The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Investigation into Patterns of Epigenetic Regulation of Alzheimer's Disease.

Kirsty Alison Boden Bsc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

May 2019

Abstract

Aberrant global and gene specific DNA methylation has been identified in late onset Alzheimer's disease (LOAD) and has also been linked to its pathology. In addition, multiple loci associated with LOAD via genome wide association studies (GWAS) studies are being shown to harbour Alzheimer's disease (AD) associated differential methylation in the LOAD brain. One aim of this thesis was to investigate if any differential methylation could be identified within the promoter regions of LOAD associated loci using leukocyte DNA, both for regions already identified as being aberrantly methylated in AD and those that have not been. This included the promoter regions of the genes *INPP5D, SIRT1, HLA-DRB1/5, SORL1* and *PTK2B*. Of these genes only promoter methylation of *INPP5D* and *SORL1* were identified in LOAD blood samples.

Work was also conducted to investigate the promoter methylation status of the genes S*IRT1, TREM2, ABCA7, MEF2C* and *PTK2ƀ* in sporadic early onset AD samples (sEOAD). The aim being to identify if any LOAD associated differential methylation also occurs in this other sub-group of AD. No differential methylation was identified at the promoter regions of any of these genes in either sEOAD blood (leukocyte) or brain (cortex) tissue. However, significant hypomethylation of a CpG site located upstream of the *MEF2C* promoter was identified in the blood of one sEOAD patient, potentially representing the presence of an AD associated epi-allele. Interestingly a region containing seven CpG sites, located within the *RIN3* 3'UTR, was also found to be hypomethylated in sEOAD leukocyte samples. Both results indicated the importance of non-promoter CpG site methylation in AD.

A whole genome bisulphite sequencing (WGBS) study was also conducted to identify differential methylation occurring within LOAD at differing sages of disease progression. For this cerebellum DNA was used from a moderate (braak stage IV) and severe (braak stage VI) patient and data was compared to published control data. This experiment identified a significant amount of differentially methylated loci in the severe LOAD suffer when compared to the moderate patient sample and twenty two genes were identified as being associated with this aberrant methylation.

Interestingly the majority of differentially methylated cytosines were located within non-CpG dincleotides, suggesting a potential role for non-CpG site methylation in LOAD. In addition, most differential methylation identified was either downstream of the promoter or intragenically located, again suggesting an important role for non-promoter methylation in AD.

Acknowledgements

Firstly I would like to thank my supervisors, Dr Andrew Bottley, Professor Graham Seymour and Professor Kevin Morgan for all of their support and guidance throughout. I would particularly like to thank Andy for his help and constant positivity, which was sometimes very much needed! Thanks also goes to Dr Keeley Brookes and all of the other members of the Human Genetics Group who have had input into the preparation of the DNA contained within the Human DNA Bank at the University of Nottingham or who have answered my panicked emails at any point over the past few years. Of course I would also like to thank the patients who generously donated their tissue, without which this thesis would not be possible. I would also like to thank the BBSRC for funding my project.

I also am extremely grateful to Ping Sun Guiping and all of the other members of the Seymour lab for all of their technical support throughout the project. I would also like to thank Nathan Archer for pretty much always having an answer as to why an experiment might not have been successful and for also helping me to realise that when it does not go to plan, it is ok because science works like that sometimes!

My unreserved appreciation also goes to my friends and family who have supported me throughout the duration of my academic career so far. To my mum who has always been there to remind me that 'you can only do your best' and to my sisters, Jo, Emma and Sarah for always answering the phone when needed or providing a bottle of wine, or two, when the occasion called for it . I would also like to thank Mel and Toni who have made the last couple of years much more bearable and have always been more than willing to provide words of encouragement when needed!

Table of contents

List of figures

List of tables

List of abbreviations

^{18}FDG	Fluorodeoxyglucose
5.10 MTHF	Methylenetetrahydrofolate
5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5hmC	5-hydroxymethlycytosine
5mC	5-methylcytosine
5-MTHF	5-methylenetetrahydrofolate
ABCA7	ATP Binding Cassette Subfamily A Member 7)
AD	Alzheimer's disease
ADAM10	ADAM Metallopeptidase Domain 10
ADGRA1	Adhesion G protein-coupled receptor a1
AEP	Asparaginyl endopeptidase
AICD	APP intracellular domain
AID	Activation induced deaminase
ALS	Amyotrophic lateral sclerosis
ANK1	Ankyrin 1
ANOVA	Analyses of variance
APOBEC1	Apolipoprotein B mRNA-editing enzyme 1
APOE	apolipoprotein E
APP	Amyliod precursor protein
APS	Adenosine 5' phosphosulfate
Arc	Activity-Regulated Cytoskeletal protein
ATP	Adenosine triphosphate
Aβ	Abeta protein
BACE1	β-site APP cleaving enzyme

BBB	Blood brain barrier
BDCA2	Plasmacytoid dendritic cell-specific type II C-type lectin
BDNF	Brain derived neurotrophic factor
BIN1	Bridging integrator 1
BLAST	Basic Local Alignment Search Tool
BSP	Bisulphite sequencing PCR
CASS4	Cas scaffolding protein family member 4
CD2AP	CD2 associated protein
CD33	Sialic acid binding Ig-like lectin 3
CDH23	Cadherin related 23
CELF1	CUGBP elav-Like family member 1
CFS	Cerebral spinal fluid
CGIs	CpG island
ChrM	Mitochondrial chromosome
CNN2	Calponin 2
CpG	Cytosines adjacent to guanines on a strand of DNA
CR	Circadian rhythms
CR1	Complement component (3b/4b) receptor 1
CR1L	Complement C3b/C4b Receptor 1 Like
CTBP2	C-Terminal Binding Protein 2
CTF	C-terminal membrane bound APP fragment
DAP10	DNAX-activating protein 10
DAP12	DNAX-activating protein 12
DAPK1	Death-associated protein kinase
dATPαS	Deoxyadenosine alpha-thio triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DIP2A	Disco interacting protein 2 homolog A

DML	Differentially methylated loci
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferases
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMT2	DNA (cytosine-5)-methyltransferase 2
DNMT3a	DNA (cytosine-5)-methyltransferase 3a
DNMT3b	DNA (cytosine-5)-methyltransferase 3b
dNTPs	Deoxynucleotide triphosphates
dsDNA	Double stranded deoxyribonucleic acid
DSG2	Desmoglein 2
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
ELISA	Enzyme-linked immunosorbent assay
EOAD	Early onset Alzheimer's disease
ER	Endoplasmic reticulum
ERK	Extracellular receptor kinase
fAD	Familial Alzheimer's disease
Fas2	Fasciclin2
FERMT2	Fermitin family member 2
fMRI	Functional MRI
FTD	Frontotemporal dementia
GABRB3	Gamma-aminobutyrica acid type A receptor beta 3 subunit
GAS1	Growth arrest-specific 1
GCF	Granulocyte chemotactic factor
GPCRs	G protein coupled receptors
GSK3	Glycogen synthase kinase-3
GSK3β	Glycogen synthase kinase 3

GTP	Guanosine-5'-triphosphate
GWAS	Genome-wide association studies
HCN2	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated ion channel 2
HCY	homocysteine
HDAC	Histone deacetylases
HDACs	Histone deacatylases
HLA-DRB1	HLA class II histocompatibility antigen, DRB1
HLA-DRB5	HLA class II histocompatibility antigen, DRB5
IFNγ	Interferon gamma
INPP5D	Inositol polyphosphate-5-phosphatase D
IRF4	Encoding Interferon Regulatory Factor 4
IRX1	Iroquois homeobox protein 1
ITAM	Immunoreceptor tyrosine-based activation motif
ITIM	Immunoreceptor tyrosine-based inhibition motif
JNK	Jun-N-terminal kinase
KSR1P1	Kinase Suppressor Of Ras 1 Pseudogene 1
LGMN	Legumain
lncRNA	Long non-coding RNA
LOAD	Late onset Alzheimer's disease
LRP1	Lipoprotein receptor-related protein 1
MALDI-TOF	Matrix assisted laser desorption/ionization-time of flight mass spectrometry
MAPK	Mitogen activated protein kinase
MAPT	Microtubule associated protein tau
MAT	Methionine adenosyltransferase
MBD1	Methyl-CpG binding domain containing protein 1
MBD2	Methyl-CpG binding domain containing protein 2
MBD3	Methyl-CpG binding domain containing protein 3

MBDs	Methyl-CpG binding domain containing proteins
MCI	Mild cognitive impairment
MDRE	Methylation dependant restriction enzyme
MeCP1	Methyl CpG binding protein 1
MeCP2	Methyl CpG binding protein 2
MEF2	Myocyte enhancer factor
MEF2C	Myocyte enhancer factor 2C
MET	Methionine
MFG	Middle frontal gyrus
MMSE	Mini-mental state exam
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MS	Methionine-synthases
MS4A4E	Membrane Spanning 4-Domains A4
MS4A6A	Membrane Spanning 4-Domains A6
MSP	Methylation specific PCR
MSRE	Methylation sensitive restriction enzymes
MTG	Middle temporal gyrus
MTHFR	Methylenetetrahydrofolate reductase
MT-N5D	Mitochondrial gene encoding NDAH dehydrogenase core subunit 5
NAD_+	Nicotinamide adenine dinucleotide
NAV1	Neuron Navigator 1
ND	Non-disease
NDMAR	N-methyl-D-aspartate receptor
NDUFS5	NADH: Ubiquinone oxidoreductase subunit S5
NFT	Neurofibrillary tangles
NHD	Nasu-Hakola disease
NLGN1	Neuroligin 1

NMDARs	*N*-methyl-D-aspartate receptors
NME8	Thioredoxin domain-containing protein 3
NT	Non-treated
OD	Optical density
PCR	Polymerase chain reaction
pDCs	Plasmacytoid dendritic cells
PET	Positron emission tomography
PI-3K	Phosphatidylinositol a3-kinase
PICALM	Phosphatidylinositol binding clathrin assembly protein
PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
PIP2	Phosphatidylinositol 4,5-bisphosphate
PLCγ	Phospholipase C gamma
PP2A	Protein phosphatase 2A
PrP^C	Cellular prion protein
PSD-95	Postsynaptic density (95 kda) protein
PSEN1	Presenilin 1
PSEN2	Presenilin 2
PTK	Phosphotyrosine kinase
PTK2B	Protein tyrosine kinase 2 beta
QC	Quality control
RHBDF2	Rhomboid 5 homolog 2
RIN3	Rab 5 binding protein which acts a stabilizer of small GTPase rab5
RNA	Ribonucleic acid
ROCK1	Rho associated coiled-coil containing protein kinase 1
RRBS	Reduced representation bisulphite sequencing
S.E.M	Standard error of mean
S100A2	S100 calcium binding protein A2
sAD	Sporadic Alzheimer's disease

SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SEC14L1	SEC14-like protein 1- protein
sEOAD	Sporadic early onset Alzheimer's disease
SERPINF1	Serpin family F member 1
SERPINF2	Serpin family F member 2
SH2	*Src homology 2*
SHIP1	SH2 domain containing inositol-5-phosphatase
SHMT	Hydroxymethltransferase,
SIRT1	Sirtuin 1
SIRT1	Sirtuin1
SLC24A4	Solute carrier family 24 (sodium/potassium/calcium exchanger, member 4)
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
SORBS3	Sorbin and SH3 domain containing the 3 cell-adhesion protein
SORCS2	Sortilin Related VPS10 Domain Containing Receptor 2
SORL1	Sortilin related receptor 1
SP1	Specificity factor 1
SPARC	Secreted protein acidic and rich in cysteine protein
SPARCL1	Secreted protein acidic and rich in cysteine-like protein 1
ssDNA	Single stranded deoxyribonucleic acid
T	Treated
TAE	Tris acetate Ethylene diamine tetraacetic acid
TBC1D22A	TBC1 domain family member 22A
TET	Ten-eleven translocases
TET1	Ten-eleven translocase 1
TF	Transcription factor

THF	Tetrahydrofolate
TMEM259	Transmembrane protein 259
TNF-α	Tumour necrosis factor alpha
TREM2	Triggering receptor expressed on myeloid cells 2
TRIM29	Tripartite Motif Containing 29
TSPs	Thrombospondins
TSS	Transcription start site
TYROBP	Tyrosine kinase binding protein
UBAC1	UBA Domain Containing 1
UBAC1	Ubiquitin-associated domain-containing protein 1
UQCRC1	Ubiquinol-cytochrome C reductase core protein I
UTR	Untranslated region
WGBS	Whole genome bisulphite sequencing
ZCWPW1	Zinc finger CW-type and PWWP domain containing 1

1. Introduction

1.1 What is Alzheimer's disease?

Alzheimer's disease (AD) is a progressive neurodegenerative disease which leads to loss of cognitive function and the development of dementia. It commonly manifests as loss of memory with changes in mood and personality and is the most common cause of dementia worldwide, causing 60-70% of cases (Qazi *et al*., 2017). An estimated 36 million people are living with dementia worldwide, and this figure is set to double every 20 years leading to 115 million new sufferers by 2050. This dramatic increase is the result of an ageing population; age being a highly significant risk factor for the development of the AD (Bekris *et al*., 2010). It is concerning that in high income countries as few as 20-50% of Alzheimer's cases are diagnosed, with this figure being dramatically reduced in low and middle income countries. Therefore AD represents one of the most important health, social and economic crisis facing researchers in the 21st century (Takahashi *et al*., 2017).

There are two main types of AD. Both sub-groups share common pathological features, making age of onset the defining symptom for the two different types of disease (Cacace *et al*., 2016). Early onset Alzheimer's disease (EOAD) is the least common type of AD, representing around 10% of AD cases. EOAD is defined as affecting people under the age of 65 years and 5%–10% of EOAD cases represent a type of AD which is inherited in an autosomal dominant manner, with disease causation lying in single gene mutations (Antonell *et al.,* 2013, Kunkle *et al*., 2017, Harvey *et al*., 2003)**.** Mutations in the genes *APP*, *PSEN1* and *PSEN2*, which all encode proteins involved in the amyloid cascade, are the most common causally associated mutations (Minati *et al*., 2009, van der Flier *et al*., 2011, Cacace *et al*., 2016)**.** These mutations account for 60-70% of inherited EOAD (Loy *et al*., 2014, Kunkle *et al.,* 2017). However most EOAD cases are not explained by genetic mutation and therefore represent a type of sporadic disease known as sporadic early onset AD (sEOAD) (Jarmolowicz *et al*., 2015).

In contrast, late onset Alzheimer's disease (LOAD), also called sporadic Alzheimer's disease (sAD), represents 95% of cases and is commonly defined as affecting people

over the age of 65 (Minati *et al*., 2009). The risk of developing LOAD doubles every five years after the age of 65, therefore age is possibly the most significant risk factor for LOAD (Ferri *et al*., 2005). LOAD is not fully explained by genetic variation, therefore other mechanisms which either alter gene activity or loss of function for genes which link to AD pathogenisis may explain disease susceptibility (Minati *et al*., 2009).

1.2 Amyloid and tau involvement in AD

Pathological features of AD include the deposition of senile plaques and neurofibrillary tangles (NFTs) within the brain (Hardy, 2006, Forstl and Kurz, 1999). Senile plaques consist of amyloid beta (Aβ) and are deposited outside of cells, while NFT contain hyperphosphoryated tau protein and are intracellular (Braak and Braak, 1991).

1.2.1 Amyloid cascade

The amyloid precursor protein (APP) is a transmembrane protein that can be processed by two mutually exclusive pathways: the amyloidogenic and non-amyloidogenic pathway. The first can result in the deposition of the senile plaques seen in AD.

In the non-amyloidogenic pathway, APP is cleaved by α-secretase (ADAM10) within the Aβ domain, producing an APPα soluble peptide and a c-terminal membrane bound APP fragment (CTF). Cleavage by γ-secretase (presenilin) then produces a P3 fragment, which is degraded, and a cytoplasmic APP intracellular domain (AICD) (Haass et al., 2012).

In the amyloidogenic pathway APP is first cleaved by β-secretase (BACE1), also producing a membrane bound APP C-terminal fragment (CTF). This is then cleaved by γ-secretase to produce an Aβ peptide and also a AICD fragment (Haass *et al*., 2012) (Figure 1.1). Multiple isoforms of Aβ can be produced, depending on the cut site of γ-secretase, the most common isoforms produced are Aβ(40) and Aβ(42).

Aβ(42) is the most prone to oligomerization and the subsequent formation of plaques (El-Agnaf *et al*., 2000). However it has been shown that the ratio of Aβ(40) and Aβ(42) is also important in driving plaque formation (Kuperstein *et al*., 2010).

AICD, which is produced in both the amyloidogenic and non-amyloidogenic pathways has been shown to induce degradation of Aβ by acting as a transcription factor which drives activation of *neprilysin* expression. *Neprilysin* is an Aβ degrading enzyme (Pardossi-Piquard *et al*., 2005). Despite AICD being produced by both processing pathways AICD produced in by the amyloidogenic pathway is most likely responsible for the majority of nuclear signalling. In this pathway APP is cleaved by β-secretase following its endocytosis from the cell surface and localization to the endocytic pathway (Rajendran *et al*., 2006). Whereas α-secretase acts on APP located on the cell surface (Sisodia, 1992). Therefore subsequent cleavage by γ-secretase can produce AICD localised to the endosomal system or the edge of the cell. AICD located in the endosomal system is responsible for the majority of AICD nuclear signalling and thus AICD produced during Aβ production drives Aβ degradation (Goodger *et al*., 2009).

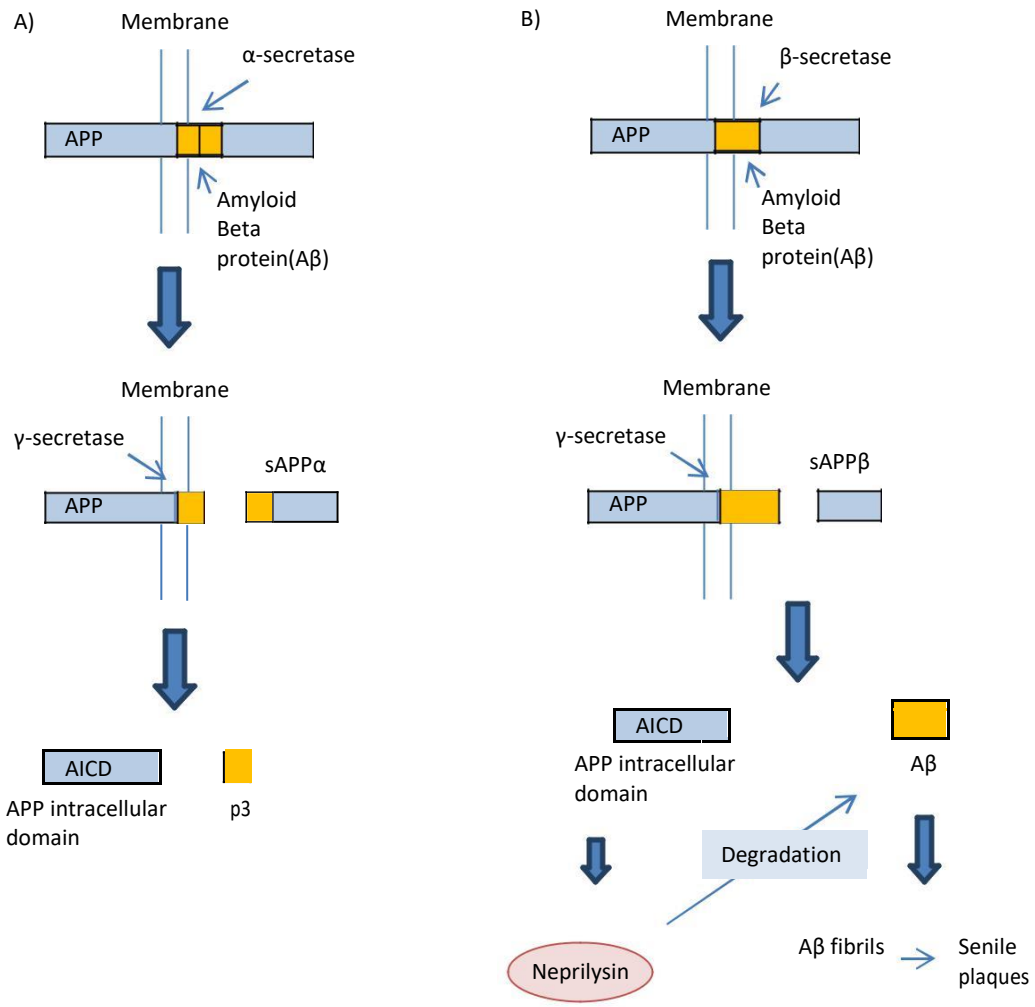


Figure 1.1: Processes involved in plaque formation in AD. A) Non-amyloidogenic pathway, APP is cleaved by α-secretase followed by γ-secretase resulting in a P3 fragment and AICD. B) Amyloidogenic pathway, APP cleavage by β-secretase and γ-secretase results in the production of Aβ which can result in plaque formation. AICD is also produced which activates *neprilysin* expression, leading to Aβ degradation. Image adapted from (Wang *et al*. 2013) and (Haass and Selkoe, 2007)

Aβ plaques are thought to drive AD pathology by disrupting synaptic function, which then leads to memory deficits (Beyreuther and Masters, 1995), and eventually causes neuronal death (Yankner *et al*., 1990). Plaques are also thought to cause activation of immune cells such as microglia, which results in an immune response causing further neuronal damage (Combs, 2009).

The involvement of amyloid in the progression and pathology of AD is supported by evidence that genetic mutations of the *APP* gene and presenilin genes *PSEN1* and *PSEN2,* which encode catalytic subunits of γ-secretase resulting in generation of Aβ, cause inheritable forms of the disease (Bettens *et al*., 2013). All three mutations result in an increased production of pathogenic Aβ peptide. However sporadic forms of AD are not fully explained by genetic mutations alone and while amyloid plaques are a pathological feature of this disease many genes and pathways outside of the amyloid cascade have been implicated in disease progression. One hypothesis is that Aβ deposition is capable of initiating the disease but alone is not sufficient to result in AD, instead aberrant Aβ may be an early pathological feature of AD drives further perturbation to AD-associated pathways, which then go on to drive disease progression (Musiek and Holtzman, 2015).

Intriguingly, Aβ has also been shown to be neuro-protective (Giuffrida *et al*., 2009). Further roles for Aβ have been suggested in the removal of harmful substances; Aβ is thought to bind to harmful substances, such as metal ions, and present them to macrophages for subsequent removal or destruction (Robinson and Bishop, 2002). Aβ is also thought to have an antimicrobial role (Soscia *et al*., 2010). This suggests that Aβ at high physiological concentration is pathogenic but at low concentrations could be involved in normal neuronal function and synaptic transmission (Puzzo and Arancio, 2013).

1.2.2 Tau and neurofibrillary tangles

Intracellular NFTs are made up hyperphosphorylated tau protein, also named microtubule associated protein tau (MAPT). Normally hyper-phosphorylated tau acts as a microtubule associated protein and is involved in their function and stability.

However, in AD tau phosphorylation is increased, leading to its disassociating from microtubules followed by aggregation into NFTs. NFTs cause disruption of the cytoskeleton followed by neuronal death (Lovestone and Reynolds, 1997).

Evidence suggests that Aβ deposition results in tau phosphorylation, accumulation and its deposition at synapses (Bilousova *et al*., 2016). It can be argued that Aβ causes a disruption in oxidative stress and ionic homeostasis which results in increased kinase activity (Sorrentino and Bonavita, 2007). Aβ induced increase in oxidative stress causes activation of p38 mitogen activated protein kinase (MAPK), a kinase responsible for tau phosphorylation (Giraldo *et al*., 2014). Aβ has also been shown to activate other kinases that act on tau, such as glycogen synthase kinase-3 (GSK3), jun-N-terminal kinase (JNK) and extracellular receptor kinase (ERK) (Lee *et al*., 2005).

Aβ and tau mediated mechanisms that drive AD pathology have been shown to converge at the synapse and it seems likely that Aβ driven toxicity has a role in activating tau mediated toxicity in AD. The molecular mechanisms by which Aβ causes AD pathology at neuronal synapses are many and diverse, all of which result in disruptions in the function of receptors and ultimately the synapse itself (Nisbet *et al*., 2015). One such mechanism is through disruption of the *N*-methyl-D-aspartate receptors (NMDARs), increased Aβ results in a decreased number of these receptors at synapses, due to its binding to α7 nicotinic receptors, resulting in retraction of the synapse due to spine shrinkage (Nisbet *et al*., 2015).

Increased Aβ levels at the synapse also results in increase Ca^{+2} influx by two different mechanisms. First, by its binding to lipid rafts which results in a hole within the membrane leading to Ca^{+2} influx (Kawarabayashi *et al*., 2004, Nisbet *et al*., 2015). Second, excess Aβ can result in extrasynaptic NMDARs being activated resulting in further Ca^{2+} influx. Ca^{2+} influx results in the activation of intracellular kinases, which in turn results in tau phosphorylation causing its aggregation into NFTs. Importantly phosphorylation of tau also results in its binding to Fyn and Fyn-tau aggregates, prior to migration to the dendritic spine.

At the dendritic spine Fyn acts to phosphorylate NMDARs. This results in the interaction of receptors with the postsynaptic density (95 kda) protein (PSD-95), which drives Aβ toxicity. Aβ can also interact with the cell surface receptor cellular prion protein (PrP^C) directly: this interaction activates Src kinase Fyn resulting in NMDAR phosphorylation and PSD-95 recruitment. This ultimately leads to NMDAR internalisation, again driving Aβ associated toxicity (reviewed in (Nisbet *et al*., 2015)) (Figure 1.2).

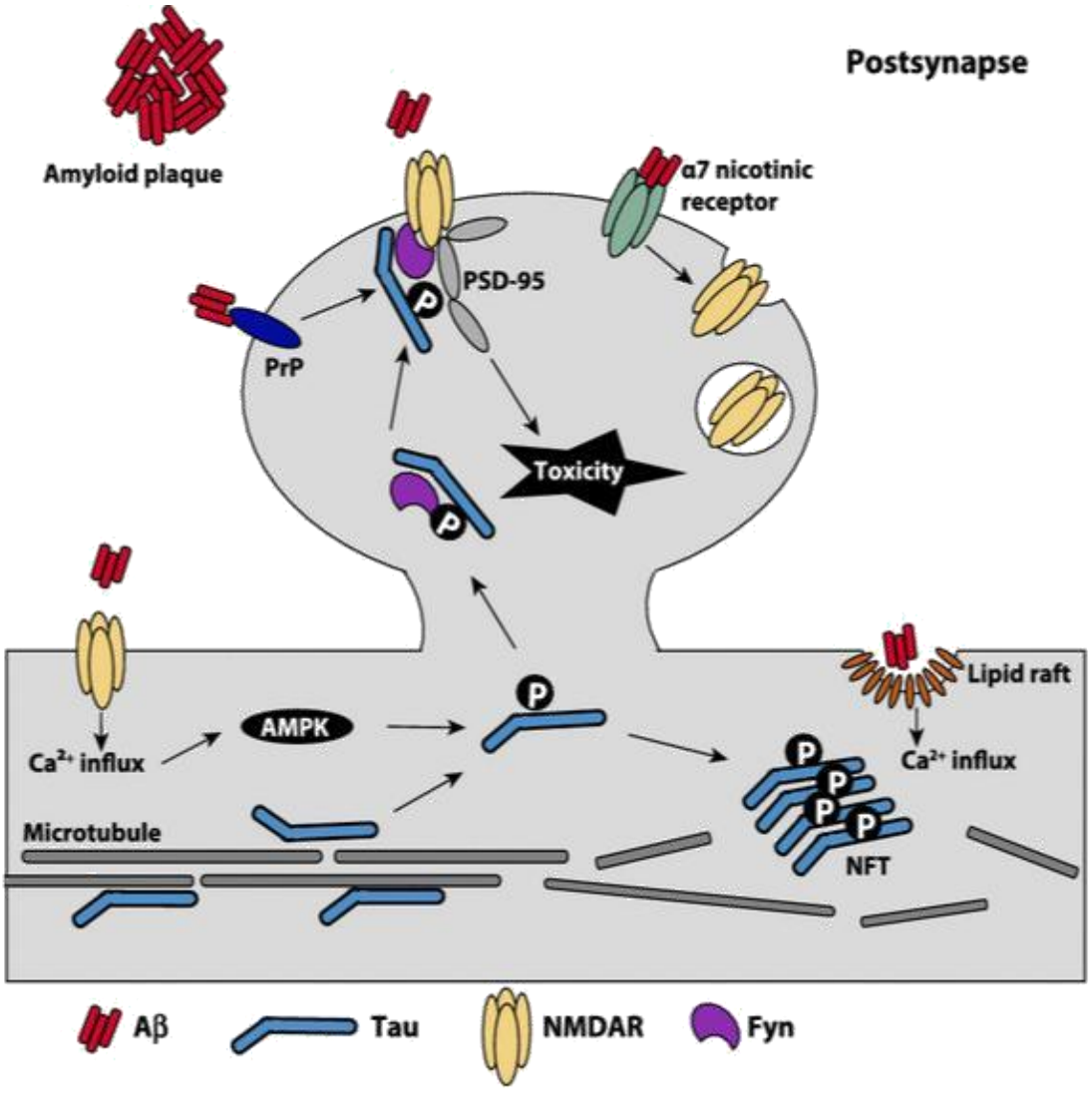


Figure 1.2: Aβ and tau mediated toxicity at the synapse. Showing Aβ and tau mediated toxicity at the synapse and how Aβ and tau driven processes interact to result in AD toxicity. Adapted from (Nisbet *et al*., 2015).

Figure 1.2 describes how Aβ accumulation may drive the activation of tau through phosphorylation, resulting in AD pathology. In contrast with this hypothesis, it has been proposed that NFTs may be deposited prior to plaque formation and tau might be responsible for driving Aβ production and toxicity within the AD brain (Braak and Del Tredici, 2013, Li *et al*., 2015). It has also been suggested that dendritically located tau mediates early Aβ toxicity at the synapse (Ittner *et al*., 2010, Guerrero-Munoz *et al*., 2015).

The importance of tau in driving AD pathogenesis is also apparent due to the fact that the amount of NFTs within a patient's brain correlates well with disease severity (Giannakopoulos *et al*., 2003), whereas the presence of plaques does not, such that multiple plaques can be present in a symptomless individual (Perez-Nievas *et al*., 2013). However, since aberrant tau activity alone results in frontotemporal dementia (FTD) rather than AD (Goedert and Jakes, 2005), it is likely that both amyloid and tau play roles in AD, possibly by collectively contributing to synaptic dysfunction and other AD pathologies (Spires-Jones and Hyman, 2014).

1.3 Genetics and Alzheimer's disease

Genetic mutations are known to have a role in the occurrence of both EOAD and LOAD. However, only a small amount of EOAD cases are inherited in an autosomal dominant manner, most cases of EOAD are sporadic (sEOAD) and have unknown genetic causation (Jarmolowicz *et al*., 2015). In the case of inherited familial EOAD single gene mutations are accountable for disease development. These mutations lead to perturbations in amyloid processing resulting in increased Aβ production and increase the ratio of Aβ42 to Aβ40. Mutation in the genes *APP, PSEN1* and *PSEN2* (components of γ-secretase) are responsible for familial AD. *PSEN1* is the gene that is most commonly mutated in EOAD, it accounts for 18-50% of autosomal dominantly inherited cases (Bekris *et al*., 2010, Cacace *et al*., 2016, Cruchaga *et al.,* 2018).

In contrast to EOAD, only the apolipoprotein E *(APOE) ε4* allele seems to carry a significant risk for the development of LOAD. The *APOE* locus contains three allele's (*ε2, ε3* and *ε4)* and variation is determined by two missense mutations. When

expressed, each allele results in one of three protein isoforms E2, E3 and E4 (Mahley, 1988). The *ε4* allele influences both predispositions to LOAD development as well as age of onset (Seshadri *et al*., 1995, Corder *et a*l., 1993). Carriers of one *ε4* allele confers approximately a three times increased risk of developing LOAD, whereas two copies result in a twelve fold risk. *ε4* heterozygotes also develop the disease 10-12 years earlier than those with no *ε4* alleles (Corder *et al*., 1993). However, the presence of the *ε2* allele confers protection from amyloid plaque and NFT formation (Nagy *et al*., 1995). Interestingly, the *ε4* allele also increases likelihood of developing sporadic and familial EOAD as well as LOAD (Guerreiro *et al*., 2012, Karch *et al*., 2014).

How *APOE* alleles affect LOAD development is not well understood. However a link has been found between *APOE* alleles and amyloid deposition within the brain (Bales *et al*., 1999, Ramanan *et al*., 2013), determination of whether Aβ oligomerizes and forms toxic plaques and Aβ clearance (Yu *et al*., 2014a, El Haj *et al.*, 2016, Castellano *et al*., 2011). *APOE* is also involved in transportation of oligomeric Aβ to synapses, thus driving its toxicity (Koffie *et al*., 2012).

APOE ε4 has also been suggested to drive LOAD pathology independently of its effects on amyloidogenesis by driving tau hyperphosphorylation and thus NFT deposition, cytoskeletal dysfunction and neurotoxicity (Harris *et al*., 2003, El Haj *et al*., 2016). *APOE ε4* has also been suggested to be involved in: driving inflammation and immune response, being neurotoxic, deregulating lipid metabolism and causing mitochondrial dysfunction (Keene *et al*., 2011, Mahley and Huang, 2012, Hauser *et al*., 2011, Gibson *et al*., 2000). Many carriers of this allele do not go on to develop LOAD, less than 50% of people that carry two *ε4* alleles will go on to develop the disease before they die (Liddell *et al*., 2001, Karch *et al*., 2014). This suggests that the development of LOAD is much more complex than simple mendelian single gene inheritance, rather representin a complex polygenic condition with environmental factors also strongly influencing disease risk (Bekris *et al*., 2010).

The recent development and utilisation of large scale association studies (GWAS) has led to the implication of multiple new gene loci in AD, as well as confirming the

involvement of previously identified genes. A large meta-analysis of previous performed GWAS has recently led to the identification of eleven new Alzheimer's susceptibility loci and confirmed association of eight loci. Along with *APOE* this means that that it was possible to identify twenty loci associated with LOAD (Lambert *et al*., 2013b, Karch and Goate, 2015, Rigde *et al*., 2013, Ridge *et al*., 2016). However, since this study a further nine loci have been identified, making a total of twenty nine (Bertram and Tanzi, 2019).

These studies result in the identification of common genetic variants which only altered the chances of developing LOAD in a small way, however, it is likely that multiple rare genetic variants could also explain disease risk (Karch and Goate, 2015, Cuyvers and Sleegers, 2016). An example of such would be the AD risk variant *TREM2* that has been identified, which is only in 1% of population; other rare variants have been identified in *APP* and Unc-5 Netrin Receptor C (*UNC5C*) (Guerreiro *et al*., 2012, Rigde *et al*., 2016, Cruchaga *et al*., 2014, Jonsson *et al.,* 2012*,* Cuyvers and Sleegers, 2016). It is most likely that both common and rare genetic variants influence disease susceptibility and may even occur at the same loci and gene-gene interaction are also likely important in explaining AD risk (Rigde *et al*., 2016, singleton *et al*., 2011, Kilpinen and Barrett, 2013).

However, considered together, the known AD associated variants account for 30.62% of the population attributed risk (PAR) of LOAD, thus there is a large missing proportion of disease causation yet to be explained (Ridge *et al.,* 2016*,* Lambert *et al*., 2013b). It is likely that environment has a substantial influence on the development of AD, however, further genetic analysis is needed to identify genetic variants causing AD, this would include the use of next generation sequencing (NGS) and investigation into both common and rare variants, with structural variants also needing to be investigated (Rigde *et al*., 2016, Cuyvers and Sleegers, 2016).

Despite the limitations of conventional genetic analysis in explaining all cases of AD, GWAS have been extremely productive in identifying new biological pathways and also confirming previously reported pathways that may contribute to LOAD

pathogenesis. Meta-analysis studies have identified genes involved in APP trafficking and metabolism, tau functionality and also inflammation and immunity, all of which have been previously implicated in LOAD. In addition genes encoding proteins involved in endocytosis, cholesterol metabolism, intercellular signalling, cell adhesion and pre-mRNA editing have also been implicated in AD (Lambert *et al*., 2013b, Cuyvers and Sleegers, 2016). While not accounting for disease risk in its entirety, these studies produce exciting new insights into the biology of LOAD and also may lead to the identification of new therapeutic targets (Lambert *et al*., 2013b, Medway and Morgan, 2014, Karch and Goate, 2015, Cuyvers and Sleegers, 2016).

1.4 What is epigenetics and DNA methylation?

The missing explanation of LOAD risk implies that environmental factors may contribute to disease development. Since the environment can drive epigenetic changes to DNA, it is likely that epigenetic changes may play a role in driving the development of LOAD. Epigenetic aberration can be defined as dynamic changes to the genetic material that result in alterations in gene expression, without an underlying alteration to the DNA sequence. The most well-known epigenetic mark is DNA methylation, but other epigenetic regulation includes changes to histone modifications, histone variants, RNA methylation and none coding RNAs (Fraga, 2009, Wang *et al*., 2013). Epigenetic variation can be influenced by an individual's environment, by diet, chemical exposure, proteins, hormone levels and drug exposure. This allows for a system whereby an individual's life events can affect gene expression and thus disease susceptibility (Abdolmaleky *et al*., 2004).

DNA methylation involves the addition of methyl groups to the 5^{th} carbon of cytosines within the DNA. DNA methyltransferases (DNMTs) are responsible for this action. DNMTs use S-adenosylmethionine (SAM) as a methyl donor to catalyse the transfer of a methyl group to a single stranded piece of DNA (Gonzalgo and Jones, 1997, Abdolmaleky *et al*., 2004) (Figure 1.3). Four DNMTs exist, DNMT1, DNMT2,

DNMT3a and DNMT3b; DNMT1 are responsible for the maintenance of methylation in mammalian cells through cellular DNA replication (Espada *et al*., 2004).

Cytosine methylation commonly occurs at 5'CpG'3, locations in the DNA sequence where a cytosine residue is followed by a guanine; the p indicates the phosphodiester bond that links two bases. CpG's are particularly common in the promoter region of genes and make up CpG islands (CGIs), which are short sections of DNA that are CpG site rich (Deaton and Bird, 2011). Within promoters, methylation of cytosines located within CpG islands leads to transcriptional repression (Mehler, 2008, Graff and Mansuy, 2008, Schubeler, 2015). This transcriptional repression can be direct, due to methylation preventing the binding of transcription factors needed for transcription initiation. Repression can however also be indirect. This involves the recruitment of repressor proteins or Methyl-CpG binding domain (MBD) containing proteins of which there are four, MeCP2, MBD1, MBD2, and MBD3 (Abdolmaleky *et al*., 2004)**.** DNMTs may also cause transcriptional repression through a transcription repression domain, which can interact with histone deacetylases (HDACs). Histone deacetylation results in chromatin structures which inhibit gene expression (Fuks *et al*., 2000)**.**

Proteins can also bind to MBD containing proteins and form repressive complexes. One such example is MeCP1 which binds MBD2, and acts to maintain methylation. MBD2 and MeCP1 also form part of a larger complex which contains HDACs and proteins that are capable of nucleosome remodelling, both of which lead to further transcriptional repression (Graff and Mansuy, 2008, Feng and Zhang, 2001)**.** MECP2 however is a MBD containing protein that binds directly to methylated DNA, where it can then recruit HDACs also resulting in transcriptional repression (Nan *et al*., 1998)**.**

Previously DNA methylation has been associated with transcriptional repression alone, however Ziller *et al.* (2013) showed that the location of the methylation can be important for the affect it will have on the transcript structure, it has also been demonstrated that intragenic methylation is important for modulation of alternative

spicing and methylation located within the gene body has been associated with an increased gene expression (Manakea *et al*., 2013, Varley *et al*., 2013).

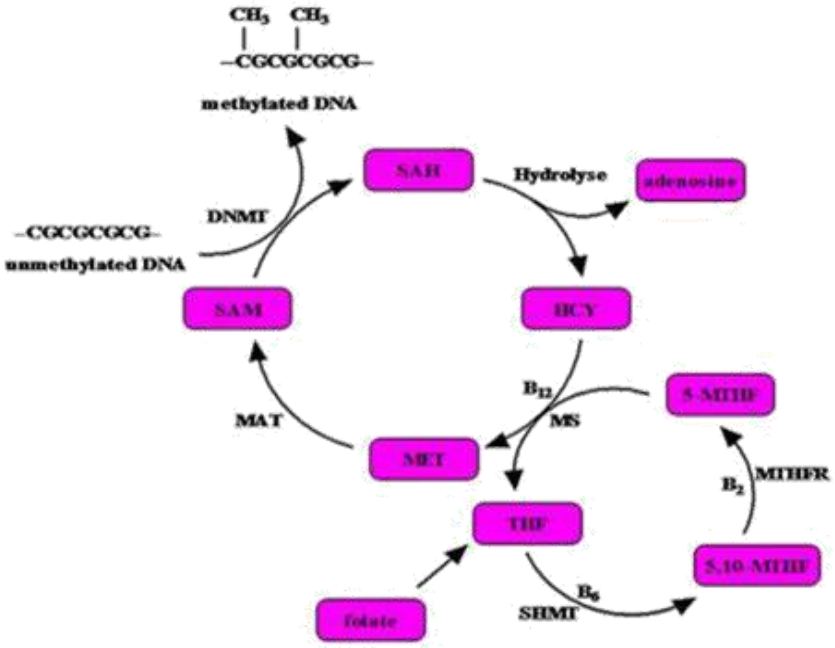


Figure 1.3: Folate, Methionine and homocysteine metabolism in DNA methylation. (DNMTs) methlytransferases, (SAM) S-adenosylmethionine, (SAH) S-adenosylhomocysteine, (HCY) homocysteine, (5-MTHF) 5-methylenetetrahydrofolate, (MS)methionine-synthases, (MET) methionine, (MAT) Methionine adenosyltransferase , (THF) tetrahydrofolate, (5.10 MTHF), methylenetetrahydrofolate, (SHMT) hydroxymethltransferase, (MTHFR) methylenetetrahydrofolate reductase. Image extracted from Wang *et al.* (2013).

Folate, methionine and homocysteine metabolism are critical to the process of DNA methylation, (Figure 1.3). S-adenosylhomocysteine (SAH) is produced when DNMTs use SAM as a methyl donor for DNA methylation. The ratio between SAM and SAH is known as methylation potential and DNA methylation is highly dependent on this ratio. SAH levels are controlled by its hydrolysation to homocycteine (HCY) and

adenosine. 5-methylenetetrahydrofolate (5-MTHF) acts as a methyl donor and cobalamin-dependant methionine-synthases (MS) as the catalyst in a methylation reaction that converts HCY to methionine (MET). Methionine adenosyltransferase (MAT) then catalyses the reaction producing SAM from methionine. As well as methionine metabolism being important for DNA methylation, folate and B-vitamins are also important in the cycle. Folate is used to produce tetrahydrofolate (THF) which is then converted to 5.10 methylenetetrahydrofolate (5.10 MTHF) by the B6 dependant enzyme serine hydroxymethyltransferase (SHMT). 5,10 MTHF is then converted to 5.MTHF by the vitamin B2 dependant enzyme methylenetetrahydrofolate reductase (MTHFR). MTHFR is then available for the use as a methyl donor in conversion of HCY to MET (Chouliaras *et al*., 2010).

DNA de-methylation is performed by ten-eleven translocases (TET). These enzymes convert 5-methylcytosine (5mC) into 5-hydroxymethlycytosine (5hmC) in an oxidation reaction (Tahiliani *et al*., 2009)**.** The effect of 5hmC on transcription regulation is largely still elusive although both transcriptional repression and activation have been suggested (Coppieters *et al*., 2014, Valinluck *et al*., 2004)**.** Further oxidation of 5hmC can produce two other cytosine derivatives 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), these may also have a role in regulating gene expression (Ito *et al*., 2011). Following oxidation, a deamination reaction performed by activation induced deaminase (AID) or apolipoprotein B mRNA-editing enzyme 1 (APOBEC1) results in a T/G mismatch (Morgan *et al*., 2004). This mismatch is then repaired by glycosylases which sees the DNA converted back to its non-methylated form (Morgan *et al*., 2004).

1.5 DNA methylation and AD

Epigenetic involvement in AD seems likely and has also been demonstrated as a risk factor for other diseases such as cancer (Wajed *et al.,* 2001). Multiple genes and many different pathways have been implicated in AD risk and since epigenetic regulation may control the expression of multiple genes simultaneously, it is a good hypothesis that the genes associated with LOAD may be co-ordinated through

epigenetic regulation. In addition genes previously not associated with LOAD are also being identified in epigenetic studies and thus are also likely to be important in disease pathology (Mastroeni *et al*., 2010).

1.5.1 Global methylation and hydroxymethylation in AD

Multiple lines of evidence also suggest that epigenetic alterations may drive LOAD. In the mid-nineties LOAD patients were shown to have decrease in S-adenosyl methionine (SAM) and S-adenosylhomocysteine (SAH), suggesting altered DNA methylation occurs in AD (Morrison *et al*., 1996). Global changes such as hypomethylation and hypo-hydroxymethylation have also been reported in specific region of the brain and also for specific LOAD associated genes (Condliffe *et al*., 2014, Chouliaras *et al*., 2013, Wang *et al*., 2008, Frasquet *et al*., 2018).

Intriguingly changes in methylation/demethylation present during LOAD progression could be detectable and maybe initiated before an individual acquires AD symptoms (Bradley-Whitman and Lovell, 2013). Studies on monozygotic twin pairs discordant for AD also provide evidence that environmental factors may influence disease risk, possibly by causing epigenetic changes that drive disease progression (Chouliaras *et al*., 2013, Mastroeni *et al*., 2009, Konki *et al*., 2018).

Although evidence suggests that epigenetic modifications may be involved in driving LOAD, studies of global methylation levels in LOAD samples have revealed contradictory results. Condliffe *et al*. (2014) failed to identify any difference between 5mC in the entorhinal cortex and cerebellum of LOAD patients vs controls, and Lashley *et al.* (2015) failed to identify any changes in 5mC or 5hmC in the entorhinal cortex. This was despite previous studies showing differences in DNA methylation between these regions in the AD brain (Bakulski *et al*., 2012). 5mC was shown to be decreased in the entorhinal cortex of AD patients (Chouliaras *et al*., 2013, Mastroeni *et al*., 2010), but not the cerebellum (Mastroeni *et al*., 2010). 5mC has also been shown to be decreased in the frontal cortex and hippocampal regions of an AD patient from one of a monozygotic twin pair discordant for AD (Mastroeni *et al*., 2009, Chouliaras *et al*., 2013).

However, it should also be noted that discrepancies between these results may be explained by small experimental sample sizes and different sample cohorts (Condliffe *et al*., 2014), or a less likely explanation is that the differing results are due to different tissue processing and immune-staining methods used (Coppieters *et al*., 2014). Another plausible explanation could be that different brain regions are investigated and the presence of 5mC and 5hmC may differ between brain regions. In this model a disease affected tissue may show one profile while surrounding tissue shows a different 'response' profile which is due to the pathological conditions of the nearby cells. Therefore the same stressor can result in two different effect profiles (Coppieters *et al*., 2014).

A more recent study also suggests that the degree of AD pathology present in the individual at the time of sampling may also be important for global changes in methylation and hydroxymethylation. Ellison *et al.* (2017) investigated both epigenetic marks in several brain regions in sample taken at varies stages of disease. The regions harbouring more extensive AD pathology where found to have global changes in both 5mC and 5hmC in early disease, which became comparable to control samples in late stage disease (Ellison *et al.,* 2017).

One study conducted using peripheral blood mononuclear cells resulted in the identification of global hypermethylation in LOAD samples, suggesting global blood methylation might represent an attractive biomarker of AD (Di Francesco *et al*., 2015). The use of blood DNA does present its own challenges however. Blood samples are also heterogeneous for cell type and therefore cell specific changes in methylation might be mis- interpreted (Backland *et al*., 2015). In addition, blood cells turnover much faster then neuronal cells and are also not the primary site of AD related pathology (Wu *et al*., 2012). However, despite these challenges studies are successfully identifying AD-associated aberrant methylation patterns in blood that have also been linked to pathological features of AD such as cognitive decline (Xu *et al*, 2018; Madrid *et al*, 2018, Mercorio *et al*, 2018).

The cause of the changes in methylation observed in LOAD remains elusive, however, it has been suggested that it may be caused by a decrease in the levels of DNA methyltransferase 1 (DNMT1) or a loss of nuclear transport. This may be caused by down regulation of proteins responsible for DNMT1 transport into and out of the nucleus (Casillas *et al*., 2003, Mastroeni *et al*., 2013)**.** Other gene specific studies have also found that those genes responsible for DNA methylation are hypermethylated in AD, which might also contribute to loss of methylation (Wang *et al*., 2008). In contrast Di Francesco *et al*. (2015) found increased expression of *DNMT1* which resulted in global hypermethylation in LOAD samples.

Evidence also exists that suggests that alterations in another epigenetic mark, 5hmC, might also drive disease progression. A GWAS revealed genetic variation in *TET1*, the enzyme responsible for oxidation of 5mC to 5hmC, can increase LOAD risk (Morgan *et al*., 2008)**.** Using immuno-fluorescent staining, 5-hydoxymethylcytosine was shown to be globally decreased in the entorhinal cortex and cerebellum of AD patients compared with controls (Condliffe *et al*., 2014). This result is concurrent with recent findings that 5hmC is globally decreased in hippocampal glial cells of LOAD cases compared with controls and also in the hippocampus of a LOAD patient that forms part of a monozygotic twin pair discordant for the disease (Chouliaras *et al*., 2013).

However, other studies using immunostaining of 5hmC and 5mC in tissue samples from the human middle frontal gyrus (MFG) and middle temporal gyrus (MTG) showed an increase in both 5mC and 5hmC epigenetic marks in LOAD patients when compared with age matched controls (Coppieters *et al*., 2014). This finding was in agreement with a previous study that investigated changes in the LOAD prefrontal cortex; a study that also showed global hypermethylation in this region (Rao *et al*., 2012). In addition, while decreased hydroxymethylation was observed in the hippocampus, no difference was identified in either the cortex or cerebellum using a mouse model of AD (Coppieters *et al*., 2014). Coppieters *et al*. (2014) did, however, find cell type specific changes in methylation and hydroxymethylation. Astrocytes and microglia cells were found to have reduced 5mC and 5hmC in LOAD but, in contrast, neurones were found to have increased 5mC and 5hmC.

Condliffe *et al*. (2014) also identified the presence of the 5-fC and 5-caC cytosine modifications in the brain and found a difference in the amount of 5-fC between the two brain regions investigated, however, they failed to identify a significant difference between LOAD cases and controls. This may have been due to a low sample size, suggesting further investigation could reveal significant differences in these two newly identified cytosine modifications in LOAD and lead to newly associated epigenetic mechanisms or marks.

1.5.2 Gene specific differential methylation and AD

The studies described show global changes in methylation and hydroxymethylation are likely to be of importance in AD, however, in order to identify specific pathways and functional epigenetic changes occurring in AD it is also necessary to investigate gene specific alterations. Multiple genes that have been associated with LOAD by genetic studies, such as through GWAS, are being identified as harbouring AD specific differential methylation. This includes the genes *SORL1, ABCA7, TREM2 SLC24A4*, and *BIN1* (Yu *et al.,* 2014, Smith *et al.,* 2016). This suggested that both genetics and epigenetic regulation of these genes have potential importance in driving disease pathology.

Interestingly, epigenome wide association studies (EWAS) are also identifying differential methylation located within or near genes that have no previous genetic association with AD. Watson *et al.,* (2016) identified 476 AD associated differentially methylated regions in a recent EWAS study, however of these only fifteen were located near known GWAS SNPs. Other EWAS studies have also identified Methylation in genes not previously associated with AD, including the genes *ankyrin 1* (*ANK1*), *homeobox A* (*HOXA), wingless-type MMTV integration site family member B (WNT5B)* and *AT-Rich Interactive Domain-Containing Protein 5B* (*ARID5B*) (Lunnon *et al*., 2014, De Jager *et al*., 2014, Smith *et al*., 2018, Smith *et al*., 2019, Watson *et al*., 2016). This provides further evidence for the importance of DNA methylation in AD

and also allows the identification of novel pathways important in disease pathology (Watson *et al*., 2016).

Another factor for consideration is whether methylation quantitative trait loci (mQTLs) may have a role in AD. mQTL's are QTLs which have control over DNA methylation and therefore may represent a mechanism linking AD associated genetic and epigenetic alterations (McRae *et al.,* 2018). However, a recent study has shown that mQTLs are mostly found in pseudogenes and therefore could be interesting in single sample testing but analysis of whole methylomes would reduce the importance of mQTLs (McRae *et al.,* 2018). Further to this it should also be considered that AD associated genes, whether these be linked by genetic or epigenetic studies, still have clinical relevance to the individual being tested.

1.5.2.1 LOAD Associated Aberrant Methylation of Genes in the Amyloid Pathway and Tau Processing

Genes in the amyloid pathway have been extensively associated with AD pathology and have been genetically associated with disease. Hyper and hypo methylation of genes located within this pathway has also been identified. This includes the promoters of genes involved in APP processing and Aβ production. The *APP* promoter was found to be hypomethylated in LOAD patients (West *et al*., 1995). The *APP* promoter was also found to be hypomethylated in a subsequent study investigating people over seventy and LOAD patients. Non-demented people over the age of seventy and LOAD sufferers less than seventy years of age were found to have hypomethylation of the *APP* promoter. This hypomethylation may lead to increased APP expression resulting in greater Aβ production (Tohgi *et al*., 1999). In support of this, hypomethylation was also observed in the *APP* gene of AD patients using leukocyte DNA, this also correlated with increased gene expression (Hou *et al*., 2013).

However, other studies have produced conflicting results. Two studies failed to find any difference in the *APP* promoter methylation in total homogenates of post-

mortem LOAD brain tissue or lymphocytes (Barrachina and Ferrer, 2009, Wang *et al*., 2008) and a further study failed to identify any LOAD associated methylation at the *PSEN1* promoter (Carboni *et al*., 2015). Furthermore, a more recent study looked at the methylation present at the promoter regions of *APP, PSEN1* and *PSEN2* in LOAD and significant LOAD associated hypermethylation was identified in all three genes interestingly, this study identified this LOAD associated methylation in peripheral blood rather than in brain tissue (Piaceri *et al*., 2015).

The reduction of tau expression in LOAD has also been associated with hypomethylation of the granulocyte chemotactic factor (GCF) within the *TAU* promoter, this leads to transcriptional repression. Within the same promoter a second site which binds specificity factor 1 (SP1) has been found to be hypermethylated. Since SP1 is a transcriptional activator, this epigenetic mark also contributes to *TAU* down regulation (Tohgi *et al*., 1999). However other studies failed to find any LOAD associated change in methylation of the *MAPT* promoter (Barrachina and Ferrer, 2009).

Significant variation in *PSEN1* promoter methylation was found between individuals, which may contribute towards LOAD predisposition (Wang *et al*., 2008). The *PSEN1* promoter is also reportedly hypomethylated *in vitro*, resulting in increased APP cleavage and Aβ production (Fuso *et al*., 2006). However other studies have failed to produce results suggesting *PSEN1* promoter methylation variation association with LOAD (Siegmund *et al.,* 2007). However, since homogenates of specific brain regions were used in many of the previously mentioned studies these results need to be interpreted with caution. Use of homogenates can result in small but significant changes in the methylation status of specific brain regions and cell types being missed (Barrachina and Ferrer, 2009). Most studies also only focus on a small region of genomic sequence and thus can miss important variation in regulatory sites within the genes being analysed outside of the region under investigation (Siegmund *et al*., 2007). The work conducted as part of this thesis aims to resolve these issues by focusing on leukocyte and cortex DNA.

Leukocytes have been proven to cross the blood brain barrier in response to chemokine secretion by CNS cells (Eugenin and Berman, 2003). In AD chemokines are secreted by astrocytes and the level of chemokines in both CSF and brain tissues are changed (Liu *et al*., 2014). It is therefore plausible that the inflammation caused by AD, resulting in increased chemokine secretion would allow leukocyte trafficking thought the BBB and thus these cells will experience the AD brain environment and alter methylation depending on the AD pathology present. This seems likely as studies are beginning to link methylation status of AD associated genes to AD pathology and disease progression (Hou *et al*., 2013).

The aforementioned genes all encode proteins which are involved in APP processing or relate to tau pathology, however, genes specifically involved in DNA methylation have also been found to show promoter methylation variation between LOAD and non-LOAD samples. Wang *et al*. (2008) found positions in the *MTHFR* gene promoter are hypermethylated in AD and the *DNMT1* promoter shows significant inter-individual variation, suggesting a role in AD predisposition.

Further evidence of a role of DNA methylation in AD is suggested by induced hyperhomocysteinemia. This is a condition associated with LOAD, which causes imbalance in the SAM/SAH ratio. This imbalance resulted in hypomethylation of *PSEN1* and *BACE* leading to an increase in Aβ deposition in mice over expressing *APP*. This result may suggest that a decrease in DNA methylation caused by SAH/SAM imbalance, resulted in hypomethylation of *APP, BACE* and *PSEN1* thus causing their enhanced expression, and therefore increase in Aβ deposition (Fuso *et al.*, 2008).

1.5.2.2 Other genes containing aberrant methylation associated with LOAD

Other studies have used brain DNA to identify differential methylation of specific genes in AD. Prefrontal cortex samples where used in an Illumina Human Methylation450 beadset assay to identify differentially methylated CpGs in 708 subjects. While this technique covered a large number of CpGs, it was not genome

wide and therefore could not identify all differentially methylated regions in the AD subjects used within the study (De Jager *et al*., 2014). However they did identify 71 CpGs that were differentially methylated in LOAD samples, 12 of which were validated using a smaller independent sample and using a different measure of AD pathology (Braak staging rather than the burden of neuritic amyloid plaques), two of these CpG's were within the AD associated loci *BIN1* and *ABCA7* (De Jager *et al*., 2014). Both of which have been previously associated with AD though genetic variants (Lambert *et al*., 2013b). They also investigated if changes were a cause or effect of AD by looking for association in non-demented patients who displayed amyloid burden. It was found that differential methylation was pre-symptomatic and thus changes could represent early drivers of the disease or be an early epi-mutation (De Jager *et al*., 2014).

RNA expression of genes associated with these 12 CpGs was used to functionally validate the changes. It was found that changes in expression of seven genes were annotated to AD. These genes where: *ANK1, CDH23, DIP2A, RHBDF2, RPL13, SERPINF1* and *SERPINF2,* these could all be linked to pathways involving other AD susceptibility genes (De Jager *et al.,* 2014).

ANK1 has also been found to be differentially methylated in cortical brain regions but not blood DNA in another study (Lunnon *et al*., 2014). This study also identified genes with differentially methylated regions (DMRs) that could be identified in AD whole blood DNA. The study found DMRs close to the genes *DAPK1* (death-associated protein kinase 1), *GAS1* (growth arrest-specific 1) and *NDUFS5* (NADH: ubiquinone oxidoreductase subunit S5). *GAS1* has a role in APP processing and both *DAPK1* and *NDUFS5* have been previously implicated in AD however importantly *ANK1* has not been genetically associated with AD (Lunnon *et al*., 2014, Watson *et al.,* 2016).

Other genes have been identified using EWAS studies that harbour AD associated methylation in genes with no previous genetic association with LOAD. A study by Watson *et al.,* (2016) identified a number of differentially methylated regions in the superior temporal gyrus of Alzheimer's sufferers, of those identified (475 DMRs) only

fifteen associated with genes previously implicated in AD by genetic studies. In support of this, recent studies have revealed AD associated DMRs in the genes *HOXA, WNT5B* and *ARID5B*, which have not been previously associated with AD by current genetic studies (Smith *et al*., 2018, Smith *et al*., 2019, Watson *et al*., 2016).

Other genes harbouring differential methylation have been identified by using peripheral blood. These studies were recently reviewed by Frasquet *et al*., (2018). LOAD associated methylation has been identified in *APOE, APP, BACE1, BDNF, DNMT1, DNMT3A, DNMT3B, LINE-1, MTHFR, PIN1, PSEN1* and *SIRT1* using peripheral blood DNA. These genes potentially have important roles in driving AD pathology when aberrantly methylated.

The genes brain *derived neurotrophic factor (BDNF)* and *peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (PIN1)* were the two genes identified as most commonly reported as harbouring AD associated differential methylation by Frasquet *et al*., (2018). A DMR in the promoter of the *BDNF* gene was found to be hypermethylated in AD peripheral blood by two studies using Asian populations (Chang *et al*., 2014, Nagata *et al*., 2015). However, a third study failed to reproduce this result in a Caucasian population (Carboni *et al*., 2015). Again two studies also identified LOAD specific hypomethylation of the *PIN1* promoter using peripheral blood mononuclear cells (Ferri *et al*., 2016, D'Addario *et al*., 2017).

Other genes identified by Frasquet *et al*., (2018) included *APOE, APP, BACE1*, *DNMT1, DNMT3A, DNMT3B, LINE-1, MTHFR, PSEN1*. The methylation of these genes has been previously discussed. However *Sirtuin 1* (*SIRT1*) has also been identified using leukocyte DNA from AD and non-demented individuals (Hou *et al*., 2013, Frasquet *et al*., 2018). Unfortunately this study lacked information about the methylation status of the promoter in the AD brain and a more recent study failed to identify methylation at the *SIRT1* promoter in LOAD peripheral blood (Carboni *et al*., 2015). While studies using blood DNA seem to produce conflicting results some studies demonstrate that differential methylation of specific genes can be observed in AD blood.

One issue with using blood to investigate DMRs in AD is that information about brain pathology is not usually known. Some clinical markers can be used to assess disease severity, such as cognitive function and quantitative imaging techniques, however it is not possible to assess severity of AD using Braak scoring or neuritic plaque burden as it is when using post-mortem brain DNA, making it more difficult to be sure of disease progression (De Jager *et al*., 2014).

However although there is mounting evidence that suggests epigenetic modifications are linked to LOAD, it is important to consider whether any epi-mutations identified initiate pathogenesis, drive pathogenesis once established or are merely a consequence of the disease. Although previous research has attempted to investigate if any altered methylation identified is driving AD or a result of the disease (De Jager *et al*., 2014), they could not conclude that any epi-mutations were causal. Other studies struggle to even identify when changes occur in disease progression due to smaller sample sizes.

A lot of studies are limited to only looking at change in late stage LOAD compared with aged matched controls. This can be problematic due to epigenetic changes occurring due to aging seeming significant. A notable example are the genes *SORBS3* (sorbin and SH3 domain containing the 3 cell-adhesion protein) a gene encoding a cell adhesion molecule expressed on microglia and neurones and *S100A2* (S100 calcium binding protein A2), encoding a calcium binding protein of the S100 family, both have been identified as harbouring LOAD specific methylation (Siegmund *et al*., 2007, Urdinguio et al., 2009). It has been suggested that the deregulation of these two genes may result in the synaptic impairments present in AD (Urdinguio *et al*., 2009).

However the loss of methylation in these two genes is also seen to a lesser extent in aging. Therefore in this instance LOAD may represent an accelerated form of aging manifesting in the epigenome and those mechanisms in which AD may be underlying this change in methylation. Thus the altered methylation may be a result of AD pathology rather than a driver of it.

It should also be considered that epigenetic changes resulting in change in expression of one gene can affect the methylation of the promoter region of a different gene. One example of this is that hypomethylation of *APP,* which causes Aβ accumulation, results in oxidative stress which drives methylation of the *neprilysin* promoter, suppressing its expression. Since neprilysin is responsible for Aβ degradation, this causes a cycle leading to increased Aβ accumulation which results in further methylation of the *neprilysin* promoter and so on; thus driving Aβ pathology (Chen *et al*., 2009). While this represents an interesting hypothesis, a recent study failed to identify differential methylation in the *neprilysin* promoter in AD using neuronal nuclei cells, perhaps suggesting a need for further study into this (Nagata *et al.,* 2018).

Amyloid induced inflammation has also been proposed to enhance methylation of the *neuroligin 1 (NLGN1)* promoter. NLGN1 is a postsynaptic protein involved in synaptic function and plasticity within the brain. Therefore amyloid induced inflammation could drive down regulation of *NLGN1* leading to amyloid induced memory deficit and aberrant synaptic functionality (Bie *et al*., 2014).

Alternatively aberrant methylation of other genes might result in increased Aβ production. For example global hypomethylation, which is a result of increase Aβ, can result in increased expression of caspase-3 and tumour necrosis factor (TNF-α). This can then cause an increase in the production of Aβ, thus creating a cycle of Aβ production driving AD pathology (Qazi *et al*., 2017).

It is also important to note that many studies only use brain tissue from selected brain regions and only assess specific loci. Therefore in order to achieve a more comprehensive picture, further studies should collect genome-wide data and an atlas of the whole methylome from each region is needed (Siegmund *et al*., 2007).

Specific genes have also been found to harbour AD associated differential hydroxymethylation in the LOAD brain, specific genes were also shown to have altered expression in LOAD and a *Drosophila* model was used to show that the alteration resulted in AD pathology (Bernstein *et al*., 2016). A study conducted using

a mouse model of AD also resulted in the identification of AD associated gene specific and tissue specific hydroxymethylation in AD (Shu *et al*., 2016).

1.6 Diagnosis of AD

Currently a definitive diagnosis of LOAD can only be achieved post-mortem by the identification of amyloid plaques and NFTs in the brain (Yin *et al*., 2009). Before death AD is diagnosed by the detection of neuropathological hallmarks. Such hallmarks include memory loss and decreased cognitive ability, neuronal loss and shrinkage of specific brain regions and also extracellular Aβ deposition and intracellular NFTs within the brain (Hampel *et al*., 2008).

Assessment of memory and cognitive ability is carried out using the mini-mental state exam (MMSE), as well as other neurophyscometric assessments. The MMSE is a quick method of assessment, an individual can be tested in 5-10 minutes (Burns, 1998). However, one large disadvantage to paper based cognitive testing is that ethnicity and levels of education can influence results. Also stress, diet and taking the test in an unfamiliar location can all influence the results. An important point to consider is that the stress of taking the test can make a person's AD linked poor performance worse, therefore making them appear further along in the disease progression than the actual levels of dementia present (Ng *et al*., 2007, Castro-Costa *et al*., 2008). In addition, some patients experiencing MCI do not go on to develop dementia, only around 35% of patients develop AD within three years of mild cognitive impairment (MCI) diagnosis, and after ten years most patients will continue to be dementia free (Herukka *et al*., 2017, Mitchell and Shiri-Feshki, 2009).

Medical imaging such as magnetic resonance imaging (MRI) and functional MRI can be used to assess structural changes within the brain and the activation of brain regions during cognitive tasks. Positron emission tomography (PET) is another imaging technique that allows assessment of cortical metabolism by measuring either fluorodeoxyglucose (^{18}FDG) uptake or the use of Pittsburg Compound B (PiB-PET)

which is an agent use for amyloid imaging. AD patients show a characteristic decrease in uptake of ^{18}FDG and using this technique a conversion of mild cognitive impairment MCI to AD can be predicted with an accuracy of 80%. PiB-PET can also be used to show increased Aβ in the brain, which can also be used to diagnose AD (Herukka *et al*., 2017). However PET along with many of the medical imaging techniques, are not ideal methods for diagnosis due to their high cost and also because they are not high throughput. Therefore it is not possible to efficiently and quickly screen large numbers of individuals (Hampel *et al*., 2008).

AD has been defined as a disease with multiple stages of progression; it is believed that the underlying biological effectors that cause pathology occur several years prior to symptom presentation. Accurate testing and diagnosis of AD before symptoms present, is essential for intervention at the earliest stages of the disease. However imaging and neurophyscometric assessments may not be sensitive enough to detect subtle changes occurring prior to MCI. Protein or genetic biomarkers represent a promising avenue of investigation; ideally an early biomarker or panel of biomarkers found within bodily fluid such as blood or CSF (cerebral spinal fluid) would accurately diagnose different stages of AD progression. This could allow for earlier diagnosis, possibly allowing more effective treatment, as some drugs are thought to be more beneficial when used to treat earlier stages of the disease (Lista *et al*., 2014).

Identification of biomarkers present earlier in the disease may also allow for identification of previously unknown pathways involved in AD, this may result in new therapeutic targets (Richens *et al*., 2014). Identification of an earlier stage biomarker will also allow for better testing of AD drugs. At the moment it is difficult to recruit early stage patients into drug trails so drugs are not being tested on earlier stages of the disease when modifications may be most beneficial (Henriksen *et al*., 2014).

Currently protein biomarkers are used in the diagnosis of AD. Aβ42, total tau protein and levels of phosphorylated tau can be measured in CSF and their levels are all hallmarks of AD. When considered together these three proteins have high diagnostic accuracy (Kang *et al*., 2013).

While CSF biomarkers are useful for diagnosing AD the major limitation of using CSF as a diagnostic fluid is that extraction of CSF requires an invasive lumbar puncture and can only currently be used in late stage diagnosis. This is not favourable to patients. A more accepted and less invasive test would be a blood test which used blood biomarkers to predict AD (Demartini *et al*., 2014).

Currently only T-tau plasma levels can act as a definitive blood based biomarker for AD and there are multiple problems with the identification of AD specific proteins within the blood (Olsson *et al*., 2016). AD is a homogeneous disease meaning that different patients will have different pathologies and thus may present with different biomarkers. For example many patients will present with no increase in Aβ whereas non disease elderly people will show signs of Aβ accumulation (Yang *et al*., 2012). A further problem is that people with AD are commonly elderly and therefore often have other disorders that affect the blood proteome. Therefore it is a requirement to identify which biomarkers are the result of AD and which are those that change due to other disorders.

Ideally biomarkers should be brain specific proteins as this would prevent results from being skewed by proteins produced by other organs or tissues. However it is not feasible to use brain proteins/molecules because most are localised and impeded by the blood brain barrier (BBB). A further complication is that the blood proteome is extremely dynamic and influenced by many physiological factors such as food intake, also variation can exist over a time period as short as a day. Further problems with proteomic based biomarker identification are reviewed in (Henriksen *et al*., 2014, Ghidoni *et al*., 2013).

Another potential problem is the use of healthy controls in the identification of biomarkers. Since the pathology of AD begins years before symptoms present themselves, it is extremely difficult to ascertain if a control is truly a non-diseased patient or if a patient has AD and is just yet to present with symptoms (Henriksen *et al*., 2014). Therefore studies need to be longitudinal and include samples from AD and non-AD patients which are tested over a long period of time, ideally up until

death. This will allow for some determination of whether controls are truly not diseased.

As described previously, many genetic variants have been identified as being implicated in LOAD via GWAS. However, these loci only contribute very slightly to a person's risk of developing AD. Thus there is a huge amount of heritability unaccounted for. However, as also previously motioned, many genes that have a historical association with AD such as *APP, PSEN1, PSEN2* and others have been found to have differentially methylated regions in LOAD patients. Therefore investigations into epigenetic variation within the newly associated LOAD genes may provide some insight into their pathogenic role in LOAD and how epigenetic alterations may influence this.

1.7 Introduction to the project

GWAS studies have resulted in the identification of many genes that harbour genetic variants resulting in both increase risk of LOAD development and protection from the disease. However, these genetic variants do not explain the full risk of AD (Lambert *et al*., 2013b, Rigde *et al*., 2013, Cuyvers and Sleegers, 2016). Taken together, only 61% of disease is explained by known genetic variation. Therefore a large portion of missing phenotypic variance exists (Lambert *et al*., 2013b, Cuyvers and Sleegers, 2016). Since epigenetic changes to the DNA can be driven by environmental factors it is plausible that epigenetic changes may play a role in driving the development of LOAD (Cuyvers and Sleegers, 2016).

The DNA used within this project was collected from both blood leukocytes and brain tissue (cortex). Currently many studies investigating methylation of AD associated genes have used brain tissue. However, the use of AD blood derived DNA has lead to the identification of some gene specific changes in promoter methylation associated with AD (Lunnon *et al*., 2014, Hou *et al*., 2013, reviewed by Frasquet *et al*., 2018). A notable exception however failed to reproduce results observed from brain tissue in blood (Wang *et al*., 2008, Lunnon *et al*., 2014).

Leukocyte DNA represents a good source of DNA for investigating biomarkers of AD because these cells can cross the blood brain barrier (BBB) and thus will alter their gene expression depending on the environment which they encounter in the brain (Hickey, 1991). Blood is also much easier to obtain as it can be taken from living patients.

If differences in methylation can be identified in leukocyte DNA then these changes can then be investigated in brain tissue derived DNA. A change identified in both tissues or even just in blood could result in biomarker identification. However it would be beneficial to compare changes in leukocyte and brain tissue methylation to identify if methylation of gene promoters in leukocytes is reflective of those in brain tissue. This could be possible as it has been shown that leucocytes are capable of crossing the BBB, especially in inflammatory conditions, such as those present due to AD pathology (Eugenin and Berman, 2003, Liu *et al*., 2014). In addition studies have identified DNA methylation in leukocyte DNA that associated with the severity of AD brain pathology present and epigenetic mutation present in the brain DNA have been identified in leukocyte DNA (Stenvinkel *et al*., 2007, Hou *et al*., 2013).

The work completed in this thesis had multiple aims the first being to investigate the methylation status of the promoters of the following newly associated loci *ABCA7, CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/HLA-DRB1*, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4/RIN3*, *SORL1,* and *ZCWPW1* in LOAD leukocyte DNA . However, only loci containing CpG islands upstream of the promoter were investigated.

