21, 2019).

De Jager, Philip L et al. 2014. "Alzheimer's Disease: Early Alterations in Brain DNA Methylation at ANK1, BIN1, RHBDF2 and Other Loci." *Nature Neuroscience* 17: 1156. https://doi.org/10.1038/nn.3786.

de-Almada, B. V.P. et al. 2012. "Protective Effect of the APOE-E3 Allele in Alzheimer's Disease." *Brazilian Journal of Medical and Biological Research* 45(1): 8–12.

Deaton, Aimée M., and Adrian Bird. 2011. "CpG Islands and the Regulation of Transcription." *Genes and Development* 25(10): 1010–22.

Devanand, Davangere P. et al. (2012) 'MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease', *NeuroImage*. doi: 10.1016/j.neuroimage.2012.01.075.

Ellison, Elizabeth M., Melissa A. Bradley-Whitman, and Mark A. Lovell. 2017. "Single-Base Resolution Mapping of 5-Hydroxymethylcytosine Modifications in Hippocampus of Alzheimer's Disease Subjects." *Journal of Molecular Neuroscience* 63(2): 185–97.

Foraker, Jessica et al. 2015. "The APOE Gene Is Differentially Methylated in Alzheimer's Disease." *Journal of Alzheimer's Disease* 48(3): 745–55.

Gasparoni, Gilles et al. 2018. "DNA Methylation Analysis on Purified Neurons and Glia Dissects Age and Alzheimer's Disease-Specific Changes in the Human Cortex." *Epigenetics and Chromatin* 11(1).

Gatz, Margaret et al. 2006. "Role of Genes and Environments for Explaining Alzheimer Disease." *Archives of General Psychiatry* 63(2): 168–74.

Giri, Mohan, Man Zhang, and Yang Lü. 2016. "Genes Associated with Alzheimer's Disease: An Overview and Current Status." *Clinical Interventions in Aging* 11: 665–81.

Guerreiro, Rita et al. 2013. "TREM2 Variants in Alzheimer's Disease." *New England Journal of Medicine* 368(2): 117–27.

Guintivano, Jerry, Martin J. Aryee, and Zachary A. Kaminsky. 2013. "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(3): 290–302.

Hane, F.T., B.Y. Lee, and Z. Leonenko. 2017. "Recent Progress in Alzheimer's Disease Research, Part 1: Pathology." *Journal of Alzheimer's Disease* 57(1).

Hane, Francis T., Brenda Y. Lee, and Zoya Leonenko. 2017. "Recent Progress in Alzheimer's Disease Research, Part 1: Pathology." *Journal of Alzheimer's Disease* 57(1): 1–28.

Hannon, Eilis et al. (2015) 'Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci', *Nature Neuroscience*, 19(1), pp. 48–54. doi: 10.1038/nn.4182.
Hannon, Eilis et al. (2018) 'Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2018.09.007.

Hannon, Eilis et al. 2015. "Methylation QTLs in the Developing Brain and Their Enrichment in Schizophrenia Risk Loci." *Nature Neuroscience* 19(1): 48–54.

Herholz, Karl, and Klaus Ebmeier. 2011. "Clinical Amyloid Imaging in Alzheimer's Disease." *The Lancet Neurology* 10(7): 667–70.

Herrmann, Nathan, Serge Gauthier, Neli Boneva, and Ole Michael Lemming. 2013. "A Randomized, Double-Blind, Placebo-Controlled Trial of Memantine in a Behaviorally Enriched Sample of Patients with Moderate-to-Severe Alzheimer's Disease." *International Psychogeriatrics* 25(6): 919–27.

Hogan, David B. 2015. "Re: Long-Term Efficacy and Toxicity of Cholinesterase Inhibitors in the Treatment of Alzheimer Disease." *Canadian journal of psychiatry. Revue canadienne de psychiatrie* 60(7): 338.

Huang, Yu Wen Alvin, Bo Zhou, Marius Wernig, and Thomas C. Südhof. 2017. "ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and Aβ Secretion." *Cell* 168(3): 427-441.e21.

Huijbers, Willem et al. (2014) 'Amyloid deposition is linked to aberrant entorhinal activity among cognitively normal older adults', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.3579-13.2014.

Huynh, Rose Ann, and Chandra Mohan. 2017. "Alzheimer's Disease: Biomarkers in the Genome, Blood, and Cerebrospinal Fluid." *Frontiers in Neurology* 8(MAR).