Interestingly the methylation statuses of 28 AD associated loci were investigated in another study. This investigation included *HLA-DRB5, PTK2B, SORL1, SLC24A4, DSG2, INPP5D, MEF2C, NME8, ZCWPW1, CELF1, FERMT2*, and *CASS4* (Yu *et al.,* 2014b). It was found that in the brain, methylation of *SORL1, ABCA7, HLA-DRB5, SLC24A4*, and *BIN1* could be associated with LOAD pathology. Expression of SORL1 and ABCA7 was also found to be associated with paired helical filament tau tangle density and Aβ overload was associated with over expression of BIN1 (Yu *et al*., 2014b). An additional study also identified LOAD associated differential methylation within regions of *ABCA7, PTK2B, MEF2C, SORL1, CELF1* and *FERMT2* (Humphries *et al*., 2015). One aim

of the project was to determine if the differential methylation of these genes observed in the brain could be reproduced using leukocyte DNA.

The second aim of the project was to investigate if any genes identified as being differentially methylated within LOAD are also altered in sEOAD samples. Since both cortex and leukocyte DNA was available for sEOAD patients and controls it was also possible to investigate the methylation status of specific genic regions in both tissue types. Therefore allowing assessment of whether blood methylation is reflective of that occurring in the sEOAD brain. Genes that have been shown to be differentially methylated in LOAD were investigated using these samples.

A third aim of the work described in this thesis whole genome methylation profiling via whole genome bisulphite sequencing (WGBS), the objective of this approach was to identify patterns of aberrant methylation across the whole genome occurring during different disease stages. For this experiment LOAD cerebellum samples were used, which represented either moderate or severe LOAD cases. The whole genome sequencing for these two samples was then compared to a published control sample to identify disease specific differentially methylated sites.

1.8 Methods used for investigating differential DNA methylation

In order to achieve the aims of the project multiple methods were used to investigate methylation within both LOAD and sEOAD samples. During the work conducted in this thesis four methods were used to investigate single CpG site methylation. Of these, three methods used were targeted and only one was genome wide. The targeted methods used were: PCR amplification following digestion using methylation specific restriction enzyme McrBC and PCR of bisulphite treated DNA followed by either pyrosequencing or cloning, colony selection and Sanger sequencing. Each method for identifying differential methylation is described in more detail in the main methods section 2 and also in each results chapter.

1.9 Conclusions

The following work described in this thesis used four of the methods described to investigated methylation in both LOAD and sEOAD samples. In the case of LOAD, the promoter regions of LOAD associated genes where targeted using the bisulphite cloning methods, the McrBC enzymatic method and pyrosequencing using blood samples. In the case of sEOAD both blood (leukocyte) and brain (cortex) samples were used to also investigate methylation in LOAD associated genes. In addition to these methods WGBS was used to identify global differential methylation in LOAD cerebellum DNA.

2. Materials and Methods

2.1. Alzheimer's disease DNA samples

DNA samples were obtained from the Alzheimer's Research UK Consortium DNA Bank, a resource curated by the University of Nottingham and were either extracted from leukocytes (of LOAD and sEOAD patients and controls) or brain (sEOAD patients and controls) using phenol chloroform extraction followed by DNA quality and quantity assessment via gel electrophoresis and NanoDrop™ 3300 spectrometer respectively. LOAD samples were used in: the investigation of *PTK2β* using bisulphite sequencing and cloning, as template DNA in McrBC PCRs and also in the pyrosequencing of the *SIRT1* target. All other pyrosequencing assays were performed using DNA samples taken from patients suffering from an early onset form of LOAD, sporadic early onset Alzheimer's disease (sEOAD). This type of disease represents a more aggressive form of LOAD. All experiments also included the use of DNA samples taken from non-diseased elderly to act as a control. An overview of the details for the samples use is shown in table 2.1, more extensive details for each sample are provided in appendix 1.

The sEOAD DNA samples used in this study were genetically tested by the Alzheimer's Research UK Consortium DNA Bank and none of the samples contained FAD mutations. In addition, samples used were received with informed consent and experimental procedures were approved by the local ethics committee, Nottingham Research Ethics Committee 2 (REC reference 187 04/Q2404/130). All experimental procedures were conducted in accordance with approved guidelines.

Table 2.1: Details of AD samples used.

Group	Sex M/F	Age at Death	Age at onset	Age at sampling
LOAD Blood	2/5	N/A	71.67 (SD 2.52)	82.25 (SD 6.4)
sEOAD blood	17//8	57.95 (SD 3.28)	47.2 (SD 2.28)	48.6 (SD 2.98)
sEOAD Brain (Cortex)	4/10	59.35 (SD 6.95)	48.85 (SD 3.32)	59.35 (SD 6.95)
Control Blood	6/13	N/A	N/A	82.8 (SD 6.95)
Control Brain	6/4	84.2 (SD 3.55)	N/A	84.2 (SD 3.55)

Table 2.1: Table shows information about age at onset, sampling and death for all of the LOAD and sEOAD samples used during the work described in this thesis.

2.2 Global DNA methylation and hydroxymethylation enzyme-linked immunosorbent assay (ELISA)

2.2.1 DNA preparation

Prior to the use of DNA in the global methylation assay, samples were tested for quality and concentration of the DNA using a NanoDrop™ spectrophotometer. At least 1µL of DNA was loaded onto the NanoDrop™ to assess purity of the DNA, samples scoring over 1.8 at A260/280 were deemed suitable for further use. 100ng of each DNA sample was then used in the global methylation assay, whereas 200ng of DNA was used in the hydroxymethylation assay.

2.2.2 Global methylation and hydroxymethylation ELISA conjugative plate preparation

Global methylation and hydroxymethylation was investigated in LOAD and control samples using the Methylamp ™ Global DNA Methylation and Hydroxymethlyation Quantification kits (Epigenetek Group Inc., New York, NY, USA).

The kit uses antibodies to detect either global methylation or hydroxymethylation within each sample tested. DNA is initially bound to a material with a high affinity for DNA. Methylation or hydroxymethylation is then detected using antibodies. First a capture antibody is added to the DNA containing well, this binds to methylation or hydroxymethylation within the samples. Next a detection antibody is bound to the initial capture antibody. Enhancer solution is then added and finally methylation is detected by the addition of a developing solution which results in a varying amount of colour production, which is measure at OD 450nm. This reading is proportionate to the global levels of methylation or hydroxymethylation present in each DNA sample.

The plates were prepared first by the addition of 80µl of binding solution to each well of the plate to be used. Next, in the instance of the global methylation plate, 100ng of each sample of DNA was added to the designated well, for the hydroxymethylation plate 200ng of DNA was then added to each well. The plates were then rotated gently from side to side in order to mix the DNA sample and binding solution. The plate was then covered and incubated for 90 minutes at 37 °C to allow DNA binding. Following this the binding solution was carefully removed from each well and the well was washed three times with 150µl of 1X wash buffer.

Methylated and hydroxymethylated DNA were then assayed by the addition of the capture antibody. The capture antibody was first diluted with wash buffer (1:1000), 50µl of diluted antibody was then added to each well followed by washing of each well three times with 150µl of 1X wash buffer. 50µl of a 1:2000 dilution (with wash buffer) of detection antibody was then added to each well followed by the washing of each well four times with 150µl of 1X wash buffer. 50µl of a 1:5000 dilution ofenhancer solution was then added to each well and the plate was incubated for 30

minutes at room temperature. The enhancer solution was then removed and each well was washed a total of five times with 150µl of 1X wash buffer.

2.2.3 Detecting methylation and hydroxymethylation

The proportion of methylated DNA was then measured by the use of a Wallac Victor II plate reader (PerkinElmer) prior to this 100µl of developer solution was added to each well and the plate was incubated, in the dark, at room temperature for ten minutes to allow colour to develop. 100µl of stop solution was then added to each well to stop the colour reaction. After one-two minutes the reaction had stopped and the colour of the samples turned from blue to yellow. The absorbance for each sample was then read using a micro-plate reader at 450nm.

As described in chapter 3.2.2 a standard curve was not generated as part of this work, this meant that an exact qualification of methylation could not be obtained. However in a comparison of the absorbance between samples was used as an estimation of methylation present.

2.3 Methylation analysis by bisulphite cloning

2.3.1 Bisulphite conversion of DNA

The process of bisulphite conversion is explained in depth in section 1.8.2. Briefly conversion of genomic DNA with bisulphite treatment allows the identification of methylation at cytosine residues because unmethylated cytosines are deaminated into uracil following treatment of the DNA, subsequent PCR and DNA sequencing result in the unmethylated cytosine being visible as a thymine residue. Methylated cytosines are protected from deamination and therefore remain as cytosines following PCR and are visible as cytosine residues after sequencing.

In order to bisulphite convert the DNA samples prior to cloning the EpiTect DNA Bisulphite Kit (Qiagen, Germany) was used. 500ng of genomic DNA was converted for each patient sample used, following the manufacturer's instructions. Initially 85µl of bisulphite solution and 20µl of DNA and RNase free water solution (with 500ng of genomic DNA) were added to a 200µl PCR tube. 35µl of DNA protect buffer was then added and the sample was mixed briefly, under the correct conditions for successful bisulphite conversion (such as pH) the sample then turned from green toblue. The bisulphite conversion was carried out using a Techne® TC412/4000 thermal cycler (Techne Inc, Burlington, NJ), and the program detailed below in table 2.2.

Table 2.2: Thermal cycler procedure used for bisulphite conversation

Step	Time (min)	Temperature (°C)
Denaturation	5	95
Incubation	25	60
Denaturation	5	95
Incubation	85	60
Deneturation	5	95
Incubation	175	60
Hold	Indefinite (including over night)	20

2.3.2 Clean up and elution of bisulphite converted DNA

Following bisulphite conversion, samples were briefly vortexed to mix and then centrifuged at maximum speed for ten seconds. The samples were then transferred into new 1.5ml microcentrifuge tubes where 560µl of buffer BL was added containing 10 µg/ml of carrier RNA, followed again by brief vortexing and centrifugation at maximum speed (note all centrifugation occurred at maximum speed unless stated otherwise). The content of each microcentrifuge tubes was then transferred into EpiTect spin columns (Qiagen, Germany) and centrifuged for one minute. The flow through was then discarded and the columns were washed with 500µl of wash buffer, again columns were then centrifuged for one minute. The flow

through was discarded once more and 500µl of buffer BD was added to each column. The tubes were then incubated at room temperature for 15 minutes. Following incubation the columns were again centrifuged for one minute and flow through was discarded. The columns were then washed twice with 500µl of wash buffer, as previously described, and the spin columns were then placed into new 1.5ml centrifuge tubes and incubated at 56°C for five minutes to allow evaporation of any remaining liquid. The columns were then transferred to new sterile 1.5 ml microcentrifuge tubes. 20µl of elution buffer was then added to the centre of the membrane of each column and the columns were then centrifuged at 15,000 x g (12,000 rpm) for one minute to elute the bisulphite converted DNA. The bisulphite converted DNA was then the stored at -20°C until use.

2.3.3 Bisulphite PCR for cloning and sequencing of protein tyrosine kinase 2 beta

2.3.3.1 Primer design

To identify a CGI within the *PTK2β* promoter, the gene was located on the UCSC genome browser (University of California, Santa Cruz, CA, USA (www.genome.ucsc.edu/index)) and the nearest CGI located upstream of the transcription start site (TSS) was selected for analysis. The DNA sequence of the CpG island was then obtained using the Ensemble Genome Browser (http://www.ensembl.org/index.html) and primers were designed for the McrBC experiment using the Kismeth: bisulphite primer design website (http://katahdin.mssm.edu/kismeth/primer_design.pl). The primers designed are shown in section 3.2.3.2 and table 3.2. Since PCR DNA template was bisulphite treated prior to amplification, primers were designed to contain degenerate bases. The template strand would either contain a T or a C depending on the methylation status of the original cytosine molecule, since all non CpG cytosines would become Ts due to the bisulphite treatment. Therefore, degenerate bases were included in the primers used. In forward primers Ys were used which represented both C and T residues and in reverse primers R was used to represent A or G residues.

2.3.3.2 Bisulphite Sequencing PCR

PTK2β PCR was completed using a Techne® TC412/4000 thermal cycler (Techne Inc, Burlington, NJ). Primers were used to amplify the template DNA using the PCR reaction cycles: 95°C (5 minutes), 50 cycles (94°C (50 seconds), 54°C (30 seconds), 72°C (60 seconds)), 72°C (ten minutes). Each PCR reaction solution consisted of 4µl bisulphite treated DNA, 20ng of each primer, 4 µl RNAse free water, and 10 µl of Platinum® Blue PCR SuperMix (Invitrogen), resulting in a total volume of 20 µl.

5µl of PCR product was ran on a 2% agarose gel using a TAE buffer and to confirm PCR product. DNA was stained using ethidium bromide.

2.3.3.3 PCR clean up prior to cloning

Following identification of PCR product on an agarose gel the remaining PCR product was cleaned up prior to cloning using the GenElutePCR purification kit (Sigma-Aldrich) following manufacturer's instructions. Firstly spin columns were prepared by the addition of preparation solution to the GenElute minispin column (Sigma-Aldrich), followed by brief centrifugation at 12,000 x g for 30 seconds. The flow through was then discarded and binding solution was added to each column at an amount of five times binding solution to one times PCR product volume. Again the column was centrifuged at maximum speed for one minute and the flow through was discarded. The product was then bound to the spin column membrane. The column was then washed by the application of 500µl of wash solution to each column followed by centrifugation at maximum speed for one minute. The flow through was again discarded and the columns were spun at maximum speed for a further two minutes to remove any residual wash solution from the column. The cleaned up PCR product was then eluted from the column membrane by the addition of 50µl of elution solution followed by one minute incubation at room temperature. Lastly the spin column was placed into a new sterile 1.5 ml microcentrifuge tube and spun at maximum speed for one minute. The PCR product was then either used immediately for cloning or stored at -20 °C until use.

This process resulted in the removal of excess primers and other PCR components such as polymerase and nucleotides which could interfere with the cloning procedure.

2.3.3.4 Cloning

The purified PCR product was cloned into One Shot® TOP10 Chemically Competent E. coli (Invitrogen) cells following ligation into the pCR™4-TOPO® Vector (Invitrogen) using the manufacturer's instructions.

Initially a mix of 4µl of PCR product, 1µl salt solution, 1µl of TOPO® Vector (Invitrogen) and water to make up volume to 6µl, was incubated at room temperature for five minutes, this allowed the PCR product to become inserted into the vector. The pCR™4- TOPO® (Invitrogen) construct was then cloned into the One Shot® TOP10 Chemically Competent E. coli cells (Invitrogen).

2µl of the pCR™4- TOPO® construct (Invitrogen) was added to one vial of One Shot® chemically competent E. coli (Invitrogen) and mixed gently, followed by incubation at -4°C for 30 minutes. The cells were then immediately transferred to a water bath at 42°C for heat shock and then placed back on ice. Next 250µl of room temperature S.O.C medium (Invitrogen) was added to each vial of cells, and the vials were incubated for one hour at 37°C on a 200 rpm shaker. 10-50µl of the cells were then spread onto pre-made selective plates, LB plates containing 50 µg/mL kanamycin which had been warmed to 37°C and the plates were then incubated over night at 37°C. Several plates were made for each vial of transformed cells, ranging from 10µl to 50µl of cell suspension. This was to ensure that the correct concentration of cells to produce well-spaced colonies was achieved.

Multiple colonies were selected from each plate for plasmid purification and subsequent sequencing. Colonies were selected and cultured at 37°C overnight in LB media containing 50µg/mL kanamycin. Plasmid DNA was extracted using the PureLink

Quick Plasmid Miniprep kit (Invitrogen). Firstly, colonies were selected and grown overnight at 37°C in LB media. 1-5ml of the resultant culture was then centrifuged at 13,000 rpm and the liquid media was then removed, to harvest the cells. Following this step, the pellet was re-suspended in re-suspension buffer (250µl), and lysis buffer (250µ) was then added. The tube was then mixed by inversion and left at room temperature for five minutes. 350µl of precipitation buffer was added and the tube was mixed until a homogeneous solution was achieved. The tubes were then centrifuged at >12,000 x g for ten minutes. The remaining supernatant was transferred to a column and washed twice with wash buffer. This involved the addition of either 500µl or 750µl of wash buffer to the column followed by centrifugation at >12,000 x g for one minute. The resultant flow-through was then discarded. To elute the DNA from the columns, the columns were transferred to a new sterile 1.5ml microcentrifuge tube which was incubated at room temperature following the addition of 75µl of TE buffer. The columns were then centrifuged at greater then 12,000 x g for two minutes to recover the plasmid DNA. The columns were then discarded and the DNA was stored at -20°C until use.

To sequence T7 and T3 primers were used for sequencing of 5µl of purified plasmid DNA. Sequencing was carried out by Eurofins MWG Operon. Sequencing results were aligned using BioEdit Sequence Alignment Editor software.

2.4 Using McrBC treatment to identify DNA methylation

McrBC enzymatic treatment was used to assess the methylation status of CGI regions located upstream of the TSS for a number of AD associated genes. The exact process is described in more detail in chapter 3.1, but briefly McrBC is a methylation specific restriction enzyme which cuts at methylated CpG residues within DNA. Therefore PCR after digestion can be used to assess methylation of a specific region following McrBC treatment. PCR primers were designed to cover a region which contained CpG sites, PCRs were then performed using both McrBC treated and non-treated DNA. The PCR

products could then be compared; if methylation was present in the sample then the PCR using McrBC treated DNA template would produce fewer PCR products then that using a non-treated template.

2.4.1 McrBC primer design

Prior to McrBC experiments CpG rich regions were identified and PCR primers were designed to cover these regions. CpG rich islands within the promoter regions of each gene (*PTK2β, SORL1* and *INPP5D*), and an intragenic CGI in the case of HLA-DRB1/5, were identified using the UCSC genome browser (University of California, Santa Cruz, CA, USA (www.genome.ucsc.edu/index)). The sequence of this region was then found using Ensembl (http://www.ensembl.org/index.html*)* and primers covering multiple CpG sites were then designed using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/*).*

PCR primers were designed for regions within the AD associated genes *PTK2β, INPP5D, SORL1* and *HLA-DRB1/5* (for details of primers see appendix 2). In addition primers were also designed to cover regions within genes shown to be differentially methylated in AD by other published methylation studies. Initially primers where taken from published papers for the genes *APOE, TFAM, MTHFRP, BACE* (Wang *et al*., 2008)*, MAPT, PSEN1* (Barrachina and Ferrer, 2009) and *APP* (Barrachina and Ferrer, 2009, Tohgi *et al*., 1999). However the primers reported in these studies did not prove to be effective when tested and extensive optimisation may have been required to produce an adequate PCR product, this could have been due to the use of different reagents and equipment.

However rather than performing time consuming optimisation of these primers the UCSC genome browser was used to identify the region investigated in each paper and the previously mentioned method was used to design new primers, regions covered are shown in table 2.3 below. Primer sequences are detailed in appendix 2.

Table 2.3: Regions investigated in AD differentially methylated genes. The table includes details of each gene region that have been found to contain aberrant methylation in LOAD.

Gene	Paper investigating the gene	Region investigated by paper and methylation status in AD compared with control	Genomic region investigated by paper	Genomic region investigated using McrBC
APOE	Wang *et al*., (2008)	Promoter (hypomethylated) 3'UTR (inter-individual variation of methylation)	GRCh38-chr19-44905658-44906146 (488bp)	GRCh38-chr19-44905658-44906146 (488bp)
TFAM	Wang *et al*., (2008)	Promoter (inter-individual variation of methylation)	GRCh38-chr10-58385094-58385094 (325bp)	GRCh38-chr10-58384929-58305033(104bp)
MAPT	Barrachina *et al.,* (2008), Iwata *et al.,* (2013)	Promoter (hypo-methylation)	Promoter sequence	GRCh38-chr17-891797-892118 (321bp)

As another control, primers were designed to amplify the *GABRB3* promoter region. This region has previously been used as a methylation control (Maunakea et al., 2010). Since the region targeted contains no CpG sites it would be expected that McrBC could not cut in this region and thus a PCR using a McrBC treated DNA

template should produce the same amount of product as that of an untreated sample of DNA.

PCRs were performed containing both the primers amplifying the gene of interest and the control gene. Multiple sets of control primers were used so that products of multiplex PCR would be distinguishable from each other. The primers used are shown in table 2.4.

Table 2.4: GABRB3 primers and PCR product length. Tables shows details of the primers used in each to amplify the control region of *GABRB3.*

Gene	Primer name	Primers	Product length (bp)
GABRB3	GABRB3_P1_F1 GABRB3_P1_R1-	CCTGCAACTTTACTGAATTTAGC GGAATCTCACTTTCACCACTGG	206
GABRB3	GABRB3_P3_F1_406 GABRB3_P3_R1_406	TTTTCCCTACCCCTACCCAGT AATGCAGGCACATTTCCTGTTG	406

2.4.2 MrcBC treatment

Prior to McrBC treatment of the DNA samples the samples were tested using a NanoDrop™ to assess quality and concentration of the DNA as described in section 2.3.1. Following this each set of primers were tested and optimised in order to produce a visible PCR product with an input of 10ng of DNA. Once PCRs were successful 250ng of genomic DNA was treated with McrBC for each AD and control sample investigated. This was achieved using the following protocol. 250ng of genomic DNA was treated with 20 units of McrBC enzyme (New England Biolabs) in a 50µl reaction containing: 250ng genomic DNA, 20 units McrBC enzyme, 1X NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 @ 25°C), supplemented with 200 µg/ml BSA and 1 mM GTP and incubated at 37°C overnight.

This mixture was then heated to 67°C for twenty minutes to inactivate the McrBC enzyme. DNA was then stored at -4 °C until use.

2.4.3 PCR and agarose gel electrophoresis

PCR reactions were completed using both McrBC treated DNA and non-treated DNA as templates. Both types of PCR reaction consisted of the components shown in table 2.5, and had a final volume of 20µl.

Table 2.5: Components of McrBC PCR reactions. Table shows the McrBC PCR components and the amount of each included in each PCR mix.

PCR component	Amount (µl)	Concentration/ Units
DNA	0.4 (untreated DNA) 2 (McrBC treated DNA)	10ng of DNA
Platinum® Blue PCR SuperMix	10	0.22 units (10µl of 22U/ml)
Forward Primer	0.5	5ng (10ng/µl concentration)
Reverse Primer	0.5	5ng (10ng/µl concentration)
RNAse free water	variable	

The PCR program then used was as follows: cycles were: 95℃ (5 minutes), 35 cycles (94°C (50 seconds), 59°C (30 seconds), 72°C (60 seconds)), 72°C (10 minutes), PCR was conducted using a Techne® TC412/4000 thermal cycler (Techne Inc, Burlington, NJ). PCR products were then ran on a 2% agarose gel stained with ethidium bromide to assess successful amplification.

2.4.4 Computer methods

Band density quantification was undertaken using Genesync imaging software. This software provided a relative number representative of the amount of PCR product

visualised on the gel following amplification. This information was then used to compare the efficiency of PCRs using McrBC treated and non-treated DNA templates. Microsoft Excel was used to compare the two values and to plot the results. Two tailed T-test were also performed using Microsoft Excel and all outliers were excluded from the analysis, this was common throughout all statistical analysis performed as part of this thesis. Outliers were defined as values that were greater than 15 times the interquartile range away from the mean. T-test was used as it is a robust method for comparing small sample sizes and is less dependent on the mean when compared to a Mann-Whitney U test. All normality testing was conducted using the Kolmogorov-Smimov Test of Normality (shown in supplementary figures), however for data sets of less the five normality testing and covariant analysis was not performed due to the statistical limitations associated with sample size.

2.5 Methylation analysis by pyrosequencing

2.5.1 Bisulphite treatment of DNA

Prior to pyrosequencing DNA was bisulphite converted, as previously described (section 2.4.1). 500ng of genomic DNA was bisulphite converted using the EpiTect DNA Bisulphite Kit (Qiagen, Germany) following the manufacture's instructions.

2.5.2 Pyrosequencing primer design

Prior to pyrosequencing primer design CGIs located within the targeted genes were identified using the UCSC genome browser (University of California, Santa Cruz, CA, USA (www.genome.ucsc.edu/index), with the exception of the regions targeted for *SIRT1* and *TREM2,* for these two genes primers were designed for the regions targeted by either Hou *et al*. (2013) or Smith *et al*. (2016).

Pyrosequencing PCR Primers were designed using the program PyroMark Assay Design 2.0 (Qiagen, Germany) and obtained from Eurofins MWG. The primer sets contained either a forward or reverse primer that was biotinylated at the 5' end. This allowed capture of the PCR product on Streptavidin coated Sepharose High

Performance beads (GE Healthcare, UK) during preparation for pyrosequencing. Each set of primers was given a score out of a hundred and were colour coded. The colour blue meant excellent primers for that target. Primers could also be yellow, orange or red (with red being the least effective primer sets).

A detailed report is also available for each primer set. This report includes a diagram of primer binding sites on the target sequence, information about the size of the PCR product, the melting temperatures of the primers and details of the target sequence. The sequence to analyse is important because this sequence is used to determine the nucleotide dispensation order during the pyrosequencing reaction. This report also gives details about and potential problems with the primer sets such as chances of mis-priming, secondary structure formation, and self-complementarity. All of these factors were taken into consideration when choosing primer pairs. In all instances the primer with the highest cumulative score was chosen for use in the pyrosequencing reaction, if two or more sets of primers were used then the second best primers were also selected.

A full list of pyrosequencing primers used during work completed in this thesis can be found in appendix 3.

2.5.3 Pyrosequencing PCR

Initially the PCR mix used for pyrosequencing PCR reactions was as described for the McrBC PCR reactions using the Platinum Blue Supermix (Invitrogen, UK) (table 2.5). Using this PCR kit multiple optimisation steps were taken to result in PCR product, these steps were, step down, gradient and nested PCRs were used as well as extensive optimisation of primer amount and DNA amount required for each PCR. The experiments conducted led to the discovery that the optimum amount of each primer to use in PCRs was 0.25pmol and 0.1µl of bisulphite converted DNA (500ng eluted in 20µl of elution buffer).

However, the use of two round of PCR resulted in significant problems with negative contamination and inefficient PCR reactions. This resulted in the use of the PryoMark PCR kit (Qiagen, Germany) which is specifically designed for use in pyrosequencing PCRs. PCRs using *SIRT* primers and all other candidate genes were completed using the PyroMark PCR kit (Qiagen, Germany) exclusively. These PCR solutions were conducted following manufacturer's instructions and contained 12.5µl PyroMark PCR Mastermix 2x, 2.5µl CoralLoad Concentrate 10x, 0.25ng of each PCR primer and 10-20ng of bisulphite converted DNA, made up to 25µl with RNase-free water and did not require the use of two rounds of PCR to generate sufficient PCR product for sequencing.

PCR programmes used also varied depending on the primer set, PCR program details are shown in appendix 4. All PCRs were completed using either the Techne® TC412/4000 thermal cycler (Techne Inc, Burlington, NJ) or the Master cycle gradient machine (Eppendorf, Germany). For most gene candidates a step down and gradient PCR was initially performed to assess optimum annealing temperature. In these cases a temperature of around 5°C less than the average of the melting temperature of the two primers was chosen as a mid value for the gradient. Following gradient PCR the temperature responsible for the strongest band was chosen for further optimisation, this included the addition of more cycles if the band appeared as faint, the number of cycles used ranged from 35-50.

Prior to analysis by pyrosequencing all PCR products were assessed on a 2% agarose gel and stained with ethidium bromide solution. PCR products were run on the gel with 5µl of 1kb HyperLadder (Bioline). Gels were visualised using a transilluminator (GeneSys) and only PCR products showing strong specific bands, with no primer dimer bands present on the gel, were used for subsequent pyrosequencing assays.

2.5.4 Pyrosequencing assay design

Pyrosequencing assays were designed using the PyroMark Q24 software (Qiagen, Germany). The sequence to analyse was taken from the sequence given for each primer set by the primer design software. This was then used to create an assay.

2.5.5 Pyrosequencing process

Initial pyrosequencing was carried out using the manufacture's protocol. Specifically, the biotynylated PCR product was immobilized using Streptavidin coated Sepharose High Performance beads (GE Healthcare, UK). To achieve this 15µl of PCR product, 2µl of beads, 40µl of binding buffer (Qiagen, Germany) and 23µl of high-purity water (either Milli-Q (18.2 MΩ·cm) or RNase free water) (total volume of 80µl) were added to a new 200µl tube. The tube was then agitated at 1400 rpm for ten minutes.

The PyroMark Q24 vacuum workstation was then used to produce single stranded PCR product with the sequencing primer annealed, this was then used for sequencing. First the bound PCR product was washed with 70% ethanol, and denatured using denaturation solution, by lowering the vacuum tool, with captured PCR product held under suction, to each solution for five and ten seconds respectively. The then unbound strand was washed away, by application of vacuum tool to wash solution for ten seconds under suction, the single strand which was still bound to the beads and the vacuum remained bound while one strand was washed away. Binding buffer, denaturation solution and wash buffer were all also purchased from Qiagen.

Sequencing primer was then annealed to the still attached strand of DNA. The vacuum was switched off to allow the strand to dissociate from the vacuum. This was then incubated in a 0.3µM solution of specific sequencing primer (diluted with annealing buffer (Qiagen, Germany)) on a Qiagen Q24 pyrosequencing plate, with gentle agitation. Following this the plate was transferred to a preheated heating block using PyroMark Q24 Plate Holders (Qiagen) and incubated at 80 °C for two minutes followed by cooling at room temperature for a further five minutes, to allow the sequencing primer to anneal to the single stranded PCR product template.

Pyrosequencing was then carried out using the Pyromark Q24 Pyrosequencer (Qiagen, Germany). A PyroMark Q24 Cartridge was filled with Pyromark Gold Q24 reagents, following the instructions given by the pyrosequencing software for that specific assay. The Pyromark Gold Q24 reagents included lyphillised enzyme and

substrate mix which were reconstituted prior to use by the addition of high-purity water followed by gentle mixing. They were then allowed to reach room temperature prior to use. Enzyme mix (DNA polymerase, ATP sulfurylase, luciferase and apyrase), substrate mix (adenosine 5' phosphosulfate and luceferin) and dNTPs (dATPαS, dCTP, dTTP and dGTP) were then added to the specific wells of the PyroMark Q24 Cartridge in the amounts defined by the pre-designed assay. The cartridge was then placed into the pyrosequencer with the label facing forward. Pyrosequencing assays were then run and data was transferred to a USB device.

Following multiple attempts at pyrosequencing some modifications were introduced to this procedure in order to improve the pyrosequencing outcome. Firstly more PCR product, Streptavidin coated Sepharose High Performance beads (GE Healthcare, UK) and binding buffer were used to improve the likelihood of PCR product being bound to beads for subsequent stages of pyrosequencing, therefore increasing the sequencing quality. The pyrosequencing mixture therefore instead contained 20-25µl of PCR product, 4µl of beads and 51-56µl of binding buffer. In addition to this the concentration of sequencing primer used was also increased to 0.5µM to increase annealing of the sequencing primer to PCR template.

2.5.6 Serial pyrosequencing

In the case of serial pyrosequencing, used to cover a larger area of the *ABCA7* region investigated, multiple sequencing primers were used. An initial round of pyrosequencing was carried out as described above however following this the initial sequencing primer was removed from the PCR template and a second sequencing primer was annealed. This was achieved by the addition of a further 20µl of binding buffer to each well of the pyrosequencing plate, followed by pipetting the solution several times to resuspend the streptavidin coated sepharose beads. The PCR template was then denatured using the workflow described above to produce a single stranded template. The second sequencing primer was then annealed to the single stranded template as also described above (Tost et al., 2006).

A further advantage of serial pyrosequencing is that it allows the use of residual components remaining within the pyrosequencing cartridge following pyrosequencing. After the initial round of pyrosequencing about 60µl of each substance (enzyme mix, substrate mix and dNTPs) remain in each well of the cartridge, therefore rather than washing the cartridge and adding new reagents, the wells could be topped up to the volume required for the next round of sequencing. This is much more cost effective then completing two independent rounds of pyrosequencing and also reduced the time required for multiple rounds of sequencing.

2.5.7 Pyrosequencing results analysis

Following pyrosequencing, results were analysed using the PyroMark Q24 software. The software provides a score for each CpG site covered by each assay and also an idea about the assay as a whole. Runs are scores as blue (pass), yellow (check) or red (fail). All failed pyrosequencing runs were repeated. In addition at least two technical reps were performed for each sample, unless otherwise stated.

An internal control for the efficiency of the bisulphite conversion of each sample was also included in every pyrosequencing assay. Prior to the pyrosequencing run at least one non-CpG cytosine was chosen to be analysed in the assay. Since these are not expected to be methylated all of these should be converted during the bisulphite treatment. The amount of conversion was measure during pyrosequencing and conversion of less the 95% would result in a failed assay, this would have resulted in repeat bisulphate conversion of the sample, it should be noted that no samples tested failed this threshold quality control test.

2.5.8 Statistical analysis

Statistical analysis was performed using Microsoft excel or Genstat 18 statistical packages (Genstat work was conducted by Dr Jim Craigon, University of Nottingham).

In the instance of McrBC data and some pyrosequencing results two tailed T-tests were performed to compare the averages of two sets of data (following removal of outliers). In most cases this would be the average methylation found at each CpG site investigated in control compared to AD samples. Although this was also used to investigate methylation across the whole CpG containing region investigated for each gene). A P value of <0.05 was deemed to represent a significant difference between the average of two data sets. Kolmogorov-Smirnov Test of Normality calculation were performed on all data sets where n is greater than five; this and other tests for normality are for data sets smaller than this. All data was shown to be normally distributed.

For *RIN3* data one-way analyses of variance were used to test the blood and brain tissue data for differences between the Alzheimer and control groups of subjects. This was undertaken for each individual CpG site tested in each tissue, each of the cytosines were analysed separately as were the blood and brain tissue data sets. When multiple values were recorded for the same cytosine from the same subject the values were averaged to produce one value per person for each cytosine. These averages were then analysed using the Analysis of Variance routines within the Genstat18 statistical package.

In the case of the *MEF2C* epi-varaint the residuals from each analysis were plotted to check the assumptions of Normality and homogeneity of residuals, and also to highlight possible rare variants. For rare variants, the probability of obtaining residual values as, or more, extreme than the one observed was calculated from the standard Normal distribution curve. To achieve this, the residual deviation between the individual outlier and the mean of its group; and the overall standard deviation from the rerun analysis were used as input to the Genstat Normal probability distribution routine.

Testing for co-variants: The pathology of EOAD i.e. a potential age spread from 30-65 year of age (see www.nia.nih.gov/) presents a number of challenges for identifying appropriate control samples and for the identification and exclusion of co- variants.

Samples may be age matched, however this does not control for disease i.e. until those individuals reach age 65 it is not possible to exclude the possibility that individuals within the control group have pre-symptomatic EOAD. In order to exclude disease as a possibility samples over the age of 65 were chosen. The control group used for the study was therefore statistically different older age group (65+), this violates the rules of ANCOVA (or other co-variant tests) (Figure 2.1). However given there is a >20 year age spread in the AD age group, it was therefore possible to test and exclude age as a co-variant by testing within the EOAD group. This experimental design is also independent of condition and therefore allows for the identification of candidate epi-SNPs.

In the instance of epi-SNPs co-variants become less relevant. The hypothesis in this thesis is that by testing genes with published associations with disease progression, any epigenetic change may drive the pathogenesis of the disease. Therefore although it may not be possible to ascertain the cause of different methylation profiles, there is a strong rational for assuming that this difference in methylation is relevant to disease state.

Although co-variant analysis between groups was not included as part of the experimental design, the analysis was performed and the effect on P-values are shown in the supplementary figures.

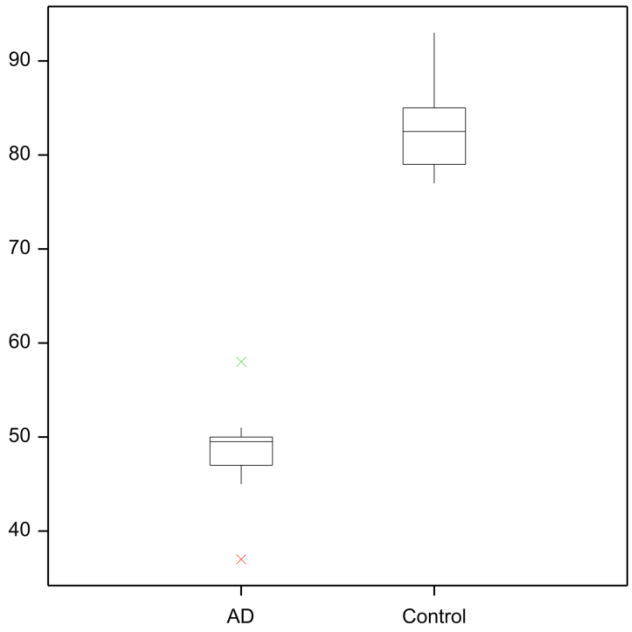


Figure 2.1: Ages of Control and AD samples used. Box plots showing difference in ages of AD and control samples used.

2.6 Whole genome bisulphite sequencing (WGBS)

2.6.1 Sample information

Two samples of LOAD cerebellum DNA were used for WGBS experiments. The DNA was obtained from the BDR (Brains for Dementia Research) and came from two LOAD suffers with differing severity of disease. One sample was used from a patient presenting with moderate AD (Braak stage IV) and one patient suffering with severe disease (Braak stage VI). Details of each sample are shown in appendix 1.

The WGBS results for these two samples were then compared to each other as well as data for a cerebellum control sample. Control data was obtained from the NCBI sequence read archive (SRA) (https://www.ncbi.nlm.nih.gov/sra). Specifically WGBS of this sample was carried out by Professor Kun Zhang of the Integrative Genomics Laboratory, Bioengineering, University of California, San Diego during project number PRJNA315194 and the sample used was SAMN04557041, SRS1341238 (N37-CRBL), *GEO* accession*:* GSM2088203.

2.6.2 Bisulphite conversion

Bisulphite conversion of 100ng of DNA from each sample was completed prior to WGBS library preparation using the EpiTect Bisulphite Conversion Kit (Qiagen, Germany) using the protocol described in section 2.4.1. Clean up of bisulphite converted DNA was also completed following the method described in section 2.4.2.

2.6.3 Library preparation

Library preparation following bisulphite conversion was carried out using the Truseq DNA methylation library preparation kit (Illumina) following the manufactured instructions. This kit was chosen as it allows the use of a much smaller initial DNA input then other available kits. The method used includes bisulphite conversion of DNA prior to library preparation; this means only 50-100ng of input DNA is needed for this initial conversion. With alternative kits bisulphite conversion commonly occurs after the DNA library has been constructed. This results in loss of library content due to DNA degradation caused by the bisulphite treatment.

Following bisulphite conversion of the DNA the now ssDNA template is used for the addition of adapter sequences which are later needed for cluster generation during sequencing. This also means that, unlike with other methods, no fragmentation of the DNA is needed prior to the addition of adapter sequences and methylated adaptors are also not required. The WGBS library preparation method is described briefly in figure 2.2.

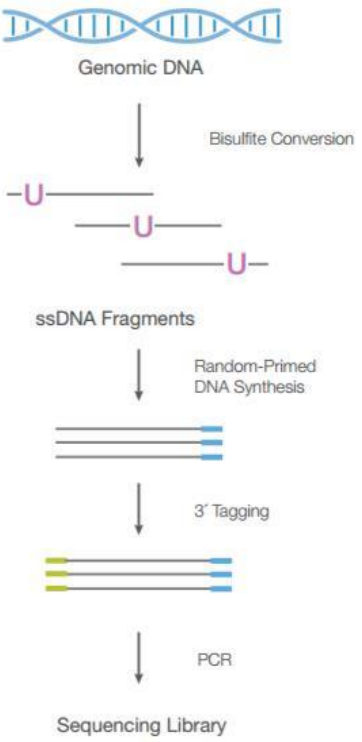


Figure 2.2: Process of using the Truseq DNA methylation library preparation kit for WGBS library preparation. Image obtained from (www.illumina.com).

2.6.3.1 Anneal synthesis primer

Following bisulphite conversion the final bisulphite converted DNA was eluted in 9µl of elution buffer. The quantity of the now bisulphite converted ssDNA was then assessed using a NanoDrop™ and the 50ng of the recovered DNA was used to complete the library preparation.

The first step in library preparation was to anneal the DNA Synthesis Primer to the ssDNA template. Initially the DNA Synthesis Primer was thawed on ice followed by inversion to mix and brief centrifugation at maximum speed. 9µl of ssDNA template was then combined with 2µl of DNA Synthesis Primer, in a separate PCR tube. The PCR tube was then transferred to a thermal cycler (Techne Inc, Burlington, NJ) and heated to 95°C for five minutes followed by incubation on ice.

2.6.3.2 Synthesis of DNA

Once the DNA Synthesis Primer had been annealed to the ssDNA template the template was then used to synthesise DNA which was tagged with a specific sequence resulting from the random hexamers added during the annealing of the synthesis primers.

For this stage of the library preparation a master mix was initially made containing: 4µl of TruSeq DNA Methyl PreMix, 0.5µl of 100mM DTT and 0.5µl of TruSeq DNA Methyl Pol per sample (total of 5µl of master mix per samples of DNA). All regents were thawed on ice prior to use.

5µl of the master mix was added to each DNA sample, on ice, to make a final volume of 16µl. The samples were then mixed by pipetting and placed in a thermal cycler. The following program was then run:

preheat lid

25°C for 5 minutes

42°C for 30 minutes

37°C for 2 minutes

Hold at 4°C

The PCR tubes were then removed from the thermal cycler and 1µl of exonuclease I was added to each tube and pipetted to mix. The tubes were then placed back into the thermal cycler and the following program was run:

preheat lid

37°C for 10 minutes

95°C for 3 minutes

25°C for 2 minutes and hold

2.6.3.3 Tagging of the DNA

Following DNA synthesis the DNA was then di-tagged. The 3' end of the synthesised

DNA was extended and a complimentary sequence was added resulting in di-tagged DNA with known sequence at both ends (3' and 5').

Initially a mastermix was made containing 7.5µl of TruSeq DNA Methyl Term Tag PreMix and 0.5µl of DNA Polymerase per sample. Prior to master mix preparation both regents were thawed on ice, inverted to mix and then centrifuged briefly. 8µl of master mix was then added to each DNA sample, and mixed using pipetting. The tubes were then placed back into the thermocycler and the following program was ran:

Preheated lid

25°C for 30 minutes

95°C for 3 minutes

Hold at 4°C

2.6.3.4 Tagged DNA Clean up

The now di-tagged DNA was then purified using 1.6x concentration of AMPure XP beads (Beckman Coulter, Bucks, UK). Prior to the clean up the beads were allowed to reach room temperature for at least 30 minutes. Following this 40µl of AMPure XP beads (Beckman Coulter, Bucks, UK), were added to each sample and mixed using pipetting. The mix was then transferred to a new 1.5ml microcentrifuge tube and left at room temperature for five minutes. The tube was then placed into a magnetic rack until the liquid became clear. This clear liquid was then removed and discarded. The remaining beads were then washed with 200µl of freshly prepared 80% ethanol followed by incubation at room temperature for one minute on the magnetic rack. Again the liquid was then discarded. This procedure was repeated twice. Following the second wash all ethanol was removed from the tube and the tube was centrifuged briefly and placed back onto the magnetic rack for one minute. All remaining ethanol was then removed from the sample using a P10 pipette. The tube was then air dried on the magnetic rack for three minutes. The tubes were then removed from the magnetic rack and 24.5µl of nuclease free water was added to

each sample. The samples were then pipetted to re-suspend the beads. The tubes were then incubated at room temperature for two minutes and placed back onto the magnetic rack until the liquid was again clear (approximately five minutes). 22.5µl of the supernatant, containing the di-tagged DNA, was transferred to a new PCR tube and placed on ice.

2.6.3.5 PCR amplification of the library

PCR was then used to amplify the library and also to create the second strand of DNA. This step also involved the addition of adapter sequences to each end of the DNA strand. In the case of the libraries made during this thesis index primers were not added or used. Instead the TruSeq DNA Methyl Reverse primer was used to create non-indexed libraries.

Prior to use all regents were thawed on ice, mixed by inversion and centrifuged briefly. To each PCR tube containing the di-tagged DNA the following was added: 25µl of FailSafe PreMix E, 1µl of TruSeq DNA Methyl Forward, 1µl of TruSeq DNA Methyl Reverse and 0.5µl (1.25U) of FailSafe PCR Enzyme Mix (Epicentre). The total volume of each sample was then 50µl.

The PCR tubes were then placed back into the thermal cycler and the following program was ran:

preheated lid

95°C for 1 minute

10 cycles of:

- 95°C for 30 seconds
- 55°C for 30 seconds
- 68°C for 3 minutes
- 68°C for 7 minutes

Hold at 4°C

2.6.3.6 Library Clean Up

The library was then purified using a 1.0x concentration of Agencourt AMPure XP beads (Beckman Coulter, Bucks, UK). This resulted in the removal of primer dimmers that may have occurred during PCR amplification.

50µl of AMPure beads were added to each PCR tube and mixed using pipetting. The entire volume of the PCR tube was then transferred to a new 1.5ml microcentifuge tube and incubated for five minutes at room temperature. The tubes were then placed on a magnetic rack until the liquid became clear (approx five minutes). The supernatant was then discarded and the beads were washed twice with 80% ethanol as described in section 2.7.3.4. The tubes were then centrifuged briefly and allowed to stand on the magnetic rack for a further minute. A P10 pipette was then used to remove any remaining liquid and the tubes were air dried for three minutes on the magnetic rack. 20µl of nuclease free water was then added to each tube and the beads were re-suspended. The tubes were then incubated at room temperature for a further two minutes and placed back onto the magnetic rack until the liquid became clear (approximately five minutes). The TruSeq DNA methylation library containing supernatant was then transferred to a new PCR tube which was stored at -25°C to -15°C until use.

2.6.3.7 Quality Control Check of Libraries

The quality and quantity of the WGBS libraries was assessed by Polar Genomics (New York) prior to sequencing. DNA concentration was calculated using a Qubit dsDNA high sensitivity assay and average fragment size and was calculated using a Agilent 2200 Tape Station.

2.6.4 Whole genome bisulphite sequencing

Sequencing of WGBS libraries was performed by Polar Genomics (New York). Each whole genome sequencing sample was sequenced using one lane of an Illumina HiSeq 2500 (Illumina, UK) using pair end 2X 150 BP sequencing. A 20% PhiX spike in was added to the library prior to sequencing to increase library diversity and therefore increase cluster generation, improving sequencing.

2.6.5 Bioinformatic analysis

The data analysis was performed by Joanna Moreton from The Advanced Data Analysis Centre (ADAC) at the University of Nottingham.

The reads were trimmed based on quality and adapter contamination using Trim Galore (v0.4.3) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with a quality score cut-off of 20. FastQC (v0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was then run to assess the quality of the trimmed reads. The human reference genome was downloaded from Ensembl (release 88) (Aken et al., 2017). The sequencing analysis was performed using Bismark (v0.17.0) (Krueger et al, 2011). In the first step, the genome was bisulphite converted and then Bowtie 2 (v2.3.0) (Langmead and Salzberg, 2012), was used to align the trimmed reads to the genome. Next, deduplication was performed by Bismark to filter out reads with the same orientation and position. In the final Bismark step, the methylation information was extracted, and used as input for 'DSS-single' (Bioconductor package 'DSS').

For the differential methylation loci (DML) analysis, all methylation calls with methylation percentage greater than 50%, and coverage greater than 10, were used. These filters were applied to all samples before using as input to the tool "DSS-single" to find DMLs (P-value threshold of 0.01/0.001) (Wu et al., 2015). Finally, 'bedtools closest' (v2.26.0) was used to find the gene(s) nearest to the methylation calls (Quinlan and Hall, 2010).

Chapter 3- Investigation into differential methylation within the promoter regions of LOAD associated genes using LOAD leukocyte samples

3.1 Introduction

Thus far the causation of LOAD cannot be explained by genetics alone and there is mounting evidence that epigenetic modifications, such as DNA methylation, may have a role to play in LOAD susceptibility. It is therefore logical to investigate if any differential methylation could be observed at loci genetically associated with this disease i.e. the hypothesis is the contributing factor is not genetic but rather an epigenetic influence on LOAD associated gene expression. The work described in this chapter therefore aimed to use both bisulphite sequencing and enzymatic methods to investigate if differential methylation could be detected in the promoter regions of the GWAS identified genes *CASS4, CELF1, FERMT2, HLA-DRB5/HLA-DRB1, INPP5D, MEF2C, NME8, PTK2ϐ,SLC244A4/RIN3, SORL1* and *ZCWPW1* (Lambert *et al*., 2013).

In the experiments described in this chapter, DNA extracted from LOAD blood (leukocyte) samples were used throughout. Previously published experiments commonly focus on the use of brain tissue; however, leukocyte DNA was used in the following experiments as blood is easily accessible whereas brain tissue has to be collected post mortem. It is well established that LOAD specific differential methylation can be identified using blood samples (Hou *et al*., 2013, Chang *et al*., 2014, Lunnon *et al*., 2014). The working hypothesis is that blood cells encounter the AD brain environment and then adapts the epigenetic regulation of genes in response. Epigenetic changes detectable in blood, such as DNA methylation, might therefore provide insight into the processes occurring within the AD brain (Hickey, 1991; Weiss *et al*., 1998; Callahan and Ransohoff, 2004; Takeshita and Ransohoff, 2013).

The work described in this chapter aimed to identify differential methylation in LOAD blood samples that could be potentially used as a biomarker of disease. The exploitation of peripheral blood cell methylation as a biomarker has been investigated with some success for a range of disparate diseases and conditions including cancer and cardiovascular disease (reviewed by Dong and Ren, 2018; Cardona-Monzonis *et al*, 2018). Also importantly researchers have also identified changes in peripheral blood specific methylation in response to neurological diseases such as Parkinson's disease (see Wang *et al*, 2019), confirming a link between diseases afflicting the brain and blood borne methylation. Previous studies have also investigated the applicability of methylation as a biomarker for Alzheimer's disease (Hou *et al*, 2013, Chang *et al*., 2014, Lunnon *et al*., 2014) and it is therefore unsurprising that blood borne marks for this disease are gaining some attention (e.g.Kobayashi *et al*, 2016).

Significant changes in blood specific methylation linked to AD have been validated using a range of techniques including pyrosequencing (Xu *et al*, 2018; Madrid *et al*, 2018) and may link to pathological features of disease such as cognitive decline (Mercorio *et al*, 2018), therefore outputs from this chapter may have applications as markers of AD.

Any differential methylation identified in blood has potential value for use as a biomarker regardless of its ability to reflect brain methylation. Therefore while it was not possible to compare the results presented within this chapter to brain tissue data, due to not having access to this type of LOAD sample, there is value in identifying differential methylation in the blood of LOAD patients.

However one issue with using blood samples is fast turnover of cells, co-morbidities in the elderly and medication. Backland *et al.* (2015) showed that some medicine can induce changes in methylation, however such effects likely require combinations of medicines and the resultant methylation profiles are complex and medication specific i.e genes linked to p-glycoprotein activity and therefore unlikely to either effect AD related genes or skew group wide averages (Ward-Caviness *et al.,* 2017).

Leukocytes also rapidly turnover, however this may result in a quicker response to factors that influence DNA methylation than cells that turn over more slowly. As discussed by Wu *et al.* (2012) in this instance DNA methylation may or may not identify DNA methylation errors directly indicative of any disease but may instead have value as biomarkers for the biological processes that influence DNA methylation.

Three experimental techniques were used to generate the data described in this chapter. Initially a quantitative ELISA assay was performed on LOAD blood samples to investigate aggregate levels of genome wide methylation. The kit utilised two antibodies. The first has high affinity to methylated DNA and is initially used to capture any methylated DNA. A second detection antibody is then introduced which binds to the capture antibody and will emit a colour when exposed to a developing solution. The amount of methylated DNA present in the sample is proportional to the optical density (OD) measured following colour development. This experiment was conducted to eliminate the possibility that any differential methylation identified is due to disease induced global AD non-specific changes in the gross aggregates of methylation rather than gene specific effects.

AD may induce changes in leukeocyte cell types e.g. monocytes relative to neutrophil levels (see Shad *et al*, 2013). It should be noted that each cell type may possess type specific methylation and therefore levels determined via ELISA may reflect this shift in cell type content driven by AD. However a detectable shift in disease linked methylation via cell type would further support the rationale for global methylation profiling as a marker of disease. Recent research has suggested that changes in disease linked levels of leukocyte sub-cell methylation levels may be harness as a predictive marker for other conditions such as cancer (Koestler *et al*, 2017).

In addition to undertaking the quantification of the global levels of methylation via an enzyme based immuno-assay, both bisulphite sequencing and an enzymatic digestion method were also employed.

Bisulphite sequencing was used to investigate methylation within the promoter region of the gene *PTK2B* (encoding protein tyrosine kinase 2β) in LOAD leukocyte samples using bisulphite cloning. In this method sample DNA was first bisulphite

treated; to convert any unmethylated cytosines to uracil. The targeted region in the converted DNA was then amplified via polymerase chain reaction (PCR) which in this case was a selected sequence of the *PTK2β* promoter (see figure 3.3). The subsequent PCR product was then cloned into a vector, and transformed into competent cells. These cells were then cultured to form colonies. Each colony was from a cell containing a vector carrying only one PCR product. Therefore multiple colonies could be selected and sequenced to assess multiple amplifications, giving quantitative methylation data at CpG sites for each individual.

The *PTK2β* promoter was assayed using this method. *PTK2β* has previously been associated with LOAD development via GWAS (Lambert *et al*., 2013). *PTK2B* encodes an intercellular kinase, a protein involved in mitogen activated protein kinase (MAPK) activity. The MAPK pathway has functions in regulating neuronal activity following neuropeptide activation and may have a role in long term potentiation and memory formation (Lev *et al*., 1995). *PTK2β* might also have a role in cell proliferation and survival (Otero *et al*., 2009). Therefore it is conceivable that altered *PTK2B* expression or activity could be a driving factor of LOAD.

Following the use of the bisulphite cloning methods for analysis of *PTK2β,* work progressed to the use of a more high-through-put and less laborious enzymatic method. The promoter CGI regions of other LOAD associated genes were investigated using the restriction endonuclease McrBC. McrBC is a DNA cleaving endocuclease which requires GTP and originates from *Escherichia coli* K-12. The enzymes cleavage site contains two variably spaces R^mC recognition elements, these sites can be on either both stands or just one strand of the DNA. Cleavage occurs near one of these recognition sites and is enacted when the translocating McrBC complex encounters a barrier i.e methylation (Stewart *et al*., 2000). Cleavage therefore occurs at methylated cytosines within a CpG configuration.

After McrBC treatment, DNA is then amplified via PCR. Methylation is detected when there is less than expected PCR product after treatment. This is due to the relative loss of template compared to control after cleavage when methylation is present.

McrBC was used to investigate the promoter regions of, *INPP5D, PTK2B, SORL1* and an intragenic region of *HLA-DRB5* and *HLA-DRB1*. An effort was made to design assays which targeted the genes *CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/HLA-DRB1*, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4/RIN3*, *SORL1* and *ZCWPW1,* which have all been genetically associated with LOAD (Lambert *et al*., 2013). However due to lack of upstream CGIs and primer efficiency it was not possible to generate data for all of the targeted gene.

In addition to *HLA-DRB5/HLA-DRB1*, *INPP5D, PTK2B, SORL1,* areas within the promoter regions of the genes apolipoprotein E (*APOE),* microtubule associated protein tau *(MAPT)* and mitochondrial transcription factor A *(TFAM)* were also investigated using the McrBC enzymatic assay. These genes, and specific target areas, were chosen due to previous studies showing differential methylation at these regions in LOAD (Wang *et al*., 2008, Barrachina and Ferrer, 2009, Iwata *et al*., 2014, Tohgi *et al*., 1999b). These published studies reported small but significant AD associated methylation changes across individual CpG sites within the promoter regions of these genes.

Wang *et al*. (2008) identified LOAD associated hypomethylation of a region within the *APOE* promoter and a hypermethylated region within the genes 3'UTR using LOAD brain samples. In the same study hypomethylation was also identified in the LOAD brain for a region targeted within the promoter of *TFAM*. Both genes also showed extensive inter-individual variation in LOAD. The promoter regions identified by Wang *et al*. (2008) were both targeted using McrBC treatment during the experiments described in this thesis chapter.

A region within the promoter of *MAPT* was also included in the McrBC experiment. The region chosen had also been previously shown to be hypomethylated in the LOAD brain using pyrosequencing (Iwata *et al*., 2014). This study identified a small but significant methylation difference at five CpG sites within the area targeted.

Previous studies mainly used brain tissue; part of the study aim was to investigate if changes observed in brain tissue could also be observed in LOAD blood. The investigation into methylation within the mentioned control genes would allow

elucidation of whether aberrant methylation reported by published studies in AD brain tissue is also observable in LOAD blood DNA.

3.2 Chapter aims:

- Use of a quantitative ELISA assay to investigate aggregate levels of genome wide methylation using LOAD blood samples
- Use of bisulphite cloning and sequencing to assess methylation of the *PTK2β* promoter in LOAD blood samples.
- Use the methylation specific restriction enzyme McrBC to assess methylation within previously investigated gene regions using LOAD leukocyte samples.

3.3 Methods

3.3.1 Alzheimer's disease DNA samples

All samples used for the work conducted in this chapter were collected from LOAD blood samples. DNA was extracted as described in section main methods. Table 3.1 below shows details of the samples used in this chapter.

Table 3.1: McrBC experiment sample details. Details of the samples used for the experiments described in this chapter are included.

DNA ID	Centre	Gender	Age at Death	Age at Onset	Age at Sampling	Diagnosis
N169	Nottingham	M	56			Control
N158	Nottingham	F	72			Control
N160	Nottingham	M	73			Control
N166	Nottingham	F	37			Control
M670	Manchester	F	87	83		Control
AD249	Nottingham	F	84			Confirmed AD
M543	Manchester	F	73	69		Confirmed AD
M604	Manchester	F	86	74		Confirmed AD
M646	Manchester	F				Confirmed AD
M673	Manchester	M	86			Confirmed AD
M767	Manchester	M		72		Probable AD

3.3.2 Global Methylation and Hydroxymethylation Assay

Details of the global method used to investigate both methylation and hydroxymethylation is described in detail in section 2.2. Briefly, In order to test if global levels of methylation significantly differ between LOAD and control leukocyte DNA, a quantitative ELISA assay was used to measure overall levels of methylation for each patient sample. Global methylation (figure 3.2) was investigated in LOAD and control samples (n=5 and n=5 for both LOAD and controls respectively) using the Methylamp ™ Global DNA Methylation and Hydroxymethylation Quantification kits. Either methylation or hydroxymethylation is recognised by an antibody which is then used to detect methylation or hydroxymethylation. The amount of either is proportional to the intensity of the OD measured.

Usually methylation standards are used to produce a standard curve following the completion of this assay. The standard curve allows for absolute quantification of amounts. In this instance a direct comparison was made in order to determine comparative differences between the LOAD and control samples and data interpretation was limited to this comparison. However, in future experiments it may be desirable to generate a standard curve in order to quantify exact methylation and hydroxymethylation for each group.

3.3.3 Bisulphite sequencing of PTK2β

3.3.3.1 Primer design

Bisulphite sequencing was used during this chapter in an attempt to investigate methylation with a region of *PTK2β*. Prior to this primers were designed to cover the region of *PTK2β* shown in figure 3.3.

3.3.3.2 Primers, target region and CpG numbers

Three sets of primers, amplifying the CpG rich region upstream of the transcription start site of the *PTK2β* gene were designed and tested for use in subsequence cloning. The primer design process is described in section 2.3.3.1. Three primer sets were designed; these are shown in table 3.2.

Table 3.2: Details of *PTK2β* Primer sequences used for bisulphite PCR and cloning. Showing primer sets used and the location of the regions targeted in the *PTK2β* gene. F –forward primer, R-reverse primer.

Primer set	Primer name
1	Ptk2b_f1- GGGGAGGAGAGYAGYAGGGGTGTGG
	Ptk2b_r1- TCCAAAAATARAACCCCTCTTCCCC
2	Ptk2b_f2- AGGGGTGTGGTTAAYAAYTYAGAGG
	Ptk2b_r2- CCCCTARCCTCCCCCARCCTCCCC
3	Ptk2b_f3- TAAYAAYTYAGAGGAGGAGGGAGAA
	Ptk2b_f3- CCCCCARCCTCCCCTRTCTRAATAT

The three sets of primers were then tested using a PCR reaction (details of this can be found in section 2.3.3.2) containing DNA from a LOAD patient and from a non-disease control. Unfortunately, primer sets 1 and 2 failed to produce any observable product. Primer set three did however produce a PCR product, therefore these primers where used to complete the PCR reactions prior to cloning. Initially three DNA samples were used: 166 (control), 646 (AD), AD249 (AD). The products from these PCRs were purified and cloned into One Shot® TOP10 Chemically Competent E. coli cells, more detail can be found in sections 2.3.3.3 and 2.3.3.4. Successfully transformed colonies were then selected and plasmid DNA was purified and sent for sequencing. The PCR products were also sent for sequencing to assess if the PCR was amplifying the correct genomic sequence.

3.3.3.3 Sequencing Results

Sequencing showed that the PCR product was either of a quality too poor to analyse or the product was highly degraded, which could have also caused the sequencing to fail. The PCR may have failed to amplify sufficient product due to inefficiency of the primers. The PCRs resulted in a faint bands when ran on a gel suggesting low yield of PCR product. Therefore, suggesting that in this instance, inefficiency of the primers was expected to be the reason for low product yield. Only a limited optimisation of the PCR protocol was attempted. This included a gradient and step-down PCR. However generating a strong PCR product proved difficult. Further efforts may have increase PCR yield and improve sequencing quality.

A second plausible explanation for the amplification failure is that the bisulphite sequencing failed or too little product was used in the PCR reaction. It seems unlikely that the bisulphite treatment was the cause due to the fact that the kit has been validated by previous and later experiments. However, the amount of bisulphite treated template DNA required could also be further optimised. The PCR may also have failed due to the PCR target being too large. Since bisulphite treatment

fragments the DNA a smaller PCR product is more ideal. This may be corrected by designing PCR primers which cover a smaller region.

The vector sequencing result showed that some of the vectors also contained no insert and had possibly just re-ligated. This may have been because the PCR product amount was too low for the ligation to be successful.

However further optimisation of this assay would have been time consuming and therefore, moving forward higher throughput methods were used.

3.3.4 McrBC Experiment Optimisation

3.3.4.1 McrBC Primer design and assay optimisation

McrBC was used to investigate the promoter regions of, *INPP5D, PTK2B, SORL1* and an intragenic region of *HLA-DRB5* and *HLA-DRB1.* An effort was made to design assays which targeted the genes *CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/HLA-DRB1*, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4*, *SORL1* and *ZCWPW1, which* have all been genetically associated with LOAD (Lambert *et al*., 2013). However due to lack of upstream CGIs and primer efficiency it was not possible to generate data for *CASS4*, *CELF1*, *FERMT2, NME8* and *ZCWPW1.*

Primers where designed to target CpG islands (CGIs) found upstream of the transcription start sites (TSS) of the genes *INPP5D, PTK2B, SORL1, SLC24A4, DSG2P* and an intragenic region of both *HLA-DRB5* and *HLA-DRB1* and regions reported to be differentially methylated in LOAD by previous studies (see section 2.4.1 table 2.4). For *the* genes *INPP5D, PTK2B, SORL1, SLC24A4, DSG2P* and an intragenic region of both *HLA-DRB5* and *HLA-DRB1* the CGIs were identified using the UCSC Genome browser and primers were then designed using Primer 3, further details can be found in section 2.4.1. Details of the regions covered for all genes are shown in table 3.1 as are the primers which were found to effectively produce a PCR product.

Table 3.3: Gene regions target by the McrBC enzymatic method. Showing the genomic regions targeted for each AD associated gene.

Gene	**CGI location in GRCh37**	**Target region size (bp)**	**No of CpGs covered**	**Primers worked**
HLA-DRB5/HLA-DRB1 Exon 2/9 and exon 3 respectively, CGI no 33)	Chr6*:* 32489738-32490132	214	18	Yes
INPP5D,	Chr2: 2:233,925,076-233,925,325	180	12	Yes
PTK2β	Chr8: 27168750-27169100	229	21	Yes
SORL1	Chr11: 121322900-121323350	268	31	Yes
SLC24A4	Chr14: 92789506-92790719	984	117	No
DSG2P	Chr18: 29,077,561-29,078,530	841	82	No
APOE	Chr19- 44905658- 44906146	488	17	Yes
TFAM	Chr10- 58384929- 58385033	104	5	Yes
MAPT	Chr17- 891797-892118	321	9	Yes

PCRs were conducted using each set of primers (detailed in appendix 2) using either McrBC treated DNA or non-treated DNA as a template. DNA from one LOAD patient and one ND (non-diseased control) were used initially. Only the primers for *SORL1, HLA-DRB1/5, INPP5D* and *PTK2β* produced a visible PCR product. PCRs were also successful for primers designed for *APOE, PSEN1* and *MAPT* (images not included).

Following successful optimisation of the primers for these seven genes, four LOAD genetic targets and three control genes, each set of primers were used in assays using five LOAD and five control (ND) DNA samples. Each assay containing PCRs using McrBC treated DNA templates and untreated template DNA. The PCR products were then ran on a 2 % agarose gel. As shown (figure 3.1) PCR products were ran on the gel in order of PCR templates: five non-treated LOAD templates, five non-treated non diseased (ND) control templates, five McrBC treated AD templates and five ND McrBC treated templates. Figure 3.1A shows the LOAD and ND samples details, for example the AD sample used in track number two was sample number 604, the sample DNA was used as template in the PCR whose product is shown in lane 12, however the DNA used in this PCR was McrBC treated prior to use as a PCR template. Similarly the same DNA sample, though McrBC treated and non-treated, was used as template in the PCRs whose product is shown in tracks 3 and 13.

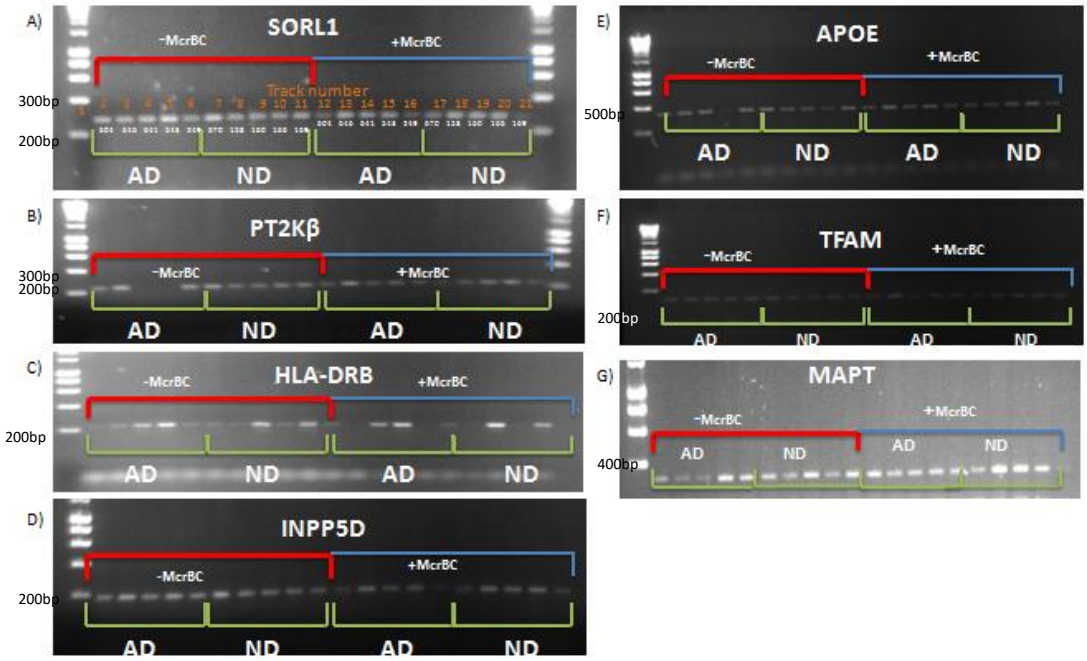


Figure 3.1: PCR results for PCRs using five AD and five ND McrBC treated and untreated DNA templates. A) *SORL1*. B) *PTK2B*. C) *HLA-DRB*. D) *INPP5D*. E) *APOE*. F) *TFAM* and G) *MAPT*. AD (LOAD patient), ND (non-diseased control).

It was very difficult to observe any noticeable reduction in PCR product between PCRs using McrBC non-treated DNA template and those using McrBC treated template. Therefore to move forward GeneSync software was used to attain band intensity for each PCR product. This value was presented as a figure for the raw volume of product of each PCR. The raw volume could then be used to estimate the amount of DNA produced by the PCR by using the raw volume value given for the nearest ladder band. Since each band of the ladder represents a known amount of DNA the raw volume given on a particular gel would represent the raw volume value corresponding to the amount of DNA in the ladder band. The amount of DNA produced for each PCR was calculated using this method, this allowed for a quantitative comparison of McrBC treated and untreated PCR results.

3.3.4.2 **GABRB3** ***Control Primers***

Primers were also designed and optimised for a CpG less region of the *GABRB3* promoter (further details can be found in section 2.4.1 (main methods)). This region was amplified using both McrBC treated and non-treated template, as expected no

difference in PCR product was found between the two assays, showing that digestion would only occur in the presence of CpG sites.

In future assays it would be desirable to develop both a viable positive and negative control for McrBC enzymatic activity to be included as part of the overall assay design. For example the enzyme Msp1 could be used as this enzyme would digest unmethylated but not methylated DNA. A comparison can then be made between PCRs using template digested by the two enzymes. In this instance, the assay was designed to identify regions of interested as part of a high throughput initial screen e.g. for further investigation rather than stand alone data. A direct comparison between McrBC digested templates was performed and described in this chapter.

3.3.5 Statistics

Statistics performed in this chapter are described in more detail in section 2.4.4. Briefly, following the exclusion of outliers from the data, two tailed T-test were used to compare data sets.

3.3 Results

3.3.1 Whole genome aggregate methylation in LOAD patient DNA

Changes in global methylation were investigated first, prior to testing promoter specific methylation in LOAD associated genes. This was necessary in order to exclude the possibility that any gene specific methylation identified is merely the result of global hyper or hypomethylation associated with disease. Global methylation and hydroxymethylation was measured using the method described in section 3.3.2.

No significant difference in global methylation or hydroxymethylation (figure 3.2) could be detected in LOAD when compared to controls. This data suggested that any regional difference in methylation or hydroxymethylation for the genes tested using targeted methods is unlikely to be the result of any LOAD specific process driving either global hyper or hypomethylation.

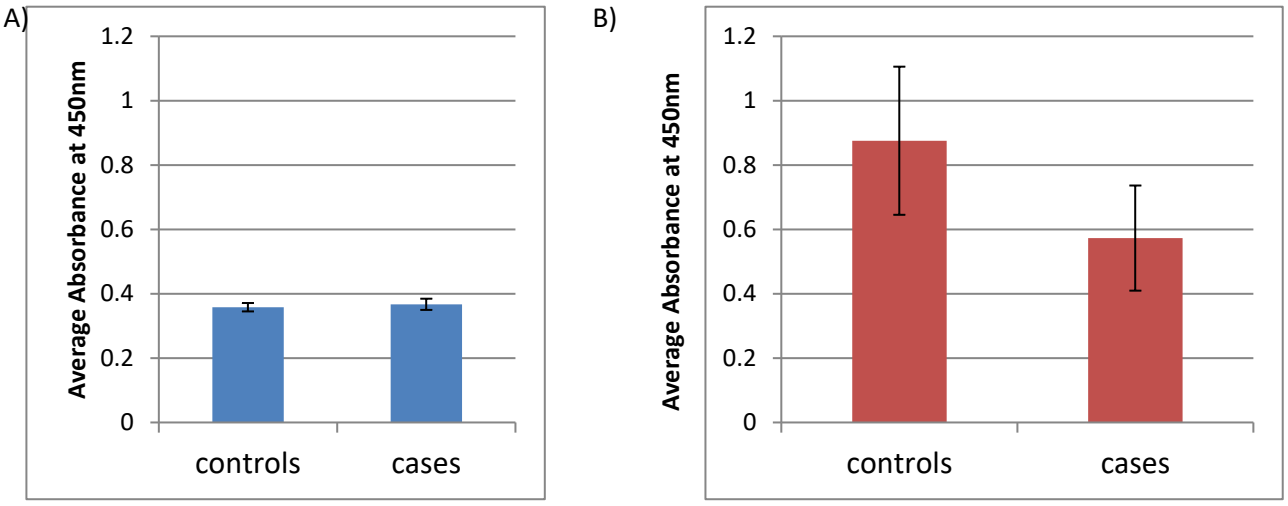


Figure 3.2: No significant difference in global DNA methylation or hydroxymethylation was observed between the AD samples tested and the control samples (for both n=5). A) Shows average absorbance for the AD and control samples tested using methylation assay (n=5). B) Shows average absorbance for the AD and control samples tested using hydroxymethylation assay (n=4).No significant difference was observed for either. Error bars shows standard error of the mean. P value 0.69 for methylation and 0.33 respectively.

3.3.2 Bisulphite sequencing of *protein tyrosine kinase 2 beta (PTK2B)*

Methylation of a region found within the promoter region of the AD associated gene *PTK2β* was investigated using the bisulphite sequencing method described in section (3.3.3).

This region was chosen due to it being located in an interesting position within the *PTK2β* gene, as shown in figure 3.3. The assay was designed to cover a sequence within a CGI which spans a potential splice site within the *PTK2β* gene. Since CGI DNA methylation has been shown to have an effect on the splicing of AD associated genes the region was identified as a potentially interesting target for investigation (Humphries *et al*., 2015).

The CGI targeted was also located on a region of gene which contained the histone mark H3K27Ac (figure 3.3). This histone mark is associated with active transcription and might suggest the presence of an active enhancer; these also made the region an interesting site for investigate in the context of LOAD.

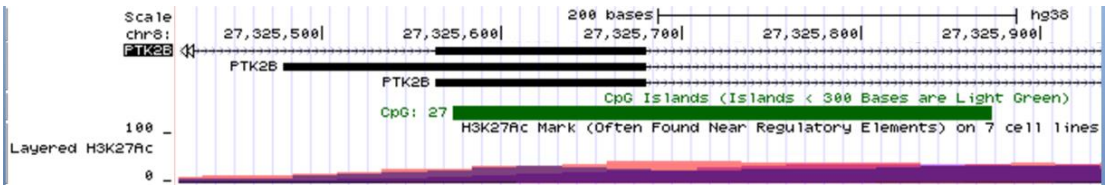


Figure 3.3: PTK2β region targeted by bisulphite cloning. Image taken from the UCSC genome browser shows the targeted CGI located within *PTK2B*. The CGI is also located within a region identified as containing the histone mark H3K27Ac.

Unfortunately, sequencing analysis did reveal the amplification of the correct region. Therefore cloning proved to be unsuccessful and it was not possible to analyses the targeted region.

It was clear that in order to successfully complete bisulphite sequencing of the *PTK2β* gene extensive optimisation of both PCR and cloning would be required. However,

since this method was time consuming and expensive it was decided that the project would continue using high-throughput enzymatic methods for detecting changes in DNA methylation in genes associated with LOAD. While methods using restriction enzymes are less quantitative than bisulphite sequencing, it is a much faster method, which allows for the investigation of many genes simultaneously.

3.3.3 Using McrBC to identify differential methylation in LOAD associated genes

Following the unsuccessful efforts of bisulphite cloning further work was conducted using the McrBC enzymatic method to elucidate any differential methylation within LOAD blood samples. Due to the high throughput potential of this method regions located within upstream CGI's where also indentified within the AD associated genes *INPP5D, SORL1, SLC24A4*, *DSG2P* and an intragenic region of both *HLA-DRB5* and *HLA-DRB1.* Details of assay design and the regions covered can be found in section 3.2.4. Despite optimisation attempts, the assays for *SLC24A4* and *DSG2* failed (table 3.3). Therefore results are presented for the other four genes only. Regions within the genes *APOE, PSEN1* and *MAPT* were also successfully investigated.

3.3.3.2 GABRB3 (gamma-aminobutyrica acid type A receptor beta 3 subunit) Control

A region of the gene *GABRB3* was also chosen as a control due to the absence of CpG sites where methylation may accrue. This region has previously been used in published research as a methylation control (Maunakea *et al*., 2010). Primer sequences used were as described by Maunakea *et al*., (2010) and were taken from the paper and optimised. Amplification generated a PCR product of size 206bp. An additional set of primers was designed, covering the same region, which produced a PCR product of 406bp in size.

Five LOAD and five control DNA samples (McrBC treated and untreated) were used in PCRs containing *GABRB3* primers and the results were visualised using gel electrophoresis (shown in figure 3.4A). The amount of DNA produced in each PCR was

calculated using the previously described method and an average was calculated for each group (LOAD non-treated, LOAD treated, ND treated and ND non-treated), figure 3.4B shows that no significant reduction in PCR product was observed for the PCRs which were completed using McrBC treated DNA for either AD samples or controls, P values calculated using two tail student T-Tests, AD T vs AD NT P= 0.51, ND T vs ND NT P=0.27, all T vs all NT P=0.23 . Therefore it was concluded, as expected, that the McrBC exhibited no endonuclease activity in this unmethylated target region.

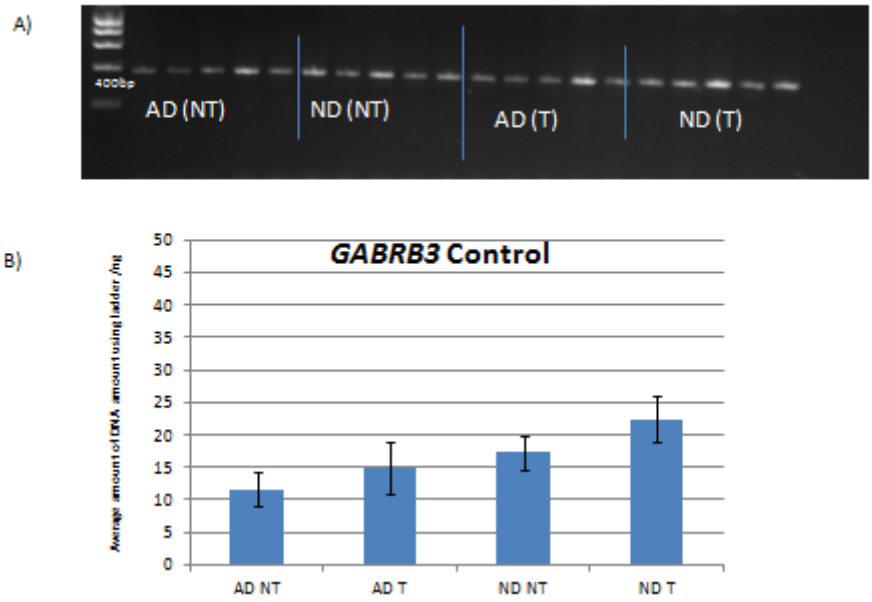


Figure 3.4: GABRB3 CpG less control showed no reduction in PCR product when McrBC treated DNA was used as a PCR product. A) Results of gel using LOAD and ND McrBC treated (T) and non- treated (NT) DNA template. B) Graph showing average amount of PCR product for each group. Error bars show S.E.M. n=5 for LOAD and control samples.

3.3.3.4 No significant LOAD associated methylation within APOE *(apolipoprotein E),* TFAM *(mitochondrial transcription factor A) and* MAPT *(microtubule associated protein tau) was identified*

No significant difference in PCR product was observed between either LOAD non-treated and treated DNA or control DNA (figure 3.5). For *APOE* LOAD treated vs not treated two tail student T-test P= 0.106, ND treated vs non treated P=0.08. For *TFAM* LOAD treated vs not treated P= 0.16, ND treated vs non treated P=0.9. For *MAPT* LOAD treated vs not treated P=0.48, ND treated vs non treated P=0.2.

There was also no significant difference observed between PCRs amplifying McrBC treated template and non-treated template for either genes *TFAM* or *MAPT* suggesting that sites targeted are not methylated in the instance of, *TFAM* P= 0.2 and *MAPT* P=0.45. However, there was a significant reduction in PCR product observed for the *APOE* region investigated P= 0.005 (figure 3.5).

These results suggested that in the areas targeted in this study, and for those individuals tested, there was no significant difference in methylation in LOAD samples when compared to non-diseased controls (*TFAM* or *MAPT)*. However a region of the *APOE* gene is identifiably methylated in both the AD and control samples.

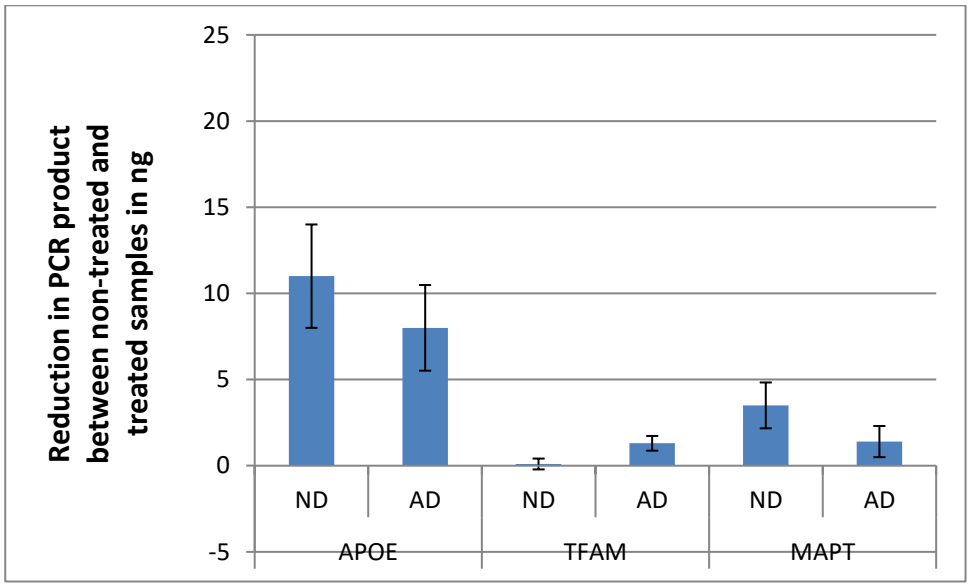


Figure 3.5: No significant difference between LOAD and control in the samples tested for *APOE, TFAM* or *MAPT*. The graph shows the difference in PCR product amount calculated between PCRs using non-treated and McrBC treated template for both the control and LOAD samples. No significant difference was found between the two groups. Bars represent S.E.M. For *APOE* ND n=3, AD n=4. For *TFAM* NT n=5, AD n=5. For *MAPT* ND= 5, AD = 4. **p=<0.01. Two tailed student T-test.

3.3.3.5 GWAS Target Gene Results- Primer efficiency

Originally the study aimed to investigate methylation within the promoter regions of the GWAS identified genes *CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/HLA-*

DRB1, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4*, *SORL1* and *ZCWPW1* . However, due to a lack of CGIs and primer inefficiency it was only possible to gather data for regions investigated in the promoter regions of *HLA-DRB, SORL1, INPP5D* and *PTK2β*. As in the instance of *TFAM, APOE* and *MAPT* it was possible to successfully amplify McrBC treated and non-treated DNA templates. For the *SORL1* and *INPP5D* assays the PCRs products were also clear. Cumulatively it was therefore possible to collect data for five LOAD and five ND samples and compare differences between PCR products from PCRs using McrBC treated and non-treated template. The PCR reaction amplifying regions of *PTK2β* were also reasonably efficient and it was possible to collect data for three LOAD and five ND control samples. However, unfortunately PCRs amplifying regions of *HLA-DRB* were not efficient and it proved difficult to generate enough PCR product to analyse. Therefore for this region, data for only two LOAD and two ND control samples are quantified.

3.3.3.6 PTK2β *(protein tyrosine kinase 2 beta) and* HLA-DRB1/5 *(HLA class II histocompatibility antigen) are not methylated at the regions investigated*

PTK2β and *HLA-DRB1/5* showed no significant difference in amplified product when using McrBC treated and non-treated template DNA. The average amount of PCR product observed is for each is shown in figure 3.6. For *HLA-DRB1/5* the average amount of product observed for treated and non-treated PCRs was 9.43 and 9.42ng of DNA, no significant difference between treated and non-treated PCR product was found, P=0.39. In addition for *HLA-DRB1/5* no significant difference was observed in PCR product when LOAD treated and non-treated products where compared, or when non-diseased treated and non-treated products where compared. LOAD treated vs non-treated P= 0.59, for ND treated vs non-treated P=0.16.

Data indicated a similar pattern for *PTK2β*, figure 3.6. A reduction in product was observed for PCRs completed using McrBC treated template. However this difference was not proven to be significant using a two-tailed student T-test, P=0.059. This might suggested that repetition with a greater number of samples is needed. However for *PTK2β* when the products of PCRs using LOAD treated and non-treated template and

non-diseased treated and non-treated template were compared no significant difference in PCR product was observed for LOAD and control samples, P= 0.26 and P=0.16 respectively.

3.3.3.7 SORL1 *(sortilin related receptor 1) and* INPP5D *(inositol polyphosphate-5-phosphatase D) results indicate methylation in the region investigated in LOAD samples*

In the instance of *SORL1* a reduction in PCR product was observed upon comparison of McrBC treated and non-treated samples. This reduction was also shown to be significant using a two tailed student T-test P= 0.017 (Figure 3.6). This result suggested that the region investigated was possibly methylated within the samples tested.

In order to ascertain if the methylation was LOAD specific the products of PCRs using LOAD non-treated and McrBC treated template where compared. A significant reduction of 41.9% was observed, P= 0.029. Interestingly no significant difference between products of PCRs using non-treated and McrBC treated template was observed for non-diseased control samples, P= 0.3. This could suggest that the methylation occurring at the region investigated was LOAD specific, suggesting possible hypermethylation of this region in the LOAD samples investigated.

A region of *INPP5D* also showed a significant reduction of PCR product of 50% for the PCR using McrBC treated DNA template (P= 0.0009). Suggesting methylation was also occurring in the investigated region.

INPP5D results also indicated that the methylation occurring might be LOAD specific. The amount of PCR product for PCRs using LOAD non-treated and McrBC treated template were compared. A reduction of 64% was observed for the PCRs using McrBC treated template, P= 0.0029. However when the PCR product of non-treated and McrBC treated PCRs for non-disease samples were compared no significant difference in PCR product was observed, P= 0.16.

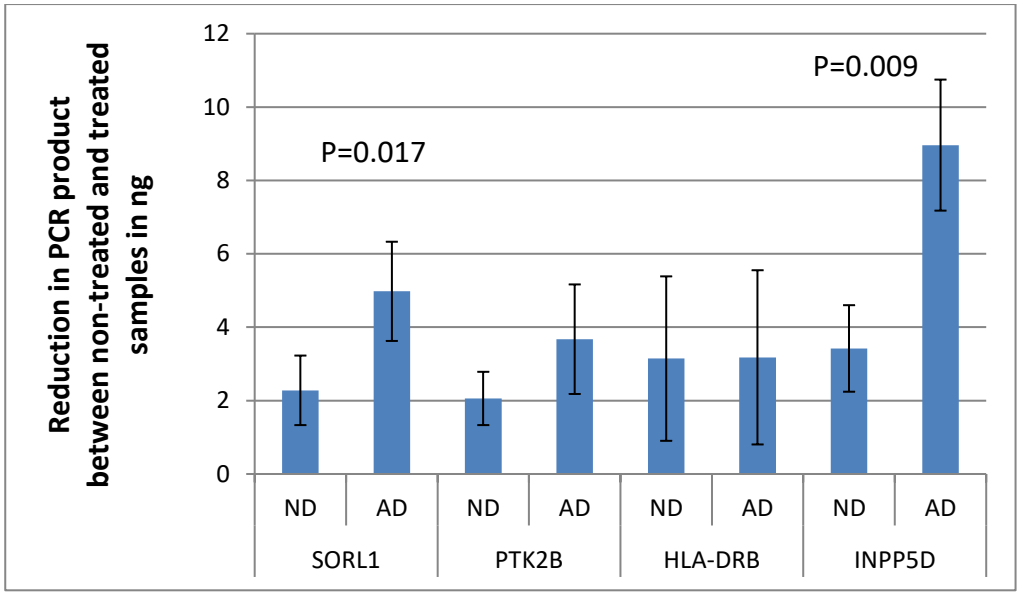


Figure 3.6: Graph showing candidate gene results: Graph shows the difference in PCR product amount calculated between PCRs using non-treated and McrBC treated template for both the control and LOAD samples. ND= non- diseased control, AD= LOAD sample, T= McrBC treated, NT = McrBC non-treated. P values calculated using a two tail student T-test. Error bars represent S.E.M. For *SORL1* and *INPP5D* n=5 for all groups. For *HLA-DRB* ND NT, ND T, AD T and AD NT n=2. For *PTK2β* ND NT, ND T n=5, AD T and AD NT n=3. *SORL1* and *INPP5D* NT n=10 and NT n=10, for *HLA-DRB* NT n=4 and T =4 and for *PTK2β* NT n=8 and T n=8.

3.4 Discussion

3.4.1 Global aggregate levels of methylation

Initially global methylation was investigated in LOAD and control samples using a quantitative ELISA method. No significant difference in global methylation was observed between LOAD and control samples. One thing to consider is that a low number of samples were tested and therefore there may be a power issue with these results. However this would more likely be a problem affecting and a difference observed between AD and controls. In future studies it may be desirable to test more samples. However, it is possible that only a few genes are differentially methylated due to AD and since the assay measures millions of CpG sites across introns, exons and promoter regions, and changes specific to the condition are likely to be masked by minor fluctuations in the epigenome.

Multiple studies using various methods have been used to investigate global methylation in LOAD. Studies have produced evidence of both global hyper and hypo methylation in LOAD brain samples (Bradley-Whitman and Lovell, 2013, Coppieters *et al*., 2014, Mastroeni *et al*., 2010, Chouliaras *et al.*, 2013, Condliffe *et al*., 2014, Bakulski *et al*., 2012). In contrast another recent study found no difference in global DNA methylation using LOAD brain samples (Lashley *et al*., 2015). In addition, other published research using LOAD blood samples did identify global hyper-methylation in LOAD peripheral blood mono-nuclear cells (Di Francesco *et al*., 2015).

These discrepancies in results make it extremely difficult to conclude the existence of consistent global methylation changes associated within LOAD. A likely reason for contradictory results is the different methods used and relatively small sample sizes assessed (Condliffe *et al*., 2014, Ellison *et al*., 2017). However another influencing factor is that often different brain regions are used and each presents with its own epigenetic profile. Cell type specific methylation has also been demonstrated in LOAD (Coppieters *et al*., 2014, Ellison *et al*., 2017).

Many studies use a range of methods to assess genome wide methylation differences, however techniques such as bisulphite sequencing do not allow distinction between methylation and hydroxymethylation (Ellison *et al*., 2017). Therefore both methylation and hydroxymethylation were investigated in this study to get a clear picture of global epigenetic changes occurring in LOAD.

No significant difference in the global presence of either epigenetic mark could be observed in the LOAD samples tested. Therefore these results suggested that no changes in global methylation or hydroxymethylation changes were occurring due to AD in the samples used in these assays.

One possible explanation for unaltered global methylation in LOAD might be the stage of disease the patient was in when the sample was taken. A recent study has shown that global methylation and hydroxymethylation changes might occur early on in the disease progression, possibly driving its progression, these changes then seemed to reverse in later stages of the disease (Ellison *et al*., 2017).

3.4.2 *MAPT* (microtubule associated protein tau), *TFAM* (mitochondrial transcription factor A) and *APOE* (apolipoprotein E)

Targets within the promoter regions of *MAPT, TFAM* and *APOE* targeted using the McrBC assay showed no significant difference in PCR product between LOAD samples and controls. This suggested that no differential methylation was occurring at the targets investigated, despite previous reports of hypomethylation within these regions (Iwata *et al*., 2014, Wang *et al*., 2008). In addition, no significant difference in methylation was observed between treated and untreated LOAD and control samples for either *MAPT* and *TFAM*. This data suggests no methylation at either region in either LOAD or control samples. However for *APOE* there was significant reduction in PCR product between McrBC treated and non-treated PCRs suggesting possible methylation at the target investigated in both control and LOAD samples, in leukocyte DNA.

It is plausible that the results generated using McrBC described in this chapter failed to reproduce those observed by Wang *et al*. (2008) and Iwata *et al*. (2014) because of the lack of experimental resolution provided by the McrBC technique. The McrBC enzymatic method does not account for variable methylation differences that might occur at individual CpGs. Both Wang *et al*. (2008) and Iwata *et al*. (2014) used methods allowing identification of individual CpG methylation difference. For *APOE* and *TFAM* Wang *et al*. (2008) identified specific CpGs as being differentially methylated (*TFAM*: CpGs 1, #6 and #14; *APOE* CpG #1 and #2). In addition Iwata *et al*. (2014) used pyrosequencing of an area containing 43 CpG sites within the *MAPT* promoter but only found significant LOAD associated hypomethylation at five of these sites. Therefore it seems likely that McrBC methods might result in omission of important individual CpG differential methylation.

Another possible reason for not reproducing the *MAPT, APOE* and *TFAM* results is that both of the previous studies identified very small, but significant, differences in methylation, the difference observed for *APOE* and *TFAM* was around 10% and for

MAPT this was even smaller at 2-4% (Iwata *et al*., 2014, Wang *et al*., 2008). Methylation differences this small might not be detectable using the McrBC enzymatic method, suggesting the need for use of higher resolution methods such as pyrosequencing.

Another explanation for the discrepancy in results would be a lack of DNA digestion by the McrBC restriction enzyme. However this seems unlikely as differential methylation was observable in *SORL1* and *INPP5D* were identified using this method.

Very small changes in methylation have been shown to be important in driving psychological diseases. One way to move forward would have been to use pyrosequencing to look at the regions targeted using the McrBC assay. This would have been capable of producing results with much higher resolution. Unfortunately it was not possible to do this due to the limitations in the amounts of LOAD blood sample DNA available.

3.4.3 Methylation in *PTK2β* (PTK2β (protein tyrosine kinase 2 beta) and *HLA-DRB1/5* (HLA class II histocompatibility antigen)

No significant difference in methylation was observed in the *PTK2β* promoter region investigated or the region investigated for *HLA-DRB1/5*.

Another study was published investigating the methylation status of AD associated genes *HLA-DRB5, PTK2B, SORL1, SLC24A4, DSG2,INPP5D, MEF2C, NME8, ZCWPW1, CELF1, FERMT2*, and *CASS* in the LOAD brain (Yu *et al*., 2014). This study included the four genes investigated using McrBC described in this chapter. Yu *et al*. (2014) failed to find any LOAD specific differential methylation within the region investigated in *PTK2β*, perhaps consistent with the findings from the McrBC experiment described in this chapter. Yu *et al*., (2014) used an array based method for their study, providing coverage of regions across the whole *PTK2β* gene. This study sought to investigate promoter specific methylation of this gene i.e. methylation correlating with likely regulation of transcription. A similar result between the two assays would therefore

not necessarily be expected. Another study did successfully identify LOAD specific methylation within the *PTK2β* locus, although a different region was investigated (Humphries *et al*., 2015).

Yu *et al*. (2014) identified LOAD specific differential methylation in brain samples within a region of *HLA-DRB5*. Forty eight CpGs were included in the study of which only eight showed differential methylation which correlated with AD pathogenesis. It is noteworthy that this significance also failed to survive multiple testing. No differential methylation was found in the leukocyte samples used in the region of *HLA-DRB* investigated using McrBC, perhaps due to the low resolution of the McrBC method or the use of a different target region. Further testing using higher resolution methods would be needed to confirm the methylation status of the *HLA-DRB1/5* region investigated using McrBC

3.4.4 *SORL1* (sortilin related receptor 1) and *INPP5D* (inositol polyphosphate-5-phosphatase D) methylation

The study conducted by Yu *et al*. (2014) also identified methylation within *SORL1* that correlated with AD pathology in the AD brain. In this study sixty nine CpG sites were investigated and again only eight showed LOAD associated methylation. In addition to this the most significantly associated CpG was located within the gene body not the promoter. In the McrBC study described in this thesis chapter, significant difference in methylation in *SORL1* was identified in the LOAD samples tested, possibly suggesting hypomethylation within the region of the *SORL1* promoter investigated. While a different region was investigated by Yu *et al*. (2014), the two studies indicated possible LOAD associated methylation occurring within both blood and brain tissue. It would have been interesting to assess the methylation status of the *SORL1* promoter in LOAD brain as well as blood.

However, in contrast to the McrBC data and Yu *et al*. (2014) other studies have failed to find a significant difference in methylation within the *SORL1* promoter when using LOAD leukocyte DNA, or any significant differential methylation within a region located in the *SORL1* gene body (Humphries *et al*., 2015, Furuya *et al*., 2012). However

hypermethylation of the *SORL1* 3'UTR in the LOAD brain, and decrease SORL1 expression in LOAD has been reported (Humphries *et al*., 2015, Scherzer *et al*., 2004).

It seems likely that *SORL1* is important in LOAD and that DNA methylation within the *SORL1* gene could be important in driving disease pathology. However, studies have yielded conflicting results as to whether differential DNA methylation occurs at the *SORL1* locus (Furuya *et al*., 2012, Yu *et al*., 2014, Humphries *et al*., 2015). The McrBC results described in this chapter suggest that differential methylation might exist in LOAD leukocyte DNA. One hypothesis is that significant hypermethylation of the *SORL1* gene may result in decreased SORL1 expression which in turn could drive AD pathology due to the function of SORL1 in trafficking APP. A reduction in SORL1 could lead to increased APP processing resulting in β-secretase cleavage and accumulation of intracellular Aβ which ultimately drives AD pathogenesis (Rogaeva *et al*., 2007). Higher resolution experiments would be needed to identify CpG site specific methylation within the *SORL1* promoter.

Interestingly Yu *et al*. (2014) failed to find any significant differential methylation in a region investigated within *INPP5D,* whereas the McrBC study described in this chapter suggested possible hypermethylation within the region investigated in *INPP5D* in the samples tested. It seems likely that INPP5D has an important role in driving AD pathology through its involvement in the immune response i.e microglia activation; and also its role in APP metabolism which is mediated through interaction with the AD associated gene CD2AP (Hollingworth *et al*., 2011). However further investigation would be required to reveal disease linked DNA methylation occurring within the regions investigated.

3.4.5 Experimental limitations

Ideally the use of more samples would have improved the results presented in this chapter. Unfortunately at the time of completing the work no further LOAD blood samples were available for testing. It would also have been interesting to compare blood data to brain data for the regions investigated. However, again unfortunately brain samples were not available for use during this study.

As mentioned above studies have produced conflicting results regarding methylation at specific gene targets in LOAD using both brain and blood tissue. This was apparent for the control genes *APOE, TFAM* and *MAPT* as well as the GWAS target genes. The difference in results found is likely caused by multiple reasons. These could possibly include:

3.4.5.1 Experimental resolution

Previous studied were conducted using techniques with higher resolution then the McrBC enzymatic method and allowed the identification of differential methylation within individual CpGs. For example Wang *et al*. (2008) identified significant differential methylation at CpG sites one and two of the *APOE* target and one, six and four of the *TFAM* target. In addition to this, of the 69 CpGs targeted within *SORL1* and 48 in *HLA-DRB5,* only eight sites for each target were found to harbour differential methylation in LOAD by Yu *et al*. (2014). Therefore because McrBC does not have this level of resolution, differential methylation of individual CpG sites within the gene targets may have been missed in this experiment. However, it is important to note that the goal of the study described in this chapter was to investigate a number of gene targets, the enzymatic method was used as it allows quick assessment of methylation at targeted regions. Any differential methylation identified could then be further investigated using methods with a higher resolution, such as pyrosequencing.

3.4.5.2 Regions investigated

Another possibility for the difference in results is that different regions of the genes are investigated in each study. For the McrBC experiment targets were chosen within the promoter regions of the genes used. However Humphries *et al*. (2014) found differential methylation within the *SORL1* 3'UTR and Yu *et al*. (2014) within the gene body, indicating the methylation within the promoter region might not be the only important methylation occurring in LOAD.

3.4.5.3 Sample difference

Another important factor to consider is that in each study is conducted using patient samples from different regions, ethnicities ect. Since epigenetic changes accumulate due to a person's lifestyle and their environment, different DNA methylation profiles may exist within the different experimental sample groups.

Another important factor to note is that blood samples are heterogeneous in nature. These experiments were conducted in order to ascertain a general overview for epigenetic profiles linked to disease observed in blood (leukocytes). It should be acknowledged that there are a number of subtypes of cell which comprise the immune system and can be detected in blood. It would be interesting in future research to profile each cell type distinct from other leukocytes. I would hypothesis that in relation to disease the 'driver' or stimulus that leads to aberrant methylation would impact on each cell leading to a similar or near identical change in methylation. However, it is known that each cell type possesses distinct profiles of expression (see Palmer *et al,* 2006) and therefore future experiments would need to be conducted to conclusively prove this.

3.4.5.4 Method limitation

One clear limitation of using the McrBC enzymatic method is the method is not quantitative. Quantification was attempted using the ladder band density and end-point PCR product template. However, ideally the method used should give an accurate prediction of percentage methylation observed. Ideally also a positive control would have been used as part of this assay, this could have included using DNA of known methylation percentage to assess the effectiveness of the McrBC enzyme. Subsequent experiments were conducted using pyrosequencing, a platform which allows for the quantification of methylation.

In addition to this ideally more samples would be needed to draw any confident conclusion from the data. It was clear that variation existed between individuals when PCR product was compared for treated and non-treated PCRs. It is likely that a much larger sample size would be needed to correct for this.

3.5 Conclusion

The experiments described in this chapter provide data describing LOAD associated increase in methylation levels for the genes *INPP5D* and *SORL1*. However the method used provides insufficient resolution to fully establish differential overall methylation within each region investigated. Therefore future experiments used the pyrosequencing platform, which allowed for high resolution information about each individual CpG site tested.

Chapter 4: Pyrosequencing to investigate DNA methylation in key AD risk genes in LOAD and sEOAD

4.1 Introduction

4.1.1 Pyrosequencing Assay Targets

This chapter describes investigation into the methylation status of the gene's *MEF2C, ABCA7, PTK2β, SIRT1, INPP5D*, *TREM2* and *RIN3* using sEOAD blood (leukocyte) and brain (cortex) samples. The results presented in this chapter is also described in Boden *et al.,* (2017), with the exception of TREM2 data. *MEF2C, ABCA7, PTK2β, SIRT1, INPP5D* and *TREM2* were chosen due to their genetic association with LOAD (Lambert *et al*., 2013b). For these gene's regions containing a high density of CpG sites (CpG islands) located upstream of the transcription start site (TSS) where investigated using pyrosequencing. An additional CpG site, upstream of the investigated CGI, was also investigated in the *MEF2C* gene. A region including seven CpG sites located within a 3' UTR CGI of the gene *RIN3* was also targeted. *RIN3* has been functionally implicated in LOAD pathology though its interaction with the AD associated gene BIN1 (Kajiho *et al*., 2003). This site was also chosen following personal communication with Professor Kevin Morgan and Dr Keeley Brookes.

As mentioned previously, a growing number of genes and biological pathways have been implicated in the development of LOAD indicating the extensive complexity of this disease (Karch and Goate, 2015). Further, the genetic association alone cannot fully explain heritability and adds to the complexity of the disease (Lambert *et al*., 2013b). It is therefore plausible that aberrant epigenetic regulation also has an important role to play in driving LOAD pathology. Specifically altered DNA methylation has been identified in multiple AD associated genes using both blood and

brain tissue (Wang *et al*., 2008, De Jager *et al*., 2014, Yu *et al*., 2014b, Chouliaras *et al*., 2013, Coppieters *et al*., 2014, Hou *et al.*, 2013, West *et al*., 1995)**.**

The work described in this chapter aimed to identify if LOAD associated genes are differentially methylated in sEOAD. The work investigated methylation of selected candidate genes in both blood and brain tissue, allowing the identification of tissue specific differences. Initially assays were designed for each gene which included the regions investigated previously using the McrBC enzymatic treatment (described in chapter 3); specifically the experiment included the promoter regions of *INPP5D, PTK2β, SORL1, SLC24A4, DSG2*, and an intragenic region of *HLADRB1/5*. However the assays for *SLC24A4, DSG2*, and *HLADRB1/5* proved unsuccessful and efforts were concentrated on the other targets. In addition to this, the promoter region of the GWAS identified gene's *ABCA7* and *MEF2C* were also included. A region of the gene *TREM2*, which was identified using exome sequencing prior to GWAS, was also included in the pyrosequencing assays (Guerriro *et al,.* 2013, Johnson *et al.,*2014). Regions of *SIRT1* and *RIN3* were also included in the investigations. SIRT1 and RIN3 have been functionally implicated in LOAD pathology (Kajiho *et al*., 2003, Tan *et al*., 2013, Furuya *et al*., 2012, Gao *et al*., 2010, Michan *et al*., 2010, Herskovits and Guarente, 2014, Wang *et al*., 2010, Min *et al*., 2010, Julien *et al*., 2009, Hou *et al*., 2013).

In most instances primers were designed to target the CGI nearest to the transcription start site of the gene. However for *TREM2* the amplified region was engineered to cover a CpG site, located 289bp upstream of the transcription start site. This chosen CpG target site was shown to be hypermethylated in the LOAD brain by another study (Smith *et al*., 2016). An additional site upstream of the investigated CGI was also investigated in the case of the *MEF2C* gene and a pyrosequencing target was chosen which covered seven CpG sites located within a 3' UTR CGI of the gene *RIN3***.**

Pyrosequencing assays were designed for the gene targets described above, using the method described in materials and method section 2.5. However some assays proved difficult to optimise at either the PCR or pyrosequencing stage, optimisation steps

included the use of several primer sets, shown table 4.2. Therefore only the results of the successful assays are described.

4.1.2 Implication of Targeted genes in AD pathology

Both *PTK2β* and *INPP5D* were investigated using the MrcBC enzymatic method (chapter 3). However for this study LOAD blood samples were used. Therefore these targets were also investigated in sEOAD samples using pyrosequencing. LOAD and sEOAD share similar pathology with genes that function in similar biological pathways which have been implicated in both types of AD. These pathways include: inflammation, calcium signalling, long term potentiation and the mitogen-activated protein kinase (MAPK) signalling pathway (Antonell *et al*., 2013). Therefore methylation profiling of *INPP5D* and *PTK2β* region in sEOAD samples was conducted to provide insight into whether LOAD epigenetic aberration was also present in sEOAD. PTK2B and INPP5D are both capable of driving AD pathology through their roles in memory formation (Antonell *et al*., 2013).

In addition to *PTK2β* and *INPP5D* a region within the *ABCA7* promoter containing eight CpG sites was also targeted. *ABCA7* (encoding ATP-binding cassette member 7) has been implicated in LOAD pathology in both genetic and epigenetic studies (Lambert *et al.*, 2013b, Hollingworth *et al.*, 2011, Humphries *et al*., 2015). Several SNPs have been identified near *ABCA7* using GWAS, leading to the identification of LOAD risk alleles, among these are rs3764650 and rs4147929 (Hollingworth *et al*., 2011). Of the twenty known loci associated with LOAD, including the new 11 identified by Lambert *et al*. (2013), *ABCA7* was found to carry the second largest risk for development of AD, following the *APOE* gene (Lambert *et al.*, 2013b).

ABCA7 may drive AD pathology through its roles in phagocytosis and therefore represents part of the host-defence system (Humphries *et al*., 2015, Jehle *et al.*, 2006). ABCA7 is also involved in modulation of amyloid processing and Aβ deposition and clearance (Wang *et al.*, 2003, Vasquez *et al*., 2013, Kim *et al*., 2013, Shulman *et al.*, 2013).

In addition to *INPP5D, PTK2β* and *ABCA7,* methylation was also investigated in the gene *MEF2C* which was also identified in the Lambert *et al.,* (2011) study. However, for this gene in addition to the CpG rich region within its promoter, a second CpG site upstream of this was also targeted.

MEF2C function also has the capability to drive AD pathology through its role in memory formation and learning (leifer *et al*., 1994, Flavell *et al.*, 2006, Cole *et al*., 2012, Tanila, 2017). MEF2 transcription factors have been shown to regulate the expression of genes involved in the innate immune response and microglia proliferation (Zhang *et al.,* 2015, Clark *et al.*, 2013, Johnson *et al.*, 2014, Mhatre *et al.*, 2015). Therefore in the AD brain it may be important for mediating immune response to AD pathology, such as Aβ plaques and tau tangles resulting in the neuro-degeneration observed in AD (Zhang *et al.,* 2015, Mhatre *et al.,* 2015). Evidence also suggests MEF2 signalling is important for APP processing and function (Burton *et al*., 2002, Camargo *et al.,* 2015) and also interestingly circadian rhythms which are often disrupted in AD (Sivachenko *et al*., 2013). Furthermore, aberrant expression of the gene is linked to several neurological disorders including fragile X syndrome, epilepsy, autism and Angelman syndrome (Rashid *et al*., 2014)

In addition to the previously described targets*,* pyrosequencing was also used to investigate regions of *TREM2, SIRT1* and *RIN3*. In the case of TREM2 one CpG site previously found to be hypermethylated in the LOAD brain by another study was targeted. This CpG site is located 289bp upstream of the *TREM2* TSS (Smith *et al*., 2016). *TREM2* encodes triggering receptor expressed on myeloid cells 2 has also has also been genetically implicated in the development of LOAD. Similarly to *ABCA7* and *INPP5D, TREM2* also functions in phagocytosis through its role in microglial cell activation which is important for apoptotic cell and Aβ deposit removal (Takahashi *et al*., 2005).

Pyrosequencing targets also included the promoter region of *SIRT1* and a region covering a 3' CGI located within the *RIN3* gene. *SIRT1* was chosen as a target due to it previously being shown to contain aberrant methylation in LOAD blood. The pyrosequencing assay described in this chapter was designed to cover a region of the

SIRT1 promoter that has previously been investigated using methylation specific PCR (MSP) and bisulphite sequencing PCR (BSP), experiments which identified hypermethylation in LOAD blood samples by (Hou *et al*., 2013). SIRT1 may be involved in attenuation of AD pathology as sirtuins have roles in oxidation, inflammation and cell survival (Bonda *et al*., 2011, Wang *et al.,* 2010). SIRT1 has been shown to be important for memory and learning and may have a role in the development of multiple neurodegenerative diseases including AD, Parkinson's disease and Huntington's disease (Gao *et al*., 2010, Michan *et al*., 2010, Herskovits and Guarente, 2014). SIRT1 may be linked to AD could be due to its ability to attenuate Aβ toxicity and its involvement in Tau mediated toxicity (Wang *et al.*, 2010, Min *et al*., 2010)

Unlike the previously described targets, *RIN3* has not been genetically associated with LOAD at present. However the gene is located on chromosome 14 and sits between the genes *SLC24A4* and *LGMN,* both of which have been identified as having LOAD associated variants through GWAS studies (Lambert *et al.*, 2013b). Both genes also encode proteins with actions that have been shown to have a role in either neuronal activity or have been directly linked to AD pathology (Basurto-Islas *et al*., 2013, Dall and Brandstetter, 2016, Larsson *et al.*, 2011, Yu *et al*., 2014b).

RIN3 is likely to have a role in AD pathology through its interaction with BIN1, a gene which has been identified as having genetic variants which associate with AD (Lambert *et al.,* 2013b, Hu *et al*., 2011, Tan *et al*., 2013). It is likely that BIN1 and RIN3 interact in the process of endocytosis particularly the early endocytic pathway, possibly driving AD pathology though APP trafficking. *RIN3* encodes rab 5 binding protein which acts as a stabilizer of small GTPase rab5, this is needed for endocytosis and transport from the plasma membrane to early endosomes. BIN1 and RIN3 were shown to interact during this process; in addition BIN1 was shown to be moved to RIN3-Rab5 containing vesicles (Kajiho *et al.*, 2003).

In the case of *RIN3* a pyrosequencing target was chosen within the 3'UTR. For *MEF2C* two pyrosequencing targets were chosen: one within a promoter CGI and a second covered a CpG site located upstream of the promoter CGI. Targets, other than those within the promoter region of these genes, were chosen because recent methylome

studies have revealed the importance of non-promoter, and non CGI CpG methylation in LOAD. Whole genome and global methylation studies are identifying non-promoter methylation in AD associated genes. For example, of the twelve genes identified as harbouring differentially methylated sites in LOAD (work described by Humphries *et al*. (2015)), eight of these sites were located within the 3'UTR of genes identified and other differentially methylated sites were found within genes body and 5'UTR regions. Other methylome studies have also identified significant non-promoter AD associated methylation (Yu *et al.*, 2014b, De Jager *et al.*, 2014, Smith *et al*., 2016)

Non-promoter CpG methylation such as that found within the 3'UTR is likely to affect disease susceptibility and progression by influencing gene expression, transcriptional elongation and splicing (Maussion *et al*., 2014, Choi *et al*., 2009, Malumbres *et al*., 1999). Specifically 3'UTR hypermethylation has been shown to decrease gene expression, similarly to promoter methylation (Maussion *et al*., 2014). Therefore 3'UTR CGIs represent an interesting site for investigation in methylome studies relating to diseases such as AD.

4.1.3 Use of sEOAD samples for Pyrosequencing Assays

sEOAD (sporadic Early onset Alzheimer's Disease) samples were used for all pyrosequencing experiments. This type of disease represents a sub-group of EOAD (early onset Alzheimer's disease) which affects individuals under the age of sixty-five years, in contrast to LOAD which occurs after 65 years of age. EOAD is the less common form of AD, representing only 10% of cases (Cacace *et al.,* 2016, Kunkle *et al.,* 2017, Pinenburg., 2010). 5-10% of EOAD cases are caused by familial genetic mutations, which are also known as fAD (familial Alzheimer's Disease) (Kunkle *et al*., 2017, Harvey *et al*., 2003). The pathogenesis of fAD is better understood and most cases (60-70%) can be attributed to the inheritance of genetic mutations within either the amyliod precursor protein (*APP*) gene or presenlin (*PSEN1* and *PSEN2*) genes (Kunkle *et al.,* 2017). However the cause of sEOAD has proven to be much more elusive. Transcriptomic studied of sEOAD and fAD brain tissue have suggested that the two diseases likely have differing causes, which eventually result in similar end stages of the disease (Antonell *et al*., 2013).

Therefore one aim of the work described in this chapter was to investigate if CpG sites shown to be differentially methylated in LOAD blood were also differentially methylated in our sEOAD samples using the pyrosequencing platform. sEOAD samples were tested to determine if differential methylation previously reported was LOAD specific or a commonality in sEOAD. Unfortunately LOAD blood and cortex samples were not available to use for the pyrosequencing assays, however sEOAD brain samples were used in order to determine if methylation status of AD blood is similar to that of the AD brain.

4.2 Chapter Aims

- Investigate LOAD associate differential methylation identified in *SIRT1* using pyrosequencing in LOAD blood samples, sEOAD blood and sEOAD cortex samples.
- Use sEOAD blood and brain samples to investigate methylation of promoter CGI's in the AD associated genes PTK2β, ABCA7, INPP5D and MEF2C and determine whether methylation observed at *INPP5D* in LOAD leukocyte DNA using the McrBC method could also be observed in sEOAD samples using pyrosequencing.
- Identify if the methylation reported at one CpG site in *TREM2* by Smith *et al*. (2015) was also present in sEOAD brain or blood tissue.
- Identify any tissue specific methylation occurring in either sEOAD brain or blood tissue for the genes targeted using pyrosequencing.
- Investigate methylation of non-promoter CpG containing regions in *RIN3* and *MEF2C*.

4.3 Methods

4.3.1 Samples Details

Detail of samples used for pyrosequencing can be found in section 2.1. Specific details of the samples used during the work described in this chapter are detailed in table 4.1.

Table 4.1: Details of AD samples used for pyrosequencing. Table shows information about age at onset, sampling and death for all of the sEOAD and control samples used during the work described in this chapter.

Group	Sex M/F	Age at Death	Age at onset	Age at sampling
sEOAD blood	17//8	57.95 (SD 3.28)	47.2 (SD 2.28)	48.6 (SD 2.98)
sEOAD Brain (Cortex)	4/10	59.35 (SD 6.95)	48.85 (SD 3.32)	59.35 (SD 6.95)
Control Blood	6/13	N/A	N/A	82.8 (SD 6.95)
Control Brain	6/4	84.2 (SD 3.55)	N/A	84.2 (SD 3.55)

4.3.2 DNA extraction

Extraction of DNA from the blood and brain samples used in this chapter is described in methods section 2.1. Briefly DNA was extracted using phenol chloroform extraction followed by assessment using either NanoDrop™3300 spectrometer or gel electrophorisis. DNA was then stored by the Alzheimer's Research UK Consortium DNA Bank at the University of Nottingham until use.

4.3.3 Pyrosequencing assay targets and assay development

Pyrosequencing was used to investigate regions located within multiple AD associated genes, initially an assay was designed to cover a region of *SIRT1* that was identified as

harbouring LOAD associated hypermethylation using leukocyte DNA by Hou *et al*. (2013). The region covered by the assay covered the two CpG sites identified as being most significantly differentially methylated in LOAD samples, table 4.1.

In the case of *SIRT1* the pyrosequencing assay was designed to cover a specific target region previously identified. This was also the case for *TREM2*. This assay was designed to cover a CpG site within the *TREM2* gene at a location which has previously been shown to be hypermethylated in LOAD (Smith *et al*., 2016).

Assays were also designed for regions of the gene's *INPP5D, ABCA7, PTK2β, SORL1, SLC24A4, DSG2* and the intragenic region of *HLA-DRB1/5* investigated using McrBC enzymatic treatment (chapter 3). For these genes the assays were designed to target CGIs found upstream of the TSS assay. Following the optimisation of these assays a further assay was developed to target a region within the 3'UTR of *RIN3*, see table 4.1. *RIN3* interestingly sits between two other LOAD associated genes, *SLC24A4* and *LGMN*.

Two regions within the *MEF2C* gene were also targeted using pyrosequencing. Initially a region located upstream of the TSS was targeted, called *MEF2C.* This region contained four CpG sites and was located within the CGI numbered 38, shown in table 4.1. This region was chosen and the pyrosequencing assay was designed as described previously for the candidate genes *PTK2β* and *ABCA7* (described in material and methods section 2.5.2).

In addition to these four promoter CpG sites, a second target was also chosen within the *MEF2C* gene. This target included a CpG site located upstream of the *MEF2C* TSS, also shown in table 4.1. This assay was named *MEF2C*(2). This target was chosen due to emerging evidence that suggests non-promoter CpG methylation could have an important role in driving disease pathology (Humphries *et al*., 2015). Non-promoter CpG sites may have important roles in regulation of transcription or transcript structure (Choi *et al.*, 2009). Therefore it seemed prudent to investigate a CpG site located upstream of the promoter CGI target.

For *TREM2* and *SIRT1* regions were defined by previous studies. However in all other cases for promoter regions the closest CGI upstream of the TSS was selected using the UCSC genome browser and pyrosequencing, primers were designed using the PyroMark Assay design software 2.0. A full list of primers and regions targeted is shown in Table 4.2.

Pyrosequencing PCRs were then optimised using gradient PCR and touch-down PCR. All PCR contamination and primer dimer formation was removed before progressing to pyrosequencing. In all cases the primers sets scoring the highest rating, which represented the best possible primer sets for the chosen target were used. USCS genome browser was used to identify the CGI's and PyroMark Assay Design 2.0 software was used to design pyrosequencing primers (process described in materials and methods 2.5.2).

Pyrosequencing targets were kept short, as pyrosequencing assay targets should ideally be kept below 150bp, and PCR products below 350bp, this prevents the formation of secondary structures, especially within the single stranded pyrosequencing template, which can affect the sequencing result (Tost and Gut, 2007).

However, in the case of *ABCA7* it was possible to include eight CpG sites within the pyrosequencing target; this is because serial pyrosequencing was used. This technique involved the use of one PCR product during pyrosequencing; however the product is then used as a template for two rounds of pyrosequencing, with a different sequencing primer being used for each round of sequencing, described in more detail in methods section 2.5.7.

Table 4.2: Pyrosequencing target regions. Showing location of regions targeted using pyrosequencing.

Gene	Position in gene	Location on GRCh37/hg 19 Assembly	CpG island number	Target sequence	Genomic co-ordinates of individual sites	Primer sequences
TREM2	Upstream of TTS	chr6:41,131,213	n/a	n/a	chr6: 41131213	Trem2_kbf1 GGGTTGGTAAGGTTTTGTATTGT Trem2_KBr1b AATCCTAACCTCTAAAAACACAACTATTC Trem2_kbs1 TTAGATTTTTTATTAGTTGTAATG
INPP5D	Promoter	chr2:233,925,148-233,148,498	18	GCCGGCCCGGCCGAGGAGGCCCACGCCCA	Chr2: 233925160 Chr2: 233925165 Chr2: 223925169 Chr2: 223925181	Inpp5d_py1f1 TGGGTTTTGGGGGTGTTT Inpp5d_py1r1b AAAAAACTCCCCTCCTTACCTATCCT Inpp5ds2 TTTTGGGGGTGTTTG
SIRT1	promoter	chr10: 69647485-69647502	101	CACTACGCCCGGCTAAT	Chr10: 69647491 Chr10: 69647495	Sirtp2f1 AGTTTTTTTAGTAGTTGGGATTATATGTA Sirtp2r1b CAAAACCAACCTAACCAACATAAA Sirtp2s1 AGTTGGGATTATATGTATATGTTA
ABCA7	promoter	chr19:1040048-1040115	43	CGCCCAATAGCAGCGTGCAGAGGCAGGGGCGTGCCCCGGCGCTGCTACCTGCGCGGGCAAGCTCACG	Chr19: 1040048 Chr19: 1040062 Chr19:1040078 Chr19: 1040085 Ch19: 1040088 Chr19: 1040100 Chr19: 1040102 Chr19: 1040114	Abca7_4_f1 GGTTAGGAGAGGTTTTTTTGTGATT Abca7_4_r1b TCCTTCTCACCTTCCAAAAACTC Abca7_4_s1 AGTAGGTTAGTGAGTG abca74_s2 GTGTAGAGGTAGGGG

PTK2B	promoter	chr8:27182721-27285249	27	CGTCCCGGCTCACCTGGCGGTGCCCGAGGAGTAGTCG	Chr8: 27183390 Chr8: 27183395 Chr8: 27183407 Chr8: 27183425 Chr8: 27183436	Ptk2b_py1f1 GAGGAGGAGGGAGAATTTAATTT Ptk2b_py1r1b AACTCCCAACTCAAATACCC Ptk2b_py1s1 ATTTGTTAGGTAGATTTATTTGTA
MEF2C(1)	promoter	chr5:88179867-88179895	38	CGAAATGAAGACAACACGGCGAGCTGCG	Chr5: 88179867 Chr5: 88179883 Chr5: 88179886 Chr5: 88179893	Mef2c_pcor_f1 GTGTTTAAAAGGGGGAAAAGTTATAAGA Mef2c_pcor_r1 ACACATACCATACCCAAACCTAATAACAAT Mef2c_pcor_s1 AGAGGAGTAGGGAGT
MEF2C(2)	promoter	chr5:88,200,070-88,200,105	n/a	TGCTTCCCTCCCCTCCCCCCTCCCGACCCCCTATG	Chr5: 88200094	Mef2c_py1f1 GATTGGATATTTTTTATTGGAATTAGTAGT Mef2c_py1r1b TATCACTAACAACCAAACCTTTATCAA Mef2c_py1s1 ATTGGAATTAGTAGTATAGGG
RIN3	3'UTR	chr14:93153018-93209772	124	CGGTCTCCGCCGAGCCGGGATCCTCAGCCGCTTCCCGCG	Chr14: 93153344 Chr14: 93153351 Chr14: 93153354 Chr14: 93153359 Chr14: 93153372 Chr14: 93153379 Chr14: 93153381	RIN3_pyf1 GGGTTTAGGGTTGTAGGTAGAGA RIN3_pyr1 AAACCCTAACCACCAATTACCATCAC RIN3_pys1 ATTGGGAATAGTAGGTTT

4.3.4 Bisulphite conversion of DNA

Prior to use in pyrosequencing PCRs DNA samples where bisulphite converted using the EpiTect DNA Bisulphite Kit (Qiagen, Germany). Manufacturer instructions were used and more detail can be found in section 2.4.2.

4.3.5 Pyrosequencing PCR Optimisation

Following the design of pyrosequencing PCR primers each primer set was optimised using several methods, with details provided in table 4.2 and section 2.5.4. Following optimisation it was decided that 0.25pmol of primer and 0.1 µl of bisulphate converted DNA was the optimal amount to use in each assay using the PyroMark PCR Kit (Qiagen, Germany). The specific PCR conditions used for each primer set can be found in appendix 4.

4.3.6 Pyrosequencing Assay design

All of the assays descried where designed using the PyroMark Q24 software (Qiagen, Germany), further details in section 2.5.5.

4.3.7 Pyrosequencing Process

The process of pyrosequencing is described in detail in section 2.5.6. Briefly following successful PCR amplification the biotinylated PCR products where immobilised on streptavidin coated Sepharose beads. Following this the PCR product was washed and denatured, the pyrosequencing sequencing primer was then annealed to the single strand of DNA left bound to the beads. Pyrosequencing was then carried out using a Q24 Pyrosequencer (Qiagen, Germany) and results were interpreted using the PyroMark Q24 software. All failed runs were repeated and each sample was used for at least two technical repetitions.

4.3.8 Assay Optimisation

Unfortunately it proved to be difficult to generate optimised pyrosequencing assays for some gene targets. It was not possible to generate pyrosequencing results for the targeted regions of *HLA-DRB 1/5, SORL1, SLC24A4* or *DSG2*. These assays likely failed for several reasons. In the case of *HLA-DRB1/5* and *SORL1* it was possible to generate a strong PCR product using the first set of primers designed, following the use of a gradient and step down PCR. However subsequent pyrosequencing failed despite this strong initial PCR product.

Similarly in the case of *SLC24A4* the first set of primers designed resulted in a good PCR product, however non-specific bands were observable on the gel. Since any non-specific PCR product can interfere with the pyrosequencing procedure, the PCR was re-optimised, resulting in a strong specific band (Tost and Gut, 2007). Unfortunately subsequent pyrosequencing also failed. In this instance a second set of primers was designed, unfortunately these also failed at the pyrosequencing stage despite the use of two different sequencing primers.

For *DSG2* the initial set of primers failed to produce adequate PCR product despite optimisation. Therefore a second set were designed, these successfully produced a suitable PCR product, which then failed at the pyrosequencing stage. Again two sequencing primers were tested for each set of PCR primers and neither resulted in usable pyrosequencing data. Since two sets of primers had been extensively optimised it seemed unlikely that pyrosequencing of this region would be successful, in addition to this if problems with PCRs persist it is unlikely that pyrosequencing will be successful (Tost and Gut, 2007). Optimisation steps taken for each pyrosequencing target are shown in table 4.3.

Table 4.3: Optimisation steps taken for pyrosequencing of LOAD associated genes.

Gene Target	Primer set	PCR Worked	Action Taken	Pyrosequencing	Action Taken
SIRT1	1	yes	Gradient PCR, touchdown PCR, PCR worked.	Yes	Used PCR product for pyrosequencing
MEF2C(1)	1	yes	Gradient PCR, touchdown PCR, PCR worked.	Yes	Used PCR product for pyrosequencing
MEF2C(2)	1	yes	Gradient PCR, touchdown PCR, PCR worked.	Yes	Used PCR product for pyrosequencing
RIN3	1	yes	Gradient PCR, touchdown PCR, PCR worked.	Yes	Used PCR product for pyrosequencing
INPP5D	1	No	Gradient PCR, touchdown PCR, PCR worked.	No	Re-designed assay to cover the anti-sense stand, ordered primer set 2
	2 (stronger score)	Yes (weak band)	Gradient PCR, touchdown PCR, strong band no secondary bands on gel.	No	Ordered new sequencing primer (seq 2). Re-optimised with sequencing primer 2, pyrosequencing worked.
TREM2	1 (designed using primer design software)	Yes (non-specific bands present)	Gradient PCR, touchdown PCR, weak band.	No	Used primers from Smith *et al*. (2016)
	2 (taken from Smith *et al*. (2016)	Yes	Gradient PCR, touchdown PCR.	Yes	Used for pyrosequencing.
ABCA7 (serial pyrosequencing)	1 set of PCR primers with two sequencing primers used for pyrosequencing	Yes	Gradient PCR, touchdown PCR.	Yes	Both sequencing primer 1 and 2 worked successfully during serial pyrosequencing.

PTK2β	1	Yes	Gradient PCR, touchdown PCR.	Yes	Used PCR product for pyrosequencing
HLA-DRB1/5	1	Yes	Gradient PCR, touchdown PCR, strong band produced.	No	All pyrosequencing attempts failed despite strong PCR product
SORL1	1	Yes	Gradient PCR, touchdown PCR, strong band produced.	No	All pyrosequencing attempts failed despite strong PCR product
SLC24A4	1	Yes (non-specific product observed)	Re-optimisation of PCR, gradient PCR, touchdown PCR, strong band produced.	No	Attempted to use a different sequencing primer however, all pyrosequencing attempts failed despite strong PCR product. Redesigned PCR primers.
	2	Yes	Gradient PCR, touchdown PCR, strong band produced.	No	All pyrosequencing failed despite the use of two different sequencing primers.
DSG2	1	No	Gradient PCR, touchdown PCR, very weak band	No	All pyrosequencing attempts failed, tried two sequencing primers. Redesigned PCR primers.
	2	Yes	Gradient PCR, touchdown PCR, visible band	No	Tried pyrosequencing with two sequencing primers, both failed.

All pyrosequencing target PCRs, prior to pyrosequencing, were extensively optimised to produce good specific PCR product; however in some instances the subsequent pyrosequencing assays were still unsuccessful. One possible reason for this is the pyrosequencing reaction not being adequate due to the assay designed. However, other possible reasons can include technical issues or problems with the pyrosequencing primers. Attempts were made to minimise the probability of these issues occurring. For example, low pyrosequencing signal, resulting in a failed assay, can be caused by a number of errors. First, the sequencing primer may not be sufficiently specific to anneal correctly to the PCR template, to avoid this issue multiple sequencing primers were used.

Faults with the pyrosequencing procedure and equipment could also result in failed assays, this could include the cartridge being blocked or problems with the heating device resulting in unsuccessful denaturation of the template prior to sequencing (Tost and Gut, 2007). To prevent this from happening the pyrosequencing cartridges where cleaned after and prior to use with high quality water and allowed to dry fully before use. The vacuum prep tool and heating device were also tested prior to every sequencing run. A positive control assay was also run along with all new assays; if this control failed pyrosequencing was repeated.

In addition to these technical steps, for all assays several controls were included. These included a template only control to ensure no secondary structures were forming which could prevent 3' end extension. Assays were also conducted which used only the sequencing primer. This was used to confirm that the sequencing primer was not forming a primer dimer. This can cause a 3' or 5' overhang, resulting in a template for extension during sequencing. A negative PCR control was also included in all assays.

Since pyrosequencing primers were designed using the PyroMark Assay Design 2.0 software and the primers selected for use were the ones which scored the highest quality score, it seemed unlikely that the use of new primer sets would rectify poor assays. Therefore rather than focusing on the optimisation of all assays, work focused on the assays that were successful, namely the assays for the *INPP5D, PTK2β, ABCA7* and *TREM2* targets. In the case of the *SIRT1, RIN3* and *MEF2C* targets, optimisation included a gradient PCR, following this the most successful temperature was selected for the assay. Initially a small amount of control and sEOAD samples were used to investigate each target, however since *RIN3* and *MEFC(2)*

showed either a significant difference in methylation in sEOAD samples or a possible epi-variant these targets were explored further and sample size was extended.

4.3.9 Statistical Assessment

Described in detail in section 2.5.9 briefly statistical analysis was performed using Microsoft excel or Genstat 18 statistical packages (Genstat work was conducted by Dr Jim Craigon, University of Nottingham). Two tailed T-tests were performed to compare the averages of methylation at each CpG site or between groups (AD and control) (value of <0.05 was used). Normality calculations were performed as described in section 2.5.9 on all data sets greater than five and all data was found to be normally distributed. Post-hoc power calculation values were undertaken by first generating either a cohen's d or hedge's g value (using Effect Size Calculator. Social Science Statistics; http://www.socsistatisics.com/effectsize/,accessed March, 2019) (see table 4.4) and then a power calculation was generated using the Real Statistics Resource Pack Software (Release 5.4). Copyright (2013-2018) Charles Zaiontz. www.real-statistics.com.

Table 4.4: Table shows post hoc power calculation results. Power values are shown for each individual CpG site investigated by each pyrosequencing assay. Vales presented are cohen's d values.

AD vs Control		Power							
Gene id.	**Tissue**	**CpG1**	**CpG2**	**CpG3**	**CpG4**	**CpG5**	**CpG6**	**CpG7**	**CpG8**
RIN	**Blood**	0.66	0.67	0.73	0.79	0.88	0.67	0.73	-
	Brain	0.34	0.38	0.26	0.36	0.37	0.39	0.43	-
SIRT1	**Blood**	0.05	0.08	-	-	-	-	-	-
	Brain	0.06	0.25	-	-	-	-	-	-
PTK2B	**Blood**	0.59	0.26	0.38	0.07	0.18	-	-	-
	Brain	0.05	0.06	0.18	0.24	0.26	-	-	-
ABCA7	**Blood**	0.15	0.34	0.05	0.11	0.06	0.11	0.05	0.06
	Brain	0.08	0.07	0.05	0.05	0.19	0.05	0.1	0.05
TREM2	**Blood**	0.16	-	-	-	-	-	-	-
	Brain	0.06	-	-	-	-	-	-	-
INPP5D	**Blood**	0.07	0.06	0.09	0.05	-	-	-	-
	Brain	0.25	0.3	0.31	0.11	-	-	-	-

As described in section 2.5.9 age was considered for covariate testing however due to the large age difference between diseased and control samples this was not deemed appropriate due to the rules of ANCOVA and other co-variant tests (see fig 2.1). However, analysis was undertaken using Genstat to fit Gender then AD vs Control and the interaction term between them in that order. Age and its interactions with gender was then fitted and treatment groups next. (A stepwise regression). Extra variation accounted for by each step is then tested in an analysis of variance. Outputs show AD vs Control effects but no interactions with Gender or age within groups, data is shown in supplementary appendix 5.

4.4 Results

4.4.1 RIN3 (ras and rab interactor 3) hypomethylation in sEOAD blood but not cortex tissue

A total of seven CpG sites were included in the *RIN3* 3'UTR pyrosequencing assay. A group wide average methylation was calculated, for each sample, across the seven CpG sites investigated. For blood tissue, significant hypomethylation was identified in the sEOAD samples for the targeted area. The average methylation in control blood was 53.6% compared to 37.3% in sEOAD blood samples. A significant reduction of methylation of 16.3% (P=0.01) (figure 4.1).

A group wide reduction in methylation was also observed in sEOAD brain tissue. Average methylation across the region in control samples was 25.9% compared to 16% in sEOAD brain samples. However this 10% reduction was not proven to be statistically significant P= 0.097 (figure 4.1).

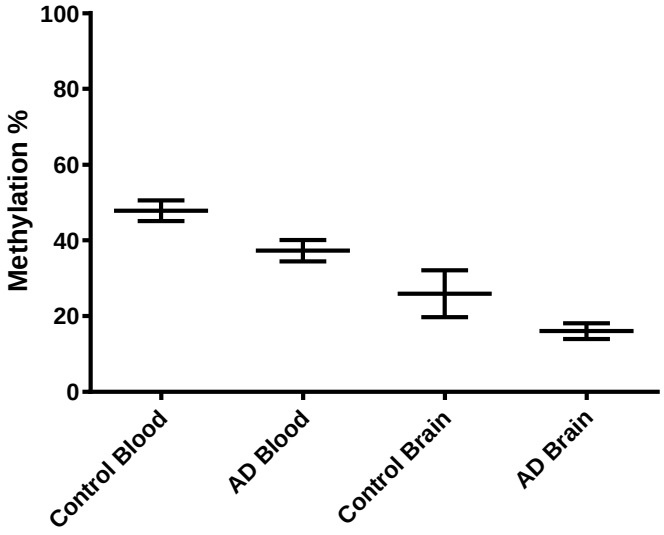


Figure 4.1: Collective group wide methylation control vs AD for *RIN3*. Graph shows average methylation across the whole region investigated in the *RIN3* 3'UTR. Control blood n= 26, AD blood n= 22, Control brain n=10, AD brain n=14. * p value <0.05.

4.4.2 RIN3 (ras and rab interactor 3) at individual CpG site resolution

In addition to the group wide significant reduction in methylation observed for the *RIN3* target region in sEOAD blood samples, significant sEOAD associated reduction in methylation was also observed at each individual CpG site (figure 4.2). For sEOAD vs control Chr14: 93153344 P= 0.019, Chr14: 93153351 P=0.018, Chr14: 93153354 P=0.012, Chr14: 93153359 P=0.009, Chr14: 93153372 P=0.002, Chr14: 93153379 P=0.018 and Chr14: 93153381 P=0.013. CpG site five also remained significant following stringent P value correction for multiple testing using the Bonferroni correction method, which calculated that a P value of less than 0.007 would be required for a confidence of 95%. This and the original P values might indicate that methylation occurs initially within the middle of the region investigated and then spreads out towards either end of the *RIN3* 3'UTR target site.

In contrast none of the CpG sites investigated were found to be significantly hypomethylated in the sEOAD brain samples, despite a reduction in methylation being seen at every CpG site in the sEOAD samples. For each CpG site P values for sEOAD vs control methylation was: Chr14: 93153344 P= 0.12, Chr14: 93153351 P=0.096, Chr14: 93153354, P=0.18, Chr14: 93153359, P=0.105, Chr14: 93153372, P=0.062, Chr14: 93153379, P=0.091 and Chr14:

93153381 P=0.076. Average methylation at each cytosine for blood and brain is shown in figure 4.2.

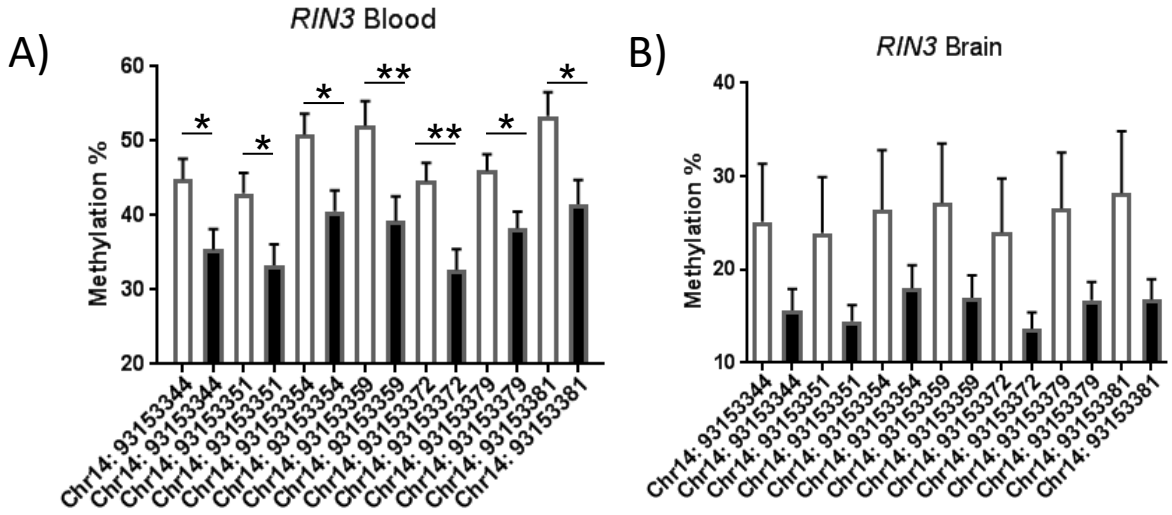


Figure 4.2: Methylation of blood and brain at each CpG investigated in *RIN3*. A) All seven CpG sites were found to be significantly hypomethylated in AD blood using a two tailed T-test, CpG5 also survived correction for multiple testing. B) No significant difference in methylation was observed in AD brain however all CpG sites had a reduction of methylation in AD. Showing average methylation found at each CpG site in the *RIN3* 3'UTR. Control shown as a white bar (blood n=26, brain n=10) and AD (blood n=22, brain n=14) shown as a black bar. Error bars represent S.E.M. *=P<0.05, **=P<0.01.

4.4.2.1 External Validation of Data

The RIN3 region identified and tested as part of this study has since been confirmed as differentially methylated in an independent LOAD (rather than EOAD) population when compared to control samples (personal communication, Dr J Tulloch, Washington State). Briefly, a small cohort of age matched AD (n=24) and control (n=24) blood samples were testing using the pyrosequencing assay described in Boden *et al*, 2017 (see appendix 6). They report that the assay was reproducible and each of the 7 CpG sites in this region was found to be consistently differentially methylated between samples.

4.4.3 The MEF2C (myocyte enhancer factor 2C) promoter was not differentially methylated in sEOAD blood or brain

The target region containing four CpG sites upstream of the *MEF2C* TSS (*MEF2C*(1)) was found not to be differentially methylated in either sEOAD blood or brain tissue, figure 4.3.

When an average methylation was calculated across the *MEF2C(1*) target region average methylation in sEOAD blood and brain was calculated as 4.8% and 5.1% respectively. Similar levels of methylation were found within control samples at this region with blood methylation being 4.1% and brain methylation being 4.4%. These methylation values were compared using a T-test to reveal no significant AD associated methylation in either tissue (blood P= 0.44, brain P= 0.42). It is important to note that the methylation averages observed where very close to the limit of detection of the pyrosequencing platform, however all samples were ran in at least duplicate and a number of controls were included in the assay, including a technical control (section 2.5.8).

In addition to this no significant difference in methylation was observed at any individual CpG site investigated in either blood or brain tissue between sEOAD and controls. For blood sEOAD vs control methylation at Chr5: 88179867 P=0.45, Chr5: 88179883 P=0.47, Chr5: 88179886 P= 0.4 and Chr5:88179893 P=0.28. For brain sEOAD vs control Chr5: 88179867 P=0.1, Chr5: 88179883 P=0.93, Chr5: 88179886 P=0.27, Chr5: 88179893 P= 0.051. Average methylation at each CpG site is shown in figure 4.3B and C.

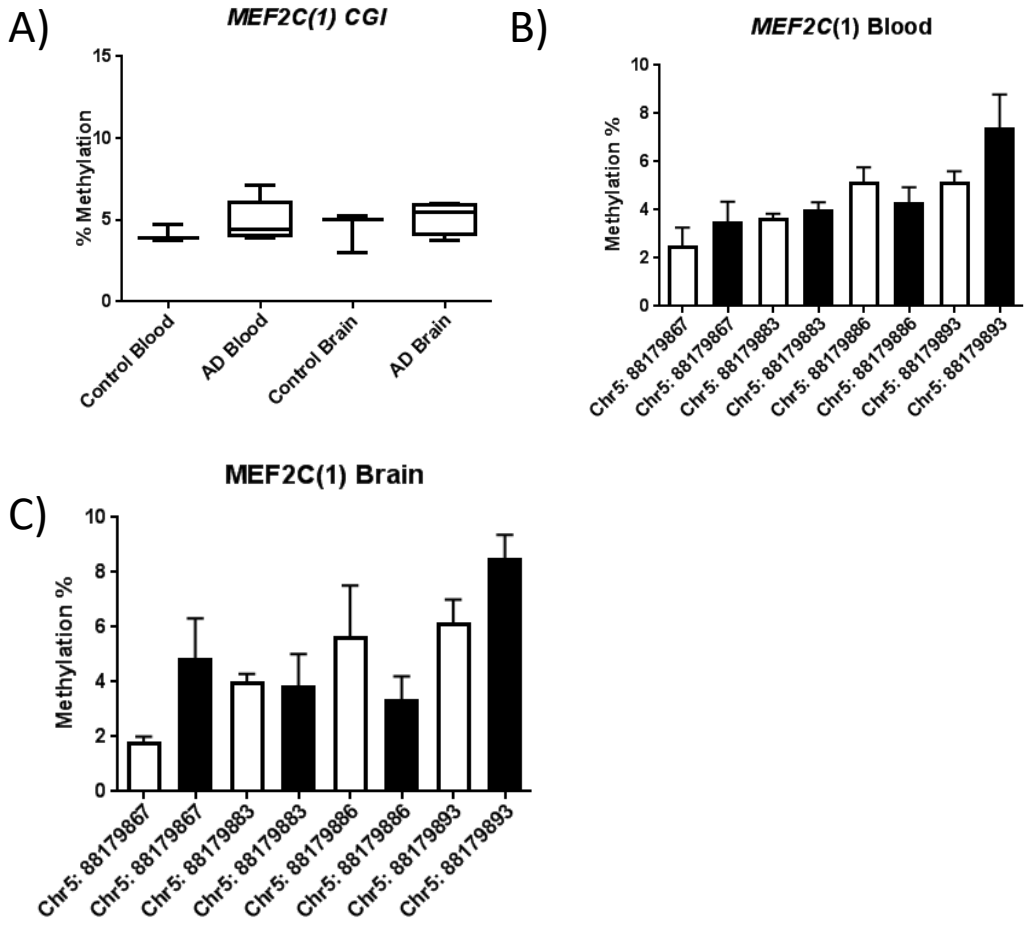


Figure 4.3: Box plots showing data representing average methylation across the promoter region investigated for *MEF2C*. No significant difference in methylation was observed between sEOAD and controls in either blood or brain tissue for region *MEF2C*(1). A, B and C) No significant difference in methylation was observed at any of the four CpG sites investigated in the *MEF2C* promoter. Bar charts show average methylation at each CpG site and error bars show S.E.M, white show control and black sEOAD. For *MEF2C*(1) control blood n=3, AD blood n=5, control brain n=3, AD brain n=4

4.4.4 MEF2C (myocyte enhancer factor 2C) Upstream CpG site might represent a sEOAD associated epi-varient

When taken as a group wide average no significant difference in methylation was associated with sEOAD in either blood or brain tissue at the *MEF2C* upstream CpG site (Chr5: 88200094) investigated (target *MEF2C*(2)). Average methylation at this site in AD blood was 91.8% and control blood was 92.3%. Similarly average methylation in sEOAD brain was 84.8% compared to 86.25% in control brain. Neither tissue showed any sEOAD associated methylation, blood

sEOAD vs control P= 0.70, brain AD vs control P=0.59. Group wide average methylation is shown in figure 4.4A.

However, a significant reduction in methylation was observed for one sEOAD patient blood sample. Patient sample M341 (taken from blood) had an average methylation of 70% at the CpG site investigated, and since the group wide average methylation at this CpG site in blood was a percentage of 92% this represented a decrease in methylation of 22%. This difference in methylation was found to be significant using a one-way analyses of variance, the likelihood of observing this results by chance was calculated as 2.0E-10.

Interestingly the brain (cortex) sample from patient M715 also showed a large hypomethylation at this CpG site when compared to the brain group wide average methylation of 85.4%. Methylation of the CpG site was found to be 70.5% in the sample, a reduction in methylation of 15%. However, while intriguing, this reduction failed to reach statistical significance.