Johnson, Keith A., Nick C. Fox, Reisa A. Sperling, and William E. Klunk. 2012. "Brain Imaging in Alzheimer Disease." *Cold Spring Harbor Perspectives in Medicine* 2(4).

Jonsson, T et al. 2013. "Variant of TREM2 Associated with the Risk of AD." *New England Journal of Medicine* 368(2): 107–16.

K Gottschalk, William, and Mirta Mihovilovic. 2016. "The Role of Upregulated APOE in Alzheimer's Disease Etiology." *Journal of Alzheimer's Disease & Parkinsonism* 06(01).

Khan, Usman A. et al. (2014) 'Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease', *Nature Neuroscience*. doi: 10.1038/nn.3606.

Knopman, David S, and Richard J Caselli. 2012. "Appraisal of Cognition in Preclinical Alzheimer's Disease: A Conceptual Review." *Neurodegenerative Disease Management* 2(2): 183–95.

Kozlenkov, Alexey et al. 2018. "A Unique Role for DNA (Hydroxy)Methylation in Epigenetic Regulation of Human Inhibitory Neurons." *Science Advances* 4(9).

Kroeze, Leonie I., Bert A. van der Reijden, and Joop H. Jansen. 2015. "5-Hydroxymethylcytosine: An Epigenetic Mark Frequently Deregulated in Cancer." *Biochimica et Biophysica Acta - Reviews on Cancer* 1855(2): 144–54.

L.W., Bonham et al. 2016. "Age-Dependent Effects of APOE E4 in Preclinical Alzheimer's Disease." *Annals of Clinical and Translational Neurology* 3(9): 668–77.

Lambert, Jean Charles et al. 2013. "Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease." *Nature Genetics* 45(12): 1452–58.

Lev Maor, Galit, Ahuvi Yearim, and Gil Ast. 2015. "The Alternative Role of DNA Methylation in Splicing Regulation." *Trends in Genetics* 31(5): 274–80.

Liu, Chia Chan, Takahisa Kanekiyo, Huaxi Xu, and Guojun Bu. 2013. "Apolipoprotein e and Alzheimer Disease: Risk, Mechanisms and Therapy." *Nature Reviews Neurology* 9(2): 106–18.

Liu, Jiaxuan et al. 2018. "DNA Methylation in the APOE Genomic Region Is Associated with Cognitive Function in African Americans." *BMC Medical Genomics* 11(1).

Lord, Jenny, Alexander J. Lu, and Carlos Cruchaga. 2014. "Identification of Rare Variants in Alzheimer's Disease." *Frontiers in Genetics* 5(OCT).

Lunnon, Katie et al. 2014. "Cross-Tissue Methylomic Profiling Strongly Implicates a Role for Cortex-Specific Deregulation of ANK1 in Alzheimer's Disease Neuropathology." *Nat Neurosci* 17(9): 1164–70.

M., Mehta, and Adem A. 2012. "New Acetylcholinesterase Inhibitors for Alzheimer's Disease." *International Journal of Alzheimer's Disease*: 728983.

Ma, C et al. 2016. "Is There a Significant Interaction Effect between Apolipoprotein E Rs405509 T/T and E4 Genotypes on Cognitive Impairment and Gray Matter Volume?" *European journal of neurology* 23(9): 1415–25. http://www.ncbi.nlm.nih.gov/pubmed/27259692 (July 11, 2019).

Ma, Yiyi et al. 2015. "Genetic Variants Modify the Effect of Age on APOE Methylation in the Genetics of Lipid Lowering Drugs and Diet Network Study." *Aging cell* 14(1): 49–59. http://www.ncbi.nlm.nih.gov/pubmed/25476875 (July 11, 2019).

Masters, Mary Clare, John C. Morris, and Catherine M. Roe. 2015. "'Noncognitive' Symptoms of Early Alzheimer Disease : A Longitudinal Analysis." *Neurology* 84(6): 617–22.

Matoulkova, Eva, Eva Michalova, Borivoj Vojtesek, and Roman Hrstka. 2012. "The Role of the 3′ Untranslated Region in Post-Transcriptional Regulation of Protein Expression in Mammalian Cells." *RNA Biology* 9(5): 563–76.

Matthews, Kevin A. et al. 2018. "Racial and Ethnic Estimates of Alzheimer's Disease and Related Dementias in the United States (2015–2060) in Adults Aged ≥65 Years." *Alzheimer's & Dementia*.