All data was calculated using a minimum of two technical replicates for each biological sample used, making technical error unlikely. Therefore the identification of hypomethylation in one blood sample might represent a rare epi-allele, associated with AD, at this CpG site. Since rare genetic differences in *MEF2C* have been associated with AD progression it is plausible that rare epigenetic variance occurring within the population might also drive sEOAD progression.

sEOAD samples M341 and M715 were also genetically tested to eradicate the possibility of any rare genetic mutations causing the appearance of the differential methylation observed in these samples. No association with the known *MEF2C* GWAS SNP (rs190982) was found in either sample tested. Suggesting that the AD associated SNP was not having an effect on the methylation observed in these samples. This work was conducted by Dr Keeley Brookes (University of Nottingham).

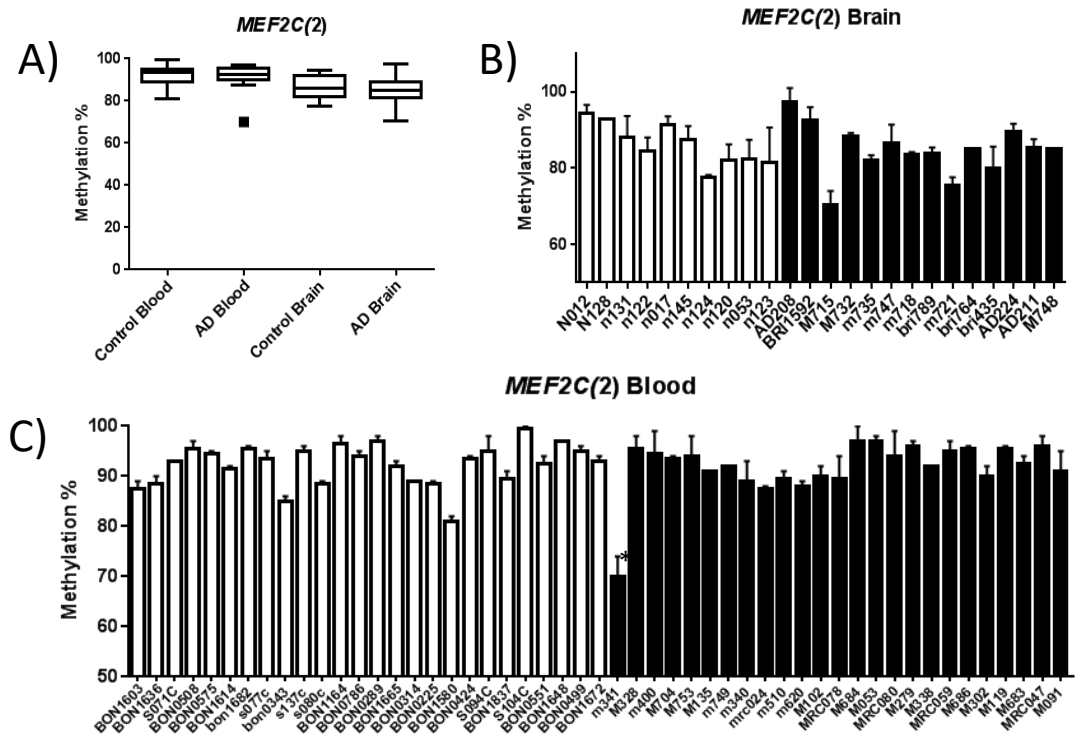


Figure 4.4: Analysis of individual methylation at the *MEF2C* upstream CpG site identified rare hypomethylation in an AD blood sample. A) Patient M341 had significant hypomethylation at the CpG site when compared to other AD and control samples tested (P=2.0E-10). B) No samples were found to be significantly hypomethylated in AD brain. Graphs show average methylation at the CpG investigated in each sample investigated (average of at least two pyrosequencing runs). Controls are shown in white and AD samples in black. Error bars represent the S.E.M. Control blood=26, AD blood n=25, control brain n=10, AD brain n=14; male brain n= 10, female brain n=14, male blood n=29, female blood n= 22.

4.4.5 No sEOAD specific methylation was observed in any of the other pyrosequencing targets

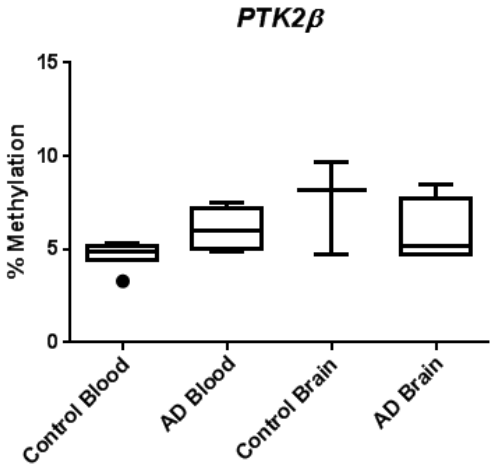
In order to assess methylation at the targets initially investigated again an average of methylation was calculated across all of CpG sites tested for each target. For all five targets investigated no significant difference in methylation was observed between sEOAD and control in brain or blood tissue (figure 4.5).

MrcBC results had indicated that the *INPP5D* promoter might harbour LOAD specific methylation in blood however, surprisingly no methylation difference was observed at the *INPP5D* promoter region targeted in either tissue using sEOAD samples together with the

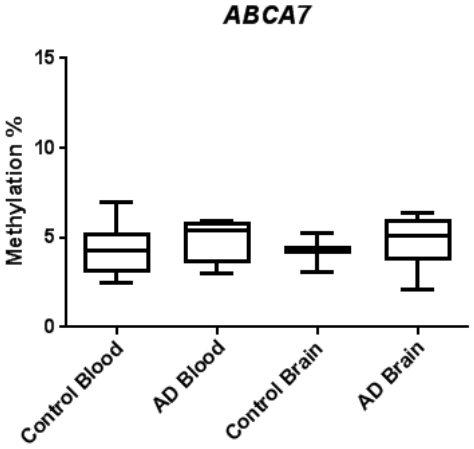
pyrosequencing platform. In addition, no significant difference in methylation was observed at the *TREM2* CpG site in either sEOAD blood or brain. This was also surprising considering Smith *et al*., (2016) identified LOAD specific hypomethylation at this site in LOAD brain tissue, however Smith *et al.,* (2016) tested a much higher number of samples and also controlled for co-varients, which likely explains this discrepancy in results. Similarly the pyrosequencing also failed to identify the any differential methylation in SIRT1, despite Hou *et al.,* (2013) reporting this in LOAD samples.

It should be noted that the average methylation observed for *PTK2β* was close to the 5% limit of the pyrosequencing platform. However, all results were taken from an average of at least two pyrosequencing runs and an internal control was included for bisulphite conversion.

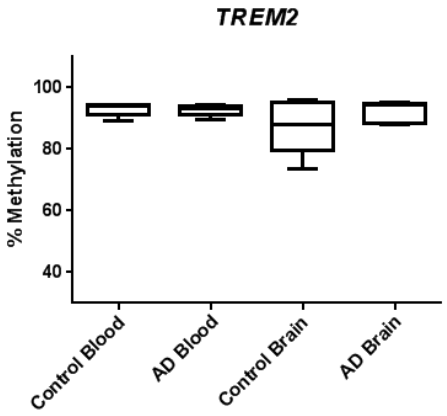
A)


PTK2β
% Methylation
15
10
5
0
Control Blood
AD Blood
Control Brain
AD Brain

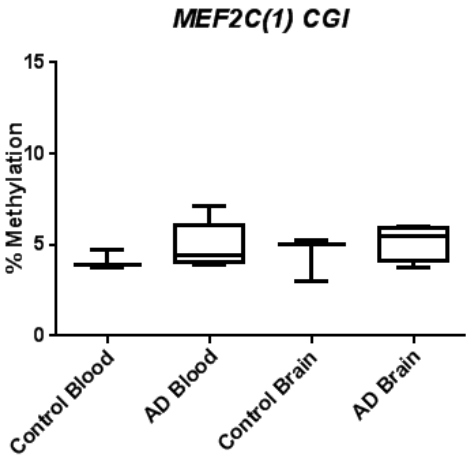
B)


ABCA7
Methylation %
15
10
5
0
Control Blood
AD Blood
Control Brain
AD Brain

C)


TREM2
% Methylation
100
80
60
40
Control Blood
AD Blood
Control Brain
AD Brain

D)


MEF2C(1) CGI
% Methylation
15
10
5
0
Control Blood
AD Blood
Control Brain
AD Brain

E)

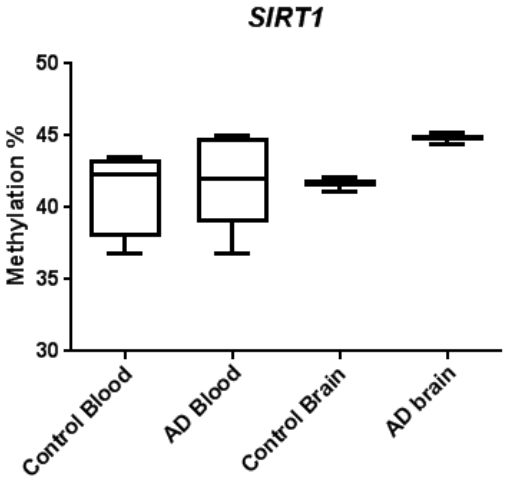

SIRT1
Methylation %
50
45
40
35
30
Control Blood
AD Blood
Control Brain
AD brain

Figure 4.5: Box plots showing data representing average methylation across the regions investigated for *PTK2β* (A), *ABCA7* (B), *TREM2* (C), *INPP5D* (D) and *SIRT1* (E). No significant difference in methylation was observed between AD and control in either blood or brain tissue.. For *PTK2β* control blood n=6, AD blood n=5, control brain n=3 and AD brain n=4. For *ABCA7* control blood n=7, AD blood n=4, control brain n=3, AD brain=6. For *TREM*2 control blood n=5, AD blood n=5, control brain n=5 and AD brain n=5. For *INPP5D* control blood n=5, AD blood n=4, control brain n=3, AD brain n=4. For *SIRT1* AD samples n=5, control blood n=4, sEOAD AD brain n=3 and control brain n=3.

4.4.6 Methylation at the resolution of individual CpG's

Methylation of each individual CpG site investigated for each target region was also compared for sEOAD and control blood and brain in the *ABCA7, INPP5D*, *PTK2β* and *SIRT1*. Since only one CpG site was covered by the *TREM2* assay, analysis was not required for this target. No significant difference was identified at any of the CpG sites investigated in *ABCA7, INPP5D*, *PTK2β* and *SIRT1* in sEOAD blood or brain (figure 4.6).

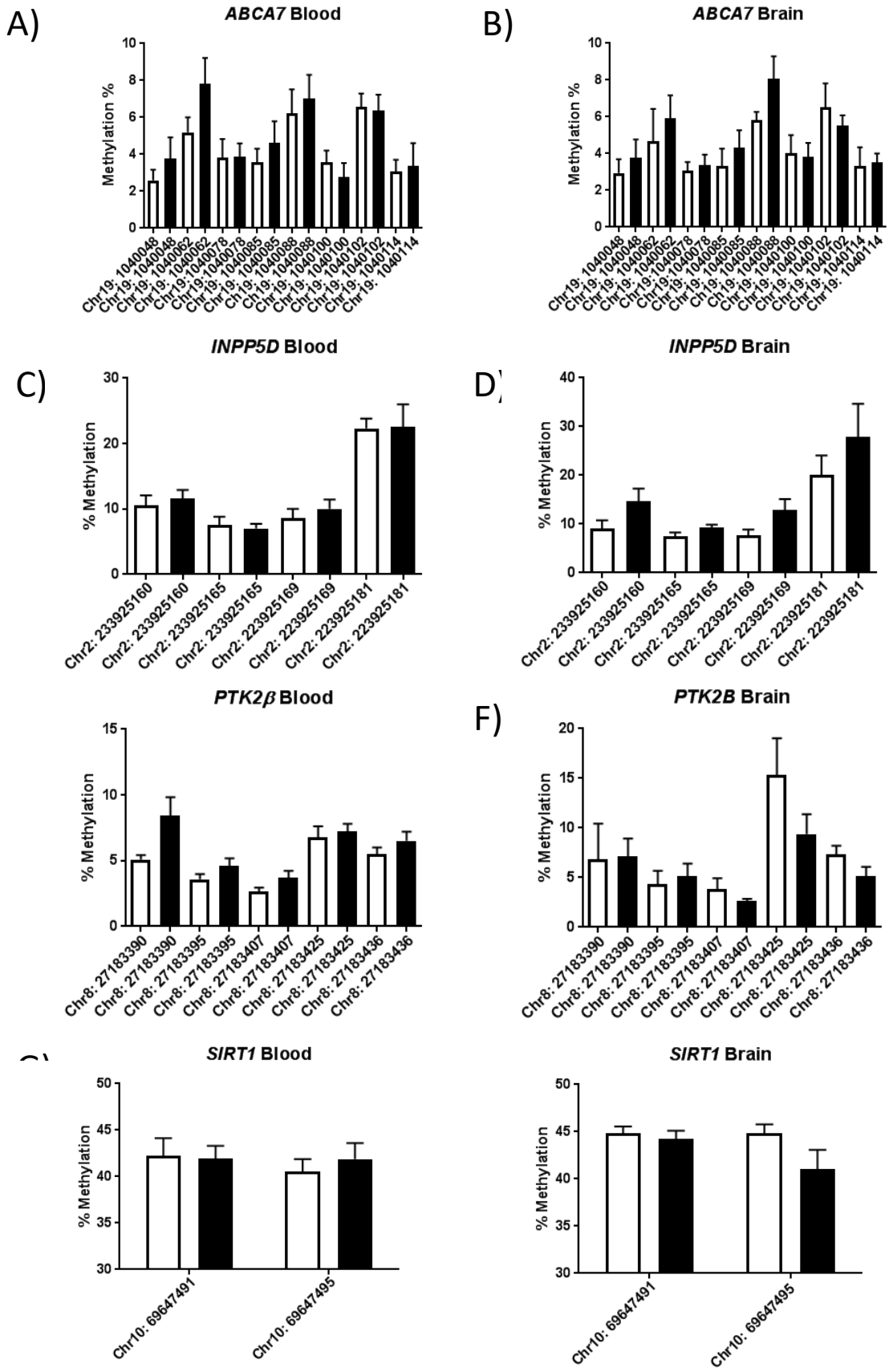

A)
ABCA7 Blood
Methylation %
Chr19: 1040048
Chr19: 1040048
Chr19: 1040062
Chr19: 1040062
Chr19: 1040078
Chr19: 1040078
Chr19: 1040085
Chr19: 1040085
Chr19: 1040088
Chr19: 1040088
Chr19: 1040100
Chr19: 1040100
Chr19: 1040102
Chr19: 1040102
Chr19: 1040114
B)
ABCA7 Brain
Methylation %
Chr19: 1040048
Chr19: 1040048
Chr19: 1040062
Chr19: 1040062
Chr19: 1040078
Chr19: 1040078
Chr19: 1040085
Chr19: 1040085
Chr19: 1040088
Chr19: 1040088
Chr19: 1040100
Chr19: 1040100
Chr19: 1040102
Chr19: 1040102
Chr19: 1040114
C)
INPP5D Blood
% Methylation
Chr2: 233925160
Chr2: 233925160
Chr2: 233925165
Chr2: 233925165
Chr2: 223925169
Chr2: 223925169
Chr2: 223925181
Chr2: 223925181
D)
INPP5D Brain
% Methylation
Chr2: 233925160
Chr2: 233925160
Chr2: 233925165
Chr2: 233925165
Chr2: 223925169
Chr2: 223925169
Chr2: 223925181
Chr2: 223925181
PTK2β Blood
% Methylation
Chr8: 27183390
Chr8: 27183390
Chr8: 27183395
Chr8: 27183395
Chr8: 27183407
Chr8: 27183407
Chr8: 27183425
Chr8: 27183425
Chr8: 27183436
Chr8: 27183436
F)
PTK2B Brain
% Methylation
Chr8: 27183390
Chr8: 27183390
Chr8: 27183395
Chr8: 27183395
Chr8: 27183407
Chr8: 27183407
Chr8: 27183425
Chr8: 27183425
Chr8: 27183436
Chr8: 27183436
SIRT1 Blood
% Methylation
Chr10: 69647491
Chr10: 69647495
SIRT1 Brain
% Methylation
Chr10: 69647491
Chr10: 69647495

Figure 4.6: DNA methylation AD brain and blood at individual CpG sites. No significant difference in methylation was observed at individual CpG resolution in any of the targets. A, C, E and G show average methylation found at each CpG investigated in *ABCA7, INPP5D , PTK2β* and *SIRT1* blood respectively and B, D,F and H show average methylation at each CpG in brain. White bars represent control average and black bars represent AD average. Error bars show S.E.M. *ABCA7* control blood n=7, AD blood n=4, control brain n=3, AD brain=6. *INPP5D* control blood n=5, AD blood n=4, control brain n=3, AD brain n=4. For *PTK2β* control blood n=6, AD blood n=5, control brain n=3 and AD brain n=4. For *SIRT1* AD samples n=5, control blood n=4, sEOAD AD brain n=3 and control brain n=3.

4.4.7 SIRT target showed differential methylation in LOAD Blood

CpG sites labelled 5 and 6 by Hou *et al*. (2013) (and subsequently Chr10: 69647491 and Chr10: 69647495 in the following report) were investigated to identify if the hypermethylation observed by Hou *et al*. (2013) could be reproduced in a separate sample of LOAD leukocyte DNA samples.

No significant difference in methylation at Chr10: 69647491 was identified using pyrosequencing, as shown in figure 4.7A. The average methylation observed at this CpG site in LOAD samples was 38.5% and in controls 41.9%. This represents a reduction in methylation in LOAD, however statically this reduction proved insignificant P= 0.16. Chr10: 69647495 also showed a reduction in methylation at this site in the LOAD samples compared to controls, a difference of 6.5%, LOAD samples had an average of 40% and control had an average of 43.5%. In addition, this reduction was found to be significant (P= 0.039) using a two tailed T-Test. It is interesting that Hou *et al*., (2013) reported hypermethylation at these sites in LOAD however the pyrosequencing results described found a slight reduction in methylation at the two sites in the LOAD samples used.

Since both sites were found to have reduced methylation in LOAD samples the average methylation across both CpG sites was calculated and compared to see if a regional methylation difference might be occurring in the samples, shown in figure 4.7B. No significant difference in methylation was observed when the region was taken as a whole, (P=0.08).

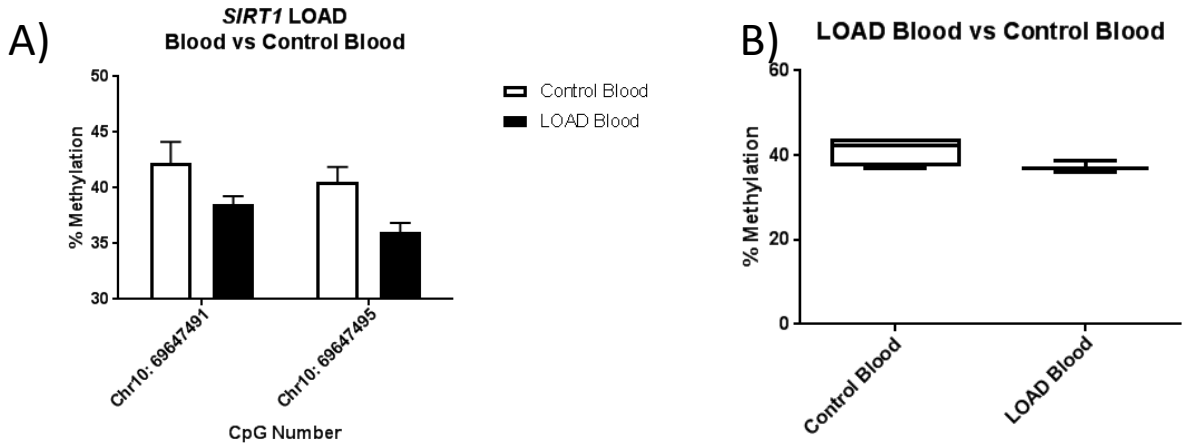


Figure 4.7: Average percentage methylation is at the two CpG sites investigated in *SIRT1* using LOAD leukocyte samples. A) Average methylation at each CpG investigated. No significant difference in methylation was observed at Chr10: 69647491 (described as site 5 by Hou et al., (2013)). Significant hypomethylation of Chr10: 69647495(site 6) was found in LOAD samples (p= 0.039). B) Boxplot showing average methylation in LOAD and control blood samples across both CpG sites. No significant difference in methylation was observed when the area was taken as a whole (p= 0.08). LOAD blood samples n=3, Control samples n=4.

These results are intriguing however it is important to note that the methylation difference observed at Chr10: 69647495 was extremely small. In addition, the sample size used was modest and ideally many more samples would be tested and a great number of samples would result in greater power, however additional samples were not available for testing. Unfortunately, due to a lack of LOAD leukocyte DNA samples, it was not possible to complete any further work investigating this region in LOAD, therefore the described assay was used to investigate methylation at the two CpG sites in sEOAD samples.

4.5 Discussion

The work described in this chapter used pyrosequencing to assess levels of methylation at regions within multiple genes that have been associated genetically and/or functionally with LOAD. Previous research has focused on gaining a comprehensive understanding of the epigenetic causation and consequence of LOAD, however very little work has been conducted focusing on sEOAD. The work of this chapter therefore aimed to identify if any sEOAD differential methylation could be identified in LOAD associated genes not only within

promoter regions but also in other genic regions, namely a CpG site upstream of the *MEF2C* promoter CGI and a region within the 3'UTR of *RIN3*.

4.5.1 RIN3 and MEF2C showed differential methylation in sEOAD

Significant hypomethylation was observed at all seven CpG sites investigated in the *RIN3* 3'UTR within sEOAD blood samples. In addition, while not statistically significant, a decrease in methylation was also observed at every CpG site investigated in sEOAD brain (cortex) samples. These results suggest that the hypomethylation occurring in sEOAD is group-wide, rather than specific to an individual, suggesting that perhaps this aberrant methylation is a consequence of the sEOAD condition rather than a driver of the disease.

The reduction observed in both tissue types might suggest that methylation levels observed in leukocytes are in fact reflective of the AD pathology which they encounter as they move through the brain. Methylation at specific sites within the gene are likely driven due to the experience of diseased tissue, rather than the epigenetic change being a spontaneous cause of disease progression.

The biological consequence of *RIN3* 3'UTR methylation would likely be instigated through its interaction with BIN1. It is plausible that 3'UTR hypomethylation in AD could drive increased expression of the RIN3 protein, since 3'UTR hypermethylation has been associated with decreased gene expression (Maussion *et al.*, 2014). One explanation would be that within the brain hypomethylation of the *RIN3* 3'UTR results in increased RIN3 expression. This is supported by the significant increase in hypomethylation observed with brain tissue when compared to blood tissue within the *RIN3* region targeted. This suggested that length and level of exposure to RIN3 is important in driving methylation levels of the *RIN3* 3'UTR.

In the case of MEF2C two regions were investigated, a region within a CGI located upstream of the TSS and a further CpG site located upstream of the initial target. Interestingly no significant sEOAD associated methylation was observed within the promoter CGI target of *MEF2C*. This is potentially surprising as aberrant methylation within the promoter regions of many genes have been associated with LOAD pathology. However, recent studies are highlighting the importance of methylation located outside of promoter regions in driving AD

risk. In addition aberrant methylation within the *MEF2C* gene body was found to be associated with LOAD pathology (Humphries *et al*., 2015).

For the second *MEF2C* region investigated, one sEOAD blood sample was found to be significantly hypomethylated when compared to the other blood samples tested, the methylation at the CpG site investigated within this sample was 22% less than the group wide methylation average. In addition, one sEOAD brain sample was also hypomethylated when compared to the other brain samples tested, this sample presented with 15% lower methylation when compared to the group wide average. However, this reduction did not prove to be statistically significant. It should also be noted that, for brain tissue, no significant difference in methylation was observed between sEOAD and control samples when the two groups were compared. This suggested that the significant hypomethylation observed might represent the presence of a rare epi-allele within the sEOAD individual tested which might act to increase their risk of developing sEOAD in a similar manner to a rare genetic polymorphism which affect genes expression. Epi-alleles represent rare epigenetic events that can occur within large populations driving substantial changes to individual phenotype (Agorio *et al*., 2017).

It could be suggested that the present of an epi-alle could be responsible for driving disease state by altering expression of pathological genes and proteins, the presence of this epigenetic modification could therefore have a large influence on an individual's chance of developing a given disease regardless of how common the variant is within a population.

MEF2C may have important functions which drive AD pathology. MEF2C deregulation due to aberrant methylation may influence the normal activity of MEF2C in memory formation, neuronal activity, synaptic function, inflammatory response and circadian rhythm. However, interestingly while disruption in circadian rhythms (CR) has been reported in AD, studies suggest that the severity of this symptom differs between AD patients (Weissova *et al*., 2016, Coogan *et al.*, 2013). MEF2 transcription factors have been shown to play an important role in regulation of CR, the drosphila ortholog Mef2 was shown to regulated circadian behaviour and neuronal modelling (Sivachenko *et al*., 2013).

A consistent alteration of CR in AD would be expected if this symptom of the disease was a common consequence of the disease pathology. However the differing severity of CR

disruption observed in AD patients might suggest that the causation may lie in patient specific aberrant methylation of *MEF2C*, therefore meaning disruption of CR in AD would vary between individuals depending on the methylation status of the *MEF2C* gene.

While interesting to hypothesise the actual impact of *MEF2C* methylation on AD pathology and specifically CR would need to be elucidated in further experiments investigating both sEOAD and LOAD samples. Further experiments would also be needed to determine if aberrant methylation is the cause of disrupted CR or whether it is further driving this LOAD symptom.

4.5.2 No significant difference in methylation was observed in sEOAD sample for the other targets

The pyrosequencing assays performed for *PTK2β, INPP5D* and *ABCA7* did not reveal any AD associated differential methylation in the regions investigated. Perhaps the most surprising was the lack of differential methylation identified in the *SIRT1* and *TREM2* targets as both were chosen due to previous studies identifying LOAD associated methylation within these regions (Hou *et al*, 2013, Smith *et al,* 2016).

The McrBC experiments described in chapter 3 investigating the same region of *INPP5D* indicated LOAD specific hypermethylation occurring in LOAD blood samples (P= 0.003). The difference in results could indicate that this hypermethylation is LOAD specific and does not occur in sEOAD samples. However it is important to note that the use of different methods might also explain the difference in results. Pyrosequencing is quantitative whereas McrBC assesses methylation across a large target region without allowing for generation of methylation data for individual CpG sites, also it is likely that a considerably higher samples size would be needed to confirm the McrBC results. Ideally LOAD samples would also have been used in the pyrosequencing assay to investigate if McrBC results could be reproduced using this other method but unfortunately LOAD samples were unavailable at this point in the study.

Interestingly other studies have also failed to find any significant difference in methylation of the *INPP5D* promoter in LOAD tissue (Yu *et al*., 2014b) and *INPP5D* was not identified as a gene which is differential expressed in the sEOAD brain using a transcriptomic study (Antonell

et al., 2013). Perhaps suggesting that despite its genetic association with LOAD and functional relevance in driving AD pathology, potentially in both LOAD and sEOAD, aberrant epigenetic regulation, at least at the regions investigated, is not driving this. Further study would be needed to confirm this.

Similarly no significant difference in methylation was observed in the targeted region of the *PTK2β* promoter between sEOAD samples and controls in either blood or brain tissue using pyrosequencing. This was also the instance following investigation into the *PTK2β* promoter using McrBC in LOAD blood (P=0.26). Possibly suggesting altered epigenetic regulation at the target region occurs neither in LOAD (blood) nor sEOAD (blood or brain). This result was surprising as *PTK2β* was shown to be over expressed in the sEOAD brain in a transcritomics study (Antonell *et al.*, 2013), perhaps indicating that a mechanism other than DNA methylation was diving this altered expression. To confirm a larger region of the *PTK2β* promoter, and perhaps other genic regions, would need to be investigated as methylation outside of the pyrosequencing target could also be responsible for altered gene expression.

The sample number used in both studies was modest and more samples would be needed to confirm this hypothesis. Further to this, studies have identified LOAD associated differential methylation within the *PTK2β* gene, however in these studies brain tissue was used (Humphries *et al*., 2015). Since the McrBC experiment showed no aberrant methylation in LOAD blood samples, it was interesting to look at *PTK2β* in both sEOAD blood and brain tissue. Humphries *et al*. (2015) identified LOAD associated differential methylation within the brain; however pyrosequencing did not result in the identification of a similar result in sEOAD brain tissue. This might indicate that the *PTK2B* aberrant methylation is LOAD specific. However it is important to note that a different region was investigated by Humphries *et al*. (2015) within the *PTK2β* locus and thus it might be that the LOAD associated methylation is also region specific, in support of this another study failed to identify any difference in methylation of *PTK2β* in LOAD (Yu *et al*., 2014b).

A region of the *ABCA7* gene was also shown to be differentially methylated in the LOAD brain by Humphries *et al*. (2015), this study also identified two other genes within the *ABCA7* locus which are differentially methylated in LOAD. In addition, the study identified LOAD associated increase in expression of the *ABCA7* locus and differential splicing. Other studies have also

reported increased *ABCA7* expression in LOAD patients, and have suggested that this over expression acts as a compensatory response to AD pathology, while AD associate genetic variants result in a reduction of expression (Shulman *et al*., 2013, Karch *et al*., 2012, Vasquez *et al*., 2013). Humphries *et al*. (2015) also suggest a hypothesis whereby the hypomethylation observed in *ABCA7* may account for the over expression of *ABCA7* protein observed in AD. However another study failed to identify any AD associated methylation in the *ABCA7* promoter (Kiyohiro *et al.,* 2017).

Pyrosequencing failed to identify any sEOAD associated methylation difference in the *ABCA7* region investigated in either blood or brain tissue. This again might suggest *ABCA7* is not regulated by DNA methylation in sEOAD as opposed to regulation in the LOAD brain. In support of this is that *ABCA7* was not identified as being differentially expressed in a transcriptomic study of the sEOAD brain (Antonell *et al*., 2013). However is should also be considered that Humphries *et al*., (2015) identified differential methylation in an intragenic CGI, therefore in order to make a successful comparison further pyrosequencing of this region would be required. Humphries *et al*. (2015) also used tissue from a different brain region, and since it is well established that tissue specific methylation occurs, it is possible that the methylation observed by Humphries *et al*. (2015) is specific to the brain region investigated.

Similarly to the three previous gene targets no AD associated methylation was observed at the CpG site investigated within the *TREM2* gene or the two CpG sites investigated in *SIRT1*. Both of these targets have been identified as harbouring LOAD specific methylation in other studies (Smith *et al*, 2016, Hou *et al,* 2013). However for *TREM2* Smith *et al*. (2016) identified LOAD specific hypermethylation at this site in LOAD brain tissue, possibly again suggesting important epigenetic alteration in LOAD but not in sEOAD. However it should be noted that again Smith *et al*. (2016) also used tissue from a different brain region in their investigations, therefore tissue specific methylation cannot be ruled out without further investigation. Importantly Smith *et al.* (2016) also included a larger sample size in their study and therefore were also able to control for co-variates such as age and gender, which could also explain differences between the two studies.

For the targets described it is also important to note that the sample size use was modest and ideally a larger study would have been conducted. However sEOAD samples are limited in

their availability and this is reflected by other studies investigating gene expression in this type of AD (Antonell *et al.,* 2013).

Importantly post hoc power calculations show that there is sufficient power in the *RIN3* experiment (e.g. CpG5 >0.8) (Table 4.3). However it is likely that for some genes tested the low sample size resulted in an underpowered experiment. It is therefore possible that for some genes, testing additional samples may result in the resolution of significant differences. It is also however possible that additional samples may also return p values of zero i.e. no variation between AD and control. It should be noted that as the hypothesis is that conditions within the brain of AD patients can impact on the epigenome of peripheral blood via hall mark characteristics of disease, power calculations may underestimate the likelihood of returning a statistically significant difference i.e. in instances where large changes in methylation result from dramatically different brain pH (see Ponto *et al*, 2014) or oxidative stress.

4.5.3 One *SIRT1* CpG site may be differentially methylated in LOAD blood

Pyrosequencing also revealed no significant difference in methylation at either CpG site investigated for *SIRT1*. However for this gene target it was possible to investigate a small number of LOAD blood samples. This revealed a small but significant difference (3.5%) in methylation of the second site investigated in LOAD blood. However, Hou *et al*. (2013) reported a large increase in methylation of approximately 30% in LOAD samples.

One reason for the difference in results may be due to the different techniques used. Hou *et al*. (2012) used bisulphite sequencing PCR. Initially PCR was used to amplify the region of the *SIRT1* gene using bisulphite treated DNA. The PCR product was then sub-cloned into the pMD18-T vector. A minimum of five clones were then sequenced to assess cytosine methylation. While cloning is a good method for gaining an idea about cytosine methylation it is time consuming and at least five or more clones are required to be analysed. To further validate their results MSP was used. This involved designing primers that annealed to methylated or unmethylated DNA after bisulphite treatment. The primers are either complimentary to a C which will be present in the methylated version of the CpG or to a T which will be present in the unmethylated product following bisulphite conversion.

These results were not quantitative, unlike pyrosequencing. This could mean that they may have missed small changes in methylation, similar to those that were observed using pyrosequencing. The use of only a restricted number of clones also might result in an inaccurate measure of DNA methylation at the sites investigated.

Another factor worth some consideration is that samples originating from different population demographics are used in each study and a very small sample was used for *SIRT1* pyrosequencing.

It is plausible that regional genetic variation may account for at least some differences in reported levels of methylation for disease related genes e.g. comparative analysis has identified differences in patterns of methylation between African-Americans and Caucasians at birth (Adkins *et al*, 2011) and that these differences may link to cancer risk in later life. Analysis of over 450,000 CpG sites in samples taken from whole blood of 573 individuals of diverse Hispanic origin identified that differential methylation between ethnic groups can at least partially be explained by the shared genetic ancestry (Galanter *et al*, 2017).

Hou *et al*. (2013) used DNA samples from 63 AD patients (AD: male, 31; female, 32; mean age, 80 ± 11 years), and 72 non-demented controls (male, 33; female, 39; mean age, 77 ± 15 years). While in this pyrosequencing study only eight have been used. In order to make any comment about the methylation status of this region using pyrosequencing a larger sample size would be needed.

Another possible reason for the difference in reported results may be that LOAD samples from different geographical regions were used. For example the samples used by Hou *et al*. (2013) were collect from individuals at Guangzhou Brain Hospital, Guangzhou Senior Hospital and Guangdong General Hospital all of which are located in Guangzhou in South China whereas the samples used in the current study where collected from patients in the UK. Therefore the two sample sets are likely to have encountered very different environments throughout their lives and also have very different diets; this may be a significant factor affecting results (Toyooka *et al*., 2003).

Other studies have failed to identify any differential DNA methylation within the *SIRT1* promoter using LOAD blood samples, supporting the *SIRT1* pyrosequencing results described in this chapter. Furuya *et al*. (2012) failed to observe any difference in *SIRT1* promoter methylation in LOAD. They investigated 22 CpGs within the *SIRT1* promoter using real time PCR and MALDI-TOF in brain and peripheral blood leukocyte DNA. They failed to see any change in methylation or gene expression. In agreement with this a second group also failed to find any disease associated differential methylation within the *SIRT1* promoter (Carboni *et al*., 2015).

However it should be noted that both of these studies did not look at methylation at the individual CpG level. Furuya *et al*. (2012) took an average methylation across the CpG sites investigated and Carboni *et al*. (2015) used methylation specific PCR. Therefore neither study would have been capable of highlighting a small but significant CpG site specific change in methylation.

The results obtained are interesting but in order to draw any conclusion many more samples of AD DNA will need to be pyrosequenced for LOAD suffers. It would also be interesting to investigate methylation of *SIRT1* within the LOAD brain, the samples used during this study were curated by the QMC and the AD consortium. Unfortunately, although they have given access to certain tissues and samples LOAD, cortex samples were requested, but unavailable.

4.6 Conclusions

Pyrosequencing failed to find any sEOAD associated methylation within *PTK2β, INPP5D, SIRT1, TREM2* and despite other studies showing differential methylation within the genes (Humphries *et al.*, 2015, Smith *et al*., 2016, Yu *et al*., 2014b, Hou *et al*., 2013) . In addition to this all of these genes have been functionally implicated in AD pathology through the identification of genetic variants and protein functionality. This makes it surprising that no differential methylation was identified.

One explanation for this result is that LOAD and sEOAD represent two distinct diseases which, although sharing pathological features, may have differing and distinct biological pathways driving disease pathology. Alternatively the biological and perhaps epigenetic alteration

driving differential expression of some genes, which results in AD pathology, might be different in the two types of Alzheimer's disease.

Alternatively, pyrosequencing may have failed to identify any sEOAD associated methylation due to the regions being investigated. In the case of *ABCA7* and *PTK2β* LOAD associated methylation was found in regions not covered in the pyrosequencing assays used in this study. Pyrosequencing is highly targeted and can therefore only cover very small regions of DNA due to the technical limitations of the assay, whereas Humphries *et al*. (2015) used a more global method which allowed identification of methylation differences across the whole *PTK2β* and *ABCA7* locus. Therefore it cannot be ruled out that the use of pyrosequencing may have resulted in important AD associated methylation either side of the targeted regions being missed.

In addition, pyrosequencing was used only to methylation profile CGIs upstream of the TSS of the genes investigated and within the promoter region. However epigenetic studies are identifying important LOAD associated methylation outside of promoter regions. Humphries *et al*. (2015) identified twenty differentially methylated CpG containing regions within the AD brain, of these twelve were located within 3'UTRs and regions were also identified within the gene body and 5'UTR, suggesting a requirement for investigation into genic regions other than the promoter CGIs in AD.

Pyrosequencing of *SIRT1* also did not reproduce the results of Hou *et al*. (2013), possible LOAD specific hypomethylation was identified in one of the CpG sites investigated. However, it is difficult to know if this result is accurate due to the limitations of the pyrosequencing platform. No significant sEOAD associated methylation was observed in either blood or brain tissue for the two *SIRT1* CpG sites investigated.

Again *SIRT1* pyrosequencing results might indicate that while LOAD and sEOAD share pathological features, each disease may have its own distinct epigenetic pattern. However further analysis of samples would be needed to confirm this.

In the case of *RIN3* further experiments will also be needed to elucidate the effect of the 3'UTR hypomethylation observed on both *RIN3* gene expression and sEOAD pathology. However, the results for both *RIN3* and *MEF2C* provide interesting insights into their potential

epigenetic deregulation in sEOAD. The aberrant methylation observed in *RIN3* likely represents a consequence of AD pathology as for this to be the case differential methylation across a large region would be expected presenting in many individuals. However in contrast the *MEF2C* epi-variant identified likely represents a driver of AD pathology, since it is specific to individual disease cases, rather than across a large number of patients.

The data presented in this chapter also highlights the importance of non-promoter methylation in causing, driving or response to AD pathology. It is also important that epigenetic variation was observed in sEOAD blood samples in this study. This allows for future testing of a much larger number of samples then may be possible if using brain tissue alone, allowing for future validation.

However an important fact to consider is that pyrosequencing using bisulphite converted DNA does not allow differentiation between methylation and hydroxymethylation at the sites investigated. This makes it impossible to determine whether the identified differentially methylated sites are in fact differentially methylated or hydroxymethylated in AD. Future work would be needed to resolve this. This could be accomplished by used OXBS-pyrosequencing (Stewart *et al.,* 2015).

Chapter 5: Whole Genome Bisulphite Sequencing of LOAD Cerebellum Samples

5.1 Introduction

As described previously in this thesis (section 1.8) many methods can and have been used to identify differential methylation in LOAD associated genes. Many of these studies use methods which cover small sections of DNA, such as pyrosequencing, to identify differential methylation within specific regions of single target genes or, in some cases multiple genes simultaneously. Mostly the promoter region is selected for investigation, since promoter methylation has been directly linked with gene expression (Mehler, 2008, Graff and Mansuy, 2008). It is becoming apparent that multiple biological pathways and numerous genes are subjected to epigenetic dysregulation in AD. Therefore, in order to fully understand the involvement of DNA methylation in the patheogenesis of AD a more global profile of differential DNA methylation is required (Qazi *et al*., 2017).

An obvious advantage of methods that allow analysis of methylation at smaller genic regions is low cost per assay. Conversely global and whole methylome sequencing is highly expensive. Pyrosequencing allows many samples to be tested, whereas global methods and certainly whole genome methods can be limited in use due to the large cost associated with each assay (Plongthongkum *et al*., 2014). However, the ability to assay many more samples using methods such as pyrosequencing, following global and whole genome investigations, provides an excellent opportunity to validate any regions identified as being differentially methylated (Plongthongkum *et al*., 2014).

Methylation assays covering large regions of the genome can be categorised into two groups, targeted and targeted assays. Targeted assays include methods such as the Illumina bead chip 450K array (Plongthongkum *et al*., 2014). Whereas targeted methods include whole genome bisulphite sequencing (WGBS); this method has the potential to simultaneously analyse methylation at all cytosine residues within the genome of the sample analysed.

The Illumina 450K bead chip assay has been used to assess methylation in multiple published methylation studies. This assay uses DNA oligomer probe which will bind to either methylated or unmethylated CpG sites. The probe either recognises a UC (unmethylated cytosine following bisulphite conversion) or a 5mC (methylated cytosine following bisulphite conversion) nucleotide. Following hybridization of the probes a fluorescent signal is used to represent methylation status at individual CpG sites (Pidsley *et al*., 2013).

This bead based assay covers approximately 480,000 CpG sites in the human genome, representing coverage of 99% of all reference sequence genes (NCBI). Approximately 96% of CGIs are covered. However, while this covers the majority of CGIs located within promoter regions, this assay also allows analysis of a small number of non-promoter CpG sites including some located within 3' and 5'UTRs and intra and intergenic regions (Slieker *et al*., 2013, Dedeurwaerder *et al*., 2011, Fan and Chi, 2016). However, the areas covered are focused in CGIs therefore the information is not generated for non-CGI CpG sites and the assay does not cover the whole methylome (Dedeurwaerder *et al*., 2011).

Illumina have however recently released the MethylationEPIC BeadChip 850K array which covers more the 850000 CpG sites in the human genome; while also cost effective, this assay also fails to provide whole methylome data, which consists of ~28 million CpGs (Lövkvist et al., 2016). The Illumina platform therefore represents ~3% of all possible sites that be differentially methylated.

Another method commonly used to assess DNA methylation is reduced representation bisulphite sequencing (RRBS). This method utilises bisulphite treatment of DNA followed by the use of specific methylation insensitive restriction enzymes to identify CpG methylation. The restriction enzymes *MspI* and *ApeKI* are commonly used (Plongthongkum *et al*., 2014). First the restriction enzymes are used to digest the genome resulting in fragments of various sizes. Fragments of a specific size (40-220bp) are then selected and sequenced following bisulphite conversion, allowing the identification of methylated cytosine residues (Meissner *et al*., 2008)

This method is restrictive as it only provides coverage of 10-20% of CpG sites within the human genome, depending on the restriction enzymes used. A problem associated with

the RRBS technique, and also others which use restriction enzyme digest, is that digestion may be incomplete, resulting in artefacts within the data set (Fan and Chi, 2016). In addition, this only allows analysis of CpG site dense regions such as CGIs, therefore important aberrant methylation in other regions will be missed (Plongthongkum *et al*., 2014, Fan and Chi, 2016). However, a major benefit to researchers is the much lower cost associated with this method making it a realistic approach for use in the analysis of multiple samples (Fan and Chi, 2016).

Unlike the previous two methods described, whole genome bisulphite sequencing has the capability of covering the entire methylome and therefore represents a truly whole genome approach. This is hugely advantageous as it allows for identification of aberrant methylation at non-CGI CpG sites as well as cytosine methylation at non CpG cytosine residues.

WGBS involves the production of a bisulphite converted DNA library prior to whole genome sequencing. Therefore unmethylated cytosines are converted to uracil and subsequently thymine, following PCR amplification of the library, whereas methylated cytosines remain as cytosine. Following sequencing the presence of a cytosine residue at a given location indicates methylation at that site. A percentage methylation at a given site is calculated based on sequencing coverage of the specific sites being analysed, sequencing coverage is a representation of the number of reads which are independent of each other but are aligned to a reference genome covering the site being investigated (Ji *et al*., 2014).

One technical issue with WGBS is that large amounts of input DNA are needed. This is because the DNA library is bisulphite converted following its preparation. This results in degradation of the library, possibly affecting sequencing results; hence a large initial DNA input is required. However, this problem has been addressed recently by the introduction of DNA bisulphite conversion prior to library preparation. Oligonucleotides specific to post conversion DNA are then used during library amplification (Plongthongkum *et al*., 2014).

As mentioned previously the major advantage of using WGBS is the ability to determine methylation across the entire methylome. WGBS is estimated to cover up to 90% of all

cytosines in the genome (Ji *et al*., 2014). This is extremely important as methylation outside of promoter regions is now being identified as having functional relevance for gene expression. Cytosine methylation within gene bodies has been shown to affect gene expression and alternative splicing is influenced by splice boundary methylation (Lister *et al*., 2009, Laurent *et al*., 2010). In addition, non-promoter CpG sites have been found to be differentially methylated in LOAD, and hypomethylation of the *RIN3* 3'UTR was identified using pyrosequencing during work completed as part of this thesis (Chapter 4) (Yu *et al*., 2014b, De Jager *et al*., 2014, Humphries *et al*., 2015, Smith *et al*., 2016).

WGBS also allows the identification of methylation at non-CpG cytosine residues, also known as CgH (H=A, C, T) sites. Methylation at these sites has been shown to be important in multiple cell types including: embryonic stem cells, brain tissue and oocytes (Lister *et al*., 2009, Shirane *et al*., 2013, Xie *et al*., 2012, Varley *et al*., 2013). However methylation of non-CpG cytosines is much rarer outside of the brain but is likely to still be functionally relevant (Patil *et al*., 2014).

Therefore, WGBS represents an excellent method for whole methylome analysis. Limited studies have utilised this method for identification of differentially methylated regions in LOAD. Most focus on use of the 450K array due to the more accessible cost of the assay. The work described in this chapter used WGBS to compare the methylomes of two AD patients and one control. Brain (cerebellum) tissue was used for WGBS from one sufferer of severe LOAD pathology and one suffer of moderate LOAD pathology, braak stages VI and IV. The results for these two samples were then compared to publicly available WGBS data for a control cerebellum sample. The intention is therefore to identify any differential methylation across the methylome for each stage of disease. The DML identified where then compared to WGBS data from a published LOAD sample in an attempt to validate the DML.

The cerebellum has to date not been considered relevant to the pathophysiology of Alzheimer's disease, with few reports of histopath samples staining for high levels of amyloid plaques or tau tangles in this region (reviewed by Jacobs et al, 2017). However recent data suggests that the cerebellum shows substantial changes in protein expression linked to AD, when compared to control tissue (Xu et al, 2019). This tissue is therefore an excellent proxy for AD induced effects in a 'toxic' neuronal environment,

rather than as tissue directly affected by damaging protein or oxidative stress related insults. It can be hypothesised that this tissue is therefore more representative of the likely epigenetic changes induced in peripheral blood cells.

5.2 Chapter aims

- To use whole genome bisulphite sequencing to map the methylome of two LOAD cerebellum samples representing two stages of disease severity.
- Compare the sequenced LOAD methylomes to published WGBS cerebellum non-disease control data to identify differentially methylated regions between controls and LOAD suffers.
- Compare the methylomes of the patient with severe LOAD to patient suffering with mild disease to potentially identify aberrant methylation occurring early in disease progression and later in the disease.
- Use a published LOAD sample to validate the DML identified.

5.3 Methods

5.3.1 Sample information

Samples were chosen from a database curated by ARUK. These samples were chosen at the outset of the experimental design process due to the additional data that could be obtained e.g. GWAS data, had wider epigenetic profiles suggested this would be necessary. DNA from cerebellum samples from two LOAD suffers were used for the WGBS described in this thesis chapter. Both samples were received from BDR (Brains for Dementia Research). Further details of these two samples are provided in table 5.1

As shown in the table the two samples were received from male donors so gender specific effects were excluded. However, the age at death, PM interval and additional pathology's cannot be excluded as possibly effecting results, however these two samples were the only samples available for use and since all differential methylation identified should be further validated, potentially by pyrosequencing analysis, these samples were used.

The WGBS control data used for comparison to these two samples was produced by Professor Kun Zhang of the Integrative Genomics Laboratory, Bioengineering, University of California, San Diego. This data was generated by pyrosequencing of cerebellum tissue obtained from a healthy 25 year old male. This data was chosen because it was the only WGBS data generated using a healthy male cerebellum sample available on the NCBI sequence read archive (SRA) (https://www.ncbi.nlm.nih.gov/sra). This sample was gender matched to the two LOAD samples used however again age could influence results, however since this was the only sample available and validation would be needed this data was used for analysis in this thesis. More details of this sample can be found in section 2.6.1.

Ideally, due to batch effects, data generated from a control sample sequenced at the same time as the LOAD samples would have been used however due to the

financial burden imposed by the WGBS method this was not possible and instead the published cerebellum control data was used.

Data generated by WGBS of a LOAD grey matter sample (female ages 81 years) was also used for comparison to the cerebellum WGBS data, with the intention of validating the DML identified. This data was generated as part of the study conducted by Sanchex-Mut *et al.* (2016), Epigenetics and Biology Program, Bellvitge Biomedical Research Institute, Barcelona, Spain. Data was obtained directly from the study authors.

Table 5.1: Details of cerebellum DNA samples used for WGBS. Table shows information about the two cerebellum samples used for WGBS during work described in this thesis chapter.

DNA_ Code	Gender	Age at death	PM_Delay	PM_Interval	Diagnosis	Braak staging	Pathology 1	Pathology 2
BK076	M	87	19	26	AD	IV	Alzheimer's disease, BNE stage IV	Aryrophilic grain disease,mild amyloid angiopathy, hippocampal sclerosis, mild small vessel disease
BK005	M	66	90	114	AD	VI	Alzheimer's Disease, Braak 6, BNE 6	TDP-43 pathology, limited to amygdala

5.3.2 Bisulphite conversion of DNA

Prior to the preparation of libraries for WGBS both LOAD DNA samples were subjected to bisulphite conversion. Details of this procedure are described in both sections 2.4.1 and 2.6.2. Briefly as previously described the EpiTect Bisulphite Conversion Kit (Qiagen Germany) was used for bisulphite conversion

of DNA and manufactures instructions were used. Clean up of the converted DNA is further described in section 2.4.2, this was also conducted following manufacture's instructions.

5.3.3 Library prep

Following bisulphite conversion of the DNA WGBS libraries were produced. This process is described in more detail in section 2.6.3. The Truseq DNA methylation library preparation it (Illumina) was used, manufacture's instructions were followed.

Briefly, 50-100ng of input DNA was used, following bisulphite conversion the DNA was single stranded, sequencing primer was added to this and random primed DNA synthesis occurred (section 2.6.3.2). Following this tags were added to the 3' and 5' ends of the synthesised DNA (section 2.6.3.3) and this now tagged DNA was cleaned up (see section 2.6.3.4) prior to amplification of the library. To amplify the library PCR was used (details in section 2.6.3.5), during this step adapter sequences were also added. Lastly the library was again clean-up, as described in section 2.6.3.6.

5.3.4 Pre-sequencing analysis

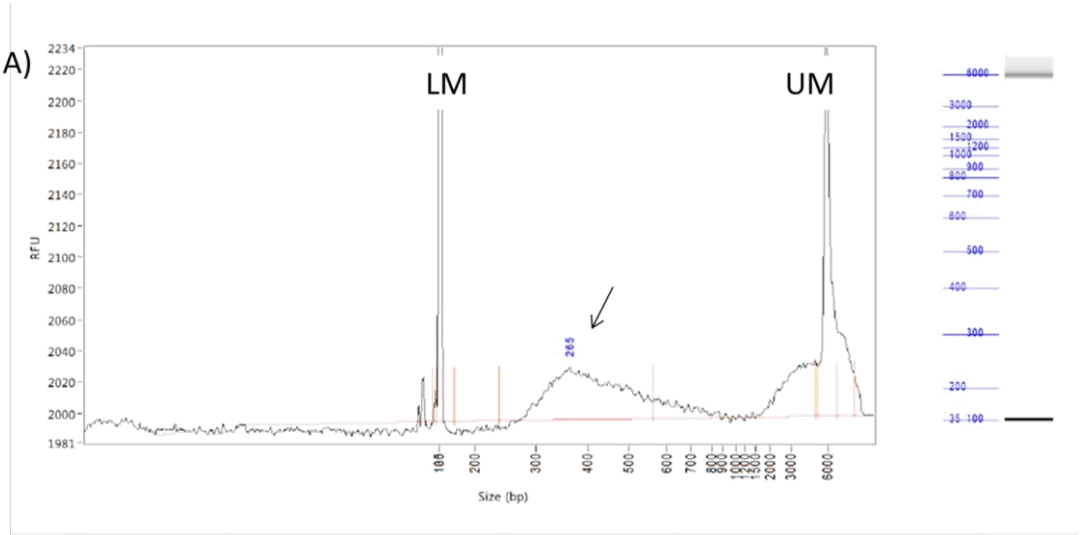
Prior to sequencing quality control (QC) assessment was performed on each preparation to ascertain the quantity and quality of the library. Firstly DNA concentration was calculated using a Qubit dsDNA high sensitivity assay and average fragment size was calculated using a Agilent 2200 Tape Station (QC was carried out by Polar Genomics, New York).

For both libraries fragments of lengths ranging from around 100-500bp were generated, with average fragment size being 288bp and 306bp for samples one and two respectively (details shown in figure 5.1). An expected fragment size following WGBS library prep with the TruSeq DNA methylation kit was therefore an average of 260-380bp. Following the removal of adapter sequences, this

resulted in a DNA insert of approximately 160bp for use in paired end sequencing. Therefore it was possible to conclude that the libraries contained fragments of a suitable size for subsequent sequencing.

However, the DNA concentration of both libraries were low, with a reading of 0.8ng/µl for sample one and 0.2ng/µl for sample two. The ideal DNA library concentration for use in sequencing would be at least 1ng/µl. Therefore, both libraries fell below this limit. A 20% PhiX spike in was therefore added to each library prior to sequencing to increase the diversity of the library. This increased cluster generation, therefore improving sequencing quality.

A)



B)

Sample	From (bp)	To (bp)	Average (bp)	Concentration ng/µl
Sample 1	140	466	305	0.8
Sample 2	175	441	288	0.2

Figure 5.1: Summary of QC analysis for each library. A) Representative bioanalyzer trace for sample one upper marker (UM) and lower marker (LM) are shown as well as the main sample peak, indicated by an arrow. B) Table showing a summary of QC information for both samples including average fragment sizes within each library and the DNA concentration of each library.

5.3.5 Sequencing of the WGBS libraries

Following QC assessment both WGBS libraries were sequenced using paired end 2X 150 BP sequencing. To allow adequate coverage for subsequent analysis each

library was loaded onto, and ran on, one lane of a Illumina HiSeq 2500 machine, sequencing was carried out by Polar Genomics, New York.

For samples one and two 3.2 and 2.1 million reads were generated by sequencing respectively, table 5.2. This resulted in an estimated 16X coverage for sample one and 10X for sample two. However for sample two a large amount of duplicate reads were present in the sequencing results, therefore a significant number of reads were removed during processing of the data. This level of duplication is likely due to the low concentration of the library. Ideally for human analysis coverage of 10-30X, or greater, is required. However recent studies have shown that 5-15X can produce satisfactory results when used to identify differentially methylated regions (Ziller *et al*., 2015). In addition a pool of input DNA was used to generate the libraries therefore it is likely that biological randomness was maintained within the sample and duplication was caused by PCR artefacts. Ideally the library preparation would have been repeated and optimised, however the library assessment was sent out of house (conducted by Polar Genomics, New York), therefore the turn-around time was extensive and limited the possibility of optimisation. Given more time several steps of the library prep protocol could have been further optimised, this would include: DNA input amount and the library amplification PCR. Therefore the sequencing reads were used for bioinformatic analysis.

For the control sample, data from three WGBS runs were available for use however two of the runs resulted in very few reads following trimming therefore the data for only the run producing the most reads was used in bioinformatic analysis. This run produced $4.6x10^7$ sequencing reads resulting in approximately 2.3X coverage. This coverage was low but since this data represented the most suitable control data available it was used for analysis. Unfortunately no other WGBS data for a cerebellum sample was available for use.

For both LOAD and the control sample the sequencing data was mapped to the human reference genome. For sample one and the control, 72% of the sequencing reads mapped to the genome. However, for sample two it was only possible to

map 46% of the reads to the genome. This could have been due to the low DNA concentration of the library and the large number of duplicates present.

Table 5.2: Sequencing reads generated for the control, sample one and sample two. Table shows the WGBS sequencing reads produces for all three samples used during work described in this chapter.

	Raw		Trimmed		Aligned		% aligned	Deduplicated		% deduplicated
Sample	Reads	Pairs	Reads	Pairs	Reads	Pairs		Reads	Pairs	
1	3.25E+08	1.62E+08	8.06E+07	1.61E+08	2.31E+08	1.16E+08	71.7%	1.53E+08	7.63E+07	65.9%
2	2.05E+08	1.03E+08	5.14E+07	1.03E+08	9.47E+07	4.74E+07	46.0%	1.14E+07	5.68E+06	12.0%
Control	4.65E+07	2.32E+07	1.16E+07	2.31E+07	3.32E+07	1.66E+07	71.6%	3.30E+07	1.65E+07	99.4%

5.3.6 Identification of differentially methylated loci

In order to identify differentially methylated loci (DML) three WGBS samples were used. This included the two AD samples sequenced as described above and the third published control sample.

In order to elucidate differentially methylated sites between these three samples initially all loci containing greater than 50% methylation were identified in each set of WGBS data. For all samples the data was filtered so that only loci containing greater than 10X coverage were identified.

In order to identify DML (differentially methylated loci) the “DSS-single” tool was used initially with a P-value threshold of 0.01 (performed by Dr Joanna Moreton (ADAC the University of Nottingham, method described in section 2.6.5)), the DSS-single from Bioconductor page DSS used estimated normalized dispersion of the data across the samples, it then converted the data points into a multivariate model. The DML’s were then called using a walds test, for this test likelyhood

ration was used to test the joint significance levels of coefficients simultaneously, this resulted p-value thresholds despite the fact that N=1. Three comparisons were made between the data sets. Sample one was compared with the control data, sample two was compared with the control data and lastly sample one data was compared to the data generated for sample two. DML were scored when a loci was found to harbour >50% methylation in one sample but not the sample to which it was compared.

The gene closest to the DML was identified and used for pathway analysis. This was completed using a literature search.

An effort was made to track the DML identified to enhancer regions. In this instance enhancer regions were defined as being located 2kb-5kb of a gene. As no unifying feature of a enhancer regions exist to allow them to be identified conclusively this area was used (Pennacchio *et al.,* 2013). No DML indentified were found to be located within enhancer regions.

The identified DML were also investigated in a further LOAD sample. The WGBS data for this sample was generated by Sanchez *et al.,* (2016) and was obtained from https://www.ncbi.nlm.nih.gov/sra/SRX534204[accn]. This sample was a grey matter sample taken from an 81 year old female; further details can be found in appendix 1. To compare this sample to the cerebellum WGBS data analysis identical to that described above was used.

5.4 Results

5.4.1 Identification of differentially methylated loci

5.4.1.1 Identification of differentially methylated loci in LOAD samples compared to controls

Both LOAD samples were compared to the control data to identify any differential methylation occurring in either stage of the disease. In the earlier stages of the

disease (represented by sample one) it would be plausible that less DML might be observable when compared to the more severe LOAD stage. The hypothesis is that any differential methylation that occurs early in disease progression is responsible for driving the disease and therefore might occur in very specific gene loci, whereas differential methylation occurring later on in disease might also occur in response to disease pathology, therefore resulting in a wider spread of aberrant methylation.

31 DML were indentified in sample one when compared to the control and 3 DML in sample two when compared to the control. The DMLs identified in sample two are shown in table 5.3.

Two of the DML identified within sample two failed to associate with any known gene. Not all genes have been assigned function or may be an intergenic region, suggesting a regulator component - yet to be determined. In addition, the third DML associated was located on the Y chromosome and encodes the pseudogene, C-Terminal Binding Protein 2. Interestingly, a recent study identified this gene as being expressed during puberty with a reduction of expression during later adulthood (Shi *et al*., 2016). Since the control sample data came from a much younger subject it might be plausible that the differential methylation observed in the severe LOAD sample represents an methylation change associated with aging rather than AD.

Table 5.3: Three DML identified in LOAD sample two when compared to the control. Table shows details of the three regions identified as being differentially methylated between sample two and the control sample

Chromosome	cytosine location	Cytosine dinucleotide site	CGI located	Percentage difference in methylation	AD Hypo/hyper	gene	Location	Gene
GL000224.1	17677	CG	N	32.027	hypo	.	Intergenic	N/A
GL000224.1	17699	CG	N	32.027	hypo	.	Intergenic	N/A
Y	56690113	CG	N	22.8583	hypo	CTBP2P1	Downstream	Encodes- C-Terminal Binding Protein 2 Pseudogene 1.

The 31 DML were identified when sample one was compared to the control data are shown in table 5.4. This included *CH17-333M13.2* which is a pseudogene of unknown function. Interestingly differentially methylated regions were identified both upstream and downstream of this gene. Other regions identified associated with two other pseudogenes, one uncharacterised and one encoding the kinase suppressor of ras 1 pseudogene 1 which has been associated with lncRNA. The genes *U6* and *CTBP2P1* were also found to harbour DML. *CTBP2P1* was also identified as associating with DML indentified when sample two was compared to the control, and as with sample two, therefore it seems plausible that the aberrant methylation observed in sample one is also age related.

Table 5.4: 35 DML identified in LOAD sample one when compared to the control. Table showing the differentially methylated regions identified by comparing the control sample to sample one, information about the nearest gene and its function are also shown.

Gene name	Gene location	Number of DMLs identified	Cytosine Chromosome location	Location of DMLs	Hypo/ Hyper (in sample one)	Gene function
CH17-333M13.2	Chr1: 143343180-143343275	9	143203799-143252186	Upstream	hypo	Unprocessed pseudogene
KSR1P1	Chr10: 42149310-42149549	1	40634300	Downstream	Hyper	Kinase Suppressor Of Ras 1 Pseudogene 1: affiliated with the lncRNA class.
CH17-342O10.1	Chr16: 34353616-34371659	13	34571940-34571969	Downstream	Hyper	Uncharacterised pseudogene.
Unknown	KI270304.1	1	930	Intergenic	Hyper	N/A
Unknown	KI270709.1	4	8482-8625	Intergenic	Hyper	N/A
U6	KI270744.1- 51009-51114	2	112020-112024	Downstream	Hypo	Encodes U6 Small Nuclear 1, which is a RNA gene, affiliated with the snRNA class.
CTBP2P1	ChrY: 56855244-56855488	1	56690113	Upstream	hypo	Encodes- C-Terminal Binding Protein 2 Pseudogene 1.

5.4.1.2 Identification of differentially methylated loci between LOAD samples.

319 DML were identified in sample two after bioinformatic comparison to sample one. The DML were found to be located on several chromosomes, these included chromosomes 1, 4, 6, 9, 19, 22, 8, 11, 17, 19 and KI270733.1 and the mitochondrial chromosome (see figure 5.2A).

The number of DMLs occurring at each type of cytosine dinucleotide was also calculated as a percentage, figure 5.2B. The highest number of DML were identified within CC and CA dinuceotides, representing 26.9% and 34.7% of DML respectively and 23.5% of differentially methylated cytosines were identified in CG residues. This result suggests importance of non-CpG site cytosine methylation in severe LOAD. The majority of the DMLs identified where also found to be hypo methylated in sample two when compared to sample 1 (see figure 5.2).

The location of the DMLs identified in relation to its closest associated gene was also calculated. 50% of DML were located downstream of the nearest associated gene while 38% and 12% were located intragenically and upstream respectively (see figure 5.2D).

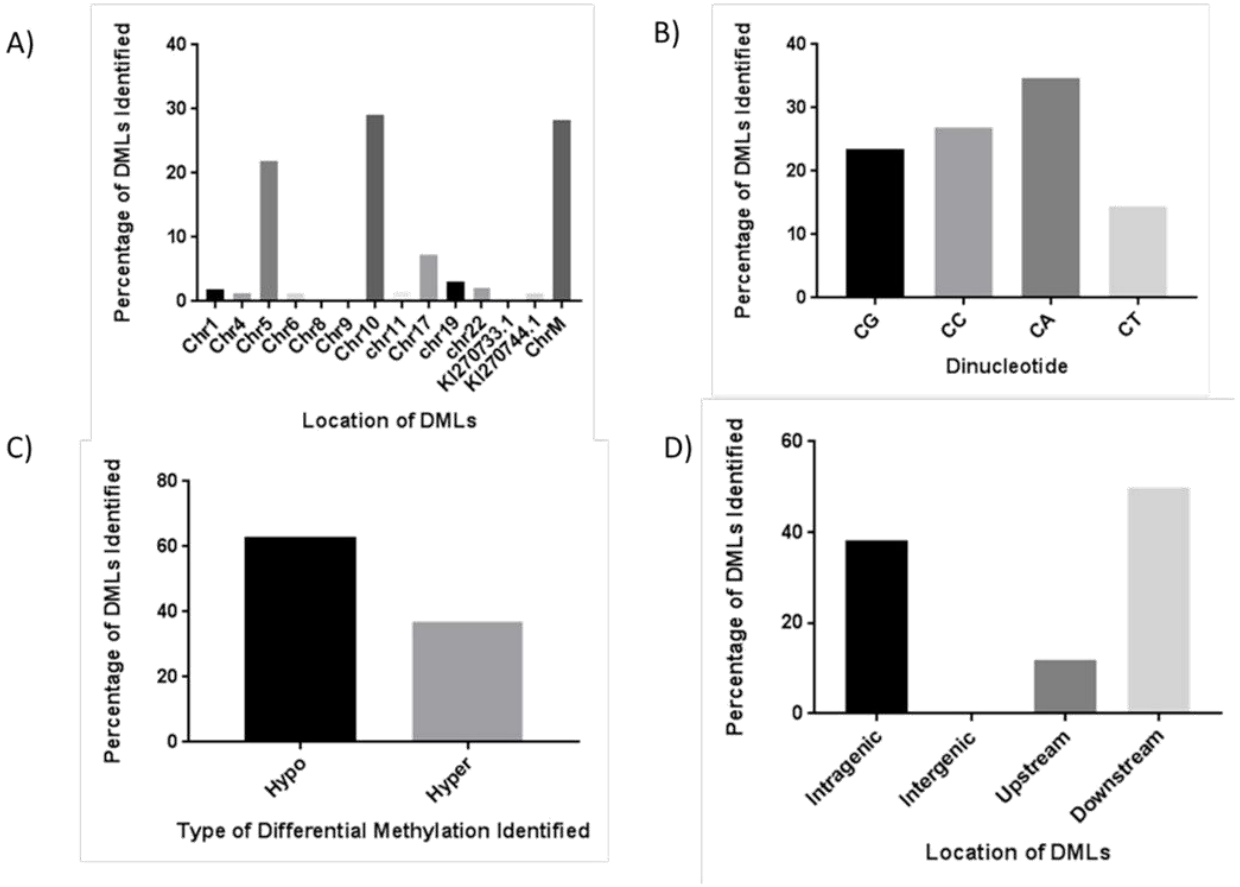


Figure 5.2: Results for sample one vs sample two. A) Chromosomal location of DMLs. B) Graph represents the percentage of DML that were identified at each cytosine dinucleotide. C) Percentage of DML found to be hyper and hypomethylated in sample two compared with sample one is shown. D) DML location within the gene associated.

In total twenty two genes were found to be associated with differentially methylated cytosines present in sample two when compared to sample one. This suggested that significant alteration in methylation occurs in specific genes in severe LOAD which are not occurring earlier in the disease.

The genes identified as harbouring differential methylation are detailed in table 5.5. The table provides information about the gene name and the number of DML associated with the gene. The chromosomal location of both the gene and the differentially methylated region is also provided. When only one cytosine site was found to be associated with a gene only the chromosomal location of this site is provided. The number of DML found within each differentially methylated region is also shown, as is whether the DML were hypo or hypermethylated in sample two when compared to sample one. The genic location of differential methylation is also provided. When multiple differentially methylated regions or sites associated with one gene the regions are separated based on whether they were found to be hypo or hypermethylated and are labelled A-E in order of ascending chromosome location.

Table 5.5: DMRs identified in sample two when compared to sample one.
Table shows the locations of DMR identifies when comparing sample two and sample one, details of the type of differential methylation and the number of DML is also included in the table.

Gene Name	DMR location on chromosome	No of DML	Hypo/hyper	DML location
AV1	201690539	1	Hyper	Intragenic
IPO9-AS1	201690539	1	Hyper	Intragenic
RP11-503C24.3/ MILLT4	168237225-168237253	2	Hyper	Intragenic
SORCS2	7363760-	4	hyper	Intragenic
	7363953			
IRF4	A) 397060, B) 396904	2	hypo	Intragenic
XX-FYM637E10_5.1	A) 131851030-131851042	A) 7,	A) hyper	A) Downstream
	B) 131851044-131851063	B) 10,	B) hypo	B) Downstream
	C) 131851064	C) 1,	C) hyper	C) Downstream
	D) 131851066-131851190	D) 65,	D) hypo	D) Downstream
	E) 131851192-131851199	E) 5.	E) hyper	E) Downstream
HCN2	A) 602114-602120	A) 2	A) hyper	A) Intragenic
	B) 602147	B) 1	B) hypo	B) Intragenic
	C) 602153-602175	C) 2	C) hyper	C) Intragenic
	D) 602181-602313	D) 5	D) hypo	D) Intragenic
TBC1D22A	47170899	1	Hyper	Intragenic
U6	11893 11917-12129	1 3	A) Hyper B) hypo	Upstream Upstream
MT-ND5	A) 13805-14008, B) 26- 122	A) 67, B) 23	A) hypo, B) hyper	A) Intragenic B) Upstream
CR1L	143254463-207732132	4	Hypo	Intargenic
CH17-333M13.2	143254463	1	Hypo	Upstream
IRX1	A) 3785006-3785179 B) 3785174 C) 3785176-3785230	A) 58 B)1 C) 8	A) Hyper B) Hypo C) Hyper	A) Downstream B) Downstream C) Downstream
UNC5A	176872515-176872550	2	Hyper	Intragenic
MIR1302-7	141874192	1	Hypo	Downstream
UBAC1	135981558	1	hyper	Downstream
ADGRA1	A) 133058015-133058045 B) 133058077-133058141 C) 133058173	A) 2 B) 2 C) 1	A) Hyper B) Hypo C) Hyper	A) Upstream B) Upstream C) Upstream
TRIM29	133058173-120085757	6	Hyper	Upstream
SEC14L1	77215944	1	Hyper	Intragenic
AC139099.6	A) 7715944-83215850 B) 83215859-83215882 C) 83215892-83215897 D) 83215906-83116000	A) 4 B) 3 C) 2 D) 13	A) Hyper B) Hypo C) Hyper D) Hypo	A) Intragenic B) Intragenic C) Intragenic D) Intragenic
pRNA	152162	1	hypo	Upstream

5.4.2 Identification of differential methylation occurring throughout disease progression

Two of the genes associated with DMLs were found to associate with differential methylation in sample one (moderate AD), when compared to the control and also sample two (severe AD) when compared to sample one (moderate AD), suggesting that epigenetic regulation is reprogrammed in these genes during disease progression. Epigenetic reprogramming could occur as a response to worsening AD pathology, for example an increase in Aβ will result in oxidative stress which will change the methylation status of the genes involved in oxidative stress response, also interestingly Aβ has been shown to drive its own production this way (Chen *et al.,* 2009). The two genes identified were *CH17-333M13.2* and *U6*.

Comparison of LOAD sample one to the control resulted in the identification of hypomethylation (<50% methylation) of nine DML upstream of the gene *CH17-333M13.2*. A further DML was also identified as being hypermethylated in sample two (>50% methylation) when compared to sample one. The methylation at each of these identified DML in each sample are described in Figure 5.3A and shown in figure 5.3B. As shown the nine upstream DMLs are hypermethylated in control but become hypomethylated in AD. The other identified DML, located closer to the gene, was hypermethylated in both the control and sample one (moderate LOAD sample) but became hypomethylated in sample two (severe LOAD). This might indicated a pattern of epigenetic regulation of the gene that results in its gradual hypomethylation, and perhaps increased expression, throughout disease progression.

In the case of the gene *U6* analysis revealed hypomethylation (<50%) of two DMLs located downstream of the gene in sample one when compared to the control. Hypermethylation of one further DMLs, and hypomethylation of three DMLs, were identified in sample two when compared to sample one, however, these were located upstream of the gene rather than downstream. Figure 5.3A describes the methylation at each DML and figure 5.3C shows the methylation at each identified DML in each LOAD sample and the control. The two

downstream DMLs were hypomethylated in both LOAD samples when compared to the control. However, the four upstream DMLs harbour the same levels of methylation in the control and sample one, and hypermethylation of the most upstream DML and hypomethylation the three remaining DMLs was only associated with the sample two and therefore severe LOAD.

A)

Gene	Gene location	No of DMLs identified	DMR Location	Location of DMLs	Hypo/ hyper
CH17-333M13.2	Chr1: 143343180-143343275	9	143203799-143252186	Upstream	Hypo
U6	KI270744.1-51009-51114	2	112020-112024.	Downstream	Hypo

B)

Gene	Gene location	No of DMLs	DMR location	Location of DMLs	Hypo/ hyper
CH17-333M13.2	Chr1: 143343180-143343275	1	143254463	Upstream	Hypo
U6	KI270744.1: 51009-51114	1	11893.	Upstream	Hyper
	KI270744.1: 51009-51114	3	11917-12129	Upstream	Hypo

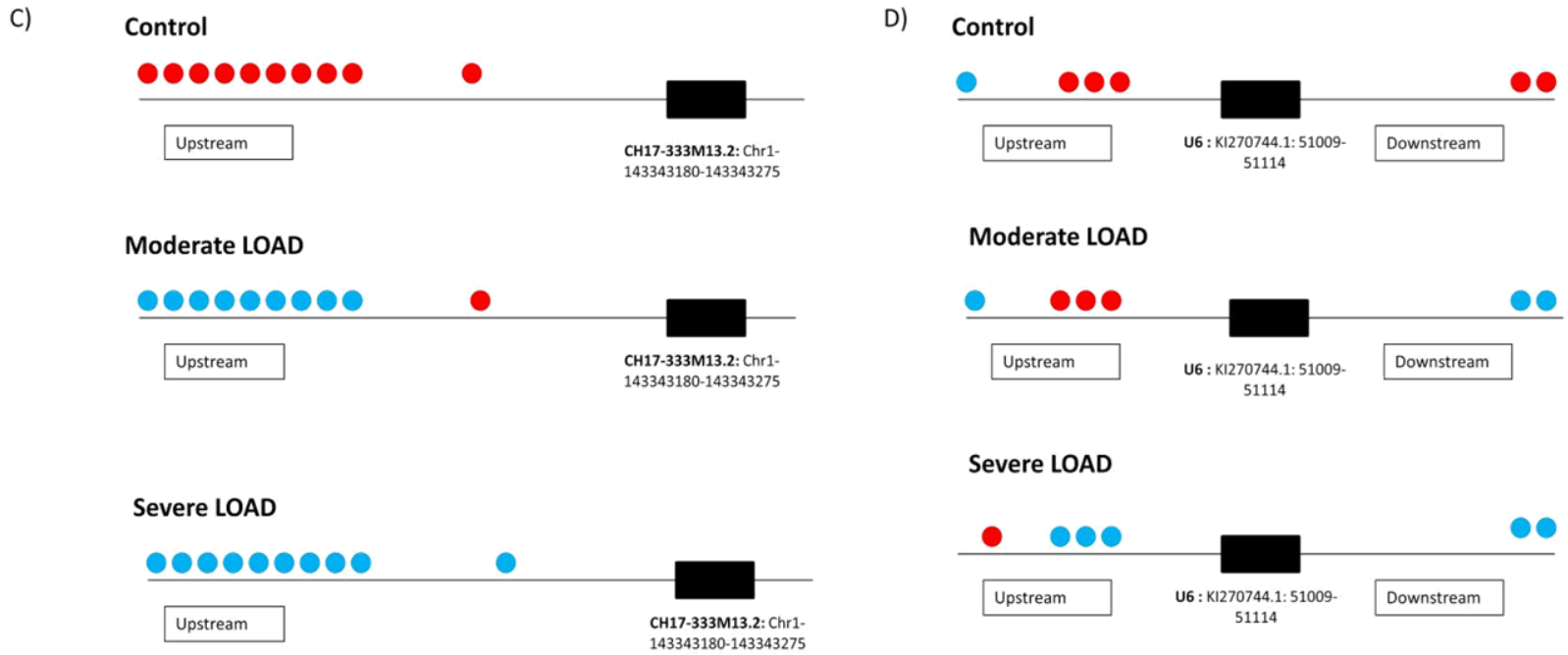


Figure 5.3: DML identified in *CH17-333M13.2* and *U6* in LOAD and control samples. A) Table showing DML identified in sample one compared to control in *CH17-333M13.2* and *U6*. B) Table showing DML identified in sample two compared to sample one in the genes *CH17-333M13.2* and *U6*. C) Imaging showing hyper and hypomethylation present at each DML in the control and LOAD samples for the gene *CH17-333M13.2*. D) Showing hyper and hypomethylation present at each DML in the control and LOAD samples for the gene *U6*. Genes are represented as a black box. Hypomethylation (<50%) is shown as a blue circle and hypermethylation (>50%) is shown as a red circle.

5.4.3 Genes associated with differential methylation in severe LOAD

In total twenty two genes were identified as being associated with differentially methylated regions or single loci in the sample representing severe LOAD. It is therefore plausible that aberrant methylation of these genes occurs in the later stages of LOAD. The functions of the genes identified are provided in table 5.6.

Table 5.6: Twenty two genes were identified as being associated with DML in severe LOAD. Tables shows details of the genes associated with the DML found when sample two was compared to sample one; the function of these genes is described.

Gene Name	Gene function
NAV1	Encodes Neuron Navigator 1- role in neuronal development and regeneration.
IPO9-AS1	IPO9 Antisense RNA 1) is an RNA Gene, and is affiliated with the non-coding RNA class.
CR1L	Complement C3b/C4b Receptor 1 Like- CR1 is a paralogue of this gene. Role in the immune response specifically in response to microbes.
CH17-333M13.2	Unprocessed pseudogene.
SORCS2	Sortilin Related VPS10 Domain Containing Receptor 2- strongly expressed in the central nervous system and important for its normal function. Role in protein transport and intracellular and intercellular signalling. Potential role in intracellular APP processing (Reitz et al., 2013).
IRX1	Iroquois homeobox protein 1- role during pattern formation during development, are particularity important in neural development.
UNC5A	Is part of the family of netrin receptors which mediated the movement of axons away from netrin-1 resulting in cell migration, particularly during development. Also have a role in cell death (Hashimoto et al., 2016).
IRF4	Encoding Interferon Regulatory Factor 4- functions as a transcription factor that regulates the interferon response to viral infection and activation of microglia and macrophages. Therefore is essential in innate and adaptive immune system function (Mamun and Liu, 2017).
RP11-503C24.3/ MILLT4	Encodes lincRNA.
MIR1302-7	MicroRNA gene encoding Has-Mir-1302-7. MicroRNAs are involved in post translational gene expression
UBAC1	Ubiquitin-associated domain-containing protein 1. Function in the ubiquitination and targeting of toxic proteins, including APP, for proteolysis (Chung et al., 2001).

XX-FYM637E10_5.1	Encodes an uncategorized gene, and is affiliated with the lncRNA class.

ADGRA1	Adhesion G protein-coupled receptor a1 (also known as G protein-coupled receptor 123)- Protein is part of the family of G-protein- coupled receptors which functions in regulation of immune response, sensory systems and development; can also influence APP processing through alteration of BACE expression.
TRIM29	Tripartite Motif Containing 29- potential role as a transcriptional regulatory factor. TRIM family protein also have role in regulating innate immune response. Other TRIM family member proteins have been associated with AD.
SEC14L1	SEC14-like protein 1- protein has a role in innate immune function in response to viruses. Also may regulated cholinergic receptors. Another role has been suggested in intracellular transport.
AC139099.6	Gene encodes LincRNA.
HCN2	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated ion channel 2. Functions in regulating pacemaker activity in the heart and brain. Associated with γ-secreatse and regulated APP processing.
TBC1D22A	TBC1 domain family member 22A. Functions in GTPase-activation of Rab family protein(s).
pRNA	Encodes a miscellaneous non-coding RNA which doesn't not fit into any know RNA category.
U6	U6 Small Nuclear 1) which is a RNA Gene, affiliated with the snRNA class. Is a component of the spliceosome.
MT-ND5	Encodes the mitochondrial gene NADH dehydrogenase core subunit 5. Functions in respiration.

Functional analysis of the nearest genes to the DML identified was completed to identify any biological pathways that might have been epigenetically altered in AD. Table 5.7 shows pathways that were identified as having had more the three genes harbouring DML. Only pathways containing three or more DMLs containing genes are shown, however other pathways identified are discussed in section 5.6.3

Table 5.7: Functional analysis of DML associated genes. The biological functions of the genes associated with DML are shown.

Biological function	Gene associated	Number of genes identified with this function
Immune response	ADGRA1, TRIM29, SEC14L1	3
CNS development or function	NAV1, SORCS2, HCN2 TBC1D22A	6
Regulatory RNA	IPO-AS1, RP11503C24.3/ MILLT4, MIR1302-7, FYM637E10_5.1, AC139099.6, U6	5
Amyliod processing	SORCS2, UBAC1, ADGRA1, HCN2	4

5.4.4 Validation of DML using a published LOAD WGBS data set

The DML identified were also investigated using an independent WGBS data set. A whole methylome published WGBS data set kindly provided by the authors of Sanchez-Mut *et al*., (2016) and was obtained from a gray matter of the dorsolateral prefrontal taken from an 81 year old female. This published data set was obtained from a different region of the brain compared to the samples used for the WGBS work described in this thesis. This provides validation for the sites identified in the cerebellum as differentially methylated in different regions of the brain.

Importantly differentially methylated regions were identified in this new LOAD sample associated with the genes *CH17-333M13.2, pRNA* and *U6*. These three genes were identified as harbouring differential methylation in the LOAD cerebellum samples (table 5.8).

In the case of *CH17-333M13.2* 9 DML were identified when sample one was compared to control and 1 DML was identified when sample one and sample two were compared. As described in section 5.5.2 the hypothesis is that this gene is hypomethylated in both moderate and severe LOAD and addition hypomethylation was observed at one CpG site in severe LOAD. Comparison

of the Sanchez- Mut *et al*., (2016) to the control sample resulted in identification of LOAD associated hypomethylation at this entire region and additional DML were found to be located upstream of the previously identified *CH17-333M13.2* 9 DML containing DMR.

The cerebellum WGBS also suggested that the gene *U6* may be differentially methylated throughout LOAD disease progression (section 5.5.2). Similarly to *CH17-333M13.2* the grey matter sample also showed a similar differential methylation to the severe LOAD cerebellum sample. In this case upstream hypomethylation was identified in both LOAD samples. The Sanchez *et al.,* (2016) sample was also shown to harbour DML that associated with the *pRNA* gene. This gene was also found to be hypomethylated upstream in the cerebellum of the severe LOAD suffer but not the moderate sufferer.

The similarity between the DMR in the grey matter sample and the severe LOAD suffers cerebellum sample could suggest that this hypomethylation occurs later in disease progression and is thus apparent in the grey matter, which is affected by LOAD pathology and in the cerebellum in later stage disease.

Table 5.8: Validation of the WGBS results using a published sample found DMR associated with three genes identified by the WGBS experiment. Table shows details of the genes identified as associating with differential methylation in the LOAD cerebellum samples and were also identified when the WGBS data from Sanch-Mut *et al.,* (2016) was compared to the control data.

Sample comparison made	**Gene Associated with DMR**	**Number of DMLs identified**	**Location of DMLs**	**Hypo/ hyper methylation**
1 vs control	CH17-333M13.2	9	Upstream	Hypo
Sample 1 vs sample 2	CH17-333M13.2	1	Upstream	Hypo
Sanchez Mut *et al.,* (2016) vs control	CH17-333M13.2	34	Upstream	Hypo
Sample 1 vs sample 2	pRNA	1	Upstream	Hypo
Sanchez Mut *et al.,* (2016) vs control	pRNA	1	Upstream	Hypo
Sample 1 vs control	U6	2	Downstream	Hyper
Sample 1 vs sample 2	U6	4	Both upstream	Hyper and hypo
Sanchez Mut *et al.,* (2016) vs control	U6	2	Upstream	Hypo

5.5 Discussion

5.5.1 Comparison of LOAD and control samples

The primary aim of the WGBS experiment was to identify any differential methylation that occurs throughout the progression of LOAD; the objective being to compare both the moderate and severe LOAD sample sequencing results to both a control data and to each other. The hypothesis was that this would result in the identification of DML and regions present at each stage of LOAD. However, unfortunately when the control and LOAD sample data was compared very few DML were identified in sample one and sample two. This was surprising as a large amount of DMLs were identified when the sample one data was compared to sample two.

It is likely that the discrepancy was caused by the inadequate genomic coverage obtained for the control data. The data generated for sample one, two and the control mapped to 71.7%, 46% and 71.6% of the genome respectively meaning that a proportion of the genome was not covered for each sample. In addition only loci containing greater then 10X coverage was used in the analysis. Therefore if a locus was identified as being differentially methylated in sample two when compared to sample one was covered >10 times in these two sample but <10 times in the control the loci would not have been included in the LOAD control comparison analysis, and given the low coverage provided by the control sample it is unlikely that many regions were covered adequately for analysis.

5.5.2 Distribution and location of DMLs throughout the genome

For the identification of DMLs a P-vale of 0.01 was used. This resulted in the identification of three DMLs in sample two when compared to the control and thirty-five DMLs in sample one when compared to the control. In addition, 319 DMLs were identified when sample two was compared to sample one.

Significant intragenic and downstream cytosine methylation was identified in sample two (severe AD) when compared to sample one (moderate AD). In

addition the majority of this methylation represented hypomethylation within the severe LOAD sample, compared to hypermethylation in the moderate LOAD sample. Also around three quarters of the DMLs were located within non-CpG cytosine residues. Only a quarter of the identified differential methylation was located within CpG dinucleotides and of this only two of the identified differentially methylated CpG were located in CGIs, (these were associated with the genes *HNC2* CGI 37 (intragenic) and *UBAC1* CGI58 (upstream)).

Non-CG methylation has been identified in both human and mouse brain samples, and has been shown to represent up to a quarter of neuronal cytosine methylation (Kinde *et al*., 2015, Guo *et al*., 2014, Sanchez-Mut *et al*., 2013). In addition, non-CpG methylation has also been shown to occur throughout the genome, specifically within the gene body, but is also enriched at non-CpG containing regions including both 5'upstream and 3'downstream regions (Lister *et al*., 2009, Guo *et al*., 2014). Therefore it might be unsurprising that so much variation was identified at these residues between the LOAD samples and that most of the differentially methylated regions identified were located either intragenically or downstream of the associated genes.

Non-CpG methylation has also been shown to result in transcriptional repression and corresponding reduction in gene expression in the brain (Guo *et al*., 2014). This repression is caused by the binding of the transcriptional repressor methyl binding domain family protein methyl-CpG binding protein 2 (MeCP2), to methylated cytosine residues within these sites. MeCP2 was previously thought to preferentially bind to methylated CpG residues; however more recent studies have revealed affinity for methylated CpH residues and specifically those in the configuration CA (Guo *et al*., 2014, Gabel *et al*., 2015). This is interesting since the majority of differentially methylated cytosines identified between the two LOAD samples were located within CA dinucleotides.

There is also evidence to suggest that CpH methylation is particularly important in the brain and for wider neuronal function. Non-CpG cytosine methylation is far more common in the brain than somatic tissue (Ziller *et al*., 2011), and mutations of MeCP2 result in the neural developmental disorder

Rett syndrome (Kinde *et al*., 2015, Guo *et al*., 2014). This suggests that the transcription repression caused by MeCP2 interaction with methylated CpH dinucleotides adds a further layer of epigenetic regulation to the neuronal genome which is important in its development and functionality. Therefore it seems plausible that aberrant methylation at these sites might be responsible for the development of neurological diseases such as LOAD.

The differences in methylation observed in severe LOAD reported in this thesis might therefore represent epigenetic modifications that alter the expression of their associated genes, driving severe LOAD pathology. Alternatively the aberrant methylation observed might also be a consequence of the pathological environment that develops as the disease progresses; the methylation of non-CpG residues may be altered in response to pathological events such as Aβ and NFT accumulation within the brain. Aβ accumulation, resulting in increased oxidative stress, have been shown to result in aberrant methylation of the promoter regions of other genes such as *neprilysin* (Chen *et al*., 2009).

For the work described the DMLs identified were not investigated as to whether they corresponded to mQTLs. This was because mQTLs are likely to be important in single sample analysis, however the inclusion of additional whole methylome data sets should minimise the impact of mQTL effects. McRae et al (2018) identified 2,916 cis and 2,025 trans DNA mQTL using methylation from whole blood measured (Illumina) however mQTLS were overrepresented in subtelomeric regions enriched with pseudogenes. It should also be noted that any difference in genes linked to AD via genetic analysis or other methods would be clinically relevant to that patient, indifferent to the cause e.g. epigenetic, genetic or mQTL.

5.5.3 Genes associated with DMLs

22 genes were identified as associating with DMLs in sample two when compared with sample one. Since the samples represented moderate and severe AD it is plausible that the differential methylation identified in sample two may be reflective of pathology which is specific to later stages of the

disease. Many of the genes identified have biological functions or are involved in pathways which have been previously linked to AD pathology and disease progression, shown in table 5.7.

The immune response represents one of the main pathways identified as being perturbed in AD by GWAS. In addition to genes which encode a protein involved in the immune response two genes encoding transcription factors which regulate expression of genes encoding immune response proteins were also identified by the WGBS study. Previously multiple genes involved in the immune response and inflammation have been identified as containing LOAD risk variants (Guerreiro and Hardy, 2014). These include *TREM2, HLA-DRB1, HLA-DRB5* and *CR1*, in addition aberrant methylation of genes with immune functions have also been identified in LOAD (Karch and Goate, 2015, Yu *et al*., 2014b, Combs, 2009, Zhang *et al*., 2015). Therefore it seems unsurprising that genes were identified by the WGBS studies described in this thesis, that have roles in the immune system and the inflammatory response.

Four of the genes identified have potential function in the regulation of APP processing. This is again unsurprising as the majority of genes that confer AD risk have roles that affected Aβ production. This includes the genes *APP, PSEN1 and PSEN2,* and *ADAM10* (Karch and Goate, 2015). Differential methylation of *APP* and other genes that encode proteins involved in the amyliod cascade have also been identified in LOAD (Iwata *et al*., 2014) (see section 1.5).

Three of the genes identified in this study (*ADGRA1, HNC2* and *UBAC1*) have also previously been associated with AD. *ADGRA1* encodes the protein adhesion G protein-coupled receptor a1. G protein coupled receptors (GPCRs) function to induce intracellular signalling and have been linked to regulation of BACE1 (β-site APP cleaving enzyme) which is the secretase responsible for the generation of Aβ peptide (Zhao *et al*., 2016). The exact affect of G protein-coupled receptors on BACE1 is not well understood but signalling pathways stimulated by GPCR activation are thought to potentially alter BACE1 expression, degradation or trafficking (Zhao *et al*., 2016). Further roles of GPCR have been suggested in α and γ-secratase activity and also Aβ degradation (Thathiah and De Strooper, 2011).

HCN2 encoding potassium/sodium hyperpolarization-activated cyclic nucleotide-gated ion channel 2 has been also been shown to have a role in the regulation of Aβ production in AD through its association with γ-secretase. Loss of HNC2 function results in a reduction of Aβ production possibly due to an effect on APP processing (Frykman *et al*., 2017). Hypermethylation of the *HCN*2 promoter region has also been identified in the LOAD brain (Sanchez-Mut *et al*., 2016).

UBAC1, encoding ubiquitin-associated domain-containing protein 1 also likely has a role in driving AD pathology through its involvement in the APP pathway. UBAC1 is involved in the removal of toxic proteins by ubiquitin-dependent proteolysis (UPP), APP is an example of such a protein. One hypothesis is that in neurodegenerative diseases such as AD the accumulation of toxic proteins is so great that the UPP process becomes overloaded resulting in accumulation of ubiquitin- positive aggregates leading to neuronal dysfunction and eventual cell death (Chung *et al*., 2001).

Two genes were also identified which encode proteins with functions in intracellular transport. *SORCS2* (Sortilin Related VPS10 Domain Containing Receptor 2) is a gene located within the vacuolar protein sorting (VPS) gene family and is of particular interest as it belongs to the same gene family as *SORL. SORL1* has been both genetically linked to AD and shown to harbour AD linked aberrant methylation (Lambert *et al*., 2013b, Yu *et al*., 2014b, Rogaeva *et al*., 2007, et al., 2013). Yu *et al*., (2014b) identified aberrant methylation and expression of *SORL1* in the LOAD brain. Both *SORL1* and *SORCS2* were found to harbour SNPs which associated with fAD (Rogaeva *et al*., 2007, Reitz *et al*., 2013). *SORC2* was also identified as containing LOAD associated CpG site hypomethylation within its promoter region in brain tissue by another WGBS study (Sanchez-Mut *et al*., 2016). SORL1, and other proteins that function in intercellular transport, may drive AD progression by influencing APP processing in the endocytic pathway (Karch and Goate, 2015).

The mitochondrial gene *MT-ND5* (encoding the protein NADH dehydrogenase core subunit 5) was also identified as being differentially methylated in the

severe LOAD patient. *MT-ND5* has previously been identified as harbouring a reduction in hydroxymethylation associated with aging in mice (Dzitoyeva *et al*., 2012). Increased expression of the MT-ND5 gene was also recently shown to associate with MCI (Lunnon *et al*., 2017). Mitochondrial dysfunction is also a common occurrence in AD and the mitochondrial cascade hypothesis states that basal mitochondrial function, which is under genetic regulation, will decline throughout aging resulting in eventual onset of AD (Devall *et al*., 2014, Swerdlow *et al*., 2010, De Jager *et al*., 2014). In addition epigenetic regulation of the mitochondrial genome and its involvement in AD has been documented (Lunnon *et al*., 2017).

A large number of genes were identified with roles in normal neuronal function and development. This was also unsurprising as many of the LOAD associated genes identified by GWAS function within the CNS system and have roles in neuronal survival (Bettens *et al*., 2013, Van Cauwenberghe *et al*., 2016). These genes was *NAV1,* which encodes neuronal navigator 1, which is involved in neuronal migration through its regulation of axon guidance. The protein associates with microtubules in neurones (Martinez-Lopez *et al*., 2005). The expression of NAV1 was found to be increased in post mortem LOAD CA4 hippocampal cells (Ho Kim *et al*., 2015). In addition, the neuronal navigator family proteins NAV2 and NAV3 have also been associated with AD and both were shown to be down regulated in the LOAD hippocampus using a mouse model (Wang *et al*., 2017, Zong *et al*., 2015).

UNC5C, encodING the uncoordinated-5c netrin receptor gene, was another gene identified with a role in neuronal function. UNC5C is needed for embryonic development of neurones through its regulation of axon guidance; however absence of the receptors ligand results in activation of cell death (Moore *et al*., 2007). A genetic mutation within the *UNC5C* gene has been linked to increased risk of LOAD (Wetzel-Smith *et al*., 2014). Neurones containing this mutation have increased vulnerability to neuronal death in response to insult. One such insult is the presence of APP which induces the death of these neurones through mediation of an intracellular death signalling cascade, netrin can bind to APP and prevent this death cascade from being activated (Hashimoto *et al*., 2016).

Perhaps more surprising was the identification of numerous genes encoding proteins involved in regulation of gene expression and also genes enocding regulatory RNA harbouring differential methylation in the severe LOAD patient. *IRF4* and *TRIM29* were identified as harbouring differential methylation in severe LOAD.

IRF4, encoding interferon regulatory factor 4 has been associated with LOAD previous to the WGBS work conducted as part of this thesis. A study using a mouse model of AD found *IRF4* to be down regulated in microglial cells when compared to controls (Orre *et al*., 2014). The study also identified differential expression of multiple other transcription regulatory genes which associated with the AD phenotype (Orre *et al*., 2014). IRF4 has also been shown induce an anti-inflammatory response within the brain and its expression has been associated with conduction of programmed cell death within the brain following stroke. The expression of *IRF4* also results in a anti-inflammatory micro-environment in the AD brain (Mamun and Liu, 2017). Therefore the reduced expression observed by Orre *et al*. (2014) may drive AD pathology through an increase in inflammation.

Intragenic hypomethylation of *IRF4* was observed in the severe LOAD sufferer. This hypomethylation may result in an increase in expression of *IRF4;* representing a response to the pro-inflammatory environment created in the later stages of LOAD by increased Aβ deposition. This would need to be validated

TRIM29 (tripartite motif-containing protein 29*)* also known as ataxia-telangiectasia group d-associated protein was another gene with regulatory function that was identified. This gene has been associated with the development of several types of cancer and has been shown to drive lung cancer cell invasion through its activation of the ERK and JNK pathways (Tang *et al.,* 2013). While *TRIM29* has yet to be genetically associated with AD deregulation of the ERK and JNK signalling pathways have been implicated in AD pathology using post-mortem brain tissue (Zhu *et al*., 2001).

The *U6* gene was also identified as being differentially methylated in severe LOAD in the WGBS study described in this thesis chapter. *U6* encodes a snRNA molecule which forms part of the spliceosome and therefore is involved with gene splicing (Wahl and Luhrmann, 2015). It therefore has the potential to cause far reaching aberrant protein expression in AD by affecting the alternative splicing of multiple genes. In addition, AD associated hypomethylation of 26 CpG residues located within the *U6* promoters were identified by another WGBS study in the LOAD brain (Sanchez-Mut *et al*., 2016). Another study also identified AD associated methylation in brain tissue at one CpG site located within 50kb of *U6* (De Jager *et al*., 2014). In addition, both microRNAs (miRNAs) and long non-coding (lncRNAs) have been shown to influence AD disease progression (Zhang, 2016).

5.6 Validation using an external LOAD WGBS data set

Three of the genes associated with differential methylation in the cerebellum samples were also identified as harbouring differential methylation in the Sanchez- Mut *et al.,* (2016) grey matter sample. These three genes were *U6, pRNA* and *CH17-333M13.2.* The aim of this part of the study was to validate the genes identified in a independent and published LOAD WGBS sample.

Historically cerebellum has commonly not been used to study epigenetic effects in AD because this brain region has been considered to not be relevant to AD pathophysiology. However the WGBS data comparison between the cerebellum and grey matter tissue did identify identical differentially methylated genes. This suggests that the cerebellum could be capable of harbouring important LOAD associated differential methylation. This seems unsurprising as AD associated pathology such amyloid plaques and tau tangles have been identified in the cerebellum of LOAD suffers; and also protein expression changes linked to AD have also been identified in the cerebellum (reviewed by Jacobs et al, 2017, Xu et al, 2019). Importantly this comparison

analysis against published data further supports the whole methylome data described above (section 5.4).

5.7 Experimental limitations and future work

While the WGBS experiment did reveal important DMR and genes associated with these regions, it is important to consider the main limitations of this study. This includes sample size, genome coverage and the control sample used. Importantly an effort was made to validate the DMRs identified using published WGBS data from another LOAD sample and three genes were also found to harbour LOAD associated differential methylation in this sample.

Due to the high cost of WGBS it was only possible to generate data for the two LOAD samples. Ideally at least three biological replicates would be needed to confidently identify any differential methylation occurs between groups. The use of only one sample does not allow for the elimination of inter-individual differential methylation that might be occurring between the two samples, variation not associated with LOAD pathology.

Another potential limitation is that the genome coverage provided by each sample was relatively low. In particular sample two and the control had very low coverage. Sequencing of sample two also resulted in a lot of duplicate reads, possibly due to the low concentration of the library. Ideally the library preparation would have been repeated for sample two prior to sequencing and a library of higher quality would have been used. Due to the time constraints this was not possible. However only regions with greater then 10X coverage were used for analysis therefore these regions would have been identified anyway.

The control sample had the lowest coverage of all three samples. This made it extremely difficult to obtain useful results when control data was compared with LOAD samples. In addition, the control sample data obtained from a non-aged matched individual. Therefore any DMLs identified between the LOAD samples and control could not be ruled out as being age associated. Ideally an age matched control sample would have been used for WGBS and compared

to the LOAD samples. However due to the cost of undertaking WGBS this was not possible.

Although every attempt was made to obtain a sample set with as broad a coverage as possible, there is a realisation that the experimental plan is unfortunately limited by material available to test. I would therefore suggest, like many published studies, that additional experimentation would be required to further explore any interesting data.

More targeted methods such as pyrosequencing should be used in future to investigate the differentially methylated regions identified in both moderate and severe LOAD patients and age matched controls, with a particular focus on the three genes that were also identified in the Sanchez Mut *et al.,* (2016) sample. The use of this method would allow investigation into the identified regions in many patient samples, resulting in confident identification of differential methylation.

The results generated in this chapter were compared to a WGBS data from a published LOAD sample (Sanchez-Mut *et al.,* 2016). This allowed some validation of data. However pyrosequencing validation would also be useful, unfortunately this was not possible during this project due to time constraints, however the data generated during this study has lead to the identification of interesting sites for further investigation by future research.

5.8 Conclusion

In conclusion, the WGBS work conducted as part of this thesis has led to the identification of some interesting differentially methylated regions in genes that should be further investigated in LOAD. However, they will need to be validated using more targeted methods and a substantially increased sample size to be confident of results. It might also be interesting to, once validated, assess if any of this differential methylation is also identifiable in late stage sEOAD, as the epigenetic deregulation might be pathology rather than disease sub-group specific.

Chapter 6: General Discussion

6.1 Aims of the thesis

Aberrant DNA methylation has been linked with the pathogenesis of multiple diseases including AD (Qazi *et al*., 2017) (section 1.5). Many studies have investigated both global methylation and gene specific methylation in LOAD and LOAD associated genes. However many gene specific studies have revealed conflicting results. This could possibly be due to the inconsistent use of brain tissue in many studies. It is known that different regions of the brain and even different cell types can harbour their own unique epigenetic profiles, making it difficult to obtain consistent results from studies which use tissue from differing brain regions. For this reason LOAD leukocyte DNA was used in this thesis to investigate methylation in loci identified as carrying LOAD risk alleles via GWAS testing (Lambert *et al*., 2013b, Humphries *et al*., 2015 Yu *et al*., 2014b, Smith *et al*., 2016).

Leukocyte DNA was also used because studies have shown AD associated methylation can be identified in peripheral blood samples using a variety of experimental methods (Di Francesco *et al*., 2015, Xu *et al,* 2018; Madrid *et al,* 2018, Mercorio *et al,* 2018, Hou *et al*., 2013, Chang *et al*., 2014, Lunnon *et al*., 2014). It is also likely that at least some of the changes identified in blood are linked to the pathogenesis or variation in methylation occurring within brain tissue (Wang *et al.,* 2019). Regardless of association between blood methylation and methylation status of the AD brain, blood methylation linked to disease status has an independent value. It is possible that any differential methylation identified in blood could represent a biomarker of AD.

Investigation into aberrant methylation in both LOAD samples and sEOAD samples was undertaken for a number of genes. To date only a few studies have investigated aberrant methylation in sEOAD and given the similar pathologies present in both forms of AD it seems a plausible hypothesis that genes identified as epigenetically deregulated in LOAD may harbour similar disease causing alteration in DNA methylation in sEOAD.

Many previous methylation studies have focused on the identification of AD associated methylation located within the promoter region of AD related genes. However, recent global and genome wide epigenetic studies have revealed that important differential methylation also occurs at non-promoter CpG sites such as in the 5' and 3' UTRs and can be found within both intra and intergenic regions (Smith *et al*., 2016, De Jager *et al*., 2014, Yu *et al*., 2014b, Humphries *et al*., 2015). In addition, studies are beginning to reveal that aberrant methylation at non-promoter CpG sites might have significant importance in regulation of gene elongation, transcriptional repression and alternative spicing (Maussion *et al*., 2014, Choi *et al*., 2009, Malumbres *et al*., 1999). Therefore work conducted in this thesis also targeted non-promoter CpG site rich regions of interest.

6.2 Targeted methods

Several targeted methods are described in this thesis and were utilised to identify differential methylation associated with both LOAD and sEOAD. In some instances identical genic regions were also investigated in both types of AD. This allowed comparison between the disease types and therefore the identification of disease type specific aberrant methylation.

Initially bisulphite cloning and the McrBC enzyme based profiling methods were used to investigate the promoter regions of genes previously associated with LOAD via GWAS studies (Lambert *et al*., 2013b). Interestingly since the work conducted in this thesis began, other studies have found that aberrant methylation within genes associated with LOAD through GWAS can be identified (Humphries *et al*., 2015, Yu *et al*., 2014b, De Jager *et al*., 2014, Hou *et al*., 2013, Smith *et al*., 2016).

McrBC enzymatic experiments identified possible hypermethylation within the promoter regions of the genes *INPP5*D and *SORL1*, both of which have been implicated in LOAD pathology or development. Although this result is interesting the McrBC method is limited by its inability to give quantitative data or detailed information about individual CpG sites. Therefore the methylation profiling work

switched to focus on the use of the pyrosequencing platform, allowing quantitative data generation for each individual CpG site assayed.

Unfortunately the pyrosequencing assay designed to cover the *SORL1* region investigated using the McrBC method failed at the pyrosequencing stage. However, it was possible to investigate the *INPP5D* promoter target in both sEOAD leukocyte and cortex DNA. The previous McrBC work had been conducted and indentified methylation using LOAD leukocyte DNA, making it somewhat surprising that no sEOAD associated DNA methylation was identified. In addition no aberrant methylation was identified in cortex tissue either, suggesting a unique epigenetic profile might exist for the two forms of AD.

This hypothesis was supported by the pyrosequencing results obtained for *TREM2* and *SIRT1*. A CpG site upstream of the *TREM2* promoter was found to be hypermethylated in the LOAD brain (Smith *et al*., 2016). *SIRT1* was also shown to contain LOAD associated hypermethylation at multiple CpG sites (Hou *et al*., 2013). In the case of these genes pyrosequencing targets were chosen that included CpG sites shown to be aberrantly methylated in LOAD by the afore mentioned published studies, these sites were then investigated using sEOAD leukocyte and cortex DNA.

For these genes no significant difference in methylation was observed between sEOAD and control samples for either tissue type tested. Interestingly however one CpG site investigated for *SIRT1* did show a small but statistically significant reduction in methylation in LOAD leukocyte samples, whereas a second CpG site investigated in this gene did not, despite both being proven to harbour AD associated hypermethylation in the study conducted by Hou *et al*. (2013). One explanation for this is that Hou *et al*. (2013) used a Chinese population for their study whereas the samples used during work conducted in this thesis were obtained from a UK population. This might indicate not only that sEOAD and LOAD represent two distinct diseases, which are driven by differing epigenetic alteration, but also that ethnicity and environment might also be important in driving disease specific DNA methylation responsible for causing progression of the same pathology.

Pyrosequencing was also used to identify differential methylation in the promoter regions of *ABCA7*, *MEF2C* and *PTK2B*. These genes were initially selected for investigation due to their genetic association with LOAD (Lambert *et al*., 2013b). However, since beginning the work conducted in this thesis, published research has identified LOAD associated aberrant methylation within these genes (Yu *et al*., 2014b, Humphries *et al*., 2015). While it should be considered that many of these studies investigated different gene regions and tissue type to those used in experiments described in this thesis, it is still interesting to note that pyrosequencing failed to identify any significant differential methylation at the regions target in these genes using sEOAD blood and brain samples. This further supports the hypothesis that LOAD and sEOAD do not harbour common epigenetic deregulation resulting in their shared pathological features. However more work would be needed to confirm this hypothesis.

As part of the work described in this thesis a limited number of genes were investigated and of these genes only small regions of sequence were targeted. This is due to the technical limitations of the McrBC enzymatic method and pyrosequencing platform. An interesting avenue for further study would be the use either global or whole genome methods to compare the entire LOAD and sEOAD methylome. This would allow the identification of differential methylation occurring in both types of AD, which both differ between the two disease types and is common to them both. A plausible hypothesis is that any disease type specific aberrant methylation identified might be a driver of disease progression whereas any common differential methylation could be a consequence of the similar pathology shared by each disease type. However, to obtain this type of information DNA samples would be required from controls and patients with varying disease severity of both LOAD and sEOAD. This makes the availability of such samples and cost major imitating factors when considering conducting an experiment of this kind.

Unfortunately no brain DNA was available for testing from LOAD patients however it would also be interesting to investigate whether the hypermethylation identified in

SORL1 and *INPP5D* blood DNA was also identifiable in the LOAD brain from matched samples.

In addition to the *MEF2C* promoter region investigated a further CpG site upstream of this region was also targeted via pyrosequencing. No significant difference in methylation at this site was found within either sEOAD blood or brain tissue when samples are considered group-wide. However, a significant reduction in methylation was observed at the site in one leukocyte sample and a non-statistically significant reduction at this site was also observed for one sEOAD brain sample. Suggesting the possible presence of an epi-allele at this site which might explain an increased risk of certain AD associated symptoms in some individual's (Agorio *et al*., 2017).

This result is interesting as similar aberrant methylation may have been previously missed by studies due to the common goal of identifying only regions or multiple CpG sites which present as 'aberrant' in the majority of AD cases. It highlights the possible importance of rare epigenetic modifications which can influence disease susceptibility in a similar way to rare genetic variations found in AD associated genes.

Unfortunately blood and brain tissue were not available from the same patient. However it would have been interesting to investigate if the hypomethylation observed at this CpG site was blood specific or was also present within the AD brain. Presence of this reduction in methylation within the brain and blood of the same individual would support the hypothesis that leukocytes alter their gene expression through methylation based on the environment they encounter as they pass through the brain.

The pyrosequencing work conducted using the 3'UTR of the *RIN3* gene as a target suggested that this might be the case. The identification of hypomethylation in the *MEF2C* upstream CpG site and not within the promoter CGI investigated suggested that non-promoter CpG methylation is important in AD. This hypothesis is supported by other studies which have identified non-promoter CpG methylation associated with LOAD in both blood and brain tissue (Humphries *et al*., 2015, Smith *et al*., 2016). Therefore the 3'UTR of *RIN3* was chosen for analysis by pyrosequencing.

Interestingly significant hypomethylation was observed in sEOAD blood samples across all seven CpG sites included in the assay and a non-statistically significant reduction in methylation was also observed across all seven CpG sites in brain tissue. This suggested the possibility that leukocyte hypomethylate the 3'UTR of *RIN3* in response to the environment that they encounter in the brain. This hypothesis is supported by the fact that significant hypomethylation of both the *MEF2C* upstream CpG sites and the *RIN3* 3'UTR are both significantly hypomethylated in brain tissue compared to blood tissue.

The results for *MEF2C* and *RIN3* are intriguing, particularly due to the fact that hypomethylation is observable in blood DNA making *RIN3* 3'UTR methylation a plausible biomarker of sEOAD. In the instance of the *MEF2C* epi-mutation it is possible that hypomethylation at this site is a driver of sEOAD progression as the mark is rare and could indicate that specific individuals carry their own unique epigenetic profile which then drive or exacerbate AD pathology. However in the instance of *RIN3* hypomethylation spans a larger region of CpG sites, and may represent aberrant methylation occuring as a result of AD pathology.

However, future work will be needed to elucidate the biological consequences of the aberrant methylation observed and described in this thesis. 3'UTR methylation has been associated with a reduction in gene expression by other studies (Maussion *et al*., 2014). Therefore one hypothesis would be that the reduction in methylation at the 3'UTR of *RIN3* might result in increased expression of *RIN3* in AD. Unfortunately gene expression studies were not undertaken as part of the work conducted in this thesis due to viable patient RNA obtained post mortem being difficult to obtain, therefore unavailable for testing as part of this study.

It would also be of interest to investigate *INPP5D* and *SORL1* expression in LOAD leukocytes as hypermethylation of promoter CpG containing region was suggested at by the McrBC experimentation. Unfortunately due to similar reasons described for sEOAD no RNA was available for expression analysis of LOAD samples either.

6.3 Whole genome methods

The WGBS experiment resulted in the identification of a substantial amount of differential methylation occurring within the severe AD patient sample. This resulted in the identification of multiple genes with functional relevance to AD pathology. These results have to be interpreted with caution as it was only possible to sequence one biological repetition for each stage of LOAD and the published control data used for comparison was not ideal. However three of genes associating with DML were also identified as harbouring differential methylation using published WGBS data from a grey mater sample (Sanchez-Mut *et al.,* 2016). Further validation would be required to confirm the differential methylation identified.

The WGBS experiment results provide interesting insights into differential methylation occurring in the later stages of LOAD pathology. The WGBS experiment identified both singular DML and also large differentially methylated regions which contained multiple DML within them. This result seems consistent with the hypothesis that within AD (perhaps both sEOAD and LOAD) specific regulatory cytosine sites are differentially methylated and drive AD pathology, such as the *MEF2C* upstream CpG site identified using the sEOAD samples and pyrosequencing. Whereas other regions containing widespread differential methylation, such as that identified in the *RIN3* 3'UTR, are a consequence of the disease pathology.

It was also interesting that the majority of differentially methylated regions identified by the WGBS experiment were located downstream of the nearest associated gene and intergenically. It would be interesting to investigate, in future work, how much of this differential methylation is occurring within 3'UTR regions. Since the only regions identified as being significantly differentially methylation in the sEOAD samples by pyrosequencing was a region located upstream of the TSS and a region within the 3'UTR of *RIN3*, all of the work conducted in this thesis suggests that non-promoter differential methylation has an important role to play in AD and therefore methylation studies should consider other genic regions.

One limitation of both the pyrosequencing and WGBS methods is that bisulphite conversion was used prior to sequence to allow identification of differential

methylation. However bisulphite conversion does not differentiate between methylation and hydroxymethylation. Therefore without further work it is not known if all of the differential methylated cytosines identified were not in fact hydroxymethylated prior to conversion. Importantly information is not available about any differences in methylation and hydroxymethylation and therefore this type is DML would not be identified.

However, in the instance of the WGBS data most of the differential methylation identified is likely to be methylation rather then hydroxymethylation. This is because most variation was identified at CpH sites and studies have shown that very little CpH hydroxymethylation occurs within the human brain (Guo *et al*., 2014). However mouse studies have identified hydroxymethyltaion in CpA sites located near active enhancer regions in neuronal cells (Mellen *et al.,* 2017). Further research would be needed to elucidate if this is also the case in humans.

For the differentially methylated regions identified using pyrosequencing hydroxymethylation could be identified in future study using Tet-assisted bisulfite sequencing (TAB-seq), alternatively this could have been used rather than pyrosequencing as an alternative method. In this method HmC sites are protected by β-glucosyltransferase (β-GT)–mediated glucosylation during treatment with mTet1 protein which results in oxidation of 5-mC to 5-caC which is then read as T during bisulphite sequencing, therefore only the 5-gmC residues will be read at Cs in bisulphite sequencing (Kinde *et al*., 2015, Yu *et al*., 2012).

Alternatively hydroxymethylation could also be identified using oxBS sequencing or pyrosequencing. In this type of sequencing an additional oxidation step is added to convert the 5hmCs to 5fC which are then converted by bisulphite treatment to uracil. A sample which is only bisulphite treated is then sequenced in parallel allowing identification of both 5mC and 5hmC (Booth *et al.,* 2012).

This use of these methods would have allowed differentiation between hydroxymethylation and methylation and the identification of additional DML. These methods were not used as a decision was made to potentially use these methods as

tools for further investigation into regions found to be harbouring DML by pyrosequencing and WGBS.

6.4 Conclusion and future work

In conclusion the work presented in this thesis has provided interesting avenues for further work into differential methylation occurring in both sEOAD and LOAD. The pyrosequencing and WGBS results suggest that methylation in non-promoter regions, including intergenic and within downstream regulatory regions, is important in AD and highlights the importance of investigation into methylation occurring outside of promoter regions. The WGBS results, while needing future validation, might also suggest that non-CpG site cytosine methylation is also important in more advanced LOAD pathology.

Although some studies have been undertaken to look at CpGs present within regions of the genome associated with genes tested in this study, as shown in this results presented in this thesis, there is not always a correlation between methylation of adjacent CpGs or between promoter and 3' genic regions. Therefore it is likely that each individual CpG or genic or intergenic region should be assessed and considered separately.

Future work should focus on validation and elucidation of the biologically consequence of the differential methylation identified. Unfortunately no gene expression analysis was performed as part of this thesis, but investigation should be carried out to identify if the differential methylation observed in the *MEF2C* upstream CpG site and the *RIN3* 3'UTR have any effect on gene expression. Once validated the differentially methylated regions identified in the severe LOAD patient cerebellum sample should also be investigated for affect on gene expression. It would also be interesting to see if any of the validated differentially methylated regions are present in LOAD blood samples or sEOAD samples.

It would be desirable to undertake RNA profiling experiments in order to correlate any variation in methylation status with change in expression. Post mortem handling

of samples is one limit to performing accurate RNA qualification. Samples are typically obtained sometime after death leading to degradation or inappropriately handled, this leads to degradation within the sample and loss of transcripts. It may be possible in future to mine published gene expression data for these genes to identify variations in expression linked to disease. However as the hypothesis would suggest that X% methylation has an impact on Y% transcript levels normalised to patient specific levels of reference gene transcripts, such analysis would be supportive of further testing rather than replacing the need for these experiments.

Further work would also be needed to exclude the possibility that additional pathologies existing within the samples used are not responsible for the aberrant methylation identified. For the pyrosequencing analysis an effort was made to use samples that did not have another form of dementia, however in order to generate a large enough data set this was not always possible (see appendix 1 for sample details). In addition for the WGBS study the samples were chosen because WGBS data sets existed for these samples and both samples had secondary pathology (see appendix 2). Therefore further work would be needed to exclude secondary pathologies from analysis.

Appendices

Appendix 1: AD sample information

Table 1: LOAD blood sample information: detail of all LOAD samples used for both McrBC and pyrosequencing experiments are included. Information about the specific samples used for each study is provided in the chapter specific methods.

DNA ID	Centre	Gender	Age at Death	Age at Onset	Age at Sampling	Diagnosis
N169	Nottingham	M			56	Control
N158	Nottingham	F			72	Control
N160	Nottingham	M			73	Control
N166	Nottingham	F			37	Control
AD249	Nottingham	F			84	LOAD
M543	Manchester	F		69	73	LOAD
M604	Manchester	F		74	86	LOAD
M646	Manchester	F		n/a	n/a	LOAD
M673	Manchester	M			86	LOAD
M767	Manchester	M		72		Probable LOAD

Table 2: sEOAD DNA sample information. Table includes all information available for the sEOAD samples used for the pyrosequencing work described in chapter 4, the specific samples used for each experiment are included in the chapter methods section. Where the number -9 is included this corresponded to a predicted 9 year disease progression, no other information was available about these samples. Some samples presented with multiple pathologies, this is discussed in chapter 6.

	SAMPLE TYPE	SAMPLE ID	CENTRE	SEX	AGE-AT-DEATH	AGE-AT-ONSET	AGE-AT SAMPLING	AGE-AT ENTRY TO STUDY	STATUS	DIAGNOSIS NOTES
1	sEOAD Blood	mrc024	NOTTINGHAM	M			47	47	AD	
2	sEOAD Blood	m749	MANCHESTER	F	-9	-9	50		PROBABLE AD	AD
3	sEOAD Blood	m340	MANCHESTER	M	45	45	37	39	CONFIRMED AD	AD (PS-1)
4	sEOAD Blood	m510	MANCHESTER	F	-9	-9	49	49	POSSIBLE AD	
5	sEOAD Blood	m620	MANCHESTER	M	-9	-9	50	-9	PROBABLE AD	AD+ALCOHOL
6	sEOAD Blood	M400	MANCHESTER	F	-9	-9	50	62	PROBABLE AD	
7	sEOAD Blood	m753	MANCHESTER	F	-9	-9	48		PROBABLE AD	MCI (MILD AD)
8	sEOAD Blood	m135	MANCHESTER	F	-9	-9	51	52	PROBABLE AD	
9	sEOAD Blood	m328	MANCHESTER	M	-9	-9	46	51	PROBABLE AD	
10	sEOAD Blood	m341	MANCHESTER	M	-9	-9	47	51	PROBABLE AD	
11	sEOAD Blood	mrc078	NOTTINGHAM	M			47	47	MCI?AD	
12	sEOAD Blood	m684	MANCHESTER	M	-9	-9	50	-9	PROBABLE AD	
13	sEOAD Blood	m053	MANCHESTER	F	-9	-9	50	55	PROBABLE AD	
14	sEOAD Blood	mrc080	NOTTINGHAM	F			45	45	MCI ?AD	
15	sEOAD Blood	m279	MANCHESTER	M	-9	-9	51	51	PROBABLE AD	
16	sEOAD Blood	m338	MANCHESTER	M	-9	-9	51	53	PROBABLE AD	
17	sEOAD Blood	mrc059	NOTTINGHAM	M			51	51	MCI?AD	
18	sEOAD Blood	m086	MANCHESTER	M	-9	68	58	61	PROBABLE AD	V EARLY / AD OR FRONTAL LOBE

										SYNDROME
19	sEOAD Blood	m302	MANCHESTER	M	-9	-9	49	53	PROBABLE AD	
20	sEOAD Blood	mrc047	NOTTINGHAM	M			46	46	AD	
21	sEOAD Blood	m091	MANCHESTER	M	-9	-9	50	54	PROBABLE AD	
22	sEOAD Blood	**M102**	MANCHESTER	M	-9	-9	52	54	PROBABLE AD	
23	sEOAD Blood	**M119**	MANCHESTER	F	-9	-9	51		PROBABLE AD	
24	sEOAD Blood	**M683**	MANCHESTER	M	-9	-9	50	-9	PROBABLE AD	
		(used for MEF2C only)								
1	sEOAD Brain	m718	MANCHESTER	F	66	52			CONFIRMED AD	
2	sEOAD Brain	m721	MANCHESTER	F	60	50			CONFIRMED AD	
3	sEOAD Brain	bri435	BRISTOL	F	54	48			AD	
4	sEOAD Brain	bri764	BRISTOL	F	65	51			AD	
5	sEOAD Brain	bri789	BRISTOL	F	67	52			AD/DLB	
6	sEOAD Brain	M715	MANCHESTER	F	61	52			CONFIRMED AD	
7	sEOAD Brain	m732	MANCHESTER	M	70	46			CONFIRMED AD	
8	sEOAD Brain	ad208	NOTTINGHAM	F	57	49			CONFIRMED AD	
9	sEOAD Brain	m747	MANCHESTER	F	53	47			CONFIRMED AD	
10	sEOAD Brain	m735	MANCHESTER	F	58	51			CONFIRMED AD	
11	sEOAD Brain	m748	MANCHESTER	M	44	40			CONFIRMED AD	
12	sEOAD Brain	bri592	BRISTOL	F	65	51			AD	
13	sEOAD Brain	ad224	NOTTINGHAM	M	57	49			CONFIRMED AD	
14	sEOAD Brain	ad211	NOTTINGHAM	M	54	46			CONFIRMED AD	
1	Control Blood	s080c	SOUTHAMPTON	F			80		CONTROL	
2	Control Blood	BON1636	BONN	F			85		CONTROL	
3	Control Blood	bon1603	BONN	F			80		CONTROL	
4	Control Blood	s071c	SOUTHAMPTON	F			83		CONTROL	

5	Control Blood	bon0508	BONN	F			90		CONTROL	
6	Control Blood	bon0575	BONN	M			81		CONTROL	
7	Control Blood	bon1682	BONN	F			84		CONTROL	
8	Control Blood	s077c	SOUTHAMPTON	M			83		CONTROL	
9	Control Blood	bon0343	BONN	F			87		CONTROL	
10	Control Blood	s137c	SOUTHAMPTON	M			83		CONTROL	
11	Control Blood	bon1614	BONN	F			79		CONTROL	
12	Control Blood	bon1164	BONN	M			78		CONTROL	
13	Control Blood	bon0786	BONN	F			90		CONTROL	
14	Control Blood	bon0289	BONN	F			87		CONTROL	
15	Control Blood	bon1665	BONN	F			78		CONTROL	
16	Control Blood	bon0314	BONN	M			93		CONTROL	
17	Control Blood	bon0225	BONN	F			78		CONTROL	
18	Control Blood	bon1580	BONN	M			81		CONTROL	
19	Control Blood	bon0424	BONN	F			84		CONTROL	
20	Control Blood	s094c	SOUTHAMPTON	M			85		CONTROL	
21	Control Blood	bon0172	BONN	M			77		CONTROL	
22	Control Blood	bon1837	BONN	M			78		CONTROL	
23	Control Blood	s104c	SOUTHAMPTON	M			82		CONTROL	
24	Control Blood	bon0551	BONN	M			90		CONTROL	
25	Control Blood	bon1648	BONN	M			80		CONTROL	
26	Control Blood	bon0499	BONN	M			77		CONTROL	
1	Control Brain	n053	NOTTINGHAM	M	81				CONTROL	
2	Control Brain	n120	NOTTINGHAM	F	84				CONTROL	
3	Control Brain	n123	NOTTINGHAM	M	82				CONTROL	
4	Control Brain	n124	NOTTINGHAM	M	90				CONTROL	

5	Control Brain	n145	NOTTINGHAM	F	83				CONTROL	
6	Control Brain	N017	NOTTINGHAM	M	85				CONTROL	
7	Control Brain	n128	NOTTINGHAM	F	85				CONTROL	
8	Control Brain	n012	NOTTINGHAM	M	79				CONTROL	
9	Control Brain	n122	NOTTINGHAM	F	83				CONTROL	
10	Control Brain	n131	NOTTINGHAM	M	82				CONTROL	

Table 3: LOAD cerebellum samples used for WGBS

DNA_ Code	Gender	Age at Death	PM_Delay	PM_Interval	Diagnosis	Braak Staging	Pathology 1	Pathology 2
BK076	M	87	19	26	AD	IV	Alzheimer's disease BNE stage IV	Aryrophilic grain disease,mild amyloid angiopathy, hippocampal sclerosis, mild small vessel disease
BK005	M	66	90	114	AD	VI	Alzheimer's disease (Braak 6, BNE 6)	TDP-43 pathology, limited to amygdala

Appendix 2: Primers used for McrBC experiment

Gene	Primers	PCR product Size (bp)
APP	App1f1_CTTGACCAGGGAATGTGTCAGTGTT	252
	App1r1_ CTCCAGGGTTGCATGCCCATGAAGC	
Hla-drb5/1	Hladrbp1f1_ CGCAAGTCCTCCTCTTGGTT	214
	Hladrbp1r1_ AGTTAAGGTTCCAGTGCCCG	
Sorl1	Sorl1p4f1_ GCGACTCCCGTTCCTATTCA	268
	Sorl1p4r1_ AAAACTGCTCACCTGTCCGT	
Ptk2b	Ptk2bp2f1_ TTCCCCTGGAACGCTGAGAG	229
	Ptk2bp2r1_ CACGGAAGCCCTACTACGC	
INPP5D	Inpp5dp2f1_ CTCGGTGGTGTGTGGGTC	180
	Inpp5dp2r1_ TATGCCCGGGAGATGGACTC	
SLC24A4	Slc24a4p1f1_ AGAAAGTTCCCGGGGAGAGC	984
	Slc24a4p1r1_ AGGAGGGGAAAACATCTCGC	
DSG2	Dsg2p1f1_ GGCCCAAAGCACAATCACAA	841
	Dsg2p1r1_ TCGCAACTTACTCTCCTGGC	
APOE promoter	APOEp1f1_ GGACACTGGGGACACCCAGTAGGTGC	320
	APOEp1r1_ CCTCTGCCCTGCTGTGCCTGGGGCAG	
TFAM	Tfamp1f1_ TAATAGATAGTTTTGTATTTAGGAT	117
	Tfamp1r1_ AAACCAAACTAAAAAACTACA	
MTHFR	Mthfrp1f1_ GCAGCATGATAAGCACAAAGTCCTGT	402
	Mthfrp1r1_ CTCTGTGCTGCTGCTGCAGGTG	
BACE	Bacep1f1_ TCCATGCTGAAAGAAAGACTGACAGA	282
	Bacep1r1_ TCAGGCCACCATAATCCAGCT	
MAPT	Maptp1f1_TGTAACTGAGTTAGCTTGCTTTAAGC3	321
	Maptp1r1_ CCTCCTGTAGTTGGAGTCTTTGTGTC	
PSEN1	Psen1p1f1_TGGGCCCAATTTATATAGGGGCTTT	193

	Psen1p1r1_TAGCTCAGGTTCCTTCCAGACCA	

Appendix 3: Primers used for pyrosequencing

Gene Target	Primers	Region of gene	CGI number	No of CpGs covered
INPP5D	Inpp5d_py1f1 TGGGTTTTGGGGGTGTTT Inpp5d_py1r1b AAAAAACTCCCCTCCTTACCTATCCT Inpp5ds2 TTTTGGGGGTGTTTG	Promoter	18	4
	Inantf1 GGGTGATGTTGTTATGGTTTTAGTA Inantsr1b CTCCCCTCCATAATATATAAATCCTAAAAA inants1 GGTTTTAGTAGGGGAT INPP5Ds2 TTTTGGGGGTGTTTG	Promoter	18	4
TREM2	Trem2_kbf1 GGGTTGGTAAGGTTTTGTATTGT Trem2_KBr1b AATCCTAACCTCTAAAAACACAACTATTC Trem2_kbs1 TTAGATTTTTTATTAGTTGTAATG	Upstream of TTS	n/a	1
	Trem2_Pf1 GAGGGTTTTGGTTTTTAAAGGTATAG Trem2_pr1b TACAAAACCTAACCCAAAAATCAC Trem2_ps1 ATTTTTGTAAGGTTGAAATTAGA		n/a	1
ABCA7 (serial pyrosequencing)	Abca7_4_f1:GGTTAGGAGAGGTTTTTTTGTGATT Abca7_4_r1b TCCTTCTCACCTTCCAAAAACTC Abca7_4_s1 AGTAGGTTAGTGAGTG abca74_s2 GTGTAGAGGTAGGGG	Promoter	43	8
PTK2β	Ptk2b_py1f1 GAGGAGGAGGGAGAATTTAATTT Ptk2b_py1r1b-AACTCCCAACTCAAATACCC Ptk2b_py1s1 ATTTGTTAGGTAGATTTATTTGTA	Promoter	27	5
HLA-DRB1/5	Hladrb_py1f1 TTGTTGTAAGAAAAGTGTTGTGGGGATA Hladrb_py1r1b-ATTAAAATCAATTAAAATTCCAATACCC	Promoter		3

	Hladrb_py1s1 GGGGAAGATATTGATAGAGA			
SORL1	Sorl1_py1f1 TTGAGGGTTGTGTATTTAGGAAAGTGG Sorl1_py1r1b-CCCTAACCAACACACCTCCTTAT Sorl1_py1s1 GTGTATTTAGGAAAGTGGTTA	Promoter		3
SLC24A4	Slc2a4a_pf1 TGGGAAGAGGTAGGTTTAGGAATATG Slc2a4a_pr1b-ACTTAAACCCCCTAAAACCAACTCCT Slc2a4a_ps1-GGTAGGTTTAGGAATATGGA Slc2a4a_ps2 GTAGGTTTAGGAATATGGAA	Promoter		3
	Slc2a4a_pf2 TTTTGGGAAGAGGTAGGTTTAGGAATA Slc2a4a_pr1b-ACTTAAACCCCCTAAAACCAACTCCT Slc2a4a_ps1-GGTAGGTTTAGGAATATGGA Slc2a4a_ps2 GTAGGTTTAGGAATATGGAA			
DSG2	Dsg2_pf1-GGGGAGGTAGGGGTTAGG Dsg2_pr1_b-ACCTACCCCACCCTTTTCCCC Dsg2_ps1-GGGGTTAGGGAGGAG Dsg2_ps2 GGGTTAGGGAGGAGT	Promoter		8
	Dsg2_pf2 AGGTAGGGGTTAGGGAGGA Dsg2_pr1_b-ACCTACCCCACCCTTTTCCCC Dsg2_ps1-GGGGTTAGGGAGGAG Dsg2_ps2 GGGTTAGGGAGGAGT			
MEF2C (1) (promoter CGI)	Mef2c_py1f1 GATTGGATATTTTTTATTGGAATTAGTAGT Mef2c_py1r1b TATCACTAACAACCAAACCTTTATCAA Mef2c_py1s1 ATTGGAATTAGTAGTATAGGG	Promoter	38	4
MEF2C (2) (upstream CpG)	Mef2c_py1f1 ATTGGATATTTTTTATTGGAATTAGTAGT Mef2c_py1r1b-TATCACTAACAACCAAACCTTTATCAA Mef2c_py1s1 ATTGGAATTAGTAGTATAGGG	Upstream CpG	n/a	1
RIN3	RIN3_pyf1 GGGTTTAGGGTTGTAGGTAGAGA	3'UTR	124	7

	RIN3_pyr1 AAACCCTAACCACCAATTACCATCAC RIN3_pys1 ATTGGGAATAGTAGGTTT			
SIRT1	Sirtp2f1 AGTTTTTTTAGTAGTTGGGATTATATGTA Sirtp2r1b CAAAACCAACCTAACCAACATAAA Sirtp2s1 AGTTGGGATTATATGTATATGTTA	Promoter	101	2
SPARCL1 PCR1	SPARCL1_P1F1: TGGTTGGTTTTAAATTTTGTTAAT SPARCL1_P1R1_b: ACCCCAAATTCTAATTATTTATATATATCT SPARCL1_P1S1: AAATGGGAATATAGTAAAATTTAT	Promoter	n/a	2
SPARCL1 PCR2	SPARCL1_P2F1: ATGAGTAGTGAGGATTAATTGATAAT SPARCL1_P2R1_b: AACACAACTCTCCACTATAACT SPARCL1_P2S1: AGATTAGATATATATAAATAATTAG	Promoter	n/a	2
SPARCL1 PCR3	SPARCL1_P3F1: AGAAGGTTAAGGTATTTGAAGTATTT SPARCL1_P3S1: TGTTAGAATAATTTTTAATTGAGTA SPARCL1_P3R1_b: AAAAACTCCATACTATTATTTCTTTCTAA	Promoter	n/a	1
SPARCL12	SPARCL1_P1F1: TGGTTGGTTTTAAATTTTGTTAAT SPARCL1_P2R1_b: AACACAACTCTCCACTATAACT sparcl1_12_f3-ATGGTTATAATGGTTGGTATTTAAGAAG sparcl1_12_R1b CCCTCAAATAAAAACACAATAAAACTTAC sparcl1p12_ns1-GGAGGTGTGTATTTATTTTTG sparcl1p12_ns2 ATGTGTTAGATATTGTGTTAATAG	Promoter	n/a	3

Appendix 4: Thermocycler cycles used for each pyrosequencing PCR assay

PCR Name	PCR Program	PCR Assay program was used for
Bisnat50	95°C (5 mins), 50 cycles (94°C (50sec), 50°C (30secs), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR1*
Bisnat54	95°C (5 mins), 50 cycles (94°C (50sec), 54°C (30secs), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR1, SPARCL1 PCR2 SPARCL1 PCR3 SPARCL12*
KAP55	95°C (5 mins), 50 cycles (94°C (50sec), 55°C (30secs (r=2, g=6), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR3*
KAP30	95°C (5 mins), 50 cycles (94°C (50sec), 58°C (30secs (r=2, g=6), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR1*
TD-PCR KBSPARC1	95°C (15 mins), 15 cycles of (96°C (30secs), 60 (-1°C per cycle), 72°C(30secs), 50 cycles (94°C (30sec), 53.6°C (30secs), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR1 SPARCL1 PCR3*
SD-SIRT	95°C (15 mins), 15 cycles of (96°C (30secs), 64 (-1°C per cycle), 72°C(30secs), 50 cycles (94°C (30sec), 53.8°C (30secs), 72°C (60secs)), 72°C (ten	*SIRT*

	mins)	
SD-ABCA7	95°C (15 mins), 15 cycles of (96°C (30secs), 64 (-1°C per cycle), 72°C(30secs), 50 cycles (94°C (30sec), 58.7°C (30secs), 72°C (60secs)), 72°C (ten mins)	*ABCA7, PTK2β, INPP5D (primer set 1), MEF2C(1), SORL1, HLA-DRB1/5.*
MEF2C(2)	95°C (15 mins), 15 cycles of (96°C (30secs), 55 (-1°C per cycle), 72°C(30secs), 50 cycles (94°C (30sec), 48°C (30secs), 72°C (60secs)), 72°C (ten mins)	*MEF2C(2)*
RIN3	95°C (15 mins), 15 cycles of (96°C (30secs), 60 (-1°C per cycle), 72°C(30secs), 50 cycles (94°C (30sec), 55°C (30secs), 72°C (60secs)), 72°C (ten mins)	*RIN3*
SPARCLP3B	95°C (15 mins), 15 cycles of (96°C (30secs), 65 (-1°C per cycle), 72°C(30secs), 45 cycles (94°C (30sec), 58°C (30secs), 72°C (60secs)), 72°C (ten mins)	*SPARCL1 PCR3*
INPP5DANTI	95°C (15 mins), 15 cycles of (96°C (30secs), 55 (-1°C per cycle), 72°C(30secs), 45 cycles (94°C (30sec), 50.2°C (30secs), 72°C (60secs)), 72°C (ten mins)	*INPP5D* (primer set 2)
TREM2	95°C (15 mins), 15 cycles of (96°C (30secs), 60 (-1°C per cycle),	*TREM2*

	72°C(30secs), 45 cycles (94°C (30sec), 55°C (30secs), 72°C (60secs)), 72°C (ten mins)	

Appendix 5: Co-variate Analysis

Shows covariate analysis performed on all pyrosequencing data.

Appendix 5.1: Blood co-variate analysis including gender and age.

```
      Genstat 64-bit Release 19.1 ( PC/Windows 7) 29 January 2019 08:16:02
Copyright 2018, VSN International Ltd.
Registered to: University of Nottingham

 ------------------------------------------

 Genstat Nineteenth Edition
 Genstat Procedure Library Release PL27.1

 ------------------------------------------

   1  SET [WORKINGDIRECTORY='D:/People''s data/Graham Seymour/Andrew
Bottley/Kirsty Boden data JanFEB2019'; DIAGNOSTIC=messages]
   2  "Data taken from file: '\
  -3  D:/People''s data/Graham Seymour/Andrew Bottley/Kirsty Boden
data JanFEB2019/Compiled Boden data Jan 2019.xlsx\
  -4  '"
   5  DELETE [REDEFINE=yes] _stitle_: TEXT _stitle_
   6  READ [PRINT=*; SETNVALUES=yes] _stitle_
  10  PRINT [IPRINT=*] _stitle_; JUST=left

Data imported from Excel file: D:\People's data\Graham Seymour\Andrew Bottley\Kirsty Boden
data JanFEB2019\Compiled Boden data Jan 2019.xlsx
 on: 29-Jan-2019 8:16:14
 taken from sheet "Rin Brain (2)", cells A2:K25

  11  DELETE [REDEFINE=yes]
AD_Brain_averages,Gender,Age_at_Sampling,CpG1Brain,\
  12
CpG2Brain,CpG3Brain,CpG4Brain,CpG5Brain,CpG6Brain,CpG7Brain,ADvsContro
l
  13  UNITS [NVALUES=*]
  14  TEXT [NVALUES=24] AD_Brain_averages
  15  READ AD_Brain_averages

     Identifier      Minimum         Mean      Maximum       Values      Missing
AD_Brain_averages                                              24            0

  19  FACTOR [MODIFY=no; NVALUES=24; LEVELS=2; LABELS=!t('F','M')\
  20  ; REFERENCE=1] Gender
  21  READ Gender; FREPRESENTATION=ordinal

 Identifier       Values      Missing       Levels
 Gender             24            0            2

  23  VARIATE [NVALUES=24] Age_at_Sampling; DECIMALS=0
  24  READ Age_at_Sampling
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
Age_at_Sampling	44.00	69.38	90.00	24	0	

```
  26  VARIATE [NVALUES=24] CpG1Brain; DECIMALS=0
  27  READ CpG1Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG1Brain	6.333	19.60	63.00	24	0	Skew

```
  30  VARIATE [NVALUES=24] CpG2Brain; DECIMALS=0
  31  READ CpG2Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG2Brain	6.000	18.41	62.00	24	0	Skew

```
  34  VARIATE [NVALUES=24] CpG3Brain; DECIMALS=0
  35  READ CpG3Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG3Brain	7.500	21.50	66.00	24	0	Skew

```
  38  VARIATE [NVALUES=24] CpG4Brain; DECIMALS=0
  39  READ CpG4Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG4Brain	5.000	21.23	64.00	24	0

```
  42  VARIATE [NVALUES=24] CpG5Brain; DECIMALS=0
  43  READ CpG5Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG5Brain	5.000	17.98	61.00	24	0	Skew

```
  46  VARIATE [NVALUES=24] CpG6Brain; DECIMALS=0
  47  READ CpG6Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG6Brain	5.000	20.80	61.00	24	0

```
  50  VARIATE [NVALUES=24] CpG7Brain; DECIMALS=0
  51  READ CpG7Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG7Brain	5.500	21.55	71.00	24	0	Skew

```
  54  FACTOR [MODIFY=no; NVALUES=24; LEVELS=2;
LABELS=!t('AD','Control')\
  55  ; REFERENCE=1] ADvsControl
  56  READ ADvsControl; FREPRESENTATION=ordinal
```

Identifier	Values	Missing	Levels
ADvsControl	24	0	2

```
  58
  59  %PostMessage 1129; 0; 10000001 "Sheet Update Completed"
  60  "General Analysis of Variance"
  61  BLOCK "No Blocking"
  62  TREATMENTS Gender*ADvsControl
  63  COVARIATE Age_at_Sampling
  64  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  65   FPROB=yes; PSE=diff] CpG1Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG1Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	653.6	653.6	3.42	0.81	0.080
ADvsControl	1	253.3	253.3	1.33	0.17	0.264
Gender.ADvsControl	1	166.8	166.8	0.87	0.97	0.362
Covariate	1	161.6	161.6	0.85		0.369
Residual	19	3631.8	191.1		0.99	
Total	23	4913.4				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	30. approx. s.e. 12.
units 24	26. approx. s.e. 12.

Covariate regressions

Variate: CpG1Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.50	0.541

Tables of means (adjusted for covariate)

Variate: CpG1Brain
Covariate: Age_at_Sampling

Grand mean 20.

Gender	F	M
	14.	28.
rep.	14	10

ADvsControl	AD	Control
	12.	31.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.4	14.7

```
  66  "General Analysis of Variance"
  67  BLOCK "No Blocking"
  68  TREATMENTS Gender*ADvsControl
  69  COVARIATE Age_at_Sampling
  70  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  71   FPROB=yes; PSE=diff] CpG2Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG2Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	564.2	564.2	3.44	0.81	0.079
ADvsControl	1	311.3	311.3	1.90	0.17	0.184
Gender.ADvsControl	1	118.4	118.4	0.72	0.97	0.406
Covariate	1	196.4	196.4	1.20		0.287
Residual	19	3113.0	163.8		1.01	
Total	23	4298.1				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22		31. approx. s.e.	11.
units 24		25. approx. s.e.	11.

Covariate regressions

Variate: CpG2Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.55	0.501

Tables of means (adjusted for covariate)

Variate: CpG2Brain
Covariate: Age_at_Sampling

Grand mean 18.

Gender	F	M
	13.	26.
rep.	14	10

ADvsControl	AD	Control
	10.	30.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	5.9	13.7

```
  72  "General Analysis of Variance"
  73  BLOCK "No Blocking"
  74  TREATMENTS Gender*ADvsControl
  75  COVARIATE Age_at_Sampling
  76  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  77   FPROB=yes; PSE=diff] CpG3Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG3Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	787.1	787.1	3.74	0.81	0.068
ADvsControl	1	260.0	260.0	1.23	0.17	0.280
Gender.ADvsControl	1	29.4	29.4	0.14	0.97	0.713
Covariate	1	172.8	172.8	0.82		0.376
Residual	19	4000.1	210.5		0.99	
Total	23	5201.8				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22 33. approx. s.e. 13.

Covariate regressions

Variate: CpG3Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.51	0.568

Tables of means (adjusted for covariate)

Variate: CpG3Brain
Covariate: Age_at_Sampling

Grand mean 22.

Gender	F	M
	15.	30.
rep.	14	10

ADvsControl	AD	Control
	14.	32.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.7	15.5

```
  78  "General Analysis of Variance"
  79  BLOCK "No Blocking"
  80  TREATMENTS Gender*ADvsControl
  81  COVARIATE Age_at_Sampling
  82  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  83   FPROB=yes; PSE=diff] CpG4Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG4Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	973.3	973.3	4.97	0.81	0.038
ADvsControl	1	286.2	286.2	1.46	0.17	0.242
Gender.ADvsControl	1	61.7	61.7	0.31	0.97	0.581
Covariate	1	169.1	169.1	0.86		0.365
Residual	19	3721.2	195.9		0.99	
Total	23	5266.4				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	27. approx. s.e. 12.
units 24	30. approx. s.e. 12.

Covariate regressions

Variate: CpG4Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.51	0.548

Tables of means (adjusted for covariate)

Variate: CpG4Brain
Covariate: Age_at_Sampling

Grand mean 21.

Gender	F	M
	14.	31.
rep.	14	10

ADvsControl	AD	Control
	13.	32.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.5	14.9

```
  84  "General Analysis of Variance"
  85  BLOCK "No Blocking"
  86  TREATMENTS Gender*ADvsControl
  87  COVARIATE Age_at_Sampling
  88  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  89   FPROB=yes; PSE=diff] CpG5Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG5Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	462.4	462.4	2.91	0.81	0.104
ADvsControl	1	306.2	306.2	1.93	0.17	0.181
Gender.ADvsControl	1	116.8	116.8	0.74	0.97	0.402
Covariate	1	158.2	158.2	1.00		0.331
Residual	19	3018.3	158.9		1.00	
Total	23	4179.8				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	30. approx. s.e. 11.
units 24	24. approx. s.e. 11.

Covariate regressions

Variate: CpG5Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.49	0.493

Tables of means (adjusted for covariate)

Variate: CpG5Brain
Covariate: Age_at_Sampling

Grand mean 18.

Gender	F	M
	13.	25.
rep.	14	10

ADvsControl	AD	Control
	10.	30.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	5.8	13.4

```
  90  "General Analysis of Variance"
  91  BLOCK "No Blocking"
  92  TREATMENTS Gender*ADvsControl
  93  COVARIATE Age_at_Sampling
  94  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  95   FPROB=yes; PSE=diff] CpG6Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG6Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	557.2	557.2	3.16	0.81	0.092
ADvsControl	1	344.0	344.0	1.95	0.17	0.179
Gender.ADvsControl	1	48.7	48.7	0.28	0.97	0.605
Covariate	1	193.5	193.5	1.10		0.308
Residual	19	3351.9	176.4		1.00	
Total	23	4511.1				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	28. approx. s.e. 12.
units 24	26. approx. s.e. 12.