Mayor, Susan. 2015. "Signs of Depression and Apathy Precede Memory Problems in Alzheimer's Disease, Study Shows." *BMJ (Online)* 350.

Morris, John C. et al. 2010. "APOE Predicts Amyloid-Beta but Not Tau Alzheimer Pathology in Cognitively Normal Aging." *Annals of Neurology* 67(1): 122–31.

Mukhin, V. N., K. I. Pavlov, and V. M. Klimenko. 2017. "Mechanisms of Neuron Loss in Alzheimer's Disease." *Neuroscience and Behavioral Physiology* 47(5): 508–16.

Naj, A.C. et al. 2014. "Age-at-Onset in Late Onset Alzheimer Disease Is Modified by Multiple Genetic Loci." *JAMA Neurol* 71(11): 1394–1404.

Niikura, Takako, Hirohisa Tajima, and Yoshiko Kita. 2006. "Neuronal Cell Death in Alzheimers Disease and a Neuroprotective Factor, Humanin." *Current Neuropharmacology* 4(2): 139–47.

Nitrini, Ricardo. 2010. "Diagnóstico Pré-Clínico Da Doença de Alzheimer: Prevenção Ou Vaticínio?" *Dement. neuropsychol* 4(4). http://demneuropsy.com.br/detalhe_artigo.asp?id=237.

Perl, Daniel P. 2010. "Neuropathology of Alzheimer's Disease." *Mount Sinai Journal of Medicine* 77(1): 32–42.

Pidsley, Ruth et al. 2016. "Critical Evaluation of the Illumina MethylationEPIC BeadChip Microarray for Whole-Genome DNA Methylation Profiling." *Genome Biology* 17(1).

R Core Team (2018). "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria." https://www.r-project.org/ (April 21, 2019).

Radmanesh, F., Devan, W. J., Anderson, C. D., Rosand, J., & Falcone, G. J. (2014). Accuracy of imputation to infer unobserved APOE epsilon alleles in genome-wide genotyping data. *European Journal of Human Genetics.* https://doi.org/10.1038/ejhg.2013.308

Reitz, Christiane, and Richard Mayeux. 2014. "Alzheimer Disease: Epidemiology, Diagnostic Criteria, Risk Factors and Biomarkers." *Biochemical Pharmacology* 88(4): 640–51.

Reitz, Christiane, Carol Brayne, and Richard Mayeux. 2011. "Epidemiology of Alzheimer Disease." *Nature Reviews Neurology.*

Ridge, Perry G. et al. 2016. "Assessment of the Genetic Variance of Late-Onset Alzheimer's Disease." *Neurobiology of aging* 41.

Sanchez-Mut, Jose V., and Johannes Gräff. 2015. "Epigenetic Alterations in Alzheimer's Disease." *Frontiers in Behavioral Neuroscience* 9.

Sanchez-Mut, Jose Vicente et al. 2014. "Promoter Hypermethylation of the Phosphatase DUSP22 Mediates PKA-Dependent TAU Phosphorylation and CREB Activation in Alzheimer's Disease." *Hippocampus* 24(4): 363–68.

Seward, M. E. et al. 2013. "Amyloid- Signals through Tau to Drive Ectopic Neuronal Cell Cycle Re-Entry in Alzheimer's Disease." *Journal of Cell Science* 126(5): 1278–86.

Shi, Yang et al. 2017. "ApoE4 Markedly Exacerbates Tau-Mediated Neurodegeneration in a Mouse Model of Tauopathy." *Nature* 549(7673): 523–27.

Smith, Alicia K. et al. 2014. "Methylation Quantitative Trait Loci (MeQTLs) Are Consistently Detected across Ancestry, Developmental Stage, and Tissue Type." *BMC Genomics* 15(1).

Smith, Rebecca G. et al. 2018. "Elevated DNA Methylation across a 48-Kb Region Spanning the HOXA Gene Cluster Is Associated with Alzheimer's Disease Neuropathology." *Alzheimer's and Dementia* 14(12): 1580–88.

Spiers, Helen et al. (2015) 'Methylomic trajectories across human fetal brain development', *Genome Research*. doi: 10.1101/gr.180273.114.

Spiers, Helen et al. 2017. "5-Hydroxymethylcytosine Is Highly Dynamic across Human Fetal Brain Development." *BMC Genomics* 18(1).