Covariate regressions

Variate: CpG6Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.54	0.520

Tables of means (adjusted for covariate)

Variate: CpG6Brain
Covariate: Age_at_Sampling

Grand mean 21.

Gender	F	M
	15.	29.
rep.	14	10

ADvsControl	AD	Control
	12.	33.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.1	14.2

```
  96  "General Analysis of Variance"
  97  BLOCK "No Blocking"
  98  TREATMENTS Gender*ADvsControl
  99  COVARIATE Age_at_Sampling
 100  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
 101   FPROB=yes; PSE=diff] CpG7Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG7Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	905.0	905.0	4.33	0.81	0.051
ADvsControl	1	343.5	343.5	1.65	0.17	0.215
Gender.ADvsControl	1	61.6	61.6	0.30	0.97	0.593
Covariate	1	175.3	175.3	0.84		0.371
Residual	19	3966.7	208.8		0.99	
Total	23	5593.0				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22 35. approx. s.e. 13.

Covariate regressions

Variate: CpG7Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.52	0.566

Tables of means (adjusted for covariate)

Variate: CpG7Brain
Covariate: Age_at_Sampling

Grand mean 22.

Gender	F	M
	15.	31.
rep.	14	10

ADvsControl	AD	Control
	13.	34.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl

rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.7	15.4

```
 102  " Unbalanced Analysis of Variance "
 103  BLOCK "No blocking"
 104  TREATMENT Gender*ADvsControl
 105  COVARIATE Age_at_Sampling
 106  DELETE [REDEFINE=yes] _ausave
 107  AUNBALANCED [PRINT=aovtable,means,screen; PSE=diff;
COMBINATIONS=present; ADJUSTMENT=marginal;\
 108   FACT=3; FPROB=yes] CpG4Brain; SAVE=_ausave
```

Screening of terms in an unbalanced design

Variate: CpG4Brain

Marginal and conditional test statistics and degrees of freedom

degrees of freedom for denominator (full model): 19

term	mtest	mdf	ctest	cdf
Gender	4.97	1	2.89	1
ADvsControl	3.54	1	1.46	1

term	mtest	mdf	ctest	cdf
Gender.ADvsControl	0.31	1	0.31	1

P-values of marginal and conditional tests

term	mprob	cprob
Gender	0.038	0.106
ADvsControl	0.075	0.242

term	mprob	cprob
Gender.ADvsControl	0.581	0.581

Analysis of an unbalanced design using Genstat regression

Variate: CpG4Brain

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Age_at_Sampling	1	224.0	224.0	1.14	0.298

+ Gender	1	973.3	973.3	4.97	0.038
+ ADvsControl	1	286.2	286.2	1.46	0.242
+ Gender.ADvsControl	1	61.7	61.7	0.31	0.581
Residual	19	3721.2	195.9		
Total	23	5266.4	229.0		

Predictions from regression model

Response variate: CpG4Brain

	Prediction
Gender	
F	16.44
M	26.67

Standard error of differences between predicted means 6.310

Predictions from regression model

Response variate: CpG4Brain

	Prediction
ADvsControl	
AD	12.86
Control	31.68

Standard error of differences between predicted means 14.86

Predictions from regression model

Response variate: CpG4Brain

	Prediction	
ADvsControl	AD	Control
Gender		
F	9.80	25.72
M	17.13	40.03

Minimum standard error of difference	8.62
Average standard error of difference	13.74
Maximum standard error of difference	18.02

```
 109  "General Analysis of Variance"
 110  BLOCK "No Blocking"
 111  TREATMENTS ADvsControl
 112  COVARIATE "No Covariate"
 113  ANOVA [PRINT=aovtable,information,means; FACT=32; CONTRASTS=7;
PCONTRASTS=7; FPROB=yes;\
 114   PSE=diff] CpG4Brain
```

Analysis of variance

Variate: CpG4Brain

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
ADvsControl	1	604.0	604.0	2.85	0.105
Residual	22	4662.4	211.9		
Total	23	5266.4			

Message: the following units have large residuals.

units 22	35. approx. s.e. 14.
units 24	37. approx. s.e. 14.

Tables of means

Variate: CpG4Brain

Grand mean 21.

ADvsControl	AD	Control
	17.	27.
rep.	14	10

Standard errors of differences of means

Table	ADvsControl
rep.	unequal
d.f.	22
s.e.d.	6.0

Appendix 5.2: Co-variant analysis within groups and gender

Genstat 64-bit Release 19.1 (PC/Windows 7) 30 January 2019 08:09:57
Copyright 2018, VSN International Ltd.
Registered to: University of Nottingham

```
   1  SET [WORKINGDIRECTORY='D:/People''s data/Graham Seymour/Andrew
Bottley/Kirsty Boden data JanFEB2019'; DIAGNOSTIC=messages]
   2  "Data taken from file: '\
  -3  D:/People''s data/Graham Seymour/Andrew Bottley/Kirsty Boden
data JanFEB2019/Compiled Boden data Jan 2019.xlsx\
  -4  '"
   5  DELETE [REDEFINE=yes] _stitle_: TEXT _stitle_
   6  READ [PRINT=*; SETNVALUES=yes] _stitle_
  10  PRINT [IPRINT=*] _stitle_; JUST=left
```

Data imported from Excel file: D:\People's data\Graham Seymour\Andrew Bottley\Kirsty Boden data JanFEB2019\Compiled Boden data Jan 2019.xlsx
on: 30-Jan-2019 8:10:44
taken from sheet "Rin blood (2)", cells A2:K49

```
  11  DELETE [REDEFINE=yes]
subject,Gender,Age_at_Sampling,CpG1Blood,CpG2Blood,\
  12  CpG3Blood,CpG4Blood,CpG5Blood,CpG6Blood,CpG7Blood,ADvsControl
  13  UNITS [NVALUES=*]
  14  TEXT [NVALUES=48] subject
  15  READ subject
```

Identifier	Minimum	Mean	Maximum	Values	Missing
subject				48	0

```
  22  FACTOR [MODIFY=no; NVALUES=48; LEVELS=2; LABELS=!t('F','M')\
  23  ; REFERENCE=1] Gender
  24  READ Gender; FREPRESENTATION=ordinal
```

Identifier	Values	Missing	Levels
Gender	48	0	2

```
  27  VARIATE [NVALUES=48] Age_at_Sampling; DECIMALS=0
  28  READ Age_at_Sampling
```

Identifier	Minimum	Mean	Maximum	Values	Missing
Age_at_Sampling	37.00	67.19	93.00	48	0

```
  31  VARIATE [NVALUES=48] CpG1Blood; DECIMALS=0
  32  READ CpG1Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG1Blood	12.00	40.53	72.00	48	0

```
  36  VARIATE [NVALUES=48] CpG2Blood; DECIMALS=0
  37  READ CpG2Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG2Blood	11.00	38.55	69.00	48	0

```
  42  VARIATE [NVALUES=48] CpG3Blood; DECIMALS=0
  43  READ CpG3Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG3Blood	16.00	46.16	76.00	48	0

```
  47  VARIATE [NVALUES=48] CpG4Blood; DECIMALS=0
  48  READ CpG4Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG4Blood	16.00	46.20	79.67	48	0

```
  52  VARIATE [NVALUES=48] CpG5Blood; DECIMALS=0
  53  READ CpG5Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG5Blood	9.000	39.13	66.33	48	0

```
  58  VARIATE [NVALUES=48] CpG6Blood; DECIMALS=0
  59  READ CpG6Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG6Blood	20.33	42.48	67.67	48	0

```
  64  VARIATE [NVALUES=48] CpG7Blood; DECIMALS=0
  65  READ CpG7Blood
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG7Blood	18.67	47.93	80.67	48	0

```
  70  FACTOR [MODIFY=no; NVALUES=48; LEVELS=2;
LABELS=!t('AD','Control')\
  71  ; REFERENCE=1] ADvsControl
  72  READ ADvsControl; FREPRESENTATION=ordinal
```

Identifier	Values	Missing	Levels
ADvsControl	48	0	2

```
  75
  76  %PostMessage 1129; 0; 10000001 "Sheet Update Completed"
  77  "General Linear Regression"
  78  MODEL CpG1Blood
  79  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
  80   FACT=9] Gender*ADvsControl*Age_at_Sampling
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Age_at_Sampling + Gender + ADvsControl + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Gender.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	2334.	333.4	1.90	0.096
Residual	40	7030.	175.8		
Total	47	9364.	199.2		

Percentage variance accounted for 11.8
Standard error of observations is estimated to be 13.3.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	3.22

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.57
15	34.	0.81
19	38.	0.40
38	44.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	-31.	133.	-0.23	0.819
Age_at_Sampling	1.37	2.71	0.51	0.616
Gender M	14.	138.	0.10	0.921
ADvsControl Control	-34.	153.	-0.22	0.826
Age_at_Sampling.Gender M	-0.30	2.82	-0.11	0.915
Age_at_Sampling.ADvsControl Control	-0.10	2.85	-0.04	0.971
Gender M .ADvsControl Control	167.	170.	0.98	0.334
Age_at_Sampling.Gender M .ADvsControl Control	-1.78	3.07	-0.58	0.565

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Age_at_Sampling	1	1168.0	1168.0	6.65	0.014
+ Gender	1	182.3	182.3	1.04	0.315
+ ADvsControl	1	11.8	11.8	0.07	0.797
+ Age_at_Sampling.Gender	1	144.1	144.1	0.82	0.371
+ Age_at_Sampling.ADvsControl	1	173.3	173.3	0.99	0.327
+ Gender.ADvsControl	1	595.2	595.2	3.39	0.073
+ Age_at_Sampling.Gender.ADvsControl	1	59.1	59.1	0.34	0.565

Residual	40	7030.3	175.8
Total	47	9364.3	199.2

```
  81  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Age_at_Sampling.Gender.ADvsControl	0.3364	1	0.34	0.565

Residual d.f. 40

```
  82  "General Linear Regression"
  83  MODEL CpG1Blood
  84  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
  85  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
  86   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	1449.	482.9	2.68	0.058
Residual	44	7916.	179.9		
Total	47	9364.	199.2		

Percentage variance accounted for 9.7
Standard error of observations is estimated to be 13.4.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	2.40

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	36.48	5.07	7.20	<.001
Gender M	-1.55	6.14	-0.25	0.801
ADvsControl Control	4.60	6.29	0.73	0.468
Gender M .ADvsControl Control	9.13	8.09	1.13	0.265

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	39.1	39.1	0.22	0.643
+ ADvsControl	1	1180.0	1180.0	6.56	0.014
+ Gender.ADvsControl	1	229.4	229.4	1.28	0.265
Residual	44	7915.7	179.9		
Total	47	9364.3	199.2		

```
  87  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	1.275	1	1.28	0.265

Residual d.f. 44

```
  88  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
  89   Age_at_Sampling
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	4	1612.	403.0	2.24	0.081
Residual	43	7752.	180.3		
Total	47	9364.	199.2		
Change	-1	-163.	163.5	0.91	0.346

Percentage variance accounted for 9.5
Standard error of observations is estimated to be 13.4.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	2.64

Message: the residuals do not appear to be random; for example, fitted values in the range 39.1 to 42.7 are consistently larger than observed values and fitted values in the range 42.7 to 47.7 are consistently smaller than observed values.

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.24

Estimates of parameters

Parameter	estimate	s.e.	t(43)	t pr.
Constant	14.2	23.9	0.59	0.555
Gender M	-1.37	6.15	-0.22	0.824
ADvsControl Control	-11.1	17.6	-0.63	0.533
Gender M .ADvsControl Control	9.54	8.11	1.18	0.246
Age_at_Sampling	0.454	0.477	0.95	0.346

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	39.1	39.1	0.22	0.644
+ ADvsControl	1	1180.0	1180.0	6.55	0.014
+ Gender.ADvsControl	1	229.4	229.4	1.27	0.266
+ Age_at_Sampling	1	163.5	163.5	0.91	0.346
Residual	43	7752.2	180.3		
Total	47	9364.3	199.2		

```
  90  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
  91   ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	5	1798.	359.7	2.00	0.099
Residual	42	7566.	180.1		
Total	47	9364.	199.2		
Change	-1	-186.	186.2	1.03	0.315

Percentage variance accounted for 9.6
Standard error of observations is estimated to be 13.4.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	2.52

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.53
19	38.	0.37
38	44.	0.31

Estimates of parameters

Parameter	estimate	s.e.	t(42)	t pr.
Constant	-16.9	38.8	-0.44	0.666
Gender M	-1.12	6.15	-0.18	0.857
ADvsControl Control	51.0	63.5	0.80	0.427
Gender M .ADvsControl Control	8.81	8.13	1.08	0.285
Age_at_Sampling	1.090	0.786	1.39	0.173
Age_at_Sampling.ADvsControl Control	-1.005	0.989	-1.02	0.315

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	39.1	39.1	0.22	0.644
+ ADvsControl	1	1180.0	1180.0	6.55	0.014
+ Gender.ADvsControl	1	229.4	229.4	1.27	0.265
+ Age_at_Sampling	1	163.5	163.5	0.91	0.346
+ Age_at_Sampling.ADvsControl	1	186.2	186.2	1.03	0.315
Residual	42	7566.0	180.1		
Total	47	9364.3	199.2		

```
  92  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
  93    Gender.Age_at_Sampling
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	6	2275.	379.1	2.19	0.063
Residual	41	7089.	172.9		
Total	47	9364.	199.2		
Change	-1	-477.	476.5	2.76	0.105

Percentage variance accounted for 13.2
Standard error of observations is estimated to be 13.1.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	3.15

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.53
19	38.	0.37
38	44.	0.46

Estimates of parameters

Parameter	estimate	s.e.	t(41)	t pr.
Constant	-98.4	62.1	-1.58	0.121
Gender M	87.6	53.8	1.63	0.111
ADvsControl Control	46.8	62.2	0.75	0.457
Gender M .ADvsControl Control	70.3	37.9	1.86	0.071
Age_at_Sampling	2.75	1.26	2.18	0.035
Age_at_Sampling.ADvsControl Control	-1.64	1.04	-1.58	0.123
Age_at_Sampling.Gender M	-1.81	1.09	-1.66	0.105

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	39.1	39.1	0.23	0.637
+ ADvsControl	1	1180.0	1180.0	6.82	0.013
+ Gender.ADvsControl	1	229.4	229.4	1.33	0.256
+ Age_at_Sampling	1	163.5	163.5	0.95	0.337
+ Age_at_Sampling.ADvsControl	1	186.2	186.2	1.08	0.305
+ Age_at_Sampling.Gender	1	476.5	476.5	2.76	0.105
Residual	41	7089.5	172.9		
Total	47	9364.3	199.2		

```
  94  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
  95    Gender.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG1Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	2334.	333.4	1.90	0.096
Residual	40	7030.	175.8		
Total	47	9364.	199.2		
Change	-1	-59.	59.1	0.34	0.565

Percentage variance accounted for 11.8
Standard error of observations is estimated to be 13.3.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	72.	3.22

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.57
15	34.	0.81
19	38.	0.40
38	44.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	-31.	133.	-0.23	0.819
Gender M	14.	138.	0.10	0.921
ADvsControl Control	-34.	153.	-0.22	0.826
Gender M .ADvsControl Control	167.	170.	0.98	0.334
Age_at_Sampling	1.37	2.71	0.51	0.616
Age_at_Sampling.ADvsControl Control	-0.10	2.85	-0.04	0.971
Age_at_Sampling.Gender M	-0.30	2.82	-0.11	0.915
Age_at_Sampling.Gender M .ADvsControl Control	-1.78	3.07	-0.58	0.565

Parameters for factors are differences compared with the reference level:

Factor	Reference level

Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	39.1	39.1	0.22	0.640
+ ADvsControl	1	1180.0	1180.0	6.71	0.013
+ Gender.ADvsControl	1	229.4	229.4	1.31	0.260
+ Age_at_Sampling	1	163.5	163.5	0.93	0.341
+ Age_at_Sampling.ADvsControl	1	186.2	186.2	1.06	0.309
+ Age_at_Sampling.Gender	1	476.5	476.5	2.71	0.107
+ Age_at_Sampling.Gender.ADvsControl	1	59.1	59.1	0.34	0.565
Residual	40	7030.3	175.8		
Total	47	9364.3	199.2		

```
  96  "General Linear Regression"
  97  MODEL CpG2Blood
  98  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
  99  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 100   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG2Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	1460.	486.6	2.62	0.062
Residual	44	8160.	185.5		
Total	47	9620.	204.7		

Percentage variance accounted for 9.4
Standard error of observations is estimated to be 13.6.

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	33.93	5.15	6.59	<.001
Gender M	-0.90	6.23	-0.14	0.886
ADvsControl Control	5.42	6.38	0.85	0.401
Gender M .ADvsControl Control	8.16	8.21	0.99	0.325

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	41.6	41.6	0.22	0.638
+ ADvsControl	1	1234.8	1234.8	6.66	0.013
+ Gender.ADvsControl	1	183.4	183.4	0.99	0.325
Residual	44	8160.0	185.5		
Total	47	9619.9	204.7		

```
 101  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	0.9892	1	0.99	0.325

Residual d.f. 44

```
 102  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 103
Age_at_Sampling+Gender.Age_at_Sampling+Gender.ADvsControl.Age_at_Sampl
ing
```

Message: term Age_at_Sampling.Gender.ADvsControl cannot be added because term Age_at_Sampling.ADvsControl is marginal to it and is not in the model.

Regression analysis

Response variate: CpG2Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	5	1769.	353.9	1.89	0.116
Residual	42	7851.	186.9		
Total	47	9620.	204.7		
Change	-2	-309.	154.7	0.83	0.444

Percentage variance accounted for 8.7
Standard error of observations is estimated to be 13.7.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	69.	2.92

Message: the residuals do not appear to be random; for example, fitted values in the range 35.5 to 43.3 are consistently larger than observed values and fitted values in the range 43.3 to 46.3 are consistently smaller than observed values.

Message: the following units have high leverage.

Unit	Response	Leverage
3	22.	0.31
38	40.	0.29

Estimates of parameters

Parameter	estimate	s.e.	t(42)	t pr.
Constant	-20.6	43.4	-0.47	0.637
Gender M	47.0	52.0	0.90	0.371
ADvsControl Control	-33.0	31.0	-1.06	0.294
Gender M .ADvsControl Control	41.9	37.0	1.13	0.264
Age_at_Sampling	1.113	0.880	1.26	0.213
Age_at_Sampling.Gender M	-0.98	1.06	-0.93	0.360

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	41.6	41.6	0.22	0.639
+ ADvsControl	1	1234.8	1234.8	6.61	0.014
+ Gender.ADvsControl	1	183.4	183.4	0.98	0.328
+ Age_at_Sampling	1	149.4	149.4	0.80	0.376
+ Age_at_Sampling.Gender	1	160.0	160.0	0.86	0.360
Residual	42	7850.6	186.9		
Total	47	9619.9	204.7		

```
 104  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 105    Gender.ADvsControl.Age_at_Sampling
```

Message: term Age_at_Sampling.Gender.ADvsControl cannot be added because term Age_at_Sampling.ADvsControl is marginal to it and is not in the model.

Regression analysis

Response variate: CpG2Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	5	1769.	353.9	1.89	0.116
Residual	42	7851.	186.9		
Total	47	9620.	204.7		
Change	0	0.	*		

Percentage variance accounted for 8.7
Standard error of observations is estimated to be 13.7.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	69.	2.92

Message: the residuals do not appear to be random; for example, fitted values in the range 35.5 to 43.3 are consistently larger than observed values and fitted values in the range 43.3 to 46.3 are consistently smaller than observed values.

Message: the following units have high leverage.

Unit	Response	Leverage
3	22.	0.31
38	40.	0.29

Estimates of parameters

Parameter	estimate	s.e.	t(42)	t pr.
Constant	-20.6	43.4	-0.47	0.637
Gender M	47.0	52.0	0.90	0.371
ADvsControl Control	-33.0	31.0	-1.06	0.294
Gender M .ADvsControl Control	41.9	37.0	1.13	0.264
Age_at_Sampling	1.113	0.880	1.26	0.213
Age_at_Sampling.Gender M	-0.98	1.06	-0.93	0.360

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	41.6	41.6	0.22	0.639
+ ADvsControl	1	1234.8	1234.8	6.61	0.014

+ Gender.ADvsControl	1	183.4	183.4	0.98	0.328
+ Age_at_Sampling	1	149.4	149.4	0.80	0.376
+ Age_at_Sampling.Gender	1	160.0	160.0	0.86	0.360
Residual	42	7850.6	186.9		
Total	47	9619.9	204.7		

```
 106  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 107   ADvsControl.Age_at_Sampling+Gender.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG2Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	2178.	311.1	1.67	0.144
Residual	40	7442.	186.0		
Total	47	9620.	204.7		
Change	-2	-409.	204.4	1.10	0.343

Percentage variance accounted for 9.1
Standard error of observations is estimated to be 13.6.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	69.	2.95

Message: the following units have high leverage.

Unit	Response	Leverage
3	22.	0.57
15	33.	0.81
19	33.	0.40
38	40.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	-22.	137.	-0.16	0.873
Gender M	5.	142.	0.04	0.971
ADvsControl Control	-31.	157.	-0.20	0.842
Gender M .ADvsControl Control	152.	175.	0.87	0.390
Age_at_Sampling	1.14	2.78	0.41	0.685
Age_at_Sampling.Gender M	-0.12	2.91	-0.04	0.969
Age_at_Sampling.ADvsControl Control				

	-0.03	2.93	-0.01	0.992
Age_at_Sampling.Gender M .ADvsControl Control	-1.69	3.16	-0.54	0.594

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	41.6	41.6	0.22	0.639
+ ADvsControl	1	1234.8	1234.8	6.64	0.014
+ Gender.ADvsControl	1	183.4	183.4	0.99	0.327
+ Age_at_Sampling	1	149.4	149.4	0.80	0.376
+ Age_at_Sampling.Gender	1	160.0	160.0	0.86	0.359
+ Age_at_Sampling.ADvsControl	1	355.1	355.1	1.91	0.175
+ Age_at_Sampling.Gender.ADvsControl	1	53.6	53.6	0.29	0.594
Residual	40	7441.9	186.0		
Total	47	9619.9	204.7		

```
 108  "General Linear Regression"
 109  MODEL CpG3Blood
 110  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
 111  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 112   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG3Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	1683.	561.0	3.02	0.040
Residual	44	8177.	185.8		
Total	47	9860.	209.8		

Percentage variance accounted for 11.4
Standard error of observations is estimated to be 13.6.

Message: the following units have large standardized residuals.

Unit	Response	Residual
23	16.	-2.36

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	40.21	5.15	7.80	<.001
Gender M	0.52	6.24	0.08	0.934
ADvsControl Control	6.70	6.39	1.05	0.300
Gender M .ADvsControl Control	7.44	8.22	0.91	0.370

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	87.0	87.0	0.47	0.497
+ ADvsControl	1	1443.7	1443.7	7.77	0.008
+ Gender.ADvsControl	1	152.4	152.4	0.82	0.370
Residual	44	8177.2	185.8		
Total	47	9860.4	209.8		

```
 113  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	0.8202	1	0.82	0.370

Residual d.f. 44

```
 114  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 115
Age_at_Sampling+Gender.Age_at_Sampling+ADvsControl.Age_at_Sampling+Gen
der.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG3Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	2712.	387.4	2.17	0.058
Residual	40	7149.	178.7		
Total	47	9860.	209.8		

Change	-4	-1029.	257.2	1.44	0.239

Percentage variance accounted for 14.8
Standard error of observations is estimated to be 13.4.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	76.	3.01

Message: the following units have high leverage.

Unit	Response	Leverage
3	27.	0.57
15	48.	0.81
19	44.	0.40
38	50.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	55.	134.	0.41	0.684
Gender M	-79.	140.	-0.56	0.576
ADvsControl Control	-109.	154.	-0.71	0.484
Gender M .ADvsControl Control	263.	172.	1.53	0.134
Age_at_Sampling	-0.30	2.73	-0.11	0.913
Age_at_Sampling.Gender M	1.63	2.85	0.57	0.571
Age_at_Sampling.ADvsControl Control	1.51	2.88	0.52	0.603
Age_at_Sampling.Gender M .ADvsControl Control	-3.75	3.09	-1.21	0.232

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	87.0	87.0	0.49	0.489
+ ADvsControl	1	1443.7	1443.7	8.08	0.007
+ Gender.ADvsControl	1	152.4	152.4	0.85	0.361
+ Age_at_Sampling	1	155.0	155.0	0.87	0.357
+ Age_at_Sampling.Gender	1	131.4	131.4	0.74	0.396
+ Age_at_Sampling.ADvsControl	1	479.5	479.5	2.68	0.109
+ Age_at_Sampling.Gender.ADvsControl	1	262.7	262.7	1.47	0.232
Residual	40	7148.5	178.7		
Total	47	9860.4	209.8		

```
 116  "General Linear Regression"
 117  MODEL CpG4Blood
 118  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
 119  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 120   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG4Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	2553.	850.9	3.28	0.030
Residual	44	11406.	259.2		
Total	47	13959.	297.0		

Percentage variance accounted for 12.7
Standard error of observations is estimated to be 16.1.

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	39.55	6.09	6.50	<.001
Gender M	-0.38	7.37	-0.05	0.959
ADvsControl Control	7.66	7.55	1.01	0.316
Gender M .ADvsControl Control	10.07	9.71	1.04	0.305

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	96.4	96.4	0.37	0.545
+ ADvsControl	1	2177.0	2177.0	8.40	0.006
+ Gender.ADvsControl	1	279.2	279.2	1.08	0.305
Residual	44	11406.2	259.2		
Total	47	13958.8	297.0		

```
 121  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	1.077	1	1.08	0.305

Residual d.f. 44

```
 122  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 123
Age_at_Sampling+Gender.Age_at_Sampling+ADvsControl.Age_at_Sampling+Gen
der.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG4Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	3591.	513.0	1.98	0.082
Residual	40	10368.	259.2		
Total	47	13959.	297.0		
Change	-4	-1039.	259.7	1.00	0.418

Percentage variance accounted for 12.7
Standard error of observations is estimated to be 16.1.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	80.	2.83

Message: the following units have high leverage.

Unit	Response	Leverage
3	23.	0.57
15	42.	0.81
19	40.	0.40
38	56.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	4.	161.	0.03	0.980
Gender M	-31.	168.	-0.18	0.857
ADvsControl Control	-78.	185.	-0.42	0.676
Gender M .ADvsControl Control	206.	207.	0.99	0.326
Age_at_Sampling	0.72	3.29	0.22	0.827
Age_at_Sampling.Gender M	0.63	3.43	0.18	0.856
Age_at_Sampling.ADvsControl Control				

	0.73	3.46	0.21	0.834
Age_at_Sampling.Gender M .ADvsControl Control				
	-2.62	3.72	-0.70	0.487

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	96.4	96.4	0.37	0.546
+ ADvsControl	1	2177.0	2177.0	8.40	0.006
+ Gender.ADvsControl	1	279.2	279.2	1.08	0.306
+ Age_at_Sampling	1	369.0	369.0	1.42	0.240
+ Age_at_Sampling.Gender	1	167.8	167.8	0.65	0.426
+ Age_at_Sampling.ADvsControl	1	374.0	374.0	1.44	0.237
+ Age_at_Sampling.Gender.ADvsControl	1	127.8	127.8	0.49	0.487
Residual	40	10367.5	259.2		
Total	47	13958.8	297.0		

```
 124  "General Linear Regression"
 125  MODEL CpG5Blood
 126  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
 127  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 128   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG5Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	2011.	670.3	4.08	0.012
Residual	44	7237.	164.5		
Total	47	9248.	196.8		

Percentage variance accounted for 16.4
Standard error of observations is estimated to be 12.8.

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	33.24	4.85	6.86	<.001

Gender M	-0.86	5.87	-0.15	0.884
ADvsControl Control	7.95	6.01	1.32	0.193
Gender M .ADvsControl Control	7.69	7.73	1.00	0.325

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	17.8	17.8	0.11	0.744
+ ADvsControl	1	1830.2	1830.2	11.13	0.002
+ Gender.ADvsControl	1	162.9	162.9	0.99	0.325
Residual	44	7236.9	164.5		
Total	47	9247.8	196.8		

```
 129  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	0.9904	1	0.99	0.325

Residual d.f. 44

```
 130  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 131
Age_at_Sampling+Gender.Age_at_Sampling+ADvsControl.Age_at_Sampling+Gen
der.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG5Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	2826.	403.8	2.52	0.031
Residual	40	6422.	160.5		
Total	47	9248.	196.8		
Change	-4	-815.	203.9	1.27	0.298

Percentage variance accounted for 18.4

Standard error of observations is estimated to be 12.7.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	66.	2.78

Message: the following units have high leverage.

Unit	Response	Leverage
3	19.	0.57
15	39.	0.81
19	34.	0.40
38	40.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	38.	127.	0.30	0.766
Gender M	-63.	132.	-0.48	0.635
ADvsControl Control	-89.	146.	-0.61	0.544
Gender M .ADvsControl Control	227.	163.	1.39	0.172
Age_at_Sampling	-0.10	2.59	-0.04	0.970
Age_at_Sampling.Gender M	1.28	2.70	0.48	0.637
Age_at_Sampling.ADvsControl Control	1.21	2.73	0.44	0.661
Age_at_Sampling.Gender M .ADvsControl Control	-3.17	2.93	-1.08	0.286

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	17.8	17.8	0.11	0.741
+ ADvsControl	1	1830.2	1830.2	11.40	0.002
+ Gender.ADvsControl	1	162.9	162.9	1.01	0.320
+ Age_at_Sampling	1	141.5	141.5	0.88	0.353
+ Age_at_Sampling.Gender	1	110.9	110.9	0.69	0.411
+ Age_at_Sampling.ADvsControl	1	375.3	375.3	2.34	0.134
+ Age_at_Sampling.Gender.ADvsControl	1	187.7	187.7	1.17	0.286
Residual	40	6421.5	160.5		
Total	47	9247.8	196.8		

```
 132  "General Linear Regression"
 133  MODEL CpG6Blood
 134  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
 135  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 136   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG6Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	1060.	353.4	3.05	0.039
Residual	44	5106.	116.1		
Total	47	6167.	131.2		

Percentage variance accounted for 11.5
Standard error of observations is estimated to be 10.8.

Message: the following units have large standardized residuals.

Unit	Response	Residual
27	68.	2.44

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	37.48	4.07	9.20	<.001
Gender M	1.17	4.93	0.24	0.814
ADvsControl Control	4.96	5.05	0.98	0.331
Gender M .ADvsControl Control	6.02	6.49	0.93	0.359

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	108.1	108.1	0.93	0.340
+ ADvsControl	1	852.2	852.2	7.34	0.010
+ Gender.ADvsControl	1	99.9	99.9	0.86	0.359
Residual	44	5106.4	116.1		
Total	47	6166.7	131.2		

```
 137  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	0.8605	1	0.86	0.359

Residual d.f. 44

```
 138  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 139
Age_at_Sampling+Gender.Age_at_Sampling+ADvsControl.Age_at_Sampling+Gen
der.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG6Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	1646.	235.1	2.08	0.068
Residual	40	4521.	113.0		
Total	47	6167.	131.2		
Change	-4	-585.	146.3	1.29	0.288

Percentage variance accounted for 13.9
Standard error of observations is estimated to be 10.6.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	65.	2.90

Message: the following units have high leverage.

Unit	Response	Leverage
3	27.	0.57
15	42.	0.81
19	41.	0.40
38	46.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	52.	106.	0.49	0.627
Gender M	-66.	111.	-0.60	0.555
ADvsControl Control	-82.	122.	-0.67	0.504
Gender M .ADvsControl Control				

	196.	137.	1.43	0.159
Age_at_Sampling	-0.30	2.17	-0.14	0.891
Age_at_Sampling.Gender M	1.38	2.27	0.61	0.546
Age_at_Sampling.ADvsControl Control	1.17	2.29	0.51	0.612
Age_at_Sampling.Gender M .ADvsControl Control	-2.86	2.46	-1.16	0.252

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	108.1	108.1	0.96	0.334
+ ADvsControl	1	852.2	852.2	7.54	0.009
+ Gender.ADvsControl	1	99.9	99.9	0.88	0.353
+ Age_at_Sampling	1	113.1	113.1	1.00	0.323
+ Age_at_Sampling.Gender	1	49.3	49.3	0.44	0.513
+ Age_at_Sampling.ADvsControl	1	270.2	270.2	2.39	0.130
+ Age_at_Sampling.Gender.ADvsControl	1	152.8	152.8	1.35	0.252
Residual	40	4521.1	113.0		
Total	47	6166.7	131.2		

```
 140  "General Linear Regression"
 141  MODEL CpG7Blood
 142  TERMS [FACT=9] Gender*ADvsControl*Age_at_Sampling
 143  FIT [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes;\
 144   FACT=9] Gender*ADvsControl
```

Regression analysis

Response variate: CpG7Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	2144.	714.7	2.86	0.047
Residual	44	10988.	249.7		
Total	47	13132.	279.4		

Percentage variance accounted for 10.6
Standard error of observations is estimated to be 15.8.

Message: the residuals do not appear to be random; for example, fitted values

in the range 41.7 to 49.1 are consistently larger than observed values and fitted values in the range 57.6 to 57.6 are consistently smaller than observed values.

Estimates of parameters

Parameter	estimate	s.e.	t(44)	t pr.
Constant	41.19	5.97	6.90	<.001
Gender M	0.49	7.23	0.07	0.947
ADvsControl Control	7.89	7.41	1.06	0.293
Gender M .ADvsControl Control	8.06	9.53	0.85	0.402

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	89.3	89.3	0.36	0.553
+ ADvsControl	1	1875.9	1875.9	7.51	0.009
+ Gender.ADvsControl	1	179.0	179.0	0.72	0.402
Residual	44	10987.9	249.7		
Total	47	13132.0	279.4		

```
 145  RWALD
```

Wald tests for dropping terms

Term	Wald statistic	d.f.	F statistic	F pr.
Gender.ADvsControl	0.7166	1	0.72	0.402

Residual d.f. 44

```
 146  ADD [PRINT=model,summary,estimates,accumulated;
CONSTANT=estimate; FPROB=yes; TPROB=yes]\
 147
Age_at_Sampling+Gender.Age_at_Sampling+ADvsControl.Age_at_Sampling+Gen
der.ADvsControl.Age_at_Sampling
```

Regression analysis

Response variate: CpG7Blood
Fitted terms: Constant + Gender + ADvsControl + Gender.ADvsControl + Age_at_Sampling + Age_at_Sampling.Gender + Age_at_Sampling.ADvsControl + Age_at_Sampling.Gender.ADvsControl

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	7	3453.	493.3	2.04	0.074
Residual	40	9679.	242.0		
Total	47	13132.	279.4		
Change	-4	-1309.	327.2	1.35	0.268

Percentage variance accounted for 13.4
Standard error of observations is estimated to be 15.6.

Message: the following units have large standardized residuals.

Unit	Response	Residual
37	81.	2.94

Message: the following units have high leverage.

Unit	Response	Leverage
3	26.	0.57
15	49.	0.81
19	46.	0.40
38	58.	0.49

Estimates of parameters

Parameter	estimate	s.e.	t(40)	t pr.
Constant	78.	156.	0.50	0.618
Gender M	-112.	162.	-0.69	0.493
ADvsControl Control	-166.	179.	-0.93	0.359
Gender M .ADvsControl Control	296.	200.	1.48	0.146
Age_at_Sampling	-0.76	3.18	-0.24	0.813
Age_at_Sampling.Gender M	2.31	3.31	0.70	0.489
Age_at_Sampling.ADvsControl Control	2.40	3.35	0.72	0.478
Age_at_Sampling.Gender M .ADvsControl Control	-4.42	3.60	-1.23	0.226

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Gender	F
ADvsControl	AD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Gender	1	89.3	89.3	0.37	0.547
+ ADvsControl	1	1875.9	1875.9	7.75	0.008
+ Gender.ADvsControl	1	179.0	179.0	0.74	0.395
+ Age_at_Sampling	1	486.1	486.1	2.01	0.164
+ Age_at_Sampling.Gender	1	131.9	131.9	0.55	0.465
+ Age_at_Sampling.ADvsControl	1	325.2	325.2	1.34	0.253

+ Age_at_Sampling.Gender.ADvsControl	1	365.6	365.6	1.51	0.226
Residual	40	9679.1	242.0		
Total	47	13132.0	279.4		

Appendix 5.3: Co-variant analysis in brain.

Genstat 64-bit Release 19.1 (PC/Windows 7) 29 January 2019 08:16:02
Copyright 2018, VSN International Ltd.
Registered to: University of Nottingham

Genstat Nineteenth Edition
Genstat Procedure Library Release PL27.1

```
   1  SET [WORKINGDIRECTORY='D:/People''s data/Graham Seymour/Andrew
Bottley/Kirsty Boden data JanFEB2019'; DIAGNOSTIC=messages]
   2  "Data taken from file: '\
  -3  D:/People''s data/Graham Seymour/Andrew Bottley/Kirsty Boden
data JanFEB2019/Compiled Boden data Jan 2019.xlsx\
  -4  '"
   5  DELETE [REDEFINE=yes] _stitle_: TEXT _stitle_
   6  READ [PRINT=*; SETNVALUES=yes] _stitle_
  10  PRINT [IPRINT=*] _stitle_; JUST=left
```

Data imported from Excel file: D:\People's data\Graham Seymour\Andrew Bottley\Kirsty Boden data JanFEB2019\Compiled Boden data Jan 2019.xlsx
on: 29-Jan-2019 8:16:14
taken from sheet "Rin Brain (2)", cells A2:K25

```
  11  DELETE [REDEFINE=yes]
AD_Brain_averages,Gender,Age_at_Sampling,CpG1Brain,\
  12
CpG2Brain,CpG3Brain,CpG4Brain,CpG5Brain,CpG6Brain,CpG7Brain,ADvsContro
l
  13  UNITS [NVALUES=*]
  14  TEXT [NVALUES=24] AD_Brain_averages
  15  READ AD_Brain_averages
```

Identifier	Minimum	Mean	Maximum	Values	Missing
AD_Brain_averages				24	0

```
  19  FACTOR [MODIFY=no; NVALUES=24; LEVELS=2; LABELS=!t('F','M')\
  20  ; REFERENCE=1] Gender
  21  READ Gender; FREPRESENTATION=ordinal
```

Identifier	Values	Missing	Levels
Gender	24	0	2

```
  23  VARIATE [NVALUES=24] Age_at_Sampling; DECIMALS=0
  24  READ Age_at_Sampling
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
Age_at_Sampling	44.00	69.38	90.00	24	0	

```
  26  VARIATE [NVALUES=24] CpG1Brain; DECIMALS=0
  27  READ CpG1Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG1Brain	6.333	19.60	63.00	24	0	Skew

```
  30  VARIATE [NVALUES=24] CpG2Brain; DECIMALS=0
  31  READ CpG2Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG2Brain	6.000	18.41	62.00	24	0	Skew

```
  34  VARIATE [NVALUES=24] CpG3Brain; DECIMALS=0
  35  READ CpG3Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG3Brain	7.500	21.50	66.00	24	0	Skew

```
  38  VARIATE [NVALUES=24] CpG4Brain; DECIMALS=0
  39  READ CpG4Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG4Brain	5.000	21.23	64.00	24	0

```
  42  VARIATE [NVALUES=24] CpG5Brain; DECIMALS=0
  43  READ CpG5Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG5Brain	5.000	17.98	61.00	24	0	Skew

```
  46  VARIATE [NVALUES=24] CpG6Brain; DECIMALS=0
  47  READ CpG6Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing
CpG6Brain	5.000	20.80	61.00	24	0

```
  50  VARIATE [NVALUES=24] CpG7Brain; DECIMALS=0
  51  READ CpG7Brain
```

Identifier	Minimum	Mean	Maximum	Values	Missing	
CpG7Brain	5.500	21.55	71.00	24	0	Skew

```
  54  FACTOR [MODIFY=no; NVALUES=24; LEVELS=2;
LABELS=!t('AD','Control')\
  55  ; REFERENCE=1] ADvsControl
  56  READ ADvsControl; FREPRESENTATION=ordinal
```

Identifier	Values	Missing	Levels
ADvsControl	24	0	2

```
  58
  59  %PostMessage 1129; 0; 10000001 "Sheet Update Completed"
  60  "General Analysis of Variance"
  61  BLOCK "No Blocking"
  62  TREATMENTS Gender*ADvsControl
  63  COVARIATE Age_at_Sampling
  64  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  65   FPROB=yes; PSE=diff] CpG1Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG1Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	653.6	653.6	3.42	0.81	0.080
ADvsControl	1	253.3	253.3	1.33	0.17	0.264
Gender.ADvsControl	1	166.8	166.8	0.87	0.97	0.362
Covariate	1	161.6	161.6	0.85		0.369
Residual	19	3631.8	191.1		0.99	
Total	23	4913.4				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22		30. approx. s.e.	12.
units 24		26. approx. s.e.	12.

Covariate regressions

Variate: CpG1Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.50	0.541

Tables of means (adjusted for covariate)

Variate: CpG1Brain

Covariate: Age_at_Sampling

Grand mean 20.

Gender	F	M
	14.	28.
rep.	14	10

ADvsControl	AD	Control
	12.	31.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.4	14.7

```
  66  "General Analysis of Variance"
  67  BLOCK "No Blocking"
  68  TREATMENTS Gender*ADvsControl
  69  COVARIATE Age_at_Sampling
  70  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  71   FPROB=yes; PSE=diff] CpG2Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG2Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	564.2	564.2	3.44	0.81	0.079
ADvsControl	1	311.3	311.3	1.90	0.17	0.184
Gender.ADvsControl	1	118.4	118.4	0.72	0.97	0.406
Covariate	1	196.4	196.4	1.20		0.287
Residual	19	3113.0	163.8		1.01	
Total	23	4298.1				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22		31. approx. s.e.	11.
units 24		25. approx. s.e.	11.

Covariate regressions

Variate: CpG2Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.55	0.501

Tables of means (adjusted for covariate)

Variate: CpG2Brain
Covariate: Age_at_Sampling

Grand mean 18.

Gender	F	M
	13.	26.
rep.	14	10

ADvsControl	AD	Control
	10.	30.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	5.9	13.7

```
  72  "General Analysis of Variance"
  73  BLOCK "No Blocking"
  74  TREATMENTS Gender*ADvsControl
  75  COVARIATE Age_at_Sampling
  76  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  77   FPROB=yes; PSE=diff] CpG3Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG3Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	787.1	787.1	3.74	0.81	0.068
ADvsControl	1	260.0	260.0	1.23	0.17	0.280
Gender.ADvsControl	1	29.4	29.4	0.14	0.97	0.713
Covariate	1	172.8	172.8	0.82		0.376
Residual	19	4000.1	210.5		0.99	
Total	23	5201.8				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22 33. approx. s.e. 13.

Covariate regressions

Variate: CpG3Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.51	0.568

Tables of means (adjusted for covariate)

Variate: CpG3Brain
Covariate: Age_at_Sampling

Grand mean 22.

Gender	F	M
	15.	30.
rep.	14	10

ADvsControl	AD	Control
	14.	32.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.7	15.5

```
  78  "General Analysis of Variance"
  79  BLOCK "No Blocking"
  80  TREATMENTS Gender*ADvsControl
  81  COVARIATE Age_at_Sampling
  82  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  83   FPROB=yes; PSE=diff] CpG4Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG4Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	973.3	973.3	4.97	0.81	0.038
ADvsControl	1	286.2	286.2	1.46	0.17	0.242
Gender.ADvsControl	1	61.7	61.7	0.31	0.97	0.581
Covariate	1	169.1	169.1	0.86		0.365
Residual	19	3721.2	195.9		0.99	
Total	23	5266.4				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	27. approx. s.e. 12.
units 24	30. approx. s.e. 12.

Covariate regressions

Variate: CpG4Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.51	0.548

Tables of means (adjusted for covariate)

Variate: CpG4Brain
Covariate: Age_at_Sampling

Grand mean 21.

Gender	F	M
	14.	31.
rep.	14	10

ADvsControl	AD	Control
	13.	32.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.5	14.9

```
  84  "General Analysis of Variance"
  85  BLOCK "No Blocking"
  86  TREATMENTS Gender*ADvsControl
  87  COVARIATE Age_at_Sampling
  88  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  89   FPROB=yes; PSE=diff] CpG5Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG5Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	462.4	462.4	2.91	0.81	0.104
ADvsControl	1	306.2	306.2	1.93	0.17	0.181
Gender.ADvsControl	1	116.8	116.8	0.74	0.97	0.402
Covariate	1	158.2	158.2	1.00		0.331
Residual	19	3018.3	158.9		1.00	
Total	23	4179.8				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	30. approx. s.e. 11.
units 24	24. approx. s.e. 11.

Covariate regressions

Variate: CpG5Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.49	0.493

Tables of means (adjusted for covariate)

Variate: CpG5Brain
Covariate: Age_at_Sampling

Grand mean 18.

Gender	F	M
	13.	25.
rep.	14	10

ADvsControl	AD	Control
	10.	30.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	5.8	13.4

```
  90  "General Analysis of Variance"
  91  BLOCK "No Blocking"
  92  TREATMENTS Gender*ADvsControl
  93  COVARIATE Age_at_Sampling
  94  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
  95   FPROB=yes; PSE=diff] CpG6Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG6Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	557.2	557.2	3.16	0.81	0.092
ADvsControl	1	344.0	344.0	1.95	0.17	0.179
Gender.ADvsControl	1	48.7	48.7	0.28	0.97	0.605
Covariate	1	193.5	193.5	1.10		0.308
Residual	19	3351.9	176.4		1.00	
Total	23	4511.1				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22	28. approx. s.e. 12.
units 24	26. approx. s.e. 12.

Covariate regressions

Variate: CpG6Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.54	0.520

Tables of means (adjusted for covariate)

Variate: CpG6Brain
Covariate: Age_at_Sampling

Grand mean 21.

Gender	F	M
	15.	29.
rep.	14	10

ADvsControl	AD	Control
	12.	33.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl
rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.1	14.2

```
  96  "General Analysis of Variance"
  97  BLOCK "No Blocking"
  98  TREATMENTS Gender*ADvsControl
  99  COVARIATE Age_at_Sampling
 100  ANOVA [PRINT=aovtable,information,means,covariate; FACT=32;
CONTRASTS=7; PCONTRASTS=7;\
 101   FPROB=yes; PSE=diff] CpG7Brain
```

Message: non-orthogonality between treatment terms. The effects (printed or used to calculate means), the efficiency factor and the sum of squares for each treatment term are for that term eliminating previous terms in the TREATMENT formula (as well as covariates)and ignoring subsequent terms.

Analysis of variance (adjusted for covariate)

Variate: CpG7Brain
Covariate: Age_at_Sampling

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Gender	1	905.0	905.0	4.33	0.81	0.051
ADvsControl	1	343.5	343.5	1.65	0.17	0.215
Gender.ADvsControl	1	61.6	61.6	0.30	0.97	0.593
Covariate	1	175.3	175.3	0.84		0.371
Residual	19	3966.7	208.8		0.99	
Total	23	5593.0				

Information summary

Model term	e.f.	non-orthogonal terms
ADvsControl	0.901	Gender

Message: the following units have large residuals.

units 22 35. approx. s.e. 13.

Covariate regressions

Variate: CpG7Brain

Covariate	coefficient	s.e.
Age_at_Sampling	-0.52	0.566

Tables of means (adjusted for covariate)

Variate: CpG7Brain
Covariate: Age_at_Sampling

Grand mean 22.

Gender	F	M
	15.	31.
rep.	14	10

ADvsControl	AD	Control
	13.	34.
rep.	14	10

Message: table of means for Gender.ADvsControl cannot be calculated (contains mutually non-orthogonal components).

Standard errors of differences of means

Table	Gender	ADvsControl

rep.	unequal	unequal
d.f.	19	19
s.e.d.	6.7	15.4

```
 102  " Unbalanced Analysis of Variance "
 103  BLOCK "No blocking"
 104  TREATMENT Gender*ADvsControl
 105  COVARIATE Age_at_Sampling
 106  DELETE [REDEFINE=yes] _ausave
 107  AUNBALANCED [PRINT=aovtable,means,screen; PSE=diff;
COMBINATIONS=present; ADJUSTMENT=marginal;\
 108   FACT=3; FPROB=yes] CpG4Brain; SAVE=_ausave
```

Screening of terms in an unbalanced design

Variate: CpG4Brain

Marginal and conditional test statistics and degrees of freedom

degrees of freedom for denominator (full model): 19

term	mtest	mdf	ctest	cdf
Gender	4.97	1	2.89	1
ADvsControl	3.54	1	1.46	1

term	mtest	mdf	ctest	cdf
Gender.ADvsControl	0.31	1	0.31	1

P-values of marginal and conditional tests

term	mprob	cprob
Gender	0.038	0.106
ADvsControl	0.075	0.242

term	mprob	cprob
Gender.ADvsControl	0.581	0.581

Analysis of an unbalanced design using Genstat regression

Variate: CpG4Brain

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Age_at_Sampling	1	224.0	224.0	1.14	0.298

+ Gender	1	973.3	973.3	4.97	0.038
+ ADvsControl	1	286.2	286.2	1.46	0.242
+ Gender.ADvsControl	1	61.7	61.7	0.31	0.581
Residual	19	3721.2	195.9		
Total	23	5266.4	229.0		

Predictions from regression model

Response variate: CpG4Brain

	Prediction
Gender	
F	16.44
M	26.67

Standard error of differences between predicted means 6.310

Predictions from regression model

Response variate: CpG4Brain

	Prediction
ADvsControl	
AD	12.86
Control	31.68

Standard error of differences between predicted means 14.86

Predictions from regression model

Response variate: CpG4Brain

	Prediction	
ADvsControl	AD	Control
Gender		
F	9.80	25.72
M	17.13	40.03

Minimum standard error of difference	8.62
Average standard error of difference	13.74
Maximum standard error of difference	18.02

```
 109  "General Analysis of Variance"
 110  BLOCK "No Blocking"
 111  TREATMENTS ADvsControl
 112  COVARIATE "No Covariate"
 113  ANOVA [PRINT=aovtable,information,means; FACT=32; CONTRASTS=7;
PCONTRASTS=7; FPROB=yes;\
 114   PSE=diff] CpG4Brain
```

Analysis of variance

Variate: CpG4Brain

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
ADvsControl	1	604.0	604.0	2.85	0.105
Residual	22	4662.4	211.9		
Total	23	5266.4			

Message: the following units have large residuals.

units 22	35. approx. s.e. 14.
units 24	37. approx. s.e. 14.

Tables of means

Variate: CpG4Brain

Grand mean 21.

ADvsControl	AD	Control
	17.	27.
rep.	14	10

Standard errors of differences of means

Table	ADvsControl
rep.	unequal
d.f.	22
s.e.d.	6.0

Appendix 6: Professional internships for PhD students reflection form

Name of Organization

The University of Nottingham Intellectual Property Commercialisation Office

Details of Placement

Please describe your main activities during the placement

During the three month internship I worked alongside members of the IP commercialisation office to develop commercialization strategies and market research reports for intellectual property emerging from the research priority area 'regenerative medicine and stem cells' . The work involved completing extensive research into new technologies and products being developed by researchers at the University of Nottingham using both meetings with inventors and internet searches. I was then tasked with identifying reasonable methods for protecting any intellectual property, such as through patents. This involved discussion with licensing executives, patent attorneys and lawyers. Following the identification of a promising technology I then went onto identify companies that might be potentially interested in licensing the technology and developing it further. In some cases I also identified funding options for ideas that required further development. I then collated all of the information I had collected into large comemrcialisation strategy reports.

Placement Achievements

Please detail all outcomes from the placement, including any publications, presentations given and reports written etc.

As mentioned above, part of my role during the work placement was to write commercialization strategy reports for promising emerging technologies. As part of this report I conducted extensive research into potentially commercially relevant research occurring at the University. I also identified possible sources of funding to extend work if needed, performed IP due diligence and identified the type of products that would be interesting for big pharma, mid pharma and small companies. The report also included extensive market research into the regenerative medicine and stem cell market in the UK, USA, Australia and Asia and the regulatory requirements needed for products emerging from the research area in order for them to be commercially viable.

Skill development

Has this Placement helped you developed any new skills or enhanced your previous skill set?

During the placement I developed multiple skills. During my time there I have working as part of a large team which included licensing executives, patent attorneys, business developers, funding specialists and lawyers. Therefore I developed both my team work skills and also my communication skills. This also allowed me to further my interpersonal skills. In addition, I was managing multiple projects simultaneously. This therefore improved both my time management and organizational skills.

Future Work

Has this Placement influenced your future career aspirations? If so, in what way?

(150-200 words)

The placement made me aware of many other career options open to me involving innovation and scientific discovery outside of the lab. It also gave me a unique opportunity to appreciate how important commercialization of research is. I am very glad that I now have this awareness regardless of whether this informs my future career directly.

Appendix 6: Published work

Journal of Alzheimer's Disease Reports 1 (2017) 97–108
DOI 10.3233/ADR-170015
IOS Press

Methylation Profiling RIN3 and MEF2C Identifies Epigenetic Marks Associated with Sporadic Early Onset Alzheimer's Disease

Kirsty A. Boden[a], Imelda S. Barber[b], Naomi Clement[b], Tulsi Patel[b], Tamar Guetta-Baranes[b], Keeley J. Brookes[b], Sally Chappell[b], Jim Craigon[a], Natalie H. Chapman[a], ARUK Consortium, Kevin Morgan[b], Graham B. Seymour[a] and Andrew Bottley[a,*]
[a]*School of Biosciences, University of Nottingham, Nottingham, UK*
[b]*Schools of Life Sciences, University of Nottingham, Nottingham, UK*

Accepted 24 July 2017

Abstract. A number of genetic loci associate with early onset Alzheimer's disease (EOAD); however, the drivers of this disease remains enigmatic. Genome wide association and *in vivo* modeling have shown that loss-of-function, e.g., ABCA7, reduced levels of SIRT1 and MEFF2C, or increased levels of PTK2þ confer risk or link to the pathogenies. It is known that DNA methylation can profoundly affect gene expression and can impact on the composition of the proteome; therefore, the aim of this study is to assess if genes associated with sporadic EOAD (sEOAD) are differentially methylated. Epi-profiles of DNA extracted from blood and cortex were compared using a pyrosequencing platform. We identified significant group-wide hypomethylation in AD blood when compared to controls for 7 CpGs located within the 3'UTR of RIN3 (CpG1 $p = 0.019$, CpG2 $p = 0.018$, CpG3 $p = 0.012$, CpG4 $p = 0.009$, CpG5 $p = 0.002$, CpG6 $p = 0.018$, and CpG7 $p = 0.013$, respectively; AD/Control $n = 22/26$; Male/Female $n = 27/21$). Observed effects were not gender specific. No group wide significant differences were found in the promoter methylation of *PTK2*þ, *ABCA7*, *SIRT1*, or *MEF2C*, genes known to associate with late onset AD. A rare and significant difference in methylation was observed for one CpG located upstream of the *MEF2C* promoter in one AD individual only (22% reduction in methylation, $p = 2.0E-10$; Control $n = 26$, AD $n = 25$, Male/Female $n = 29/22$). It is plausible aberrant methylation may mark sEOAD in blood and may manifest in some individuals as rare epi-variants for genes linked to sEOAD.

Keywords: Alzheimer's disease, epigenetics, methylation, sporadic early onset

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative condition and the leading cause of dementia among the elderly [1]. AD can segment into two classifications depending on age at diagnosis, raising the possibility of different initiators of disease for each group. Late onset AD (LOAD) is the most prevalent form, a condition which affects those over the age of 65 and accounts for approximately 90 to 95% of all those diagnosed with AD. Early onset AD (EOAD) typically affects those under the age of 65 and is much rarer, accounting for just 5% to 10% of all AD cases [2].

The genetic explanation for LOAD is complex with a growing number of different genes implicated, reviewed by [3], yet none in isolation fully accounts for disease susceptibility. EOAD however can be sub-divided into two groups, familial and sporadic. Familial AD (fAD) is thought to be easier to explain, driven by mutations within a growing list of different genes, including but not limited to amyloid protein precursor (*APP*) or presenilin (*PSEN1*

*Correspondence to: Andrew Bottley, School of Biosciences, University of Nottingham, Nottingham NG7 2UH, UK. E-mail: Andrew.bottley@nottingham.ac.uk.

ISSN 2542-4823/17/$35.00

and *PSEN2*) genes [2, 4]. Sporadic EOAD (sEOAD), although like fAD occurs in individuals under the age of 65, has proven much harder to pin to any one gene or pathway. Comparative transcriptomic analysis of sEOAD and fAD brain tissue identified over 3000 differentially expressed genes [5], suggesting sEOAD and fAD may be distinct neurodegenerative processes.

An increasing number of genes link with LOAD, an association identified through comparative genome wide association data analysis (GWAS) [6]. These new studies link numerous pathways to pathogenesis, correlating genes and cellular processes not previously considered relevant [6]. Current models suggest LOAD may be a complex multifactorial condition driven by a number of different genes located within independent pathways [3]. Although each genetic variant links to disease, frequency within the wider population can vary; even the most common LOAD risk APOE ε4 allele only presents in 40% of LOAD patients. Diversity in the frequency of genetic drivers of AD may be considered surprising given the similarities in the pathology of the disease collectively among AD sufferers. One plausible hypothesis is that these genes could be differentially regulated within groups of AD suffers, existing within the population as variants at an epigenetic or regulatory level. Some data exists to support this notion. Differences in DNA methylation of AD associated genes have been identified in DNA samples taken from both the blood and the brain of patients diagnosed with LOAD [7–12]. Notably *APP* has been found to be differentially methylated within the promoter region in LOAD, while sirtuin1 (*SIRT1*) has been shown to be differential methylated in a Chinese AD population [13–15]. Little, however, is currently known about the epigenetic status of genes that associate with sEOAD.

Stable changes in DNA methylation can be induced by a person's lifestyle or environment and may progressively accumulate over many years. Levels of methylation may direct gene expression leading to changes in the levels of proteins, which may cause, drive, or exacerbate AD disease symptoms [16, 17]. Initial epi-profiling already undertaken using LOAD samples suggests this hypothesis may be correct for later onset AD [7]; however, little work to date has been undertaken to investigate any putative epigenetic link with sEOAD. We therefore sought to investigate if those LOAD genes genetically associated with AD occur as epi-variants in sEOAD.

Genes previously associated with LOAD through GWAS profiling are likely candidates for epi-profiling with variation encoded by methylation. Genes already profiled using genetic association with LOAD include but are not limited to *PTK2*þ, *ABCA7*, and *MEF2C* [18]. These three genes are all implicated in a range of pathways associated with AD pathology. PTK2þ functions in memory formation and cell proliferation and survival [19, 20], ABCA7 has a role in regulating AþPP processing and inhibiting amyloid-þ [21–23], while MEF2C is a transcription factor involved in preventing excessive synapse formation [24].

Another gene, a candidate target of interest for both genetic and epigenetic analysis, is *SIRT1*. This gene is functionally implicated in AD disease pathology in a number of studies [25–28]. Disease linked differential methylation of a non-coding region of *SIRT1* has already been demonstrated in peripheral blood leukocytes of a Chinese LOAD population [13], therefore a comparative approach investigating sEOAD could be informative. Previous research by Hou et al. [13] identified two CpGs outside of the promoter CpG Island (CGI) to be significantly hypermethylation in LOAD samples; this suggests that epi-profiling should not just be restricted to CpG repeats, exploring the wider gene region may reveal interesting information around regulation linked to disease.

There is some evidence to suggest that investigating regulatory features other than the promoter region could be productive. A 3'UTR CGI found within the *RIN3* gene is located on chromosome 14 and sits between two AD associated genes *SLC24A4* and *LGMN* [18]. *RIN3* is interesting in isolation as this gene associates with AD pathology through its interaction with *BIN1*, a gene that shows the second most significant LOAD score after *APOE* [29]. Further it has also been found to be differentially methylated in LOAD [8]. It is likely that BIN1 and RIN3 interact in the process of endocytosis negatively effecting amyloid trafficking [30, 31]. Since 3' UTR methylation has recently emerged as an important epigenetic mark influencing gene expression, transcriptional elongation and splicing it is a valid hypothesis that this CGI may represent an informative target for epi-profiling [32–34].

To date most transcriptomic or proteomic analysis is conducted using brain tissue. While profiles obtained from brain material may be informative about the direct epigenetic consequences of AD pathology, this tissue it is not a viable source for the identification of new biomarkers. Also it is well

established that brain methylation can differ significantly between the different regions of the brain [12]. Additional technical limitations include a restriction on test subject numbers through limited availability of tissue. Interestingly leukocytes are thought to acquire epigenetic markers linked to disease as they transit through the afflicted tissue. In limited studies to date, peripheral blood DNA methylation has been shown to be representative of brain methylation [35]. Blood based epigenetic biomarkers may be a useful addition to the AD diagnostic tool kit available, while also informing on the fundamental biology of sEOAD.

This study profiles comparative levels of methylation present in individuals presenting with sEOAD. Through profiling DNA extracted from both blood and brain, we have identified variation in the methylation of key genes that may link to AD and the progression of sEOAD. Differences in methylation may be consistent across many CpGs or could be restricted to a single bp.

MATERIALS AND METHODS

DNA samples

sEOAD DNA samples were obtained from the Alzheimer's Research UK Consortium DNA Bank, a resource curated by the University of Nottingham and were either extracted from leukocytes ($n = 51$) or brain (cortex) ($n = 24$). In total 40 sEOAD and 21 control samples were used. 25 blood and 14 cortex brain AD samples and 26 blood and 10 brain control samples were used (See Table 1 for further information). DNA samples used in this study have been genetically tested by Alzheimer's Research UK Consortium DNA Bank and none contained fAD mutations.

All samples used in this study were received with informed consent and experimental procedures were approved by the local ethics committee, Nottingham Research Ethics Committee 2 (REC reference 04/Q2404/130). All experimental procedures were conducted in accordance with approved guidelines.

DNA extraction

Approximately 200 mg of cortex tissue was chopped finely on dry ice then transferred to a 1.5 ml Eppendorf tube. This was then incubated overnight (~18 h, shaking at 380 rpm, 50˚C) with 500 µl AL lysis buffer and 50 µl proteinase K (both Qiagen) and 10 µl RNase A. Next, 500 µl of refrigerated phenol chloroform was added to the sample (Sigma), which was mixed by inverting before being subjected to centrifugation for 5 min at 13,000 rpm. The top phase of the resultant sample was then removed to a clean lock phase-gel 2 ml Eppendorf, and the addition of phenol chloroform, mixing and centrifugation repeated. The top phase of this was removed, and has 3 M sodium acetate (pH 5.2) added to it in a 1 :9 ratio of sodium acetate to the sample. Chilled 100% ethanol was then added, at an equal volume to the sample, to precipitate out the DNA. Following centrifugation at 13,000 rpm for 15 min, a wash was performed using 500 µl 70% ethanol, before another centrifugation step, for 15 min (13,000 rpm). The remaining ethanol was then discarded and the pellet air-dried, before resuspension in 100 µl 1xTE buffer, by heating to 50˚C for 1 h.

Bisulphite treatment

First DNA was bisulphite-treated using the EpiTect Bisulphite Conversion Kit (Qiagen, Germany), following the manufacturer's instructions. Bisulphite treatment converted any non-methylated cytosines to uracil; therefore, in the following PCR product, non-methylated cytosines became thymine, allowing pyrosequencing of the PCR product to result in identification of cytosine methylation. 500 ng of genomic DNA was bisulphite-treated for each patient sample and 10 ng of converted DNA was used in subsequent PCRs.

In silico methods

Target regions were identified using the UCSC Genome Browser (University of California, Santa Cruz, CA, USA)(http://www.genome.ucsc.edu/index).

Table 1
Shows clinical features of AD and control samples

Group	Sex M/F	Age at death	Age at onset of disease	Age at sampling

Blood	AD	17/8	57.95 (SD 3.28)	48.6 (SD 2.98)	47.2 (SD 2.28)
	Control	13/13	N/A	N/A	82.8 (SD 4.5)
Brain	AD	4/10	59.35 (SD 6.95)	48.85 (SD 3.32)	59.35 (SD 6.95)
	Control	6/4	84.2 (SD 3.55)	N/A	84.2 (SD 3.55)

Pyrosequencing primers were then designed for these regions using the PyroMark Assay Design software version 2.0 (Qiagen) and obtained from Eurofins MWG. Pyrosequencing assays were designed using the PyroMark Q24 software.

Polymerase chain reaction (PCR)

PCRs were conducted using the PyroMark PCR Kit (Qiagen, Germany) following manufactures instructions. PCRs contained 12.5 µl PyroMark PCR Mastermix 2x (Qiagen), 2.5 µl CoralLoad Concentrate 10x (Qiagen), 0.25 ng of each PCR primer, and 10 ng DNA, made up to 25 µl with RNase-free water. Details of PCR primers can be found in Supplementary Table 1; all PCR primer pairs contained a biotinylated primer to result in a biotinylated PCR product which could be captured by sepharose coated beads during pyrosequencing. PCR programs varied depending on primers used; details can be found in Supplementary Table 2. PCR product quality was assessed by running 5 µl of PCR product on a 1% agarose gel prior to pyrosequencing.

Pyrosequencing

A pyrosequencing platform Q24 pyrosequencer (Qiagen, Germany) with PyroMark Gold Q24 Reagents (Qiagen, Germany) was used to determine percentage methylation. Pyrosequencing was carried out following manufacturer's instructions/protocol. Pyrosequencing results were interpreted using the PyroMark Q24 software.

Statistical analysis

For all the methylation percentage data, one-way analyses of variance were used to test the blood and brain tissue data for differences between the AD and control groups of subjects. Each of the cytosines were analyzed separately as were the blood and brain tissue data sets. When multiple values were recorded for the same cytosine from the same subject, the values were averaged to produce one value per person for each cytosine. These averages were then analyzed using the Analysis of Variance routines within the Genstat18 statistical package.

The residuals from each analysis were plotted to check the assumptions of normality and homogeneity of residuals, and also to highlight possible rare variants. For rare variants, the probability of obtaining residual values as, or more, extreme than the one observed was calculated from the standard normal distribution curve. To achieve this, the residual deviation between the individual outlier and the mean of its group; and the overall standard deviation from the rerun analysis were used as input to the Genstat Normal probability distribution routine.

RESULTS

PTK2β, ABCA7, SIRT1, and the MEF2C promoter CGI showed no significant methylation in AD

*PTK2*þ, *ABCA7*, and *MEF2C*: Promoter regions of AD candidate genes *PTK2*þ, *ABCA7*, *MEF2C*, and *SIRT1* were chosen as targets for analysis (see Supplementary Table 1 and Supplementary Figure 1); initial gene targets were selected based on previously published GWAS risk analysis data [18] or in the case of *SIRT1*, published promoter methylation data [13]. Previous research identified *PTK2*þ, *ABCA7*, and *MEF2C* to be associated with LOAD through GWAS [18, 36]; however, to date few experiments have investigated the contribution of these genes to sEOAD. These genes may be considered prime candidates for regulation via an epigenetic mechanism; therefore, we first chose to profile methylation at these loci in blood and brain tissue taken from sEOAD patients.

For *PTK2*þ, *ABCA7*, and *MEF2C*, the CpG island closest to the start site of transcription was chosen for analysis (details in Supplementary Figures 1 and 2). In total, we profiled 5, 8, and 4 CpGs for *PTK2*þ, *ABCA7*, and *MEF2C*, respectively; however, unexpectedly, we failed to detect any significant difference in the average regional level of methylation across genes investigated in sEOAD blood or brain derived samples (Fig. 1) ($p > 0.05$). It is interesting to note that for *PTK2*þ, *ABCA7*, and *MEF2C*, the regions targeted showed relatively low average levels of methylation in cortex and blood tissue of both AD and control samples.

Given the absence of variation we observed for *PTK2*þ, *ABCA7*, and *MEF2C*, we opted to profile levels of methylation at an AD linked loci previously reported to be differentially methylated [13]. *SIRT1* associates with AD pathology through its role in attenuating Aþ toxicity and preventing tau cytotoxicity [25–28]; therefore, it is reasonable to suggest

aberrant methylation may impede this protective effect. As this previous study was undertaken using samples obtained from LOAD patients, we chose

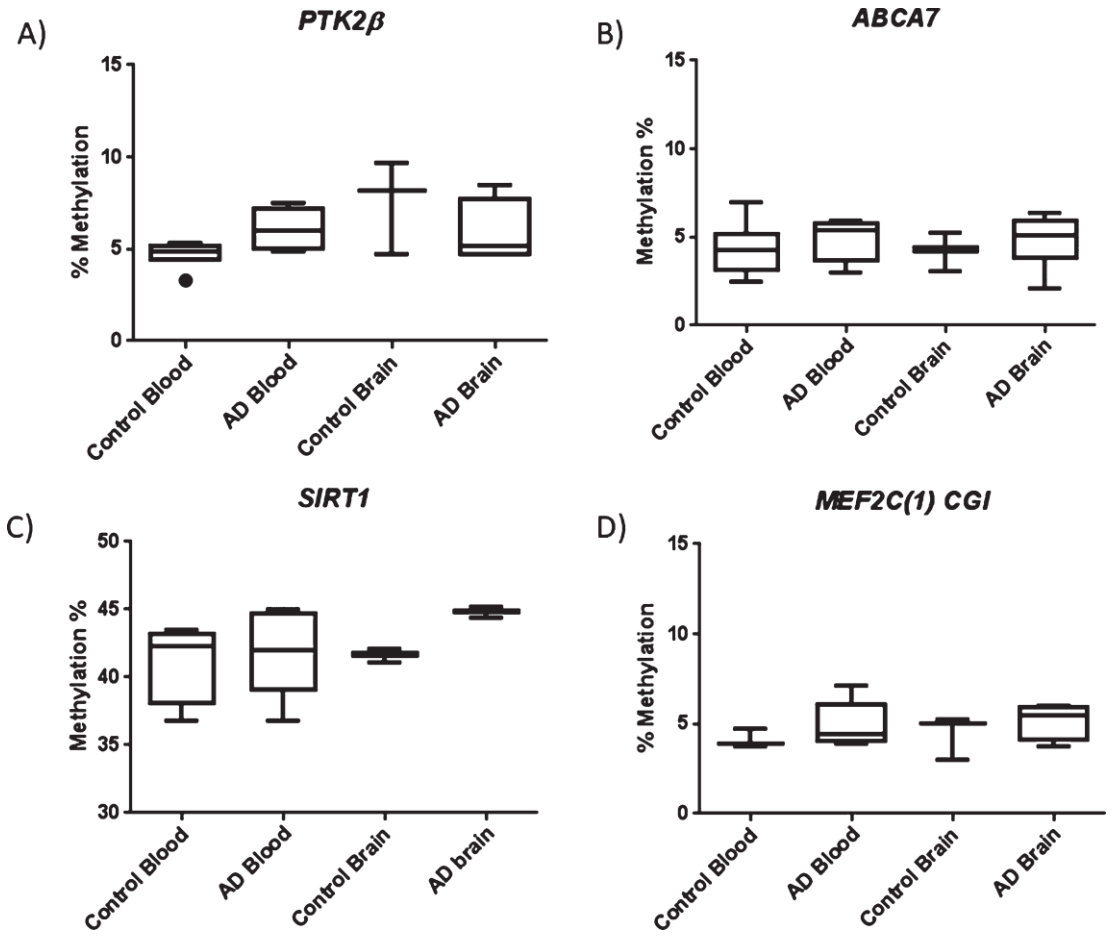


Fig. 1. Box plots showing data representing average methylation across the regions investigated for *PTK2*þ (A), *ABCA7* (B), *SIRT1* (C), and the *MEF2C* promoter CGI (D). No significant difference in methylation was observed between AD and control in either blood or brain tissue. Box plots represent median with 25th and 75th percentile as edges, the whiskers either show the lowest and highest values or extend to the first quartile minus the interquatile range multiplied by 1.5 and up to the third quartile add the interquartile range times 1.5, depending on which values are largest or smallest (for the top and bottom whiskers respectively). For *PTK2*þ control blood $n = 6$, AD blood $n = 5$, control brain $n = 3$, and AD brain $n = 4$. For *ABCA7* control blood $n = 7$, AD blood $n = 4$, control brain $n = 3$, AD brain = 6. For *SIRT1* control blood $n = 4$, AD blood $n = 5$, control brain $n = 3$, and AD brain $n = 3$. For the *MEF2C* CGI region Control blood $n = 3$, AD blood $n = 5$, Control brain $n = 3$, AD brain $n = 4$.

to investigate if the epigenetic variation previously reported could also be observed in sEOAD patients. In order to directly compare levels of methylation for our sEOAD sample library with a previously published candidate AD associated epi-loci, we profiled two CpGs reported to be hypermethylated in LOAD peripheral blood [13]. Where tested, we found no significant difference in methylation compared to controls at either CpG site in blood.

Methylation at the resolution of individual CpGs

We detected no difference in the average regional level of methylation for each of the four candidate genes tested and no significant difference in methylation was identified for any individual CpG duplet within CpG islands upstream of *ABCA7*, *SIRT1* or *MEF2C* in either tissue ($p > 0.05$; Supplementary Figure 3). However, interestingly, one specific CpG site investigated within the *PTK2*þ promoter CGI showed significant hypermethylation in AD blood ($p = 0.04$) but not AD cortex ($p = 0.47$) (Fig. 2). The average methylation found at CpG1 in AD blood was 8.4% compared to 5% in control blood. Although statistically significant, and all samples were completed in duplicate, it should be noted that this difference is relatively small [37].