Tansey, Katherine E., Darren Cameron, and Matthew J. Hill. 2018. "Genetic Risk for Alzheimer's Disease Is Concentrated in Specific Macrophage and Microglial Transcriptional Networks." *Genome Medicine* 10(1).

Tulloch, Jessica et al. 2018. "APOE DNA Methylation Is Altered in Lewy Body Dementia." *Alzheimer's and Dementia* 14(7): 889–94.

Tulloch, Jessica et al. 2018. "Glia-Specific APOE Epigenetic Changes in the Alzheimer's Disease Brain." *Brain Research* 1698: 179–86.

Varvel, N. H. et al. 2008. "A Oligomers Induce Neuronal Cell Cycle Events in Alzheimer's Disease." *Journal of Neuroscience* 28(43): 10786–93.

Vavouri, Tanya, and Ben Lehner. 2012. "Human Genes with CpG Island Promoters Have a Distinct Transcription-Associated Chromatin Organization." *Genome Biology* 13(11).

Villemagne, Victor L. et al. 2018. "Imaging Tau and Amyloid-? Proteinopathies in Alzheimer Disease and Other Conditions." *Nature Reviews Neurology* 14(4): 225–36.

Wang, Liang et al. 2016. "Evaluation of Tau Imaging in Staging Alzheimer Disease and Revealing Interactions between β-Amyloid and Tauopathy." *JAMA Neurology* 73(9): 1070–77.

Wang, Xingbin et al. 2012. "An r Package Suite for Microarray Meta-Analysis in Quality Control, Differentially Expressed Gene Analysis and Pathway Enrichment Detection." *Bioinformatics* 28(19): 2534–36.

Wu, Long, and Liqin Zhao. 2016. "ApoE2 and Alzheimer′s Disease: Time to Take a Closer Look." *Neural Regeneration Research* 11(3): 412.

Zhao, Jinying et al. 2017. "A Genome-Wide Profiling of Brain DNA Hydroxymethylation in Alzheimer's Disease." *Alzheimer's and Dementia* 13(6): 674–88.

Appendix 1

```
#1. Data organisation
## Data frames for analysis were constructed as follows:
### 1.1 Reading in data
library("wateRmelon")
setwd("/mnt/data1/Leighton/LBBF/LBB")
read.csv(file = "LBBF.dasen.betas.csv", header = TRUE, row.names = 1) ->
LBBF_betas #Beta
values of DMP analysis cohorts
read.csv(file = "LBBFpheno.csv", header = TRUE, row.names = 1) -> LBBF_pheno
#Phenotype
data of DMP analysis cohorts
LBBF_betas[c('cg01032398', 'cg04406254', 'cg05501958', 'cg06750524',
'cg08955609',
'cg12049787', 'cg14123992', 'cg16471933', 'cg18768621', 'cg18799241',
'cg19514613',
'cg21879725', 'cg26190885'), ] -> LBBF_probes
sub("X", "", colnames (LBBF_probes)) -> colnames(LBBF_probes)
t(LBBF_probes) -> LBBF_probes_reorganized
merge(LBBF_probes_reorganized, LBBF_pheno, by = "row.names") ->
LBBFpheno.2
rownames(LBBFpheno.2)<-LBBFpheno.2$Row.names
### 1.2 Inserting genotype data
LBBFpheno.2[ ,c('APOE.genotype')] -> LBBF_genotype
write.csv(LBBF_genotype, file = "LBBF_genotype.csv")
read.csv(file = "LBBF_genotype2.csv", header = TRUE) -> LBBF_genotype
cbind(LBBFpheno.2, LBBF_genotype)-> LBBFpheno.3
### 1.3 Calculating and inserting cell proportion data
setwd("/mnt/data1/reference_files/CETs/")
install.packages("cets_0.99.2.tar.gz")
library(cets)
load("cetsBrain.rda")
load("cetsDilution.rda")
modelIdx <- list(neuron = pdBrain$celltype == "N", glia = pdBrain$celltype ==
"G")
refProfile <- getReference(brain, modelIdx)
head(refProfile)
propLBB1F <- as.data.frame(estProportion(LBBF_betas, profile = refProfile))
colnames(propLBB1F)<-"cellprop"
sub("X", "", rownames (propLBB1F)) -> rownames(propLBB1F)
merge(LBBFpheno.3, propLBB1F, by="row.names")->LBBFpheno.4
LBBFpheno.4$Row.names <- NULL
rownames(LBBFpheno.4)<-LBBFpheno.4$Row.names
### 1.4 Inserting Braak stage data
LBBFpheno.2[, c('Analysis.2..Braak.Stage.')] -> LBBF_braak
write.csv(LBBF_braak, file = 'LBBF_braak.csv')
read.csv(file = 'LBBF_braak2.csv', header = TRUE) -> LBBF_braak
```

```
cbind(LBBFpheno.4, LBBF_braak)-> LBBFpheno.5
head(LBBFpheno.5)
### 1.5 Inserting diagnosis data
LBBFpheno.2[, c('Analysis.1..AD.v.CTL.')] -> LBBF_diagnosis
write.csv(LBBF_diagnosis, file = "LBBF_diagnosis.csv")
read.csv(file = 'LBBF_diagnosis2.csv', header = TRUE) -> LBBF_diagnosis
cbind(LBBFpheno.5, LBBF_diagnosis)-> LBBFpheno.6
head(LBBFpheno.6)
### 1.6 Inserting gender data
LBBFpheno.2[, c('Gender')] -> LBBF_gender
write.csv(LBBF_gender, file = "LBBF_gender.csv")
read.csv(file = 'LBBF_gender2.csv', header = TRUE) -> LBBF_gender
cbind(LBBFpheno.6, LBBF_gender) -> LBB1.F.pheno
#Complete phenotype data frame

# 2. Data QC and normalisation
## QC and normalisation was conducted as follows:
### 2.1 Loading in data
library("wateRmelon")
setwd("/mnt/data1/Leighton/LBBF/LBB")
read.csv("LBBFpheno.csv", header=T, row.names=1)->LBBFpheno
DataLBBF<-methylumIDAT(rownames(LBBFpheno),idatPath=getwd())
#Raw data
### 2.2 Quality control
#### 2.2.1 Overall check of data:
setwd("/mnt/data1/Leighton/")
tiff("test.tiff")
par(mar=c(6,2,1,1)+ 0.1)
boxplot(log(methylated(DataLBBF)), las=2, cex.axis=0.6 , cex=0.6)
dev.off()
tiff("test2.tiff")
par(mar=c(6,2,1,1)+ 0.1)
boxplot(log(unmethylated(DataLBBF)), las=2, cex.axis=0.6 , cex=0.6)
dev.off()
### 2.3 Checking for odd samples/outliers/bad data/samples:
#### 2.3.1 Sex check using X and Y chromosomes
DataLBBF@featureData@data->FD
betas(DataLBBF)-> DataLBBF.betas
merge(FD, DataLBBF.betas, by = "row.names", all = TRUE)->DataLBBFFD
t(DataLBBFFD)->DataLBBFFDt
load("/mnt/data1/Bex/probesub.rda")
probesub[probesub$CHR =="X",]->Xprobes
Fmd <- cmdscale(dist(t(exprs(DataLBBF)[rownames(fData(DataLBBF)) %in%
rownames(Xprobes),])),2)
merge(LBBFpheno, Fmd, by = "row.names", all = TRUE)->LBBFXinfo
tiff("sexcheckraw.tiff")
plot(V2~V1,LBBFXinfo,col=(LBBFXinfo$Gender))
dev.off()
### 2.4 Genotype checks using SNP probes
```

```
#### 2.4.1 Checking individual traces
plot(density(DataLBBF.betas[,1], na.rm=T), ylim=c(0,4))
#### 2.4.2 Per sample density of beta values
setwd("/mnt/data1/Leighton/")
pdf(paste("DataLBBF_traces"))
densityPlot(DataLBBF.betas)
dev.off()
#### 2.4.3 Average density of beta values
pdf(paste("DataLBBF_traces_average"))
plot(density(betas ,na.rm=T), ylim=c(0,4), col='black')
dev.off()
### 2.5 Filtering samples with >0.05 p-values in 5% of sites
pfilter(DataLBBF, perc =5)->DataLBBF.pf
### 2.6 Data Normalisation
dasen(DataLBBF.pf)->DataLBBF.pf.dasen
#The dasen function performs quantile normalisation of beta values
### 2.7 Extraction of beta values
betas(DataLBBF.pf.dasen)->LBBF.dasen.betas
write.csv(LBBF.dasen.betas, file = "LBBF.dasen.betas.csv")
#Final normalised and corrected beta values


# 3. Analysis
## Linear regression analyses for independent variables and methylation were
conducted as follows:
### 3.1 Braak stage analysis
for (i in 2:14) {
Braak_Regression <- lm(LBB1.F.pheno [, i]~Age + Gender + Cell_Proportion +
Braak_Score, data = LBBFphenotype)