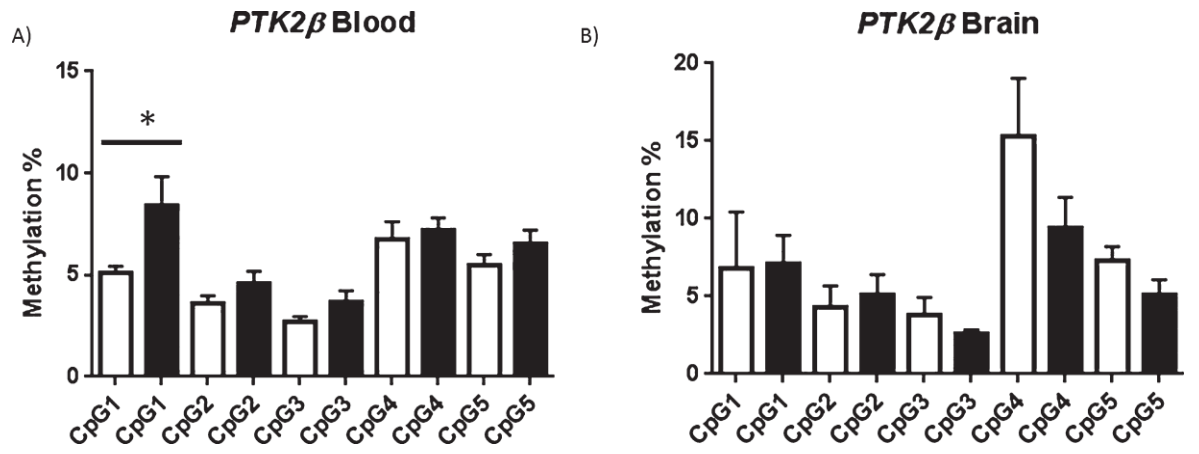


Fig. 2. *PTK2*þ shown at resolution of individual CpG: For the target region investigated in the *PTK2*þ promoter CpG1 was found to be significantly hypermethylated in AD Blood but not AD brain. Average methylation at each CpG investigated in controls (white bar) and AD (black bar) in blood (A) and brain (B) is shown, error bars represent S.E.M. $^{*}p < 0.05$ (one tailed *T*-test). For blood control $n = 6$ AD= 5, for brain control $n = 3$, AD $n = 4$.

MEF2C Epi-variant CpG upstream of promoter CGI investigated

CpGs located within the CpG island upstream of the transcription start site of *MEF2C(1)*, (see Fig. 3), show equivalent levels of methylation between AD and control samples. However, it is plausible that methylation located at other upstream CpGs outside of this location may be crucial in determining either levels of transcription or the structure of the transcript expressed, therefore a further CpG site was targeted (shown as *MEF2C(2)* in Fig. 3). As a group wide average, there was no evidence of a difference between the AD and control groups in either the blood ($p = 0.708$) or cortex ($p = 0.593$) data set (Fig. 3). Average levels of methylation were substantially higher than those of the other genes tested; average methylation of 91% in blood and 86% in brain average. No significant difference in methylation was observed between the tissue types ($p > 0.05$) (Supplementary Figure 4).

Intriguingly however, we identified a difference in methylation at this CpG in one AD patient blood sample only (Fig. 4). The probability of observing, by chance, a value as low or lower than 70 for the AD group in the blood data was calculated using one-way analyses of variance 2.0E-10. Average methylation at the CpG of 70% was recorded for the individual M341 (blood sample). This represented a 22% reduction in methylation when compared to average methylation in blood at this site in the other samples tested (total $n = 51$). Average methylation was calculated using data obtained from at least two technical repetitions in order to exclude the possibility of a technical error. It is established that rare genetic differences within this gene can be associated with sEOAD, it could be suggested that rare epigenetic variation at this site may also associate as a risk factor.

In order to eliminate the possibility of a genetic cause of the methylation observed in this experiment, samples were genotyped for the published GWAS SNP; no association with this genotype was found.

RIN3 3'UTR showed significant hypomethylation in AD blood but not AD brain

RIN3 has been proven to interact with the AD associated protein BIN1, which in turn has also been significantly associated with AD via GWAS. This gene has also been shown to be differentially methylated in LOAD with an established link to AD pathology [8, 9, 38]. We chose to investigate a CpG island identified within the 3'UTR of *RIN3* (see Supplementary Table 1 and Supplementary Figure 2). Pyrosequencing covered seven CpGs in total within this region (shown in Supplementary Figure 2) and average levels of methylation was calculated for AD blood, control blood, AD brain, and control brain (Fig. 5) across all seven CpGs tested. AD blood samples showed hypomethylation when compared to control blood (see Fig. 5).

For each of the cytosines CPG1-7 data (Fig. 6) from the blood samples, there was a significant difference

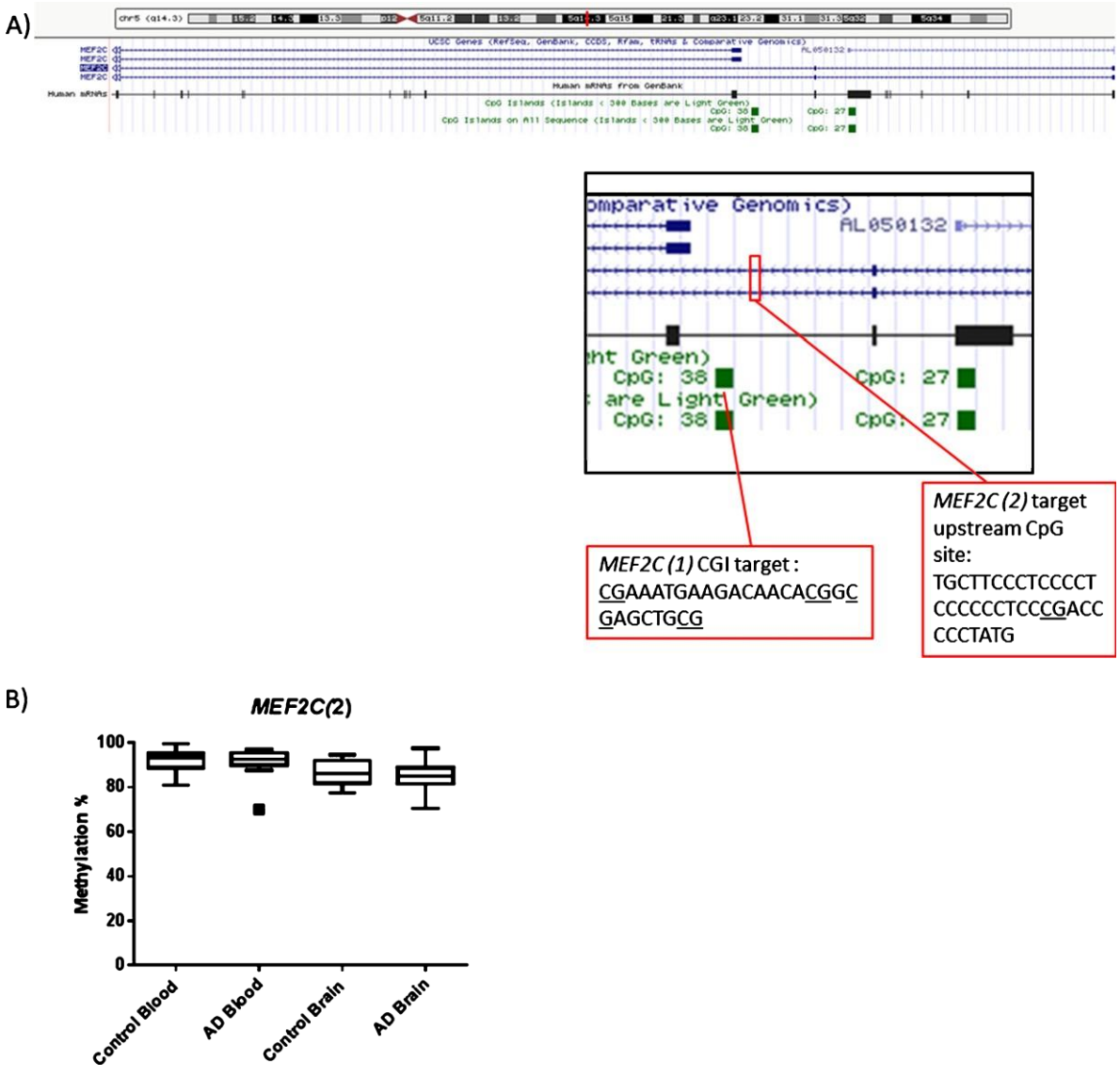


Fig. 3. (A) Diagram showing the location of *MEF2C(1)* and *MEF2C(2)* targets within the *MEF2C* gene. (B) In terms of the second region of *MEF2C* investigated no obvious collective group wide differences between Control and AD at this CpG site were observed. Box plots shows median with 25th and 75th percentile as edges, the whiskers extend to the first quartile minus the interquatile range multiplied by 1.5 and up to the third quartile add the interquartile range time 1.5. Control blood $n = 26$, AD blood $n = 25$, Control brain $n = 10$, AD brain $n = 14$.

between the AD and control group in %methylation ($p < 0.05$) (CpG1 $p = 0.019$, CpG2 $p = 0.018$, CpG3 $p = 0.012$, CpG4 $p = 0.009$, CpG5 $p = 0.002$, CpG6 $p = 0.018$ and CpG7 $p = 0.013$, respectively); the AD group being lower than the control for all CPG 1–7 and there were no extreme outliers. The p values from the analyses were also compared to Bonferroni adjusted critical p values of 0.05/7 = 0.007 to adjust for the fact that 7 CGP data-sets were analyzed separately from each individual. CpG5 retained significance ($p = 0.002$) after strict Bonferroni correction for multiple comparisons, suggesting that aberrant methylation is centered within this region and spreads outwards across neighboring CpGs.

There was no evidence ($p > 0.05$) of a difference between groups in % methylation for any of the cytosines CPG1-7 from the brain samples. Average methylation across this region in control blood was 47.8% while average methylation in AD blood was 37.29% a reduction of 10%. No significant differ-

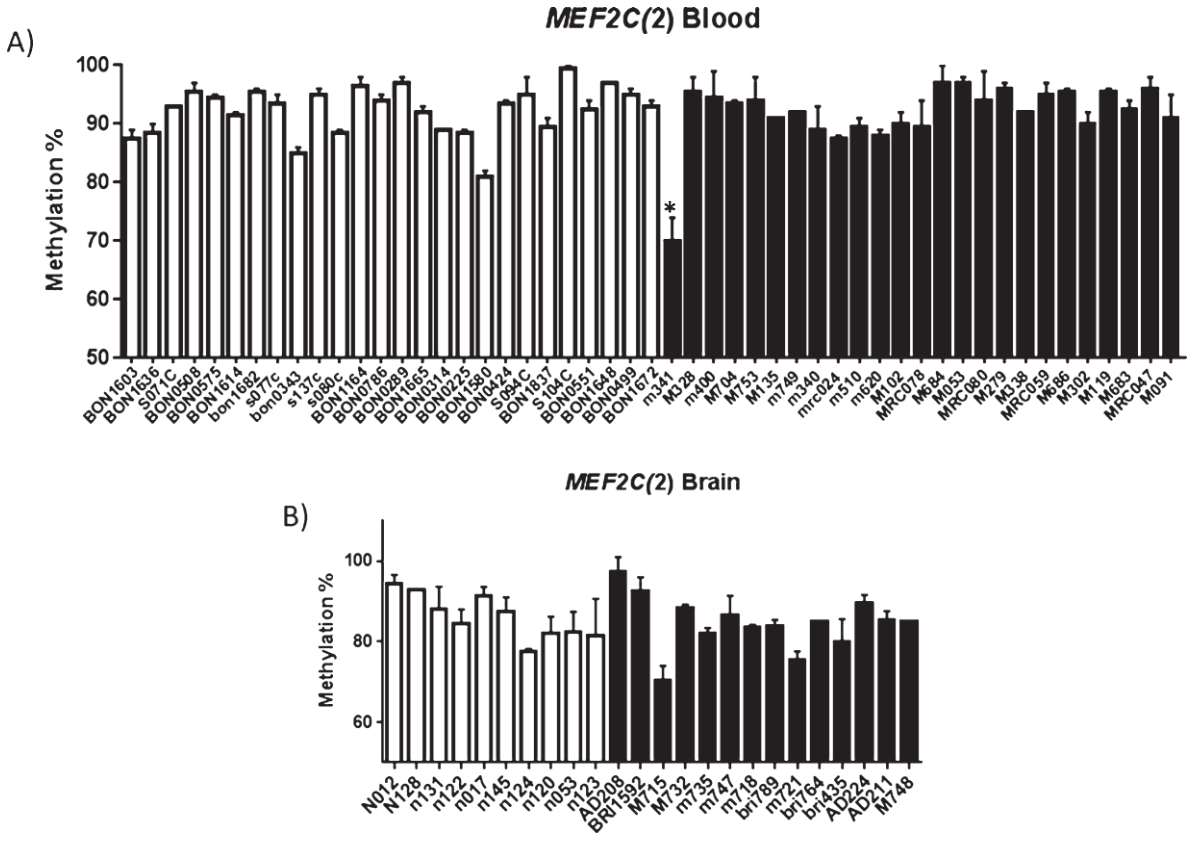


Fig. 4. (A) Methylation profiling identifies rare individual AD hypomethylation of the upstream *MEF2C* CpG site. Patient sample *M341* showed statistically significant difference in methylation (p = 2.0E-10) when compared to other blood samples (A), statistical significance was not observed for patient *M715* in brain (B). A) Shows average methylation at the CpG investigated in each sample investigated (average of at least two runs). Controls are shown in white and AD samples in black. Error bars represent the S.E.M. Control blood = 26, AD blood n = 25, control brain n = 10, AD brain n = 14; Male brain n = 10, Female brain n = 14, Male blood n = 29 0, Female blood n = 22.

ence in average methylation was detected between AD brain and control brain. No link with gender was identified for overall levels of methylation at the RIN3 site; no aggregate difference could be detected between male and female samples for either blood or brain (see Supplementary Figure 6).

Tissue specific hypomethylation was observed in AD brain

Having established that in some instances comparative regional differences in methylation may be specific to AD patients, we chose to compare relative levels of methylation between tissues tested in this study. No significant difference in methylation was observed between tissues in either control or AD samples (Supplementary Figure 5) for AD and Control *PTK2*þ (p = 0.42; p = 0.1), *ABCA7* (p = 0.43; p = 0.46), *SIRT1* (p = 0.06; p = 0.24), and *MEF2C(1)* (p = 0.33; p = 0.37) (these are the p values using one- tailed *T*-tests, two tailed are, for AD and control *PTK2*þ (p = 0.85: p = 0.2), *ABCA7* (p = 0.87; p = 0.9), *SIRT1* (p = 0.13; p = 0.49), and *MEF2C(1)* (p = 0.66;
p = 0.73). However, within the *RIN3* 3'UTR region of blood DNA was significantly hypermethylated in both AD and control blood when compared to AD and control brain (Supplementary Figure 7). Hypomethylation was also observed in brain tissue for the *MEF2C(2)* site tested (Supplementary Figure 4).

DISCUSSION

Much resource has been committed to investigating the pathogenesis of LOAD, however a

comprehensive understanding of sEOAD has thus far remained elusive; it is likely that new ways of under- standing drivers of sEOAD are required. Although

a number of genes have been identified as drivers or causal to LOAD and fAD, those genes and associated pathways do not necessarily explain the underlying causes of sEOAD. We propose that aberrant epigenetic regulation contributes at least in part to this process for some individuals. Uncontrolled or dysregulation of a number of varied gene pathways or processes has been implicated in neurodegeneration; therefore, it is plausible that aberrant regulation of one of many genes may exacerbate pathogenesis [5].

Methylation of the *SIRT1* promoter in LOAD patient samples has been previously reported by Hou et al. [13]; we therefore chose to test if equivalent methylation could also be detected in sEOAD samples. It could be suggested that due to the core similarities of disease progression between AD subtypes, any epigenetic markers which are a feature of LOAD would be matched in sEOAD. Surprisingly we found no evidence of variation in levels of methylation between sEOAD samples and controls. This suggests that the methylation directed regulation of *SIRT1* is not impaired in the process of sEOAD, and the sEOAD epigenome is distinct from LOAD in at least some instances. It is not obvious how this difference impacts on the development of sEOAD, although given the role of SIRT1 in attenuating toxicities of both tau and amyloid it is likely to have some influence on the progression of disease. Further research is required to investigate the consequence of *SIRT1* regulation in both conditions.

Within our test group, we identified that the gene *MEF2C* was significantly hypomethylated in comparison to controls for only two individuals presenting with sEOAD. As an aggregate no group-wide differences could be detected, however we observed two individuals with pronounced differences in methylation within a key regulatory region of this gene (22% and 15% lower in blood and brain, respectively). The aberrant epigenetic marks recorded for this gene in this study may represent an example of a rare epi-variant risk factors existing within the population that accentuates the risk of developing sEOAD. Changes in epialleles can occur as very rare events within large populations and can have profound effects on morphology or phenotype [39]. It is therefore a valid hypothesis to suggest that if the expression of the affected gene links with disease

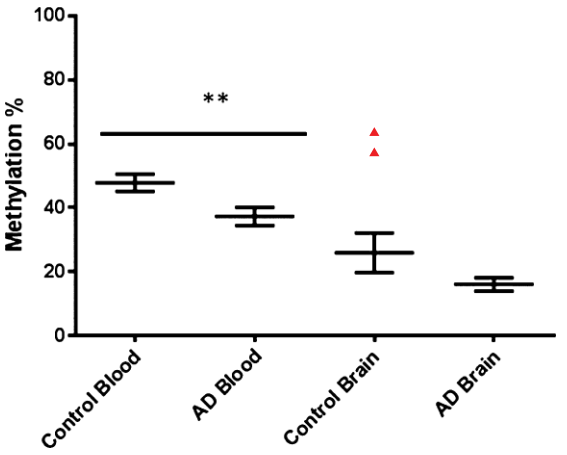


Fig. 5. Collective group wide methylation Control versus AD for *RIN3*. Graph shows average methylation across the whole region investigated in the *RIN3* 3'UTR. Control blood $n = 26$, AD blood $n = 22$, Control brain $n = 10$, AD brain $n = 14$. Control blood Male to Female $n = 12/14$, AD blood Male to Female $n = 15/7$.

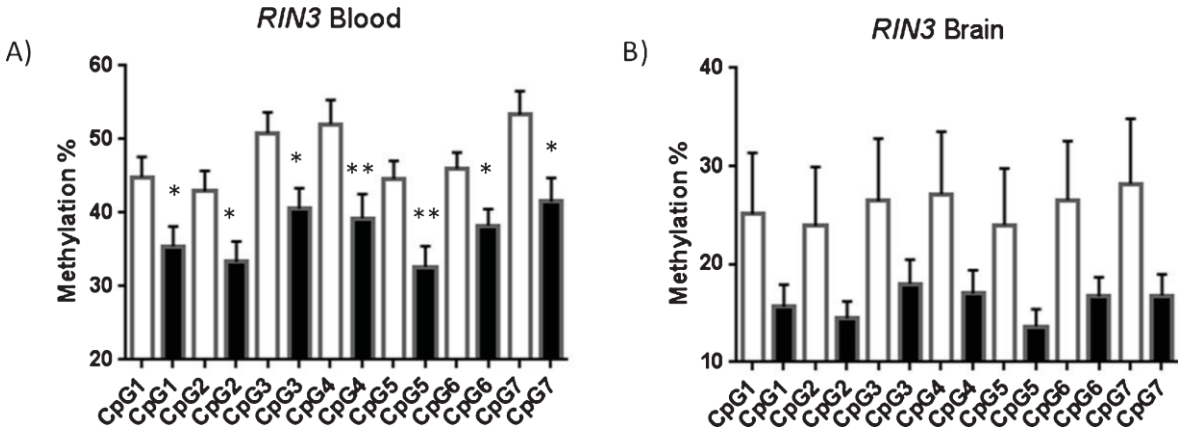


Fig. 6. RIN3 Blood and Brain; resolution of each CpG. All of the seven CpG sites investigate in RIN3 were shown to be significantly hypomethylated in AD blood (A) but not brain (B). Showing average methylation found at each CpG site in the RIN3 3'UTR. Control shown as a white bar (blood $n = 26$, brain $n = 10$) and AD (blood $n = 22$, brain $n = 14$) shown as a black bar. Error bars represent S.E.M.

pathogenesis and levels of methylation controlling expression are altered, then this change methylation can alter the risk at the level of the individual however infrequently this may occur within a population. We evaluated the possibility that methylation at this site correlates with a known genetic risk factors however upon analysis hypomethylation did not correlate to this particular genotype.

Where tested, we observed group wide differences in the levels of promoter methylation at the *RIN3* locus between sEOAD and control samples. Intriguingly statistically significant differences in the levels of methylation were not limited to one CpG, but span a number of CpGs located within this region. The group-wide nature of these epi-polymorphisms suggests that the condition of sEOAD leads to changes in methylation at the *RIN3* locus rather than being a rarer spontaneous causative driver of disease. The equivalence in both tissue samples also supports this hypothesis, e.g., it is likely that the levels of methylation identified in leukocyte DNA are reflective of AD pathology experienced as these cells transit through the brain. Methylation of these sites is therefore likely driven by the environment of the diseased tissue, rather than methylation being causative of the disease.

The consequence of changes in the levels of methylation located with the 3'UTR of *RIN3* is unclear. RIN3 is known to associate with BIN1 and links with the process of endocytosis and Aþ processing; therefore, it is plausible that any change that impacts on regulation could exacerbate the progression of AD. One hypothesis is that the AD brain ramps up the production of *RIN3* in response to the amyloid environment via demethylation of the 3'UTR region; increased methylation within 3'UTRs associates with reduced expression in reporter assay tests [33]. This is supported by the relative hypomethylation we observed in AD brain relative to blood. This may suggest a direct correlation between the length of exposure to the amyloid environment and levels of methylation for *RIN3*. Further experiments will be required to determine the consequence of 3'UTR methylation on levels of gene expression and ultimately to sEOAD pathogenesis.

Our data suggests that differences in methylation may mark the development of sEOAD. Epigenetic regulation of regions other than promoters can be aberrantly methylated in sEOAD patients promoting, initiating or driving AD pathology, suggesting that changes in epigenetic marks are a feature of the development of sEOAD. The ability to detect variations in blood allows for the testing of a large numbers of blood samples to further explore this hypothesis; an option not available if the effects manifest only in the brain epi-genome.

DETAILS FOR ARUK CONSORTIUM

David M. Mann[1], Peter Passmore[2], David Craig[2], Janet Johnston[2], Bernadette McGuinness[2], Stephen Todd[2], Reinhard Heun[3], Heike Kölsch[4], Patrick G. Kehoe[5], Emma R.L.C. Vardy[6], Nigel M. Hooper[1], Stuart Pickering-Brown[1], Julie Snowden[9], Anna Richardson[7], Matt Jones[7], David Neary[7], Jenny Harris[7], Christopher Medway[1], James Lowe[1], A. David Smith[8], Gordon Wilcock[8], Donald Warden[8], Clive Holmes[9].

[1]Institute of Brain, Behaviour and Mental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; [2]Centre for Public Health, School of Medicine, Queen's University Belfast, UK; [3]Royal Derby Hospital, Derby, UK; [4]Department of Psychiatry, University of Bonn, Bonn, Germany; [5]School of Clinical Sciences, John James Laboratories, University of Bristol, Bristol, UK; [6]Salford Royal NHS Foundation Trust; [7]Cerebral Function Unit, Greater Manchester Neurosciences Centre, Salford Royal Hospital, Stott Lane, Salford, UK; [8]University of Oxford (OPTIMA), Oxford, UK; [9]Clinical and Experimental Science, University of Southampton, Southampton, UK.

ACKNOWLEDGMENTS

This work was funded and supported by the Biotechnology and Biological Sciences Research Council (BBSRC) Doctoral Training Partnership.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/ADR-170015.

REFERENCES

[1] Bekris LM, Yu C-E, Bird TD, Tsuang DW (2010) Review article: Genetics of Alzheimer disease. *J Geriatr Psychiatry Neurol* **23**, 213-227.

[2] Minati L, Edginton T, Bruzzone MG, Giaccone G (2009) Current concepts in Alzheimer's disease: A multidisciplinary review. *Am J Alzheimers Dis Other Demen* **24**, 95-121.

[3] Van Cauwenberghe C, Van Broeckhoven C, Sleegers K (2016) The genetic landscape of Alzheimer disease: Clinical implications and perspectives. *Genet Med* **18**, 421-430.

[4] van der Flier WM, Pijnenburg YAL, Fox NC, Scheltens P (2011) Early-onset versus late-onset Alzheimer's disease: The case of the missing APOE epsilon 4 allele. *Lancet Neurol* **10**, 280-288.

[5] Antonell A, Llado A, Altirriba J, Botta-Orfila T, Balasa M, Fernandez M, Ferrer I, Sanchez-Valle R, Molinuevo JL (2013) A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. *Neurobiol Aging* **34**, 1772-1778.

[6] Karch CM, Goate AM (2015) Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry* **77**, 43-51.

[7] Wang SC, Oelze B, Schumacher A (2008) Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLos One* **3**, e2698.

[8] De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, Eaton ML, Keenan BT, Ernst J, McCabe C, Tang AN, Raj T, Replogle J, Brodeur W, Gabriel S, Chai HS, Younkin C, Younkin SG, Zou FG, Szyf M, Epstein CB, Schneider JA, Bernstein BE, Meissner A, Ertekin-Taner N, Chibnik LB, Kellis M, Mill J, Bennett DA (2014) Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* **17**, 1156-1163.

[9] Yu L, Chibnik LB, Srivastava GP, Pochet N, Yang J, Xu J, Kozubek J, Obholzer N, Leurgans SE, Schneider JA, Meissner A, De Jager PL, Bennett DA (2015) Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol* **72**, 15-24.

[10] Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, Troakes C, Al-Sarraj S, Burrage J, Macdonald R, Condliffe D, Harries LW, Katsel P, Haroutunian V, Kaminsky Z, Joachim C, Powell J, Lovestone S, Bennett DA, Schalkwyk LC, Mill J (2014) Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* **17**, 1164-1170.

[11] Chouliaras L, Mastroeni D, Delvaux E, Grover A, Kenis G, Hof PR, Steinbusch HWM, Coleman PD, Rutten BPF, van den Hove DLA (2013) Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging* **34**, 2091-2099.

[12] Coppieters N, Dieriks BV, Lill C, Faull RLM, Curtis MA, Dragunow M (2014) Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol Aging* **35**, 1334-1344.

[13] Hou YP, Chen HY, He Q, Jiang W, Luo T, Duan JH, Mu N, He YS, Wang HQ (2013) Changes in methylation patterns of multiple genes from peripheral blood leucocytes of Alzheimer's disease patients. *Acta Neuropsychiatr* **25**, 66-76.

[14] Iwata A, Nagata K, Hatsuta H, Takuma H, Bundo M, Iwamoto K, Tamaoka A, Murayama S, Saido T, Tsuji S (2014) Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum Mol Genet* **23**, 648-656.

[15] West RL, Lee JM, Maroun LE (1995) Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient. *J Mol Neurosci* **6**, 141-146.

[16] Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Transl Med* **2**, 49ra67.

[17] Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* **107** (Suppl 1), 1757-1764.

[18] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thorton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau MT, Choi SH, Reitz C, Pasquier F, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, De Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Morón FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fiévet N, Huentelman MW, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuiness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogaeva E, Gallacher J, St George-Hyslop P, Clarimon J, Lleo A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossú P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Deniz Naranjo MC, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F, European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease, Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology, Moebus S, Mecocci P, Del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannefelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley TH Jr, Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RF, Passmore P, Montine TJ, Bettens K, Rotter JI, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JS, Boerwinkle E, Riemenschneider M, Boada M, Hiltuenen M, Martin ER, Schmidt R, Rujescu D, Wang LS, Dartigues JF, Mayeux R, Tzourio C, Hofman A, Nöthen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458.

[19] Lev S, Moreno H, Martinez R, Canoll P, Peles E, Musacchio JM, Plowman GD, Rudy B, Schlessinger J (1995) Protein-tyrosine kinase PYK2 involved in Ca2+-induced regulation of ion-channel and MAP kinase functions. *Nature* **376**, 737-745.

[20] Otero K, Turnbull IR, Poliani PL, Vermi W, Cerutti E, Aoshi T, Tassi I, Takai T, Stanley SL, Miller M (2009) Macrophage

colony-stimulating factor induces the proliferation and survival of macrophages via a pathway involving DAP12 and þ-catenin. *Nat Immunol* **10**, 734-743.

[21] Tanaka N, Abe-Dohmae S, Iwamoto N, Yokoyama S (2011) Roles of ATP-binding cassette transporter A7 in cholesterol homeostasis and host defense system. *J Atheroscler Thromb* **18**, 274-281.

[22] Chan SL, Kim WS, Kwok JB, Hill AF, Cappai R, Rye KA, Garner B (2008) ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein *in vitro*. *J Neurochem* **106**, 793-804.

[23] Kim WS, Li HY, Ruberu K, Chan S, Elliott DA, Low JK, Cheng D, Karl T, Garner B (2013) Deletion of Abca7 increases cerebral amyloid-beta accumulation in the J20 mouse model of Alzheimer's disease. *J Neurosci* **33**, 4387-4394.

[24] Akhtar MW, Kim MS, Adachi M, Morris MJ, Qi XX, Richardson JA, Bassel-Duby R, Olson EN, Kavalali ET, Monteggia LM (2012) In vivo analysis of MEF2 transcription factors in synapse regulation and neuronal survival. *PLoS One* **7**, e34863.

[25] Gao J, Wang WY, Mao YW, Graff J, Guan JS, Pan L, Mak G, Kim D, Su SC, Tsai LH (2010) A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature* **466**, 1105-1109.

[26] Julien C, Tremblay C, Emond V, Lebbadi M, Salem N Jr, Bennett DA, Calon F (2009) Sirtuin 1 reduction parallels the accumulation of tau in Alzheimer disease. *J Neuropathol Exp Neurol* **68**, 48-58.

[27] Michan S, Li Y, Chou MM, Parrella E, Ge H, Long JM, Allard JS, Lewis K, Miller M, Xu W, Mervis RF, Chen J, Guerin KI, Smith LE, McBurney MW, Sinclair DA, Baudry M, de Cabo R, Longo VD (2010) SIRT1 is essential for normal cognitive function and synaptic plasticity. *J Neurosci* **30**, 9695-9707.

[28] Min SW, Cho SH, Zhou Y, Schroeder S, Haroutunian V, Seeley WW, Huang EJ, Shen Y, Masliah E, Mukherjee C, Meyers D, Cole PA, Ott M, Gan L (2010) Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron* **67**, 953-966.

[29] Tan MS, Yu JT, Tan L (2013) Bridging integrator 1 (BIN1): Form, function, and Alzheimer's disease. *Trends Mol Med* **19**, 594-603.

[30] Kajiho H, Saito K, Tsujita K, Kontani K, Araki Y, Kurosu H, Katada T (2003) RIN3: A novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J Cell Sci* **116**, 4159-4168.

[31] Treusch S, Hamamichi S, Goodman JL, Matlack KE, Chung CY, Baru V, Shulman JM, Parrado A, Bevis BJ, Valastyan JS, Han H, Lindhagen-Persson M, Reiman EM, Evans DA, Bennett DA, Olofsson A, DeJager PL, Tanzi RE, Caldwell KA, Caldwell GA, Lindquist S (2011) Functional links between Abeta toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334**, 1241-1245.

[32] Choi JK, Bae JB, Lyu J, Kim TY, Kim YJ (2009) Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol* **10**, R89.

[33] Maussion G, Yang J, Suderman M, Diallo A, Nagy C, Arnovitz M, Mechawar N, Turecki G (2014) Functional DNA methylation in a transcript specific 3ÍUTR region of TrkB associates with suicide. *Epigenetics* **9**, 1061-1070.

[34] Malumbres M, Perez de Castro I, Santos J, Fernandez Piqueras J, Pellicer A (1999) Hypermethylation of the cell cycle inhibitor p15INK4b 3Í-untranslated region interferes with its transcriptional regulation in primary lymphomas. *Oncogene* **18**, 385-396.

[35] Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA (2012) Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* **13**, R97.

[36] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-435.

[37] Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* **2**, 2265-2275.

[38] Chapuis J, Hansmannel F, Gistelinck M, Mounier A, Van Cauwenberghe C, Kolen KV, Geller F, Sottejeau Y, Harold D, Dourlen P, Grenier-Boley B, Kamatani Y, Delepine B, Demiautte F, Zelenika D, Zommer N, Hamdane M, Bellenguez C, Dartigues JF, Hauw JJ, Letronne F, Ayral AM, Sleegers K, Schellens A, Broeck LV, Engelborghs S, De Deyn PP, Vandenberghe R, O'Donovan M, Owen M, Epelbaum J, Mercken M, Karran E, Bantscheff M, Drewes G, Joberty G, Campion D, Octave JN, Berr C, Lathrop M, Callaerts P, Mann D, Williams J, Buee L, Dewachter I, Van Broeckhoven C, Amouyel P, Moechars D, Dermaut B, Lambert JC (2013) Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry* **18**, 1225-1234.

[39] Agorio A, Durand S, Fiume E, Brousse C, Gy I, Simon M, Anava S, Rechavi O, Loudet O, Camilleri C, Bouche N (2017) An arabidopsis natural epiallele maintained by a feed-forward silencing loop between histone and DNA. *PLoS Genet* **13**, e1006551.

Bibliography

ABDOLMALEKY, H. M., SMITH, C. L., FARAONE, S. V., SHAFA, R., STONE, W., GLATT, S. J. & TSUANG, M. T. 2004. Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics,* 127B**,** 51-59.

ADKINS RM, KRUSHKAL J, TYLAVSKY FA, THOMAS F. Racial differences in gene-specific DNA methylation levels are present at birth. 2011. *Birth Defects Res A Clin Mol Teratol.* 91(8), 728-36.

AGORIO, A., DURAND, S., FIUME, E., BROUSSE, C., GY, I., SIMON, M., ANAVA, S., RECHAVI, O., LOUDET, O., CAMILLERI, C. & BOUCHE, N. 2017. An Arabidopsis Natural Epiallele Maintained by a Feed-Forward Silencing Loop between Histone and DNA. *PLoS Genet,* 13**,** e1006551.

AHMED, I., TAMOUZA, R., DELORD, M., KRISHNAMOORTHY, R., TZOURIO, C., MULOT, C., NACFER, M., LAMBERT, J. C., BEAUNE, P. & LAURENT-PUIG, P. 2012. Association between Parkinson's disease and the HLA-DRB1 locus. *Movement disorders,* 27**,** 1104-1110.

AKEN, B. L., ACHUTHAN, P., AKANNI, W., AMODE, M. R., BERNSDORFF, F., BHAI, J., BILLIS, K., CARVALHO-SILVA, D., CUMMINS, C., CLAPHAM, P., GIL, L., GIRÓN, C. G., GORDON, L., HOURLIER, T., HUNT, S. E., JANACEK, S. H., JUETTEMANN, T., KEENAN, S., LAIRD, M. R., LAVIDAS, I., MAUREL, T., MCLAREN, W., MOORE, B., MURPHY, D. N., NAG, R., NEWMAN, V., NUHN, M., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALTS, B., WILDER, S. P., ZADISSA, A., KOSTADIMA, M., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., CUNNINGHAM, F., YATES, A., ZERBINO, D. R. & FLICEK, P. 2017. Ensembl 2017. *Nucleic Acids Research,* 45**,** D635-D642.

ANTONELL, A., LLADO, A., ALTIRRIBA, J., BOTTA-ORFILA, T., BALASA, M., FERNANDEZ, M., FERRER, I., SANCHEZ-VALLE, R. & MOLINUEVO, J. L. 2013. A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. *Neurobiol Aging,* 34**,** 1772-8.

BAKULSKI, K. M., DOLINOY, D. C., SARTOR, M. A., PAULSON, H. L., KONEN, J. R., LIEBERMAN, A. P., ALBIN, R. L., HU, H. & ROZEK, L. S. 2012. Genome-Wide DNA Methylation Differences Between Late-Onset Alzheimer's Disease and Cognitively Normal Controls in Human Frontal Cortex. *Journal of Alzheimers Disease,* 29**,** 571-588.

BALES, K. R., VERINA, T., CUMMINS, D. J., DU, Y., DODEL, R. C., SAURA, J., FISHMAN, C. E., DELONG, C. A., PICCARDO, P. & PETEGNIEF, V. 1999. Apolipoprotein E is essential for amyloid deposition in the APPV717F transgenic mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences,* 96**,** 15233-15238.

BAO, M., HANABUCHI, S., FACCHINETTI, V., DU, Q., BOVER, L., PLUMAS, J., CHAPEROT, L., CAO, W., QIN, J. & SUN, S.-C. 2012. CD2AP/SHIP1 complex positively regulates plasmacytoid dendritic cell receptor signaling by inhibiting the E3 ubiquitin ligase Cbl. *The Journal of Immunology,* 189**,** 786-792.

BARRACHINA, M. & FERRER, I. 2009. DNA Methylation of Alzheimer Disease and Tauopathy-Related Genes in Postmortem Brain. *Journal of Neuropathology and Experimental Neurology,* 68**,** 880-891.

BARROW, A. D. & TROWSDALE, J. 2006. You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling (vol 36, pg 7, 2006). *European Journal of Immunology,* 36**,** 2276-2276.

BARTRAM, L & TANZI, R. Alzheimer disease risk genes: 29 and counting. 2019. *Nature Reviews Neurology.* 15, 191-192.

BASURTO-ISLAS, G., GRUNDKE-IQBAL, I., TUNG, Y. C., LIU, F. & IQBAL, K. 2013. Activation of asparaginyl endopeptidase leads to Tau hyperphosphorylation in Alzheimer disease. *J Biol Chem,* 288**,** 17495-507.

BEDNARSKA-MAKARUK M, GRABAN A, SOBCZYŃSKA-MALEFORA A, HARRINGTON DJ, MITCHELL M, VOONG K, DAI L, ŁOJKOWSKA W, BOCHYŃSKA A, RYGLEWICZ D, WIŚNIEWSKA A, WEHR H. Homocysteine metabolism and the associations of global DNA methylation with selected gene polymorphisms and nutritional factors in patients with dementia. 2016. *Exp Genrontol.* 81, 83-91.

BEKRIS, L. M., YU, C.-E., BIRD, T. D. & TSUANG, D. W. 2010. Review article: genetics of Alzheimer disease. *Journal of geriatric psychiatry and neurology,* 23**,** 213-227.

BERNSTEIN, A. I., LIN, Y., STREET, R. C., LIN, L., DAI, Q., YU, L., BAO, H., GEARING, M., LAH, J. J., NELSON, P. T., HE, C., LEVEY, A. I., MULLE, J. G., DUAN, R. & JIN, P. 2016. 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity. *Hum Mol Genet,* 25**,** 2437-2450.

BETTENS, K., SLEEGERS, K. & VAN BROECKHOVEN, C. 2013. Genetic insights in Alzheimer's disease. *Lancet Neurol,* 12**,** 92-104.

BEYREUTHER, K. & MASTERS, C. L. 1995. NEURODEGENERATION AND DEMENTIA - ALZHEIMERS-DISEASE AS A MODEL. *Arzneimittel-Forschung/Drug Research,* 45-1**,** 347-350.

BIE, B. H., WU, J., YANG, H., XU, J. J. J., BROWN, D. L. & NAGUIB, M. 2014. Epigenetic suppression of neuroligin 1 underlies amyloid-induced memory deficiency. *Nature Neuroscience,* 17**,** 223-231.

BILOUSOVA, T., MILLER, C. A., POON, W. W., VINTERS, H. V., CORRADA, M., KAWAS, C., HAYDEN, E. Y., TEPLOW, D. B., GLABE, C., ALBAY, R., 3RD, COLE, G. M., TENG, E. & GYLYS, K.

H. 2016. Synaptic Amyloid-beta Oligomers Precede p-Tau and Differentiate High Pathology Control Cases. *Am J Pathol,* 186**,** 185-98.

BONDA, D. J., LEE, H. G., CAMINS, A., PALLAS, M., CASADESUS, G., SMITH, M. A. & ZHU, X. 2011. The sirtuin pathway in ageing and Alzheimer disease: mechanistic and therapeutic considerations. *Lancet Neurol,* 10**,** 275-9.

BONHAM, L. W., SIRKIS, D. W., FAN, J., APARICIO, R. E., TSE, M., RAMOS, E. M., WANG, Q., COPPOLA, G., ROSEN, H. J., MILLER, B. L. & YOKOYAMA, J. S. 2017. Identification of a rare coding variant in TREM2 in a Chinese individual with Alzheimer's disease. *Neurocase,* 23**,** 65-69.

BRAAK, H. & BRAAK, E. 1991. NEUROPATHOLOGICAL STAGING OF ALZHEIMER-RELATED CHANGES. *Acta Neuropathologica,* 82**,** 239-259.

BRAAK, H. & DEL TREDICI, K. 2013. Amyloid-β may be released from non-junctional varicosities of axons generated from abnormal tau-containing brainstem nuclei in sporadic Alzheimer’s disease: a hypothesis. *Acta neuropathologica,* 126**,** 303-306.

BRADLEY-WHITMAN, M. A. & LOVELL, M. A. 2013. Epigenetic changes in the progression of Alzheimer's disease. *Mechanisms of Ageing and Development,* 134**,** 486-495.

BRADSHAW, E. M., CHIBNIK, L. B., KEENAN, B. T., OTTOBONI, L., RAJ, T., TANG, A., ROSENKRANTZ, L. L., IMBOYWA, S., LEE, M., VON KORFF, A., MORRIS, M. C., EVANS, D. A., JOHNSON, K., SPERLING, R. A., SCHNEIDER, J. A., BENNETT, D.

& DE JAGER, P. L. 2013. CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat Neurosci,* 16**,** 848-50.

BURNS, A. 1998. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. M. Folstein, S. Folstein and P. McHugh, Journal of Psychiatric

Research (1975) 12, 189-198. Introduction. *International Journal of Geriatric Psychiatry,* 13**,** 285-285.

BURTON, T. R., DIBROV, A., KASHOUR, T. & AMARA, F. M. 2002. Anti-apoptotic wild-type Alzheimer amyloid precursor protein signaling involves the p38 mitogen-activated protein kinase/MEF2 pathway. *Brain Res Mol Brain Res,* 108**,** 102-20.

CACACE, R., SLEEGERS, K. & VAN BROECKHOVEN, C. 2016. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement,* 12**,** 733-48.

CARDONA-MONZONÍS, A. BELTRÁN-GARCÍA, J. IBAÑEZ-CABELLOS, JS, PÉREZ-MACHADO, G, MALKANI, K, SANCHIS-GOMAR, F, MENA-MOLLÁ, S, GARCÍA-GIMÉNEZ, JL. Epigenetic biomarkers in cardiovascular disease. 2018. *Journal of laboratory and precision medicine.* doi: 10.21037.

CADY, J., KOVAL, E. D., BENITEZ, B. A., ZAIDMAN, C., JOCKEL-BALSAROTTI, J., ALLRED, P., BALOH, R. H., RAVITS, J., SIMPSON, E., APPEL, S. H., PESTRONK, A., GOATE, A. M., MILLER, T. M., CRUCHAGA, C. & HARMS, M. B. 2014. The TREM2 variant p.R47H is a risk factor for sporadic amyotrophic lateral sclerosis. *JAMA Neurol,* 71**,** 449-53.

CAMARGO, L. M., ZHANG, X. D., LOERCH, P., CACERES, R. M., MARINE, S. D., UVA, P., FERRER, M., DE RINALDIS, E., STONE, D. J., MAJERCAK, J., RAY, W. J., YI-AN, C., SHEARMAN, M. S. & MIZUGUCHI, K. 2015. Correction: Pathway-Based Analysis of Genome-Wide siRNA Screens Reveals the Regulatory Landscape of App Processing. *PLoS One.* United States.

CARBONI, L., LATTANZIO, F., CANDELETTI, S., PORCELLINI, E., RASCHI, E., LICASTRO, F. & ROMUALDI, P. 2015. Peripheral leukocyte expression of the potential biomarker proteins Bdnf, Sirt1, and Psen1 is not regulated by promoter methylation in Alzheimer's disease patients. *Neurosci Lett,* 605**,** 44-8.

CASILLAS, M. A., LOPATINA, N., ANDREWS, L. G. & TOLLEFSBOL, T. O. 2003. Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. *Molecular and Cellular Biochemistry,* 252**,** 33-43.

CASTELLANO, J. M., KIM, J., STEWART, F. R., JIANG, H., DEMATTOS, R. B., PATTERSON, B. W., FAGAN, A. M., MORRIS, J. C., MAWUENYEGA, K. G., CRUCHAGA, C., GOATE, A. M., BALES, K. R., PAUL, S. M., BATEMAN, R. J. & HOLTZMAN, D. M. 2011. Human apoE isoforms differentially regulate brain amyloid-beta peptide clearance. *Sci Transl Med,* 3**,** 89ra57.

CASTELLANI CA, LAUFER BI, MELKA MG, DIEH EJ, O'REILLY RL, & SINGH SM. DNA methylation differences in monozygotic twin pairs discordant for schizophrenia identifies psychosis related genes and networks. 2015. *BMC Med Genomics.* 8,17.

CASTRO-COSTA, E., FUZIKAWA, C., UCHOA, E., FIRMO, J. O. A. & LIMA-COSTA, M. F. 2008. Norms for the mini-mental state examination - Adjustment of the cut-off point in population-based studies (evidences from the Bambui health aging study). *Arquivos De Neuro-Psiquiatria,* 66**,** 524-528.

CHAN, S. L., KIM, W. S., KWOK, J. B., HILL, A. F., CAPPAI, R., RYE, K. A. & GARNER, B. 2008. ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. *Journal of Neurochemistry,* 106**,** 793-804.

CHANG, L., WANG, Y., JI, H., DAI, D., XU, X., JIANG, D., HONG, Q., YE, H., ZHANG, X., ZHOU, X., LIU, Y., LI, J., CHEN, Z., LI, Y., ZHOU, D., ZHUO, R., ZHANG, Y., YIN, H., MAO, C., DUAN, S. & WANG, Q. 2014. Elevation of peripheral BDNF promoter methylation links to the risk of Alzheimer's disease. *PLoS One,* 9**,** e110773.

CHANG, M. Y., BOULDEN, J., KATZ, J. B., WANG, L., MEYER, T. J., SOLER, A. P., MULLER, A. J. & PRENDERGAST, G. C. 2007. Bin1 ablation increases susceptibility to cancer during aging, particularly lung cancer. *Cancer Res,* 67**,** 7605-12.

CHAPUIS, J., HANSMANNEL, F., GISTELINCK, M., MOUNIER, A., VAN CAUWENBERGHE, C., KOLEN, K. V., GELLER, F., SOTTEJEAU, Y., HAROLD, D., DOURLEN, P., GRENIER-BOLEY, B., KAMATANI, Y., DELEPINE, B., DEMIAUTTE, F., ZELENIKA, D., ZOMMER, N., HAMDANE, M.,

BELLENGUEZ, C., DARTIGUES, J. F., HAUW, J. J., LETRONNE, F., AYRAL, A. M., SLEEGERS, K., SCHELLENS, A., BROECK, L. V., ENGELBORGHS, S., DE DEYN, P. P., VANDENBERGHE, R., O'DONOVAN, M., OWEN, M., EPELBAUM, J., MERCKEN, M., KARRAN, E., BANTSCHEFF, M., DREWES, G., JOBERTY, G., CAMPION, D., OCTAVE, J. N., BERR, C., LATHROP, M., CALLAERTS, P., MANN, D., WILLIAMS, J., BUEE, L., DEWACHTER, I., VAN BROECKHOVEN, C., AMOUYEL, P., MOECHARS, D., DERMAUT, B. & LAMBERT, J. C. 2013. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry,* 18**,** 1225-34.

CHEN, K.-L., WANG, S. S.-S., YANG, Y.-Y., YUAN, R.-Y., CHEN, R.-M. & HU, C.-J. 2009. The epigenetic effects of amyloid-β< sub> 1–40</sub> on global DNA and neprilysin genes in murine cerebral endothelial cells. *Biochemical and biophysical research communications,* 378**,** 57-61.

CHO, S. H., CHEN, J. A., SAYED, F., WARD, M. E., GAO, F., NGUYEN, T. A., KRABBE, G., SOHN, P. D., LO, I., MINAMI, S., DEVIDZE, N., ZHOU, Y., COPPOLA, G. & GAN, L. 2015. SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1beta. *J Neurosci,* 35**,** 807-18.

CHOI, J. K., BAE, J. B., LYU, J., KIM, T. Y. & KIM, Y. J. 2009. Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol,* 10**,** R89.

CHOULIARAS, L., MASTROENI, D., DELVAUX, E., GROVER, A., KENIS, G., HOF, P. R., STEINBUSCH, H. W. M., COLEMAN, P. D., RUTTEN, B. P. F. & VAN DEN HOVE, D. L. A. 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of Aging,* 34**,** 2091-2099.

CHOULIARAS, L., RUTTEN, B. P. F., KENIS, G., PEERBOOMS, O., VISSER, P. J., VERHEY, F., VAN OS, J., STEINBUSCH, H. W. M. & VAN DEN HOVE, D. L. A. 2010. Epigenetic regulation in the pathophysiology of Alzheimer's disease. *Progress in Neurobiology,* 90**,** 498-510.

CHOWDHURY, S., SHEPHERD, J. D., OKUNO, H., LYFORD, G., PETRALIA, R. S., PLATH, N., KUHL, D., HUGANIR, R. L. & WORLEY, P. F. 2006. Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron,* 52**,** 445-59.

CHUNG, K. K., DAWSON, V. L. & DAWSON, T. M. 2001. The role of the ubiquitin-proteasomal pathway in Parkinson's disease and other neurodegenerative disorders. *Trends Neurosci,* 24**,** S7-14.

CLARK, R. I., TAN, S. W., PEAN, C. B., ROOSTALU, U., VIVANCOS, V., BRONDA, K., PILATOVA, M., FU, J., WALKER, D. W., BERDEAUX, R., GEISSMANN, F. & DIONNE, M. S. 2013. MEF2 is an in vivo immune-metabolic switch. *Cell,* 155**,** 435-47.

CLARKE, L. E. & BARRES, B. A. 2013. Emerging roles of astrocytes in neural circuit development (vol 14, pg 311-321, 2013). *Nature Reviews Neuroscience,* 14**,** 443-443.

COLE, C. J., MERCALDO, V., RESTIVO, L., YIU, A. P., SEKERES, M. J., HAN, J. H., VETERE, G., PEKAR, T., ROSS, P. J., NEVE, R. L., FRANKLAND, P. W. & JOSSELYN, S. A. 2012. MEF2 negatively regulates learning-induced structural plasticity and memory formation. *Nat Neurosci,* 15**,** 1255-64.

COLONNA, M. 2003. TREMs in the immune system and beyond. *Nat Rev Immunol,* 3**,** 445-53.

COLONNA, M. & WANG, Y. 2016. TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nat Rev Neurosci,* 17**,** 201-7.

COMBS, C. K. 2009. Inflammation and Microglia Actions in Alzheimer's Disease. *Journal of Neuroimmune Pharmacology,* 4**,** 380-388.

CONDLIFFE, D., WONG, A., TROAKES, C., PROITSI, P., PATEL, Y., CHOULIARAS, L., FERNANDES, C., COOPER, J., LOVESTONE, S., SCHALKWYK, L., MILL, J. & LUNNON, K. 2014. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiology of Aging,* 35**,** 1850-1854.

CONSORTIUM, I. M. S. G. & 2, W. T. C. C. C. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature,* 476**,** 214-219.

CONTE, A., PELLEGRINI, S. & TAGLIAZUCCHI, D. 2003. Synergistic protection of PC12 cells from beta-amyloid toxicity by resveratrol and catechin. *Brain Res Bull,* 62**,** 29-38.

COOGAN, A. N., SCHUTOVA, B., HUSUNG, S., FURCZYK, K., BAUNE, B. T., KROPP, P., HASSLER, F. & THOME, J. 2013. The circadian system in Alzheimer's disease: disturbances, mechanisms, and opportunities. *Biol Psychiatry,* 74**,** 333-9.

COPPIETERS, N., DIERIKS, B. V., LILL, C., FAULL, R. L. M., CURTIS, M. A. & DRAGUNOW, M. 2014. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiology of Aging,* 35**,** 1334-1344.

CORDER, E., SAUNDERS, A., STRITTMATTER, W., SCHMECHEL, D., GASKELL, P., SMALL, G., ROSES, A., HAINES, J. & PERICAK-VANCE, M. A. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science,* 261**,** 921-923.

CRUCHAGA C, KARCH CM, JIN SC, BENITEZ BA, CAI Y, GUERREIRO R, HARARI O, NORTON J, BUDDE J, BERTELSEN S, JENG AT, COOPER B, SKORUPA T, CARRELL D, LEVITCH D, HSU S, CHOI J, RYTEN M, SASSI C, BRAS J, GIBBS RJ, HERNANDEZ DG, LUPTON MK, POWELL J, FORABOSCO P, RIDGE PG, CORCORAN CD, TSCHANZ JT, NORTON MC, MUNGER RG, SCHMUTZ C, LEARY M, DEMIRCI FY, BAMNE MN, WANG X, LOPEZ OL, GANGULI M, MEDWAY C, TURTON J, LORD J, BRAAE A, BARBER I, BROWN K; ALZHEIMER'S RESEARCH UK (ARUK) CONSORTIUM, PASTOR P, LORENZO-BETANCOR O, BRKANAC Z, SCOTT E, TOPOL E, MORGAN K, ROGAEVA E, SINGLETON A, HARDY J, KAMBOH MI, GEORGE-HYSLOP PS, CAIRNS N, MORRIS JC, KAUWE JSK, GOATE AM. Rare coding variants in the phospholipase D3 gene confer riskfor Alzheimer's disease. 2014. *Nature.* 23(505), 550-554.

CUYVERS E & SLEEGERS K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. 2016. *Lancet Neurol.* 15(8), 857-868.

D'ADDARIO C, CANDIA SB, AROSIO B, DI BARTOLOMEO M, ABBATE C, CASÈ A, CANDELETTI S, ROMUALDI P, DAMANTI S, MACCARRONE M, BERGAMASCHINI L, MARI D. Transcriptional and epigenetic phenomena in peripheral blood cells of monozygotic twins discordant for Alzheimer's disease, a case report. 2017. *J Neurol Sci.* 15, 211-216.

DALL, E. & BRANDSTETTER, H. 2016. Structure and function of legumain in health and disease. *Biochimie,* 122**,** 126-50.

DAMEN, J. E., LIU, L., ROSTEN, P., HUMPHRIES, R. K., JEFFERSON, A. B., MAJERUS, P. W. & KRYSTAL, G. 1996. The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetraphosphate and phosphatidylinositol 3, 4, 5-triphosphate 5-phosphatase. *Proceedings of the National Academy of Sciences,* 93**,** 1689-1693.

DE JAGER, P. L., SRIVASTAVA, G., LUNNON, K., BURGESS, J., SCHALKWYK, L. C., YU, L., EATON, M. L., KEENAN, B. T., ERNST, J., MCCABE, C., TANG, A. N., RAJ, T., REPLOGLE, J., BRODEUR, W., GABRIEL, S., CHAI, H. S., YOUNKIN, C., YOUNKIN, S. G., ZOU, F. G., SZYF, M., EPSTEIN, C. B., SCHNEIDER, J. A., BERNSTEIN, B. E., MEISSNER, A., ERTEKIN-TANER, N.,

CHIBNIK, L. B., KELLIS, M., MILL, J. & BENNETT, D. A. 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience,* 17**,** 1156-1163.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes & development,* 25**,** 1010-1022.

DEDEURWAERDER, S., DEFRANCE, M., CALONNE, E., DENIS, H., SOTIRIOU, C. & FUKS, F. 2011. Evaluation of the Infinium Methylation 450K technology. *Epigenomics,* 3**,** 771-784.

DEMARTINI, D. R., SCHILLING, L. P., DA COSTA, J. C. & CARLINI, C. R. 2014. Alzheimer's and Parkinson's diseases: An environmental proteomic point of view. *Journal of Proteomics,* 104**,** 24-36.

DEVALL, M., MILL, J. & LUNNON, K. 2014. The mitochondrial epigenome: a role in Alzheimer's disease? *Epigenomics,* 6**,** 665-675.

DI FRANCESCO, A., AROSIO, B., FALCONI, A., MICIONI DI BONAVENTURA, M. V., KARIMI, M., MARI, D., CASATI, M., MACCARRONE, M. & D'ADDARIO, C. 2015. Global changes in DNA methylation in Alzheimer's disease peripheral blood mononuclear cells. *Brain Behav Immun,* 45**,** 139-44. DIKIC, I. 2002. CIN85/CMS family of adaptor molecules. *FEBS letters,* 529**,** 110-115.

DON, R. H., COX, P. T., WAINWRIGHT, B. J., BAKER, K. & MATTICK, J. S. 1991. TOUCHDOWN PCR TO CIRCUMVENT SPURIOUS PRIMING DURING GENE AMPLIFICATION. *Nucleic Acids Research,* 19**,** 4008-4008.

Dong, L & Ren, H. Blood-based DNA Methylation Biomarkers for Early Detection of Colorectal Cancer. 2018. *Journal Proteomics Bioinform.* 11(6), 120-126.

DZITOYEVA, S., CHEN, H. & MANEV, H. 2012. Effect of aging on 5-hydroxymethylcytosine in brain mitochondria. *Neurobiology of aging,* 33**,** 2881-2891.

EL HAJ, M., ANTOINE, P., AMOUYEL, P., LAMBERT, J. C., PASQUIER, F. & KAPOGIANNIS, D. 2016. Apolipoprotein E (APOE) epsilon4 and episodic memory decline in Alzheimer's disease: A review. *Ageing Res Rev,* 27**,** 15-22.

EL-AGNAF, O. M. A., MAHIL, D. S., PATEL, B. P. & AUSTEN, B. M. 2000. Oligomerization and toxicity of beta-amyloid-42 implicated in Alzheimer's disease. *Biochemical and Biophysical Research Communications,* 273**,** 1003-1007.

ELLISON, E. M., ABNER, E. L. & LOVELL, M. A. 2017. Multiregional analysis of global 5-methylcytosine and 5-hydroxymethylcytosine throughout the progression of Alzheimer's disease. *J Neurochem,* 140**,** 383-394.

ESPADA, J., BALLESTAR, E., FRAGA, M. F., GAREA, A. V., JUARRANZ, A., STOCKERT, J. C., ROBERTSON, K. D., FUKS, F. O. & ESTELLER, M. 2004. Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *Journal of Biological Chemistry,* 279**,** 37175-37184.

EUGENIN EA & BERMAN JW. Chemokine-dependent mechanisms of leukocyte trafficking across a model of the blood-brain barrier. 2003. *Methods.* 29(4), 351-61.

FAN, S. & CHI, W. 2016. Methods for genome-wide DNA methylation analysis in human cancer. *Brief Funct Genomics,* 15**,** 432-442.

FENG, Q. & ZHANG, Y. 2001. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes & Development,* 15**,** 827-832.

FERRI, C. P., PRINCE, M., BRAYNE, C., BRODATY, H., FRATIGLIONI, L., GANGULI, M., HALL, K., HASEGAWA, K., HENDRIE, H., HUANG, Y. Q., JORM, A., MATHERS, C., MENEZES, P. R.,

RIMMER, E., SCAZUFCA, M. & ALZHEIMERS DIS, I. 2005.Global prevalence of dementia: a Delphi consensus study. *Lancet,* 366**,** 2112-2117.

FERRI, E., AROSIO, B., D'ADDARIO, C., GALIMBERTI, D., GUSSAGO, C., PUCCI, M., CASATI, M., FENOGLIO, C., ABBATE, C., ROSSI, P. D., SCARPINI, E., MACCARRONE, M. & MARI, D. 2016. Gene promoter methylation and expression of Pin1 differ between patients with frontotemporal dementia and Alzheimer's disease. *J Neurol Sci,* 362**,** 283-6.

FLAVELL, S. W., COWAN, C. W., KIM, T. K., GREER, P. L., LIN, Y., PARADIS, S., GRIFFITH, E. C., HU, L. S., CHEN, C. & GREENBERG, M. E. 2006. Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science,* 311**,** 1008-12.

FORSTL, H. & KURZ, A. 1999. Clinical features of Alzheimer's disease. *European Archives of Psychiatry and Clinical Neuroscience,* 249**,** 288-290.

FRAGA, M. F. 2009. Genetic and epigenetic regulation of aging. *Current opinion in immunology,* 21**,** 446-453.

FRANSQUET PD, LACAZE P, SAFFERY R, MCNEIL J, WOODS R, RYAN J. Blood DNA methylation as a potential biomarker of dementia: A systematic review. 2018. *Alzheimer's Dement.* 14(1), 81-103.

FROMMER, M., MCDONALD, L. E., MILLAR, D. S., COLLIS, C. M., WATT, F., GRIGG, G. W., MOLLOY, P. L. & PAUL, C. L. 1992. A GENOMIC SEQUENCING PROTOCOL THAT YIELDS A POSITIVE DISPLAY OF 5-METHYLCYTOSINE RESIDUES IN INDIVIDUAL DNA STRANDS. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 1827-1831.

FRYKMAN, S., INOUE, M., IKEDA, A., TERANISHI, Y., KIHARA, T., LUNDGREN, J. L., YAMAMOTO, N. G., BOGDANOVIC, N., WINBLAD, B., SCHEDIN-WEISS, S. & TJERNBERG, L. O. 2017. Maturation and processing of the amyloid precursor protein is regulated by the

potassium/sodium hyperpolarization-activated cyclic nucleotide-gated ion channel 2 (HCN2). *Biochem Biophys Res Commun,* 483**,** 352-358.
FUKS, F., BURGERS, W. A., BREHM, A., HUGHES-DAVIES, L. & KOUZARIDES, T. 2000. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature Genetics,* 24**,** 88-91.

FURUYA, T. K., DA SILVA, P. N. O., PAYAO, S. L. M., RASMUSSEN, L. T., DE LABIO, R. W., BERTOLUCCI, P. H. F., BRAGA, I. L. S., CHEN, E. S., TURECKI, G., MECHAWAR, N., MILL, J. & SMITH, M. 2012. SORL1 and SIRT1 mRNA expression and promoter methylation levels in aging and Alzheimer's Disease. *Neurochemistry International,* 61**,** 973-975.

FUSO, A., NICOLIA, V., CAVALLARO, R. A., RICCERI, L., D'ANSELMI, F., COLUCCIA, P., CALAMANDREI, G. & SCARPA, S. 2008. B-vitamin deprivation induces hyperhomocysteinemia and brain< i> S</i>-adenosylhomocysteine, depletes brain< i> S</i>-adenosylmethionine, and enhances PS1 and BACE expression and amyloid-β deposition in mice. *Molecular and cellular neuroscience,* 37**,** 731-746.

FUSO, A., SEMINARA, L., CAVALLARO, R. A., D'ANSELMI, F. & SCARPA, S. 2006. S-adenosylmethionine/homocysteine cycle alterations modify DNA methylation status with consequent deregulation of PS1 and BACE and beta-amyloid production (vol 28, pg 195, 2005). *Molecular and Cellular Neuroscience,* 32**,** 419-419.

GABEL, H. W., KINDE, B. Z., STROUD, H., GILBERT, C. S., HARMIN, D. A., KASTAN, N. R., HEMBERG, M., EBERT, D. H. & GREENBERG, M. E. 2015. Disruption of DNA methylation-dependent long gene repression in Rett syndrome. *Nature,* 522**,** 89-93.

GALANTER , J, GIGNOUX, CR ,OH, SS, TORGERSON, D , PINO-YANES, M, THAKUR, N, ENG, C , HU, D, HUNTSMAN, S, FARBER, HJ, AVILA, PC, BRIGINO-BUENAVENTURA, E,LENOIR, MA, MEADE, K, SEREBRISKY, D, RODRI´GUEZ-CINTRO´ N, R, KUMAR, R, RODRI´GUEZ-SANTANA, JR, SEIBOLD, MA, BORRELL, LN, BURCHARD, EG, ZAITLEN, N. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. 2017. *Elife.* 6:e20532.

GAO, J., WANG, W. Y., MAO, Y. W., GRAFF, J., GUAN, J. S., PAN, L., MAK, G., KIM, D., SU, S. C. & TSAI, L. H. 2010. A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature,* 466**,** 1105-9.

GHIDONI, R., PATERLINI, A. & BENUSSI, L. 2013. Translational proteomics in Alzheimer's disease and related disorders. *Clinical Biochemistry,* 46**,** 480-486.

GIANNAKOPOULOS, P., HERRMANN, F., BUSSIERE, T., BOURAS, C., KÖVARI, E., PERL, D., MORRISON, J., GOLD, G. & HOF, P. 2003. Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology,* 60**,** 1495-1500.

GIBSON, G., HAROUTUNIAN, V., ZHANG, H., PARK, L., SHI, Q., LESSER, M., MOHS, R., SHEU, R. & BLASS, J. 2000. Mitochondrial damage in Alzheimer's disease varies with apolipoprotein E genotype. *Annals of neurology,* 48**,** 297-303.

GIRALDO, E., LLORET, A., FUCHSBERGER, T. & VIÑA, J. Aβ and tau toxicities in Alzheimer's are linked via oxidative stress-induced p38 activation. Protective role of vitamin E. *Redox Biology*.

GIUFFRIDA, M. L., CARACI, F., PIGNATARO, B., CATALDO, S., DE BONA, P., BRUNO, V., MOLINARO, G., PAPPALARDO, G., MESSINA, A., PALMIGIANO, A., GAROZZO, D., NICOLETTI, F., RIZZARELLI, E. & COPANI, A. 2009. beta-Amyloid Monomers Are Neuroprotective. *Journal of Neuroscience,* 29**,** 10582-10587.

GOEDERT, M. & JAKES, R. 2005. Mutations causing neurodegenerative tauopathies. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1739**,** 240-250.

GOLD, E. S., MORRISSETTE, N. S., UNDERHILL, D. M., GUO, J., BASSETTI, M. & ADEREM, A. 2000. Amphiphysin IIm, a novel amphiphysin II isoform, is required for macrophage phagocytosis. *Immunity,* 12**,** 285-92.

GOLD, E. S., SIMMONS, R. M., PETERSEN, T. W., CAMPBELL, L. A., KUO, C. C. & ADEREM, A. 2004. Amphiphysin IIm is required for survival of Chlamydia pneumoniae in macrophages. *J Exp Med,* 200**,** 581-6.

GONZALGO, M. L. & JONES, P. A. 1997. Mutagenic and epigenetic effects of DNA methylation. *Mutation Research-Reviews in Mutation Research,* 386**,** 107-118.

GOODGER, Z. V., RAJENDRAN, L., TRUTZEL, A., KOHLI, B. M., NITSCH, R. M. & KONIETZKO, U. 2009. Nuclear signaling by the APP intracellular domain occurs predominantly through the amyloidogenic processing pathway. *Journal of Cell Science,* 122**,** 3703-3714.

GRAFF, J. & MANSUY, I. A. 2008. Epigenetic codes in cognition and behaviour. *Behavioural Brain Research,* 192**,** 70-87.

GRICIUC, A., SERRANO-POZO, A., PARRADO, A. R., LESINSKI, A. N., ASSELIN, C. N., MULLIN, K., HOOLI, B., CHOI, S. H., HYMAN, B. T. & TANZI, R. E. 2013. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron,* 78**,** 631-43.

GUERREIRO, R. & HARDY, J. 2014. Genetics of Alzheimer's Disease. *Neurotherapeutics,* 11**,** 732-737.

GUERREIRO, R., WOJTAS, A., BRAS, J., CARRASQUILLO, M., ROGAEVA, E., MAJOUNIE, E., CRUCHAGA, C., SASSI, C., KAUWE, J. S., YOUNKIN, S., HAZRATI, L., COLLINGE, J., POCOCK, J., LASHLEY, T., WILLIAMS, J., LAMBERT, J. C., AMOUYEL, P., GOATE, A., RADEMAKERS, R., MORGAN, K., POWELL, J., ST GEORGE-HYSLOP, P., SINGLETON, A. & HARDY, J. 2013a. TREM2 variants in Alzheimer's disease. *N Engl J Med,* 368**,** 117-27.

GUERREIRO, R. J., GUSTAFSON, D. R. & HARDY, J. 2012. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging,* 33**,** 437-56.

GUERREIRO, R. J., LOHMANN, E., BRAS, J. M., GIBBS, J. R., ROHRER, J. D., GURUNLIAN, N., DURSUN, B., BILGIC, B., HANAGASI, H., GURVIT, H., EMRE, M., SINGLETON, A. & HARDY, J. 2013b. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol,* 70**,** 78-84.

GUERRERO-MUNOZ, M. J., GERSON, J. & CASTILLO-CARRANZA, D. L. 2015. Tau Oligomers: The Toxic Player at Synapses in Alzheimer's Disease. *Front Cell Neurosci,* 9**,** 464.

GUO, J. U., SU, Y., SHIN, J. H., SHIN, J., LI, H., XIE, B., ZHONG, C., HU, S., LE, T., FAN, G., ZHU, H., CHANG, Q., GAO, Y., MING, G.-L. & SONG, H. 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci,* 17**,** 215-222.

GUPTA, R., NAGARAJAN, A. & WAJAPEYEE, N. 2010. Advances in genome-wide DNA methylation analysis. *Biotechniques,* 49**,** III-XI.

HAASS, C., KAETHER, C., THINAKARAN, G. & SISODIA, S. 2012. Trafficking and Proteolytic Processing of APP. *Cold Spring Harbor Perspectives in Medicine,* 2.

HAASS, C. & SELKOE, D. J. 2007. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nature reviews Molecular cell biology,* 8**,** 101-112.

HAMERMAN, J. A., JARJOURA, J. R., HUMPHREY, M. B., NAKAMURA, M. C., SEAMAN, W. E. & LANIER, L. L. 2006. Cutting edge: inhibition of TLR and FcR responses in macrophages by triggering receptor expressed on myeloid cells (TREM)-2 and DAP12. *J Immunol,* 177**,** 2051-5.

HAMPEL, H., BUERGER, K., TEIPEL, S. J., BOKDE, A. L. W., ZETTERBERG, H. & BLENNOW, K. 2008. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers & Dementia,* 4**,** 38-48.

HARDY, J. 2006. A hundred years of Alzheimer's disease research. *Neuron,* 52**,** 3-13. HARRIS, F. M., BRECHT, W. J., XU, Q., TESSEUR, I., KEKONIUS, L., WYSS-CORAY, T.,FISH, J. D., MASLIAH, E., HOPKINS, P. C. & SCEARCE-LEVIE, K. 2003. Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice. *Proceedings of the National Academy of Sciences,* 100**,** 10966-10971.

HARVEY RJ, SKELTON-ROBINSON M, ROSSOR MN. The prevalence and causes of dementia in people under the age of 65 years. 2003. *J Neurol Neurosurg Psychiatry*. 74(9),1206-1209.

HASHIMOTO, Y., TOYAMA, Y., KUSAKARI, S., NAWA, M. & MATSUOKA, M. 2016. An Alzheimer Disease-linked Rare Mutation Potentiates Netrin Receptor Uncoordinated-5C-induced Signaling That Merges with Amyloid beta Precursor Protein Signaling. *J Biol Chem,* 291**,** 12282-93.

HAUSER, P. S., NARAYANASWAMI, V. & RYAN, R. O. 2011. Apolipoprotein E: from lipid transport to neurobiology. *Progress in lipid research,* 50**,** 62-74.

HENRIKSEN, K., O'BRYANT, S. E., HAMPER, H., TROJANOWSKI, J. Q., MONTINE, T. J., JEROMIN, A., BLENNOW, K., LONNEBORG, A., WYSS-CORAY, T., SOARES, H.,BAZENET, C., SJOGREN, M., HU, W., LOVESTONE, S., KARSDAL, M. A., WEINER, M. W. & BLOOD-BASED BIOMARKER INTEREST, G. 2014. The future of blood-based biomarkers for Alzheimer's disease. *Alzheimers & Dementia,* 10**,** 115-131.

HERMAN, J. G., GRAFF, J. R., MYOHANEN, S., NELKIN, B. D. & BAYLIN, S. B. 1996. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 9821-9826.

HERSKOVITS, A. Z. & GUARENTE, L. 2014. SIRT1 in neurodevelopment and brain senescence. *Neuron,* 81**,** 471-83.

HERUKKA, S. K., SIMONSEN, A. H., ANDREASEN, N., BALDEIRAS, I., BJERKE, M., BLENNOW, K., ENGELBORGHS, S., FRISONI, G. B., GABRYELEWICZ, T., GALLUZZI, S., HANDELS, R., KRAMBERGER, M. G., KULCZYNSKA, A., MOLINUEVO, J. L., MROCZKO, B., NORDBERG, A., OLIVEIRA, C. R., OTTO, M., RINNE, J. O., ROT, U., SAKA, E., SOININEN, H., STRUYFS, H., SUARDI, S., VISSER, P. J., WINBLAD, B., ZETTERBERG, H. & WALDEMAR, G. 2017. Recommendations for cerebrospinal fluid Alzheimer's disease biomarkers in the diagnostic evaluation of mild cognitive impairment. *Alzheimers Dement,* 13**,** 285-295.

HESLEGRAVE, A., HEYWOOD, W., PATERSON, R., MAGDALINOU, N., SVENSSON, J., JOHANSSON, P., OHRFELT, A., BLENNOW, K., HARDY, J., SCHOTT, J., MILLS, K.

ZETTERBERG, H. 2016. Increased cerebrospinal fluid soluble TREM2 concentration in Alzheimer's disease. *Mol Neurodegener,* 11**,** 3.

HICKEY, W. F. 1991. Migration of Hematogenous Cells Through the Blood-Brain Barrier and the Initiation of CNS Inflammation. *Brain Pathology,* 1**,** 97-105.

HO KIM, J., FRANCK, J., KANG, T., HEINSEN, H., RAVID, R., FERRER, I., HEE CHEON, M., LEE, J. Y., SHIN YOO, J., STEINBUSCH, H. W., SALZET, M., FOURNIER, I. & MOK PARK, Y. 2015. Proteome-wide characterization of signalling interactions in the hippocampal CA4/DG subfield of patients with Alzheimer's disease. *Sci Rep,* 5**,** 11138.

HOLLINGWORTH, P., HAROLD, D., SIMS, R., GERRISH, A., LAMBERT, J.-C., CARRASQUILLO, M. M., ABRAHAM, R., HAMSHERE, M. L., PAHWA, J. S. & MOSKVINA, V. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature genetics,* 43**,** 429-435.

HOU, Y. P., CHEN, H. Y., HE, Q., JIANG, W., LUO, T., DUAN, J. H., MU, N., HE, Y. S. & WANG, H. Q. 2013. Changes in methylation patterns of multiple genes from peripheral blood leucocytes of Alzheimer's disease patients. *Acta Neuropsychiatrica,* 25**,** 66-76.

HOU Y, CHEN H, HE Q, JIANG W, LUO T, DUAN J, MU N, HE Y, WANG H. Changes in methyltaion patterns of multiple genes from peripheral blood leucocytes of Alziemer's disease patients. 2013. *Acta Neuropsychiatry.* 25(2), 66-76.

HUO Z, ZHU Y, YU L, YANG J, DE JAGER P, BENNETT DA, ZHAO J. DNA methylation variability in Alzheimer's disease. 2019. *Nerobiol Aging.* 76, 35-44.

HU, X., LIOU, A. K., LEAK, R. K., XU, M., AN, C., SUENAGA, J., SHI, Y., GAO, Y., ZHENG, P. & CHEN, J. 2014. Neurobiology of microglial action in CNS injuries: receptor-mediated signaling mechanisms and functional roles. *Progress in neurobiology*.

HU, X., PICKERING, E., LIU, Y. C., HALL, S., FOURNIER, H., KATZ, E., DECHAIRO, B., JOHN, S., VAN EERDEWEGH, P. & SOARES, H. 2011. Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. *PLoS One,* 6**,** e16616.

HUANG, Q. Q., HOSSAIN, M. M., WU, K., PARAI, K., POPE, R. M. & JIN, J. P. 2008. Role of H2-calponin in regulating macrophage motility and phagocytosis. *J Biol Chem,* 283**,** 25887-99.

HUI-CHEN WU LISSETTE DELGADO-CRUZATA JULIE D. FLOM MARY PERRIN YUYAN LIAOJENNIFER S. FERRIS REGINA M. SANTELLA MARY BETH TERRY. Repetitive element DNA methylation levels in white blood cell DNA from sisters discordant for breast cancer from the New York site of the Breast Cancer Family Registry. 2012. *Carcinogenesis.* 33 (10), 1946-1952.

HUMPHRIES, C., KOHLI, M. A., WHITEHEAD, P., MASH, D. C., PERICAK-VANCE, M. A.GILBERT, J. 2015. Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci. *Molecular and Cellular Neuroscience,* 67**,** 37-45.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science,* 333**,** 1300-1303.

ITTNER, L. M., KE, Y. D., DELERUE, F., BI, M., GLADBACH, A., VAN EERSEL, J., WOLFING, H., CHIENG, B. C., CHRISTIE, M. J., NAPIER, I. A., ECKERT, A., STAUFENBIEL, M., HARDEMAN, E. & GOTZ, J. 2010. Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models. *Cell,* 142**,** 387-97.

IWATA, A., NAGATA, K., HATSUTA, H., TAKUMA, H., BUNDO, M., IWAMOTO, K., TAMAOKA, A., MURAYAMA, S., SAIDO, T. & TSUJI, S. 2014. Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum Mol Genet,* 23**,** 648-56.

JACOBS HIL, HOPKINS DA, MAYRHOFER HC, BRUNER E, VAN LEEUWEN FW, RAAIJMAKERS W, SCHMAHMANN JD. The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. 2018. 141(1), 37-47.

JANG, J. H. & SURH, Y. J. 2003. Protective effect of resveratrol on beta-amyloid-induced oxidative PC12 cell death. *Free Radic Biol Med,* 34**,** 1100-10.

JARMOLOWICZ, A. I., CHEN, H. Y. & PANEGYRES, P. K. 2015. The patterns of inheritance in early-onset dementia: Alzheimer's disease and frontotemporal dementia. *Am J Alzheimers Dis Other Demen,* 30**,** 299-306.

JAY, T. R., MILLER, C. M., CHENG, P. J., GRAHAM, L. C., BEMILLER, S., BROIHIER, M. L., XU, G., MARGEVICIUS, D., KARLO, J. C., SOUSA, G. L., COTLEUR, A. C., BUTOVSKY, O., BEKRIS, L., STAUGAITIS, S. M., LEVERENZ, J. B., PIMPLIKAR, S. W., LANDRETH, G. E., HOWELL, G. R., RANSOHOFF, R. M. & LAMB, B. T. 2015. TREM2 deficiency eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. *J Exp Med,* 212**,** 287-95.

JEHLE, A. W., GARDAI, S. J., LI, S., LINSEL-NITSCHKE, P., MORIMOTO, K., JANSSEN, W. J., VANDIVIER, R. W., WANG, N., GREENBERG, S., DALE, B. M., QIN, C., HENSON, P. M. & TALL,

A. R. 2006. ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J Cell Biol.*

JI, L. X., SASAKI, T., SUN, X. X., MA, P., LEWIS, Z. A. & SCHMITZ, R. J. 2014. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Frontiers in Genetics,* 5.

JIANG, M. H., ZHANG, Y. H., FEI, J., CHANG, X. X., FAN, W. W., QIAN, X. Q., ZHANG, T. B. & LU, D. R. 2010. Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Laboratory Investigation,* 90**,** 282-290.

JIANG, T., TAN, L., CHEN, Q., TAN, M. S., ZHOU, J. S., ZHU, X. C., LU, H., WANG, H. F., ZHANG, Y. D. & YU, J. T. 2016. A rare coding variant in TREM2 increases risk for Alzheimer's disease in Han Chinese. *Neurobiol Aging,* 42**,** 217.e1-3.

JIANG, T., TAN, L., ZHU, X. C., ZHANG, Q. Q., CAO, L., TAN, M. S., GU, L. Z., WANG, H. F., DING, Z. Z., ZHANG, Y. D. & YU, J. T. 2014. Upregulation of TREM2 ameliorates neuropathology and rescues spatial cognitive impairment in a transgenic mouse model of Alzheimer's disease. *Neuropsychopharmacology,* 39**,** 2949-62.

JIN, S. C., CARRASQUILLO, M. M., BENITEZ, B. A., SKORUPA, T., CARRELL, D., PATEL, D., LINCOLN, S., KRISHNAN, S., KACHADOORIAN, M., REITZ, C., MAYEUX, R., WINGO, T. S., LAH, J. J., LEVEY, A. I., MURRELL, J., HENDRIE, H., FOROUD, T., GRAFF-RADFORD, N. R., GOATE, A. M., CRUCHAGA, C. & ERTEKIN-TANER, N. 2015. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. *Mol Neurodegener.* London.

JONATHAN M. WEISS, SHERRY A. DOWNIE, WILLIAM D. LYMAN & AND JOAN W. BERMAN. Astrocyte-derived monocyte-chemoattractant protein-1 directs the transmigration of leukocytes across a model of the human blood-brain barrier. 1998. Journal of immunology, 161, 6896-6903.

JONSSON T, ATWAL JK, STEINBERG S, SNAEDAL J, JONSSON PV, BJORNSSON S, STEFANSSON H, SULEM P, GUDBJARTSSON D, MALONEY J, HOYTE K, GUSTAFSON A, LIU Y, LU Y, BHANGALE T, GRAHAM RR, HUTTENLOCHER J, BJORNSDOTTIR G, ANDREASSEN OA, JÖNSSON EG, PALOTIE A, BEHRENS TW, MAGNUSSON OT, KONG A, THORSTEINSDOTTIR U, WATTS RJ, STEFANSSON K. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. 2012. *Nature.* 488(7409),96-9.

JOHNSON, M. E., DELIARD, S., ZHU, F., XIA, Q., WELLS, A. D., HANKENSON, K. D. & GRANT, S. F. 2014. A ChIP-seq-defined genome-wide map of MEF2C binding reveals inflammatory pathways associated with its role in bone density determination. *Calcif Tissue Int,* 94**,** 396-402.

JONSSON, T., STEFANSSON, H., STEINBERG, S., JONSDOTTIR, I., JONSSON, P. V., SNAEDAL, J., BJORNSSON, S., HUTTENLOCHER, J., LEVEY, A. I., LAH, J. J., RUJESCU, D., HAMPEL, H., GIEGLING, I., ANDREASSEN, O. A., ENGEDAL, K., ULSTEIN, I., DJUROVIC, S., IBRAHIM-VERBAAS, C., HOFMAN, A., IKRAM, M. A., VAN DUIJN, C. M., THORSTEINSDOTTIR, U., KONG, A. & STEFANSSON, K. 2013. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med,* 368**,** 107-16.

JULIEN, C., TREMBLAY, C., EMOND, V., LEBBADI, M., SALEM, N., JR., BENNETT, D. A. CALON, F. 2009. Sirtuin 1 reduction parallels the accumulation of tau in Alzheimer disease. *J Neuropathol Exp Neurol,* 68**,** 48-58.

KAJIHO, H., SAITO, K., TSUJITA, K., KONTANI, K., ARAKI, Y., KUROSU, H. & KATADA, T. 2003. RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J Cell Sci,* 116**,** 4159-68.

KANG, J. H., KORECKA, M., TOLEDO, J. B., TROJANOWSKI, J. Q. & SHAW, L. M. 2013. Clinical Utility and Analytical Challenges in Measurement of Cerebrospinal Fluid Amyloid-beta(1-42) and tau Proteins as Alzheimer Disease Biomarkers. *Clinical Chemistry,* 59**,** 903-916.
KARCH, C. M., CRUCHAGA, C. & GOATE, A. M. 2014. Alzheimer's disease genetics:

from the bench to the clinic. *Neuron,* 83**,** 11-26.

KARCH, C. M. & GOATE, A. M. 2015. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry,* 77**,** 43-51.

KARCH, C. M., JENG, A. T., NOWOTNY, P., CADY, J., CRUCHAGA, C. & GOATE, A. M. 2012. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One,* 7**,** e50976.

KAWARABAYASHI, T., SHOJI, M., YOUNKIN, L. H., WEN-LANG, L., DICKSON, D. W., MURAKAMI, T., MATSUBARA, E., ABE, K., ASHE, K. H. & YOUNKIN, S. G. 2004. Dimeric amyloid beta protein rapidly accumulates in lipid rafts followed by apolipoprotein E and phosphorylated tau accumulation in the Tg2576 mouse model of Alzheimer's disease. *J Neurosci,* 24**,** 3801-9.

KEENE, C. D., CUDABACK, E., LI, X., MONTINE, K. S. & MONTINE, T. J. 2011. Apolipoprotein E isoforms and regulation of the innate immune response in brain of patients with Alzheimer's disease. *Current opinion in neurobiology,* 21**,** 920-928.

KILPINEN H, BARRETT JC. How next-generation sequencing is transforming complex disease genetics. 2013. *Trends Genet TIG.* 29, 23-30.

KIM, W. S., LI, H. Y., RUBERU, K., CHAN, S., ELLIOTT, D. A., LOW, J. K., CHENG, D., KARL, T. & GARNER, B. 2013. Deletion of Abca7 Increases Cerebral Amyloid-beta Accumulation in the J20 Mouse Model of Alzheimer's Disease. *Journal of Neuroscience,* 33**,** 4387-4394.

KINDE, B., GABEL, H. W., GILBERT, C. S., GRIFFITH, E. C. & GREENBERG, M. E. 2015. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A,* 112**,** 6800-6.

KINOSHITA, A., FUKUMOTO, H., SHAH, T., WHELAN, C. M., IRIZARRY, M. C. & HYMAN, B. T. 2003. Demonstration by FRET of BACE interaction with the amyloid precursor protein at the cell surface and in early endosomes. *Journal of cell science,* 116**,** 3339-3346.

KISS, C., NISHIKAWA, J., TAKADA, K., TRIVEDI, P., KLEIN, G. & SZEKELY, L. 2003. T cell leukemia 1 oncogene expression depends on the presence of Epstein-Barr virus in the virus-carrying Burkitt lymphoma lines. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 4813-4818.

KLEINBERGER, G., YAMANISHI, Y., SUAREZ-CALVET, M., CZIRR, E., LOHMANN, E., CUYVERS, E., STRUYFS, H., PETTKUS, N., WENNINGER-WEINZIERL, A., MAZAHERI, F., TAHIROVIC, S., LLEO, A., ALCOLEA, D., FORTEA, J., WILLEM, M., LAMMICH, S., MOLINUEVO, J. L., SANCHEZ-VALLE, R., ANTONELL, A., RAMIREZ, A., HENEKA, M. T., SLEEGERS, K., VAN DER ZEE, J., MARTIN, J. J.,ENGELBORGHS, S., DEMIRTAS-TATLIDEDE, A., ZETTERBERG, H., VAN BROECKHOVEN, C., GURVIT, H., WYSS-CORAY, T., HARDY, J., COLONNA, M. & HAASS, C. 2014. TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci Transl Med,* 6**,** 243ra86.

KOESTLER DC, USSET J, CHRISTENSEN BC, MARSIT CJ, KARAGAS MR, KELSEY KT, WIENCKE JK. DNA Methylation-Derived Neutrophil-to-Lymphocyte Ratio: An Epigenetic Tool to Explore Cancer Inflammation and Outcomes. 2017. *Cancer Epigemiology Biomarkers*. 26(3), 328-338.

KOBAYASH, N, SHINAGAWA, S, NAGATA, T, SHIMADA,S, SHIBATA,S, OHNUMA, T, KASANUKI, K, ARAI,H, YAMADA, H, NAKAYAMA, K, KONDO, K. Development of Biomarkers Based on DNA Methylation in the *NCAPH2*/*LMF2* Promoter Region for Diagnosis of Alzheimer's Disease and Amnesic Mild Cognitive Impairment. *PLOS One.* 0146449.

KOFFIE, R. M., HASHIMOTO, T., TAI, H.-C., KAY, K. R., SERRANO-POZO, A., JOYNER, D., HOU, S., KOPEIKINA, K. J., FROSCH, M. P. & LEE, V. M. 2012. Apolipoprotein E4 effects in Alzheimer's disease are mediated by synaptotoxic oligomeric amyloid-β. *Brain***,** aws127.

KORBIE, D. J. & MATTICK, J. S. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protocols,* 3**,** 1452-1456.

KRISTENSEN, L. S. & HANSEN, L. L. 2009. PCR-Based Methods for Detecting Single-Locus DNA Methylation Biomarkers in Cancer Diagnostics, Prognostics, and Response to Treatment. *Clinical Chemistry,* 55**,** 1471-1483.

KRUEGER F, ANDREWS SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. 2011. *Bioinformatics*. 27, 1571–1572.

KUNKLE BW, VARDARAJAN BN, NAJ AC, WHITEHEAD PL, ROLATI S, SLIFER S, CARNEY RM, CUCCARO ML, VANCE JM, GILBERT JR, WANG LS, FARRER LA, REITZ C, HAINES JL, BEECHAM GW, MARTIN ER, SCHELLENBERG GD, MAYEUX RP, PERICAK-VANCE MA. Early-Onset Alzheimer Disease and Candidate Risk Genes Involved in Endolysosomal Transport. 2017. *JAMA Neurology.* 74(9), 1113-1122.

KUMAR, R., CHATERJEE, P., SHARMA, P. K., SINGH, A. K., GUPTA, A., GILL, K., TRIPATHI, M., DEY, A. B. & DEY, S. 2013. Sirtuin1: a promising serum protein marker for early detection of Alzheimer's disease. *PLoS One,* 8**,** e61560.

KUPERSTEIN, I., BROERSEN, K., BENILOVA, I., ROZENSKI, J., JONEKHEERE, W., DEBULPAEP, M., VANDERSTEEN, A., SEGERS-NOLTEN, I., VAN DER WERF, K., SUBRAMANIAM, V., BRAEKEN, D., CALLEWAERT, G., BARTIC, C., D'HOOGE, R., MARTINS, I. C., ROUSSEAU, F., SCHYMKOWITZ, J. & DE STROOPER, B. 2010. Neurotoxicity of Alzheimer's disease A beta peptides is induced by small changes in the A beta(42) to A beta(40) ratio. *Embo Journal,* 29**,** 3408-3420.

LAFERLA, F. M., GREEN, K. N. & ODDO, S. 2007. Intracellular amyloid-β in Alzheimer's disease. *Nature Reviews Neuroscience,* 8**,** 499-509.

LAIRD, P. W. 2010. Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics,* 11**,** 191-203.

LAMBERT, J.-C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., JUN, G., DESTEFANO, A. L., BIS, J. C. & BEECHAM, G. W. 2013a. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*.

LAMBERT, J. C., HEATH, S., EVEN, G., CAMPION, D., SLEEGERS, K., HILTUNEN, M., COMBARROS, O., ZELENIKA, D., BULLIDO, M. J., TAVERNIER, B., LETENNEUR, L., BETTENS, K., BERR, C., PASQUIER, F., FIEVET, N., BARBERGER-GATEAU, P., ENGELBORGHS, S., DE DEYN, P., MATEO, I., FRANCK, A., HELISALMI, S., PORCELLINI, E., HANON, O., DE PANCORBO, M. M., LENDON, C., DUFOUIL, C., JAILLARD, C., LEVEILLARD, T., ALVAREZ, V., BOSCO, P., MANCUSO, M., PANZA, F., NACMIAS, B., BOSSU, P., PICCARDI, P., ANNONI, G., SERIPA, D., GALIMBERTI, D., HANNEQUIN, D., LICASTRO, F., SOININEN, H., RITCHIE, K., BLANCHE, H., DARTIGUES, J. F., TZOURIO, C., GUT, I., VAN BROECKHOVEN, C., ALPEROVITCH, A., LATHROP, M. & AMOUYEL, P. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet,* 41**,** 1094-9.

LAMBERT, J. C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., JUN, G., DESTEFANO, A. L., BIS, J. C., BEECHAM, G. W., GRENIER-BOLEY, B., RUSSO, G., THORNTON-WELLS, T. A., JONES, N., SMITH, A. V., CHOURAKI, V., THOMAS, C., IKRAM, M. A., ZELENIKA, D., VARDARAJAN, B. N., KAMATANI, Y., LIN, C. F., GERRISH, A., SCHMIDT, H., KUNKLE, B., DUNSTAN, M. L., RUIZ, A., BIHOREAU, M. T., CHOI, S. H., REITZ, C., PASQUIER, F., HOLLINGWORTH, P., RAMIREZ, A., HANON, O., FITZPATRICK, A. L., BUXBAUM, J. D., CAMPION, D., CRANE, P. K., BALDWIN, C., BECKER, T., GUDNASON, V., CRUCHAGA, C., CRAIG, D., AMIN, N., BERR, C., LOPEZ, O. L., DE JAGER, P. L., DERAMECOURT, V., JOHNSTON, J. A., EVANS, D., LOVESTONE, S., LETENNEUR, L., MORON, F. J., RUBINSZTEIN, D. C., EIRIKSDOTTIR, G., SLEEGERS, K., GOATE, A. M., FIEVET, N., HUENTELMAN, M. J., GILL, M., BROWN, K., KAMBOH, M. I., KELLER, L., BARBERGER-GATEAU, P., MCGUINNESS, B., LARSON, E. B., GREEN, R., MYERS, A. J., DUFOUIL, C., TODD, S., WALLON, D., LOVE, S., ROGAEVA, E., GALLACHER, J., ST GEORGE-HYSLOP, P., CLARIMON, J., LLEO, A., BAYER, A., TSUANG, D. W., YU, L., TSOLAKI, M., BOSSU, P., SPALLETTA, G., PROITSI, P., COLLINGE, J., SORBI, S.,

SANCHEZ-GARCIA, F., FOX, N. C., HARDY, J., NARANJO, M. C. D., BOSCO, P., CLARKE, R., BRAYNE, C., GALIMBERTI, D., MANCUSO, M., MATTHEWS, F., MOEBUS, S., MECOCCI, P., DEL ZOMPO, M., MAIER, W., et al. 2013b. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics,* 45**,** 1452-U206.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LARSSON, M., DUFFY D , L., ZHU, G., LIU J , Z., MACGREGOR, S., MCRAE A , F., WRIGHT M , J., STURM R , A., MACKEY D , A., MONTGOMERY G , W., MARTIN N , G. & MEDLAND S , E. 2011. GWAS Findings for Human Iris Patterns:

Associations with Variants in Genes that Influence Normal Neuronal Pattern Development. *Am J Hum Genet.*

LASHLEY, T., GAMI, P., VALIZADEH, N., LI, A., REVESZ, T. & BALAZS, R. 2015. Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol Appl Neurobiol,* 41**,** 497-506.

LAURENT, L., WONG, E., LI, G., HUYNH, T., TSIRIGOS, A., ONG, C. T., LOW, H. M., KIN SUNG, K. W., RIGOUTSOS, I., LORING, J. & WEI, C. L. 2010. Dynamic changes in the human methylome during differentiation. *Genome Res,* 20**,** 320-31.

LEE, H.-G., PERRY, G., MOREIRA, P. I., GARRETT, M. R., LIU, Q., ZHU, X., TAKEDA, A., NUNOMURA, A. & SMITH, M. A. 2005. Tau phosphorylation in Alzheimer's disease: pathogen or protector? *Trends in Molecular Medicine,* 11**,** 164-169.

LEE, Y.-J., HAN, S. B., NAM, S.-Y., OH, K.-W. & HONG, J. T. 2010. Inflammation and Alzheimer’s disease. *Archives of pharmacal research,* 33**,** 1539-1556.

LEIFER, D., GOLDEN, J. & KOWALL, N. W. 1994. Myocyte-specific enhancer binding factor 2C expression in human brain development. *Neuroscience,* 63**,** 1067-79.

LEV, S., MORENO, H., MARTINEZ, R., CANOLL, P., PELES, E., MUSACCHIO, J. M., PLOWMAN, G. D., RUDY, B. & SCHLESSINGER, J. 1995. PROTEIN-TYROSINE KINASE PYK2 INVOLVED IN CA2+-INDUCED REGULATION OF ION-CHANNEL AND MAP KINASE FUNCTIONS. *Nature,* 376**,** 737-745.

LEVINE, M & ENSOM, HH. Post Hoc Power Analysis: An Idea Whose Time Has Passed? 2017. *The Journal of Human Pharmacology and drug therapy.* 21(4).

LI, X., LEI, P., TUO, Q., AYTON, S., LI, Q. X., MOON, S., VOLITAKIS, I., LIU, R., MASTERS, C. L., FINKELSTEIN, D. I. & BUSH, A. I. 2015. Enduring Elevations of Hippocampal Amyloid Precursor Protein and Iron Are Features of beta-Amyloid Toxicity and Are Mediated by Tau. *Neurotherapeutics,* 12**,** 862-73.

LIDDELL, M., LOVESTONE, S. & OWEN, M. 2001. Genetic risk of Alzheimer's disease: advising relatives. *The British Journal of Psychiatry,* 178**,** 7-11.

LISTA, S., GARACI, F. G., EWERS, M., TEIPEL, S., ZETTERBERG, H., BLENNOW, K. & HAMPEL, H. 2014. CSF A beta 1-42 combined with neuroimaging biomarkers in the early detection, diagnosis and prediction of Alzheimer's disease. *Alzheimers & Dementia,* 10**,** 381-392.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LIU C, CUI G, ZHU M, KANG X, GUO H. Neuroinflammation in Alzheimer's disease: chemokines produced by astrocytes and chemokine receptors. 2014. *Int J Clin Exp Pathol.* 7(12), 8342, 55.

LOVESTONE, S. & REYNOLDS, C. H. 1997. The phosphorylation of tau: A critical stage in neurodevelopment and neurodegenerative processes. *Neuroscience,* 78**,** 309-324.

LOY, C. T., SCHOFIELD, P. R., TURNER, A. M. & KWOK, J. B. J. 2014. Genetics of dementia. *Lancet,* 383**,** 828-840.

LÖVKVIST C, DODD IB, SNEPPEN K, HAERTER JO. DNA methylation in human epigenomes depends on local topology of CpG sites. 2016. *Nucleic Acids Res.* 44(11), 5123,32.

LUNNON, K., KEOHANE, A., PIDSLEY, R., NEWHOUSE, S., RIDDOCH-CONTRERAS, J., THUBRON, E. B., DEVALL, M., SOININEN, H., KLOSZEWSKA, I., MECOCCI, P., TSOLAKI, M., VELLAS, B., SCHALKWYK, L., DOBSON, R., MALIK, A. N., POWELL, J., LOVESTONE, S. & HODGES, A. 2017. Mitochondrial genes are altered in blood early in Alzheimer's disease. *Neurobiol Aging,* 53**,** 36-47.

LUNNON, K., SMITH, R., HANNON, E., DE JAGER, P. L., SRIVASTAVA, G., VOLTA, M., TROAKES, C., AL-SARRAJ, S., BURRAGE, J., MACDONALD, R., CONDLIFFE, D.,

HARRIES, L. W., KATSEL, P., HAROUTUNIAN, V., KAMINSKY, Z., JOACHIM, C., POWELL, J., LOVESTONE, S., BENNETT, D. A., SCHALKWYK, L. C. & MILL, J. 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nature Neuroscience,* 17**,** 1164-1170.

LYONS, M. R., SCHWARZ, C. M. & WEST, A. E. 2012. Members of the myocyte enhancer factor 2 transcription factor family differentially regulate Bdnf transcription in response to neuronal depolarization. *J Neurosci,* 32**,** 12780-5.

MA, S. L., TANG, N. L. & LAM, L. C. 2016. Association of gene expression and methylation of UQCRC1 to the predisposition of Alzheimer's disease in a Chinese population. *J Psychiatr Res,* 76**,** 143-7.

MADRID A, HOGAN KJ, PAPALE LA, CLARK LR, ASTHANA S, JOHNSON SC, ALISCH RS. DNA Hypomethylation in Blood Links B3GALT4 and ZADH2 to Alzheimer's Disease. 2018. *J Alzheimer's Dis.* 66(3), 927-934.

MAHLEY, R. W. 1988. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science,* 240**,** 622-630.

MAHLEY, R. W. & HUANG, Y. 2012. Apolipoprotein e sets the stage: response to injury triggers neuropathology. *Neuron,* 76**,** 871-885.

MALUMBRES, M., PEREZ DE CASTRO, I., SANTOS, J., FERNANDEZ PIQUERAS, J. & PELLICER, A. 1999. Hypermethylation of the cell cycle inhibitor p15INK4b 3'-untranslated region interferes with its transcriptional regulation in primary lymphomas. *Oncogene,* 18**,** 385-96.

MAMUN, A. & LIU, F. 2017. *Neurology & Neurotherapy Open Access Journal Role of IRF4-Mediated Inflammation: Implication in Neurodegenerative Diseases Neurol Neurother Role of IRF4-Mediated Inflammation: Implication in Neurodegenerative Diseases*.

MARTINEZ-LOPEZ, M. J., ALCANTARA, S., MASCARO, C., PEREZ-BRANGULI, F., RUIZ-LOZANO, P., MAES, T., SORIANO, E. & BUESA, C. 2005. Mouse neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration. *Mol Cell Neurosci,* 28**,** 599-612.

MASTROENI, D., CHOULIARAS, L., GROVER, A., LIANG, W. S., HAUNS, K., ROGERS, J.COLEMAN, P. D. 2013. Reduced RAN Expression and Disrupted Transport between Cytoplasm and Nucleus; A Key Event in Alzheimer's Disease Pathophysiology. *Plos One,* 8.

MASTROENI, D., GROVER, A., DELVAUX, E., WHITESIDE, C., COLEMAN, P. D. & ROGERS, J. 2010. Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. *Neurobiology of Aging,* 31**,** 2025-2037.

MASTROENI, D., MCKEE, A., GROVER, A., ROGERS, J. & COLEMAN, P. D. 2009. Epigenetic Differences in Cortical Neurons from a Pair of Monozygotic Twins Discordant for Alzheimer's Disease. *Plos One,* 4.

MAUNAKEA, A. K., NAGARAJAN, R. P., BILENKY, M., BALLINGER, T. J., D'SOUZA, C., FOUSE, S. D., JOHNSON, B. E., HONG, C., NIELSEN, C., ZHAO, Y., TURECKI, G., DELANEY, A., VARHOL, R., THIESSEN, N., SHCHORS, K., HEINE, V. M., ROWITCH, D. H., XING, X., FIORE, C., SCHILLEBEECKX, M., JONES, S. J., HAUSSLER, D., MARRA, M. A., HIRST, M., WANG, T. & COSTELLO, J. F. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature,* 466**,** 253-7.

MAUSSION, G., YANG, J., SUDERMAN, M., DIALLO, A., NAGY, C., ARNOVITZ, M., MECHAWAR, N. & TURECKI, G. 2014. Functional DNA methylation in a transcript specific 3'UTR region of TrkB associates with suicide. *Epigenetics,* 9**,** 1061-70.

MAYEUX, R. & ST GEORGE-HYSLOP, P. 2009. Brain traffic: subcellular transport of the amyloid precursor protein. *Arch Neurol.* United States.

MCRAE AF, MARIONI RE, SHAH S, YANG J, POWELL JE, HARRIS SE, GIBSON J, HENDERS AK, BOWDLER L, PAINTER JN, MURPHY L, MARTIN NG, STARR JM, WRAY NR, DEARY IJ, VISSCHER PM, MONTGOMERY GW. Identification of 55,000 Replicated DNA Methylation QTL. 2018. *Sci Rep.* 8(1), 17605.

MEDWAY, C. & MORGAN, K. 2014. Review: The genetics of Alzheimer's disease; putting flesh on the bones. *Neuropathology and Applied Neurobiology,* 40**,** 97-105.

MEHLER, M. F. 2008. Epigenetic principles and mechanisms underlying nervous system functions in health and disease. *Progress in Neurobiology,* 86**,** 305-341.

MEISSNER, A., MIKKELSEN, T. S., GU, H. C., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X. L., BERNSTEIN, B. E., NUSBAUM, C., JAFFE, D. B., GNIRKE, A., JAENISCH, R. &

LANDER, E. S. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,* 454**,** 766-U91.

MELISSA K. CALLAHAN & RICHARD M. RANSOHOFF. Analysis of leukocyte extravasation across the blood-brain barrier: Conceptual and technical aspects. 2004. *Current Allergy and Asthma Reports,* 4, 65-73.

MELLÉN M, AYATA P, HEINTZ N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. 2017. *Proc Natl Acad Sci USA*. 114(37), E7812-21.

MERCORIO R, PERGOLI L, GALIMBERTI D, FAVERO C, CARUGNO M, DALLA VALLE E, BARRETTA F, CORTINI F, SCARPINI E, VALENTINA VB, PESATORI AC. PICALM Gene Methylation in Blood of Alzheimer's DiseasePatients Is Associated with Cognitive Decline. 2018. *J Alzheimer's Dis.* 65(1), 283-292.

MHATRE, S. D., TSAI, C. A., RUBIN, A. J., JAMES, M. L. & ANDREASSON, K. I. 2015. Microglial malfunction: the third rail in the development of Alzheimer's disease. *Trends Neurosci,* 38**,** 621-36.

MICHAN, S., LI, Y., CHOU, M. M., PARRELLA, E., GE, H., LONG, J. M., ALLARD, J. S., LEWIS, K., MILLER, M., XU, W., MERVIS, R. F., CHEN, J., GUERIN, K. I., SMITH, L. E., MCBURNEY, M. W., SINCLAIR, D. A., BAUDRY, M., DE CABO, R. & LONGO, V. D. 2010. SIRT1 is essential for normal cognitive function and synaptic plasticity. *J Neurosci,* 30**,** 9695-707.

MIN, S. W., CHO, S. H., ZHOU, Y., SCHROEDER, S., HAROUTUNIAN, V., SEELEY, W. W., HUANG, E. J., SHEN, Y., MASLIAH, E., MUKHERJEE, C., MEYERS, D., COLE, P. A., OTT, M. & GAN, L. 2010. Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron,* 67**,** 953-66.

MINATI, L., EDGINTON, T., BRUZZONE, M. G. & GIACCONE, G. 2009. Current Concepts in Alzheimer's Disease: A Multidisciplinary Review. *American Journal of Alzheimers Disease and Other Dementias,* 24**,** 95-121.

MITCHELL, A. J. & SHIRI-FESHKI, M. 2009. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand,* 119**,** 252-65.

MIYASHITA, A., KOIKE, A., JUN, G., WANG, L.-S., TAKAHASHI, S., MATSUBARA, E., KAWARABAYASHI, T., SHOJI, M., TOMITA, N. & ARAI, H. 2013. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. *PloS one,* 8**,** e58618.

MIZUGUCHI, M., IKEDA, K. & KIM, S. U. 1992. DIFFERENTIAL DISTRIBUTION OF CELLULAR-FORMS OF BETA-AMYLOID PRECURSOR PROTEIN IN MURINE GLIAL-CELL CULTURES. *Brain Research,* 584**,** 219-225.

MOORE, S. W., TESSIER-LAVIGNE, M. & KENNEDY, T. E. 2007. Netrins and their receptors. *Adv Exp Med Biol,* 621**,** 17-31.

MORGAN, A. R., HAMILTON, G., TURIC, D., JEHU, L., HAROLD, D., ABRAHAM, R., HOLLINGWORTH, P., MOSKVINA, V., BRAYNE, C., RUBINSZTEIN, D. C., LYNCH, A., LAWLOR, B., GILL, M., O'DONOVAN, M., POWELL, J., LOVESTONE, S., WILLIAMS, J. & OWEN, M. J. 2008. Association analysis of 528 intra-genic SNPs in a region of chromosome 10 linked to late onset Alzheimer's disease. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics,* 147B**,** 727-731.

MORGAN, H. D., DEAN, W., COKER, H. A., REIK, W. & PETERSEN-MAHRT, S. K. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues - Implications for epigenetic reprogramming. *Journal of Biological Chemistry,* 279**,** 52353-52360.

MORRISON, L. D., SMITH, D. D. & KISH, S. J. 1996. Brain S-Adenosylmethionine Levels Are Severely Decreased in Alzheimer's Disease. *Journal of neurochemistry,* 67**,** 1328-1331.

MUSIEK, E. S. & HOLTZMAN, D. M. 2015. Three Dimensions of the Amyloid Hypothesis: Time, Space, and "Wingmen". *Nat Neurosci,* 18**,** 800-6.

NAGATA K, MANO T, MURAYAMA S, SAIDO TC, IWATA A. DNA methylation level of the neprilysin promoter in Alzheimer's disease brains. 2018. *Nerosci letters.* 670, 8-13.

NAGATA T, SHINAGAWA S, NUKARIYA K, YAMADA H, NAKAYAMA K. Association between BDNF polymorphism (Val66Met) and executive function in patients with amnestic mild cognitive impairment or mild Alzheimer disease. 2012. *Dement Geriatr Cogn Disord*. 33:266-272.

NAGATA T, KOBAYASHI N, ISHII J, SHINAGAWA S, NAKAYAMA R, SHIBATA N, KUERBAN B, OHNUMA T, KONDO K, ARAI H, YAMADA H, NAKAYAMA K. Association between DNA Methylation of the BDNF Promoter Region and Clinical Presentation in Alzheimer's Disease. 2015. *Dement Geriatr Dis Extra.* 5(1), 64-73.

NAGY, Z., ESIRI, M., JOBST, K., JOHNSTON, C., LITCHFIELD, S., SIM, E. & SMITH, A. 1995. Influence of the apolipoprotein E genotype on amyloid deposition and neurofibrillary tangle formation in Alzheimer's disease. *Neuroscience,* 69**,** 757-761.

NAN, X. S., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R.

& BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-389.

NG, T. P., NITI, M., CHIAM, P. C. & KUA, E. H. 2007. Ethnic and educational differences in cognitive test performance on mini-mental state examination in Asians. *American Journal of Geriatric Psychiatry,* 15**,** 130-139.

NISBET, R. M., POLANCO, J. C., ITTNER, L. M. & GOTZ, J. 2015. Tau aggregation and its interplay with amyloid-beta. *Acta Neuropathol,* 129**,** 207-20.

OLSSON, B., LAUTNER, R., ANDREASSON, U., OHRFELT, A., PORTELIUS, E., BJERKE, M., HOLTTA, M., ROSEN, C., OLSSON, C., STROBEL, G., WU, E., DAKIN, K., PETZOLD, M., BLENNOW, K. & ZETTERBERG, H. 2016. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol,* 15**,** 673-684.

ORRE, M., KAMPHUIS, W., OSBORN, L. M., JANSEN, A. H., KOOIJMAN, L., BOSSERS, K.HOL, E. M. 2014. Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol Aging,* 35**,** 2746-60.

OTERO, K., SHINOHARA, M., ZHAO, H., CELLA, M., GILFILLAN, S., COLUCCI, A., FACCIO, R., ROSS, F. P., TEITELBAUM, S. L. & TAKAYANAGI, H. 2012. TREM2 and β-catenin regulate bone homeostasis by controlling the rate of osteoclastogenesis. *The Journal of Immunology,* 188**,** 2612-2621.

OTERO, K., TURNBULL, I. R., POLIANI, P. L., VERMI, W., CERUTTI, E., AOSHI, T., TASSI, I., TAKAI, T., STANLEY, S. L. & MILLER, M. 2009. Macrophage colony-stimulating factor induces the proliferation and survival of macrophages via a pathway involving DAP12 and β-catenin. *Nature immunology,* 10**,** 734-743.

PALONEVA, J., AUTTI, T., RAININKO, R., PARTANEN, J., SALONEN, O., PURANEN, M., HAKOLA, P. & HALTIA, M. 2001. CNS manifestations of Nasu-Hakola disease: a frontal dementia with bone cysts. *Neurology,* 56**,** 1552-8.

PALONEVA, J., MANNINEN, T., CHRISTMAN, G., HOVANES, K., MANDELIN, J., ADOLFSSON, R., BIANCHIN, M., BIRD, T., MIRANDA, R., SALMAGGI, A., TRANEBJAERG, L., KONTTINEN, Y. &

PELTONEN, L. 2002. Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *Am J Hum Genet,* 71**,** 656-62.

PALMER C, DIEHN M, ALIZADEH A, & O BROWN P. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. 2006. *Bmc Genomics.* 7,115.

PARDOSSI-PIQUARD, R., PETIT, A., KAWARAI, T., SUNYACH, C., DA COSTA, C. A., VINCENT, B., RING, S., D'ADAMIO, L., SHEN, J., MULLER, U., HYSLOP, P. S. & CHECLER, F. 2005. Presenilin-dependent transcriptional control of the A beta-degrading enzyme neprilysin by intracellular domains of beta APP and APLP. *Neuron,* 46**,** 541-554.

PATIL, V., WARD, R. L. & HESSON, L. B. 2014. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics,* 9**,** 823-8.

Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. 2013. *Nat Rev Genet.* 14(4), 288-95.

PEREZ-NIEVAS, B. G., STEIN, T. D., TAI, H.-C., DOLS-ICARDO, O., SCOTTON, T. C., BARROETA-ESPAR, I., FERNANDEZ-CARBALLO, L., DE MUNAIN, E. L., PEREZ, J. MARQUIE, M. 2013. Dissecting phenotypic traits linked to human resilience to Alzheimer's pathology. *Brain***,** awt171.

PIACERI, I., RASPANTI, B., TEDDE, A., BAGNOLI, S., SORBI, S. & NACMIAS, B. 2015. Epigenetic modifications in Alzheimer's disease: cause or effect? *J Alzheimers Dis,* 43**,** 1169-73.

PIDSLEY, R., WONG, C. C. Y., VOLTA, M., LUNNON, K., MILL, J. & SCHALKWYK, L. C. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *Bmc Genomics,* 14.

PIJNENBURG YA. Early- versus late-onset Alzheimer's disease: more than age alone. 2010. *J Alzheimers.*19(4),1401-1408.

PLONGTHONGKUM, N., DIEP, D. H. & ZHANG, K. 2014. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nature Reviews Genetics,* 15**,** 647-661.

POTTIER, C., HANNEQUIN, D., COUTANT, S., ROVELET-LECRUX, A., WALLON, D., ROUSSEAU, S., LEGALLIC, S., PAQUET, C., BOMBOIS, S. & PARIENTE, J. 2012. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Molecular psychiatry,* 17**,** 875-879.

PRONTO, L, SCHILTZ, S, JACOB, M, MENDA, Y, WEMMIE, J, MAGNOTTA, V. Brain pH and Alzheimer's Pathology. 2014. *Journal of Nuclear Medicine.* 55, 192.

PUZZO, D. & ARANCIO, O. 2013. Amyloid-beta Peptide: Dr. Jekyll or Mr. Hyde? *Journal of Alzheimers Disease,* 33**,** S111-S120.

QAZI, T. J., QUAN, Z., MIR, A. & QING, H. 2017. Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation. *Mol Neurobiol*.

QIN, W., YANG, T., HO, L., ZHAO, Z., WANG, J., CHEN, L., ZHAO, W., THIYAGARAJAN, M., MACGROGAN, D., RODGERS, J. T., PUIGSERVER, P., SADOSHIMA, J., DENG, H., PEDRINI, S., GANDY, S., SAUVE, A. A. & PASINETTI, G. M. 2006. Neuronal SIRT1 activation as a novel mechanism underlying the prevention of Alzheimer disease amyloid neuropathology by calorie restriction. *J Biol Chem,* 281**,** 21745-54.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

RAJENDRAN, L., HONSHO, M., ZAHN, T. R., KELLER, P., GEIGER, K. D., VERKADE, P. & SIMONS, K. 2006. Alzheimer's disease beta-amyloid peptides are released in association with exosomes. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 11172-11177.

RAMANAN, V., RISACHER, S., NHO, K., KIM, S., SWAMINATHAN, S., SHEN, L., FOROUD, T., HAKONARSON, H., HUENTELMAN, M. & AISEN, P. 2013. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Molecular psychiatry*.

RAO, J. S., KELESHIAN, V. L., KLEIN, S. & RAPOPORT, S. I. 2012. Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Translational Psychiatry,* 2.

RASHID, A. J., COLE, C. J. & JOSSELYN, S. A. 2014. Emerging roles for MEF2 transcription factors in memory. *Genes Brain Behav,* 13**,** 118-25.

RAYAPROLU, S., MULLEN, B., BAKER, M., LYNCH, T., FINGER, E., SEELEY, W. W., HATANPAA, K. J., LOMEN-HOERTH, C., KERTESZ, A., BIGIO, E. H., LIPPA, C., JOSEPHS, K. A., KNOPMAN, D. S., WHITE, C. L., CASELLI, R., MACKENZIE, I. R., MILLER, B. L., BOCZARSKA-JEDYNAK, M., OPALA, G., KRYGOWSKA-WAJS, A., BARCIKOWSKA, M., YOUNKIN, S. G., PETERSEN, R. C., ERTEKIN-TANER, N., UITTI, R. J., MESCHIA, J. F., BOYLAN, K. B., BOEVE, B. F., GRAFF-RADFORD, N. R., WSZOLEK, Z. K., DICKSON, D. W., RADEMAKERS, R. & ROSS, O. A. 2013. TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. *Mol Neurodegener.*

REED, K., POULIN, M. L., YAN, L. Y. & PARISSENTI, A. M. 2010. Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Analytical Biochemistry,* 397**,** 96-106.

REITZ, C., TOSTO, G., VARDARAJAN, B., ROGAEVA, E., GHANI, M., ROGERS, R. S., CONRAD, C., HAINES, J. L., PERICAK-VANCE, M. A., FALLIN, M. D., FOROUD, T., FARRER, L. A., SCHELLENBERG, G. D., GEORGE-HYSLOP, P. S. & MAYEUX, R. 2013. Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). *Transl Psychiatry,* 3**,** e256-.

RICHENS, J. L., MORGAN, K. & O'SHEA, P. 2014. Reverse engineering of Alzheimer's disease based on biomarker pathways analysis. *Neurobiology of Aging,* 35**,** 2029-2038.

RIDOLFI, E., BARONE, C., SCARPINI, E. & GALIMBERTI, D. 2013. The Role of the Innate Immune System in Alzheimer's Disease and Frontotemporal Lobar Degeneration: An Eye on Microglia. *Clinical and Developmental Immunology,* 2013.

RIDGE PG, HOYT KB, BOEHME K, MUKHERJEE S, CRANE PK, HAINES JL, MAYEUX R, FARRER LA, PERICAK-VANCE MA, SCHELLENBERG GD, KAUWE JSK. Assessment of the genetic variance of late-onset Alzheimer's disease. 2016. *Neurobiology of Aging.* 41, 200.e13-200.e20.

RIDGE, PG, MUKHERJEE, S,CRANE, PK, KAUWE, JSK. Alzheimer's Disease: Analyzing the Missing Heritability. 2013. *Plos One.* 8(11), e79771.

ROBINSON, S. R. & BISHOP, G. M. 2002. A beta as a bioflocculant: implications for the amyloid hypothesis of Alzheimer's disease. *Neurobiology of Aging,* 23**,** 1051-1072.

ROGAEVA, E., MENG, Y., LEE, J. H., GU, Y., KAWARAI, T., ZOU, F., KATAYAMA, T., BALDWIN, C. T., CHENG, R. & HASEGAWA, H. 2007. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nature genetics,* 39**,** 168-177.

ROUBROEKS JAY, SMITH RG, VAN DEN HOVE DLA, LUNNON K. Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. 2017. *J Neurochem.* 143(2), 158-170.

RUIZ, A., HEILMANN, S., BECKER, T., HERNÁNDEZ, I., WAGNER, H., THELEN, M., MAULEÓN, A., ROSENDE-ROCA, M., BELLENGUEZ, C., BIS, J. C., HAROLD, D., GERRISH, A., SIMS, R., SOTOLONGO-GRAU, O., ESPINOSA, A., ALEGRET, M., ARRIETA, J. L., LACOUR, A., LEBER, M., BECKER, J., LAFUENTE, A., RUIZ, S., VARGAS, L., RODRÍGUEZ, O., ORTEGA, G., DOMINGUEZ, M., MAYEUX, R., HAINES, J. L., PERICAK-VANCE, M. A., FARRER, L. A., SCHELLENBERG, G. D.,

CHOURAKI, V., LAUNER, L. J., VAN DUIJN, C., SESHADRI, S., ANTÚNEZ, C., BRETELER, M. M., SERRANO-RÍOS, M., JESSEN, F., TÁRRAGA, L., NÖTHEN, M. M., MAIER, W., BOADA, M. & RAMÍREZ, A. 2014. Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. *Transl Psychiatry,* 4**,** e358-.

SAAH, A. J. & HOOVER, D. R. 1997. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. *Ann Intern Med,* 126**,** 91-4.

SABRINA K. STEWART, TIFFANY J. MORRIS, PAUL GUILHAMON,HARRY BULSTRODE, MARTIN BACHMAN, SHANKAR BALASUBRAMANIAN, & STEPHAN BECK. oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. 2015. *Methods (San Diego, Calif.).* 72, 9-15.

SADRI, R. & HORNSBY, P. J. 1996. Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. *Nucleic Acids Research,* 24**,** 5058-5059.

SALMINEN, A., OJALA, J., KAUPPINEN, A., KAARNIRANTA, K. & SUURONEN, T. 2009. Inflammation in Alzheimer's disease: amyloid-β oligomers trigger innate immunity defence via pattern recognition receptors. *Progress in neurobiology,* 87**,** 181-194.

SANCHEZ-MUT, J. V., ASO, E., PANAYOTIS, N., LOTT, I., DIERSSEN, M., RABANO, A., URDINGUIO, R. G., FERNANDEZ, A. F., ASTUDILLO, A., MARTIN-SUBERO, J. I., BALINT, B., FRAGA, M. F., GOMEZ, A., GURNOT, C., ROUX, J. C., AVILA, J., HENSCH, T. K., FERRER, I. & ESTELLER, M. 2013. DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. *Brain,* 136**,** 3018-27.

SANCHEZ-MUT, J. V., HEYN, H., VIDAL, E., MORAN, S., SAYOLS, S., DELGADO-MORALES, R., SCHULTZ, M. D., ANSOLEAGA, B., GARCIA-ESPARCIA, P., PONS-ESPINAL, M., DE LAGRAN, M. M., DOPAZO, J., RABANO, A., AVILA, J., DIERSSEN, M., LOTT, I., FERRER, I., ECKER, J. R. & ESTELLER, M. 2016. Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Transl Psychiatry,* 6**,** e718-.

SCHERZER, C. R., OFFE, K., GEARING, M., REES, H. D., FANG, G., HEILMAN, C. J., SCHALLER, C., BUJO, H., LEVEY, A. I. & LAH, J. J. 2004. Loss of apolipoprotein E receptor LR11 in Alzheimer disease. *Archives of Neurology,* 61**,** 1200-1205.

SCHUBELER, D. Function and information content of DNA methylation. 2015. *Nature.* 517, 321-326.

SESHADRI, S., DRACHMAN, D. A. & LIPPA, C. F. 1995. Apolipoprotein E ε4 Allele and the Lifetime Risk of Alzheimer's Disease: What Physicians Know, and What They Should Know. *Archives of neurology,* 52**,** 1074-1079.

SHAD KF, AGHAZADEH Y, AHMAD S, KRESS B. Peripheral markers of Alzheimer's disease: surveillance of white blood cells. 2013. *Synapse*. 67 (8), 541-3.

SHI, L., ZHANG, Z. & SU, B. 2016. Sex Biased Gene Expression Profiling of Human Brains at Major Developmental Stages. *Sci Rep,* 6.

SHIRANE, K., TOH, H., KOBAYASHI, H., MIURA, F., CHIBA, H., ITO, T., KONO, T. & SASAKI, H. 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet,* 9**,** e1003439.

SHU, L., SUN, W., LI, L., XU, Z., LIN, L., XIE, P., SHEN, H., HUANG, L., XU, Q., JIN, P. & LI, X. 2016. Genome-wide alteration of 5-hydroxymenthylcytosine in a mouse model of Alzheimer's disease. *BMC Genomics.* London.

SHULMAN, J. M., CHEN, K., KEENAN, B. T., CHIBNIK, L. B., FLEISHER, A., THIYYAGURA, P., ROONTIVA, A., MCCABE, C., PATSOPOULOS, N. A., CORNEVEAUX, J. J., YU, L., HUENTELMAN, M. J., EVANS, D. A., SCHNEIDER, J. A., REIMAN, E. M., DE JAGER, P. L. & BENNETT, D. A. 2013. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. *JAMA Neurol,* 70**,** 1150-7.

SIEGMUND, K. D., CONNOR, C. M., CAMPAN, M., LONG, T. I., WEISENBERGER, D. J., BINISZKIEWICZ, D., JAENISCH, R., LAIRD, P. W. & AKBARIAN, S. 2007. DNA Methylation in the Human Cerebral Cortex Is Dynamically Regulated throughout the Life Span and Involves Differentiated Neurons. *Plos One,* 2.

SINGLETON A, HARDY J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. 2011. *Hum Mol Genet*. 20, R158-R162.

SIMS, R., VAN DER LEE, S. J., NAJ, A. C., BELLENGUEZ, C., BADARINARAYAN, N., JAKOBSDOTTIR, J., KUNKLE, B. W., BOLAND, A., RAYBOULD, R., BIS, J. C., MARTIN, E. R., GRENIER-BOLEY, B., HEILMANN-HEIMBACH, S., CHOURAKI, V., KUZMA, A. B., SLEEGERS, K., VRONSKAYA, M., RUIZ, A., GRAHAM, R. R., OLASO, R., HOFFMANN, P., GROVE, M. L., VARDARAJAN, B. N., HILTUNEN, M., NOTHEN, M. M., WHITE, C. C., HAMILTON-NELSON, K. L., EPELBAUM, J., MAIER, W., CHOI, S. H., BEECHAM, G. W., DULARY, C., HERMS, S., SMITH, A. V., FUNK, C. C., DERBOIS, C., FORSTNER, A. J., AHMAD, S., LI, H., BACQ, D., HAROLD, D., SATIZABAL, C. L., VALLADARES, O., SQUASSINA, A., THOMAS, R., BRODY, J. A., QU, L., SANCHEZ-JUAN, P., MORGAN, T., WOLTERS, F. J., ZHAO, Y., GARCIA, F. S., DENNING, N., FORNAGE, M., MALAMON, J., NARANJO, M. C. D., MAJOUNIE, E., MOSLEY, T. H., DOMBROSKI, B., WALLON, D., LUPTON, M. K., DUPUIS, J., WHITEHEAD, P., FRATIGLIONI, L., MEDWAY, C., JIAN, X., MUKHERJEE, S., KELLER, L., BROWN, K., LIN, H., CANTWELL, L. B., PANZA, F., MCGUINNESS, B., MORENO-GRAU, S., BURGESS, J. D., SOLFRIZZI, V., PROITSI, P., ADAMS, H. H., ALLEN, M., SERIPA, D., PASTOR, P., CUPPLES, L. A., PRICE, D., HANNEQUIN, D., FRANK-GARCIA, A., LEVY, D., CHAKRABARTY, P., CAFFARRA, P., GIEGLING, I., BEISER, A. S., GIEDRAITIS, V., HAMPEL, H., GARCIA, M. E., WANG, X., LANNFELT, L., MECOCCI, P., EIRIKSDOTTIR, G., CRANE, P. K., PASQUIER, F., BOCCARDI, V., et al. 2017. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet*.

SISODIA, S. S. 1992. BETA-AMYLOID PRECURSOR PROTEIN CLEAVAGE BY A MEMBRANE-BOUND PROTEASE. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 6075-6079.

SIVACHENKO, A., LI, Y., ABRUZZI, K. C. & ROSBASH, M. 2013. The transcription factor Mef2 links the Drosophila core clock to Fas2, neuronal morphology, and circadian behavior. *Neuron,* 79**,** 281-92.

SLIEKER, R. C., BOS, S. D., GOEMAN, J. J., BOVEE, J., TALENS, R. P., VAN DER BREGGEN, R., SUCHIMAN, H. E. D., LAMEIJER, E. W., PUTTER, H., VAN DEN AKKER, E. B., ZHANG, Y. J., JUKEMA, J. W., SLAGBOOM, P. E., MEULENBELT, I.

HEIJMANS, B. T. 2013. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics & Chromatin,* 6.

SMITH, A. R., SMITH, R. G., CONDLIFFE, D., HANNON, E., SCHALKWYK, L., MILL, J. & LUNNON, K. 2016. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol Aging,* 47**,** 35-40.

SMITH AR, SMITH RG, BURRAGE J, TROAKES C, AL-SARRAJ S, KALARIA RN, SLOAN C, ROBINSON AC, MILL J, LUNNON K. A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. 2019. 74, 70-76.

SMITH AR, SMITH RG, PISHVA E, HANNON E, ROUBROEKS JAY, BURRAGE J, TROAKES C, AL-SARRAJ S, SLOAN C, MILL J, VAN DEN HOVE DL, LUNNON K. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. 2019. *Clinical Epigenetics.* 11(1), 52.

SMITH RG, HANNON E, DE JAGER PL, CHIBNIK L, LOTT SJ, CONDLIFFE D, SMITH AR, HAROUTUNIAN V, TROAKES C, AL-SARRAJ S, BENNETT DA, POWELL J, LOVESTONE S, SCHALKWYK L, MILL J, LUNNON K. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's diseaseneuropathology. 2018. *Alzheimer's Dement*. 14(12), 1580-1588.

SORRENTINO, G. & BONAVITA, V. 2007. Neurodegeneration and Alzheimer's disease: the lesson from tauopathies. *Neurological Sciences,* 28**,** 63-71.

SOSCIA, S. J., KIRBY, J. E., WASHICOSKY, K. J., TUCKER, S. M., INGELSSON, M., HYMAN, B., BURTON, M. A., GOLDSTEIN, L. E., DUONG, S., TANZI, R. E. & MOIR, R. D. 2010. The Alzheimer's Disease-Associated Amyloid beta-Protein Is an Antimicrobial Peptide. *Plos One,* 5.

SPIRES-JONES, T. L. & HYMAN, B. T. 2014. The Intersection of Amyloid Beta and Tau at Synapses in Alzheimer's Disease. *Neuron,* 82**,** 756-771.

STENVINKEL, P., KARIMI, M., JOHANSSON, S., AXELSSON, J., SULIMAN, M., LINDHOLM, B., HEIMBURGER, O., BARANY, P., ALVESTRAND, A., NORDFORS, L., QURESHI, A. R., EKSTROM, T. J. & SCHALLING, M. 2007. Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease? *Journal of Internal Medicine,* 261**,** 488-499.

STEWART, F. J., PANNE, D., BICKLE, T. A. & RALEIGH, E. A. 2000. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J Mol Biol,* 298**,** 611-22.

STEWART, F. J. & RALEIGH, E. A. 1998. Dependence of McrBC cleavage on distance between recognition elements. *Biological Chemistry,* 379**,** 611-616.

SUAREZ-CALVET, M., KLEINBERGER, G., ARAQUE CABALLERO, M. A., BRENDEL, M., ROMINGER, A., ALCOLEA, D., FORTEA, J., LLEO, A., BLESA, R., GISPERT, J. D., SANCHEZ-VALLE, R., ANTONELL, A., RAMI, L., MOLINUEVO, J. L., BROSSERON, F., TRASCHUTZ, A., HENEKA, M. T., STRUYFS, H., ENGELBORGHS, S., SLEEGERS, K., VAN BROECKHOVEN, C., ZETTERBERG, H., NELLGARD, B., BLENNOW, K., CRISPIN, A., EWERS, M. & HAASS, C. 2016. sTREM2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers. *EMBO Mol Med,* 8**,** 466-76.

SWERDLOW, R. H., BURNS, J. M. & KHAN, S. M. 2010. The Alzheimer's disease mitochondrial cascade hypothesis. *J Alzheimers Dis,* 20 Suppl 2**,** S265-79.

TAHILIANI, M., KOH, K. P., SHEN, Y. H., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science,* 324**,** 930-935.

TAKAHASHI, K., ROCHFORD, C. D. & NEUMANN, H. 2005. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J Exp Med,* 201**,** 647-57.

TAKAHASHI, R. H., NAGAO, T. & GOURAS, G. K. 2017. Plaque formation and the intraneuronal accumulation of beta-amyloid in Alzheimer's disease. *Pathol Int,* 67**,** 185-193.

TAKESHITA, YUKIO, & RICHARD M RANSOHOFF. Inflammatory cell trafficking across the blood-brain barrier: chemokine regulation and in vitro models. 2012. *Immunological reviews*, 248,1, 228-39.

TAN, M. S., YU, J. T. & TAN, L. 2013. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends Mol Med,* 19**,** 594-603.

TANG, S. S., WANG, H. F., ZHANG, W., KONG, L. L., ZHENG, Z. J., TAN, M. S., TAN, C. C., WANG, Z. X., TAN, L., JIANG, T. & YU, J. T. 2016. MEF2C rs190982 polymorphism with late-onset Alzheimer's disease in Han Chinese: A replication study and meta-analyses. *Oncotarget.*

TANG, Z. P., CUI, Q. Z., DONG, Q. Z., XU, K. & WANG, E. H. 2013. Ataxia-telangiectasia group D complementing gene (ATDC) upregulates matrix metalloproteinase 9 (MMP-9) to promote lung cancer cell invasion by activating ERK and JNK pathways. *Tumour Biol,* 34**,** 2835-42.

TANILA, H. 2017. The role of BDNF in Alzheimer's disease. *Neurobiology of Disease,* 97**,** 114-118.

THATHIAH, A. & DE STROOPER, B. 2011. The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nat Rev Neurosci,* 12**,** 73-87.

TOHGI, H., UTSUGISAWA, K., NAGANE, Y., YOSHIMURA, M., GENDA, Y. & UKITSU, M. 1999. Reduction with age in methylcytosine in the promoter region− 224∼-101 of the amyloid precursor protein gene in autopsy human cortex.
Molecular brain research, 70**,** 288-292.

TOST, J., EL ABDALAOUI, H. & GUT, I. G. 2006. Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques,* 40**,** 721-2, 724, 726.

TOST, J. & GUT, I. G. 2007. DNA methylation analysis by pyrosequencing. *Nature Protocols,* 2**,** 2265-2275.

TOYOOKA, S., MARUYAMA, R., TOYOOKA, K. O., MCLERRAN, D., FENG, Z., FUKUYAMA, Y., VIRMANI, A. K., ZOCHBAUER-MULLER, S., TSUKUDA, K., SUGIO, K., SHIMIZU, N., SHIMIZU, K., LEE, H., CHEN, C. Y., FONG, K. M., GILCREASE, M., ROTH, J. A., MINNA, J. D. & GAZDAR, A. F. 2003. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. *Int J Cancer,* 103**,** 153-60.

TURNBULL, I. R., GILFILLAN, S., CELLA, M., AOSHI, T., MILLER, M., PICCIO, L., HERNANDEZ, M. & COLONNA, M. 2006. Cutting edge: TREM-2 attenuates macrophage activation. *J Immunol,* 177**,** 3520-4.

ULRICH, J. D., FINN, M. B., WANG, Y., SHEN, A., MAHAN, T. E., JIANG, H., STEWART, F. R., PICCIO, L., COLONNA, M. & HOLTZMAN, D. M. 2014. Altered microglial response to Abeta plaques in APPPS1-21 mice heterozygous for TREM2. *Mol Neurodegener,* 9**,** 20.

ULRICH, J. D. & HOLTZMAN, D. M. 2016. TREM2 Function in Alzheimer's Disease and Neurodegeneration. *ACS Chem Neurosci,* 7**,** 420-7.

ULRICH, J. D., ULLAND, T. K., COLONNA, M. & HOLTZMAN, D. M. 2017. Elucidating the Role of TREM2 in Alzheimer's Disease. *Neuron,* 94**,** 237-248.

URDINGUIO, R. G., SANCHEZ-MUT, J. V. & ESTELLER, M. 2009. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurology,* 8**,** 1056-1072.

VAFADAR-ISFAHANI, B., BALL, G., COVENEY, C., LEMETRE, C., BOOCOCK, D., MINTHON, L., HANSSON, O., MILES, A. K., JANCIAUSKIENE, S. M., WARDEN, D., SMITH, A. D., WILCOCK, G., KALSHEKER, N., REES, R., MATHAROO-BALL, B. MORGAN, K. 2012. Identification of SPARC-like 1 Protein as Part of a Biomarker Panel for Alzheimer's Disease in Cerebrospinal Fluid. *Journal of Alzheimers Disease,* 28**,** 625-636.

VARLEY, K. E, GERTZ, J., BOWLING, K. M. Dynamic DNA methylation across diverse human cell lines and tissues. 2013. *Genome Res.* 23, 555– 567.

VALINLUCK, V., TSAI, H. H., ROGSTAD, D. K., BURDZY, A., BIRD, A. & SOWERS, L. C. 2004. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Research,* 32**,** 4100-4108.

VAN CAUWENBERGHE, C., VAN BROECKHOVEN, C. & SLEEGERS, K. 2016. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med,* 18**,** 421-30.

VAN DER FLIER, W. M., PIJNENBURG, Y. A. L., FOX, N. C. & SCHELTENS, P. 2011. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE epsilon 4 allele. *Lancet Neurology,* 10**,** 280-288.

VARLEY, K. E., GERTZ, J., BOWLING, K. M., PARKER, S. L., REDDY, T. E., PAULI-BEHN, F., CROSS, M. K., WILLIAMS, B. A., STAMATOYANNOPOULOS, J. A., CRAWFORD, G. E., ABSHER, D. M., WOLD, B. J. & MYERS, R. M. 2013. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res,* 23**,** 555-67.

VASQUEZ, J. B., FARDO, D. W. & ESTUS, S. 2013. ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. *Neurosci Lett,* 556**,** 58-62.

VINGTDEUX, V., GILIBERTO, L., ZHAO, H., CHANDAKKAR, P., WU, Q., SIMON, J. E., JANLE, E. M., LOBO, J., FERRUZZI, M. G., DAVIES, P. & MARAMBAUD, P. 2010. AMP-activated protein kinase signaling activation by resveratrol modulates amyloid-beta peptide metabolism. *J Biol Chem,* 285**,** 9100-13.

WAHL, M. C. & LUHRMANN, R. 2015. SnapShot: Spliceosome Dynamics I. *Cell,* 161**,** 1474-e1.

WAJED Q, LAIRD PW, DEMEESTER T. DNA Methylation: An Alternative Pathway to Cancer. 2001. *Annals of Surgery.* 234(1), 10-20.

WATSON CT, ROUSSOS P, GARG P, HO DJ, AZAM N, KATSEL PL, HAROUTUNIAN V, SHARP AJ. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. 2016. *Genome Med.* 8(1), 5.

WANG, C, CHEN, L, YANG, Y, ZHANG, M & WONG, G. Identification of potential blood biomarkers for Parkinson's disease by gene expression and DNA methylation data integration analysis. 2019. *Clinical Epigenetics.* 11(24), 11-24.

WANG, J., FIVECOAT, H., HO, L., PAN, Y., LING, E. & PASINETTI, G. M. 2010. The role of Sirt1: at the crossroad between promotion of longevity and protection against Alzheimer's disease neuropathology. *Biochim Biophys Acta,* 1804**,** 1690-4.

WANG, J., YU, J. T., TAN, M. S., JIANG, T. & TAN, L. 2013. Epigenetic mechanisms in Alzheimer's disease: Implications for pathogenesis and therapy. *Ageing Research Reviews,* 12**,** 1024-1041.

WANG, K. S., LIU, Y., XU, C., LIU, X. & LUO, X. 2017. Family-based association analysis of NAV2 gene with the risk and age at onset of Alzheimer's disease. *J Neuroimmunol,* 310**,** 60-65.

WANG, N., LAN, D., GERBOD-GIANNONE, M., LINSEL-NITSCHKE, P., JEHLE, A. W., CHEN, W., MARTINEZ, L. O. & TALL, A. R. 2003. ATP-binding cassette transporter A7 (ABCA7) binds apolipoprotein A-I and mediates cellular phospholipid but not cholesterol efflux. *J Biol Chem,* 278**,** 42906-12.

WANG, S. C., OELZE, B. & SCHUMACHER, A. 2008. Age-Specific Epigenetic Drift in Late-Onset Alzheimer's Disease. *Plos One,* 3.

WANG, Y., CELLA, M., MALLINSON, K., ULRICH, J. D., YOUNG, K. L., ROBINETTE, M. L., GILFILLAN, S., KRISHNAN, G. M., SUDHAKAR, S., ZINSELMEYER, B. H., HOLTZMAN, D. M., CIRRITO, J. R. & COLONNA, M. 2015. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell,* 160**,** 1061-71.

WEISSOVA, K., BARTOS, A., SLADEK, M., NOVAKOVA, M. & SUMOVA, A. 2016. Moderate Changes in the Circadian System of Alzheimer's Disease Patients Detected in Their Home Environment. *PLoS One,* 11**,** e0146200.

WEST, R. L., LEE, J. M. & MAROUN, L. E. 1995. Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient. *Journal of Molecular Neuroscience,* 6**,** 141-146.

WETZEL-SMITH, M. K., HUNKAPILLER, J., BHANGALE, T. R., SRINIVASAN, K., MALONEY, J. A., ATWAL, J. K., SA, S. M., YAYLAOGLU, M. B., FOREMAN, O., ORTMANN, W., RATHORE, N., HANSEN, D. V., TESSIER-LAVIGNE, M., MAYEUX, R., PERICAK-VANCE, M., HAINES, J., FARRER,

L. A., SCHELLENBERG, G. D., GOATE, A., BEHRENS, T. W., CRUCHAGA, C., WATTS, R. J. & GRAHAM, R. R. 2014. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med,* 20**,** 1452-7.

WU, H., XU, T., FENG, H., CHEN, L., LI, B., YAO, B., QIN, Z., JIN, P. & CONNEELY, K. N. 2015. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Research,* 43**,** e141-e141.

WU, K. C. & JIN, J. P. 2008. Calponin in non-muscle cells. *Cell Biochem Biophys,* 52**,** 139-48.

XIE, W., BARR, C. L., KIM, A., YUE, F., LEE, A. Y., EUBANKS, J., DEMPSTER, E. L. & REN, 2012. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell,* 148**,** 816-31.

XU C, LIU G, JI H, CHEN W, DAI D, CHEN Z, ZHOU D, XU L, HU H, CUI W, CHANG L, ZHA Q, LI L, DUAN S, WANG Q. Elevated methylation of OPRM1 and PRL1 genes in Alzheimer's disease. 2018. *Mol Med Rep.* 18(5), 4297-4302.

XU, H. X., GREENGARD, P. & GANDY, S. 1995. REGULATED FORMATION OF GOLGI SECRETORY VESICLES CONTAINING ALZHEIMER BETA-AMYLOID PRECURSOR PROTEIN. *Journal of Biological Chemistry,* 270**,** 23243-23245.

XU J, PATASSINI S, RUSTOGI N, RIBA-GARCIA I, HALE BD, PHILLIPS AM, WALDVOGEL H, HAINES R, BRADBURY P, STEVENS A, FAULL RLM, DOWSEY AW, COOPER GJS, UNWIN RD. Regional protein expression in human Alzheimer's brain correlates with disease severity. 2019. 2, 43.

YANG, L. C., RIEVES, D. & GANLEY, C. 2012. Brain Amyloid Imaging - FDA Approval of Florbetapir F18 Injection. *New England Journal of Medicine,* 367**,** 885-887.

YANKNER, B. A., DUFFY, L. K. & KIRSCHNER, D. A. 1990. NEUROTROPHIC AND NEUROTOXIC EFFECTS OF AMYLOID BETA-PROTEIN - REVERSAL BY TACHYKININ NEUROPEPTIDES. *Science,* 250**,** 279-282.

YAMAZAKI, KIYOHIRO, YOSHINO, YUTA, MORI, TAKAAKI,YOSHIDA, TAKU, OZAKI, YUKI, SAO, TOMOKO, MORI, YOKO, OCHI, SHINICHIRO, IGA, JUN-ICHI, UENO, SHU-ICHI. Gene Expression and Methyltaion Analysis of *ABCA7* in Pateints with Alzhimer's Disease. 2017. *Journal of Amzheimer's Disease.* 57(1), 171-181.

YEH, F. L., WANG, Y., TOM, I., GONZALEZ, L. C. & SHENG, M. 2016. TREM2 Binds to Apolipoproteins, Including APOE and CLU/APOJ, and Thereby Facilitates Uptake of Amyloid-Beta by Microglia. *Neuron,* 91**,** 328-40.

YIN, G. N., LEE, H. W., CHO, J. Y. & SUK, K. 2009. Neuronal pentraxin receptor in cerebrospinal fluid as a potential biomarker for neurodegenerative diseases. *Brain Research,* 1265**,** 158-170.

YU, J.-T., TAN, L. & HARDY, J. 2014a. Apolipoprotein E in Alzheimer's Disease: An Update. *Annual review of neuroscience*.

YU, L., CHIBNIK, L. B., SRIVASTAVA, G. P., POCHET, N., YANG, J., XU, J., KOZUBEK, J., OBHOLZER, N., LEURGANS, S. E., SCHNEIDER, J. A., MEISSNER, A., DE JAGER, P. L. & BENNETT, D. A. 2014b. Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. *JAMA Neurol*.

YU, M., HON, G. C., SZULWACH, K. E., SONG, C. X., JIN, P., REN, B. & HE, C. 2012. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc,* 7**,** 2159-70.

ZHANG, B., GAITERI, C., BODEA, L. G., WANG, Z., MCELWEE, J., PODTELEZHNIKOV, A. A., ZHANG, C., XIE, T., TRAN, L., DOBRIN, R., FLUDER, E., CLURMAN, B., MELQUIST, S., NARAYANAN, M., SUVER, C., SHAH, H., MAHAJAN, M., GILLIS, T., MYSORE, J., MACDONALD, M. E., LAMB, J. R., BENNETT, D. A., MOLONY, C., STONE, D. J., GUDNASON, V., MYERS, A. J.,

SCHADT, E. E., NEUMANN, H., ZHU, J. & EMILSSON, V. 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell,* 153**,** 707-20.

ZHANG, Z. 2016. Long non-coding RNAs in Alzheimer's disease. *Curr Top Med Chem,* 16**,** 511-9.

ZHANG, Z. G., LI, Y., NG, C. T. & SONG, Y. Q. 2015. Inflammation in Alzheimer's Disease and Molecular Genetics: Recent Update. *Arch Immunol Ther Exp (Warsz),* 63**,** 333-44.

ZHAO, J., DENG, Y., JIANG, Z. & QING, H. 2016. G Protein-Coupled Receptors (GPCRs) in Alzheimer's Disease: A Focus on BACE1 Related GPCRs. *Front Aging Neurosci,* 8.

ZHU, X., CASTELLANI, R. J., TAKEDA, A., NUNOMURA, A., ATWOOD, C. S., PERRY, G. & SMITH, M. A. 2001. Differential activation of neuronal ERK, JNK/SAPK and p38 in Alzheimer disease: the 'two hit' hypothesis. *Mech Ageing Dev,* 123**,**
39-46.

ZILLER, M. J., HANSEN, K. D., MEISSNER, A. & ARYEE, M. J. 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods,* 12**,** 230-2, 1 p following 232.

ZILLER, M. J., MÜLLER, F., LIAO, J., ZHANG, Y., GU, H., BOCK, C., BOYLE, P., EPSTEIN, C. B., BERNSTEIN, B. E., LENGAUER, T., GNIRKE, A. & MEISSNER, A. 2011. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *In:* SCHÜBELER, D. (ed.) *PLoS Genet.* San Francisco, USA.

ZONG, Y., YU, P., CHENG, H., WANG, H., WANG, X., LIANG, C., ZHU, H., QIN, Y. & QIN, C. 2015. miR-29c regulates NAV3 protein expression in a transgenic mouse model of Alzheimer's disease. *Brain Res,* 1624**,** 95-102.