print(summary(Braak_Regression)$coeff[,c(1,4)])
}
### 3.2 Genotype analysis
for (i in 2:14) {
Genotype_Regression <- lm(LBB1.F.pheno [, i]~Age + Gender + Cell_Proportion +
Genotype, data = LBBFphenotype)
print(summary(Genotype_Regression)$coeff[,c(1,4)])
}
### 3.3 Diagnosis analysis
for (i in 2:14) {
Diagnosis_Regression <- lm(LBB1.F.pheno [, i]~ Age + Gender + Cell_Proportion +
Diagnosis, data = LBBFpheno.7)
print(summary(Diagnosis_Regression)$coeff[,c(1,4)])
}
### 3.4 Meta-analysis
#### 3.4.1 Fisher's test of p-values for all cohorts
library(MetaDE)
read.csv(file = "Meta_Analysis2_p.csv", header = TRUE, row.names = NULL) ->
Meta_p
#Database of all cohort p-values
as.matrix(Meta_p[, 2:53]) -> Meta_P
```

```
t(Meta_P) -> Meta_P
list(p = Meta_P, bp = NULL) -> x
MetaDE.pvalue(x, meta.method = c("Fisher"), rth = NULL, miss.tol = 0.3, asymptotic
= FALSE)
#### 3.4.2 Fisher's test of effect sizes for all cohorts
read.csv(file = "New_Meta_Analysis2_es.csv", header = TRUE, row.names = 1) ->
Meta_es
#Database of all cohort effect sizes
Meta_es[,1:52]-> Meta_es_2
t(Meta_es_2)->Meta_es_3
list(ES = Meta_es_3, Var = Probe_var, perm.ES = NULL) -> x
MetaDE.ES(x, meta.method = c("FEM", "REM"))


# 4. Example plotting
## Plot type 1: Box-and-whisker plot
tiff(file = 'STGCCBoxplot.tiff', units = 'cm', width = 50, height = 30, res = 300)
#Plotting methylation % of CG probes with diagnosis in STG cohorts
boxplot(Methylation~cgCC, xlab = 'CG Probes', ylab = 'Corrected % Methylation', col
= c('forestgreen', 'lawngreen'))
legend("topright", c("STG-AD", "STG-C"),fill=c('forestgreen','lawngreen'))
dev.off()
## Plot type 2: Manhattan plot
input <- Manhattan_Table [,c('LBB.STG.Braak.p', 'LBB.STG.Braak.GP')]
#Plotting Braak analysis p-values with genomic position of probes in London STG
cohort
print(head(input))
png(file = 'LBB.STG.Braak_plot.png')
plot(x = input$LBB.STG.Braak.GP, y = input$LBB.STG.Braak.p,
xlab = 'GP',
ylab= 'p-value',
xlim = c(45407868, 45412647),
ylim = c(0.001, 13),
)
dev.off()
## Plot type 3: Strip chart
tiff(file = 'STGBraakStripchart1.tiff', units = 'cm', width = 50, height = 30, res = 300)
#Plotting beta values of STG cohorts with Braak stage
plot(LBBFSTGStrip[[1]]~bbscoreLBBFSTG, xlab = "Braak stage", ylab =
'Methylation',
col = c('forestgreen'), ylim = c(70, 90))
abline(lm(LBBFSTGStrip[[1]]~bbscoreLBBFSTG), lty = c(1), col = 'forestgreen')
legend ("topright", c('London Brain Bank', 'Mount Sinai Brain Bank', 'Arizona Brain
Bank'), pch = c(1, 0, 2), lty = c(1,2,3), col = 'forestgreen')
points(MSSTGStrip[[1]]~bbscoreMSSTG, col = 'forestgreen', pch = 0)
abline(lm(MSSTGStrip[[1]]~bbscoreMSSTG),lty=2, col="forestgreen")
points(ArizonaSTGStrip[[1]]~bbscoreArizonaSTG, col = 'forestgreen', pch = 2)
abline(lm(ArizonaSTGStrip[[1]]~bbscoreArizonaSTG), lty=3, col="forestgreen")
dev.off()
## Plot type 4: Effect size plot
```

```r
plot.1 <- ggplot(data = MetaBraakSTG) #Plot of overall effect size of Braak stage on
methylation in STG cohorts
tiff("MetaBraakSTGggplot.tiff", units = "cm", width = 25, height = 25, res = 300)
plot.1 + geom_point(alpha = 0.3,aes(x=MAPINFO, y=LBB_A_ESTOT.100, colour =
"London"))
+
geom_point(alpha = 0.3,aes(x=MAPINFO, y=MS_A_ESTOT.100, colour = "Mount
Sinai"))+
geom_point(alpha = 0.3,aes(x=MAPINFO, y=ARIZ_A_ESTOT.100, colour =
"Arizona"))+
geom_smooth(aes(x=MAPINFO, y=LBB_A_ESTOT.100, colour = "London"),se =
FALSE) +
geom_smooth(aes(x=MAPINFO, y=MS_A_ESTOT.100, colour = "Mount Sinai"),se =
FALSE)+
geom_smooth(aes(x=MAPINFO, y=ARIZ_A_ESTOT.100, colour = "Arizona"),se =
FALSE)+
theme(legend.justification=c(1,1), legend.position=c(1,1))+ xlab("Genomic Location")
+
ylab("Methylation Effect Size") + scale_colour_discrete(name = "")
dev.off()
## Plot type 5: Scatter plot
### Adjusted models for scatter plots were generated as follows:
adjustedmodel <- function(Meth, measure1, measure2, measure3){
model <- lm(Meth ~ measure1 + measure2 + measure3)
return(residuals(model) + coefficients(model)[1])
}
#### Covariates:
LBB1.F.pheno$Age..at.death. -> Age
LBB1.F.pheno$Gender2 -> Gender
LBB1.F.pheno$cellprop -> Cell_Proportion
#### Independent variables:
LBB1.F.pheno$bbscore -> Braak_Score
LBB1.F.pheno$Diagnosis2 -> Diagnosis
LBB1.F.pheno$APOE.genotype2 -> Genotype
cg01032398 <-adjustedmodel(as.numeric(LBBFpheno.7[,2]), Age, Gender,
Cell_Proportion)
cg04406254 <- adjustedmodel(as.numeric(LBBFpheno.7[,3]), Age, Gender,
Cell_Proportion)
cg05501958 <- adjustedmodel(as.numeric(LBBFpheno.7[,4]), Age, Gender,
Cell_Proportion)
cg06750524 <- adjustedmodel(as.numeric(LBBFpheno.7[,5]), Age, Gender,
Cell_Proportion)
cg08955609 <- adjustedmodel(as.numeric(LBBFpheno.7[,6]), Age, Gender,
Cell_Proportion)
cg12049787 <- adjustedmodel(as.numeric(LBBFpheno.7[,7]), Age, Gender,
Cell_Proportion)
cg14123992 <- adjustedmodel(as.numeric(LBBFpheno.7[,8]), Age, Gender,
Cell_Proportion)
cg16471933 <- adjustedmodel(as.numeric(LBBFpheno.7[,9]), Age, Gender,
```

Cell_Proportion)
cg18768621 <- adjustedmodel(as.numeric(LBBFpheno.7[,10]), Age, Gender,
Cell_Proportion)
cg18799241 <- adjustedmodel(as.numeric(LBBFpheno.7[,11]), Age, Gender,
Cell_Proportion)
cg19514613 <- adjustedmodel(as.numeric(LBBFpheno.7[,12]), Age, Gender,
Cell_Proportion)
cg21879725 <- adjustedmodel(as.numeric(LBBFpheno.7[,13]), Age, Gender,
Cell_Proportion)
cg26190885 <- adjustedmodel(as.numeric(LBBFpheno.7[,14]), Age, Gender,
Cell_Proportion)
LBBF_Adjusted_Model <- cbind(cg01032398, cg04406254, cg05501958,
cg06750524,
cg08955609, cg12049787, cg14123992, cg16471933, cg18768621, cg18799241,
cg19514613, cg21879725, cg26190885) #Collating adjusted models into a single
frame
Scatter graphs were plotted as follows:
for (i in 1:13) {
pdf(paste("LBBF_Probe_Braak_Graph", i))
plot(LBBF_Adjusted_Model [,i]~Braak_Score, xlab = "Braak score", ylab = "Adjusted
beta values") #Plotting beta values with Braak score
abline(lm(LBBF_Adjusted_Model[, i]~Braak_Score))
dev.off()

}